

probabilités

1

licence
capes

2

master
agrégation

Jean-Yves Ouvrard

l'essentiel

en théorie des

probabilités

Jean Jacod

Philip Protter

C A S S I N I

+



PUSCULES

Pile ou Face

Emmanuel LESIGNE

Une introduction
aux théorèmes limites du
Calcul des Probabilités



probabilités

1

licence
capes

Jean-Yves Oувrard

C A S S I N I

PROBABILITÉS I

Enseignement des mathématiques

1. J.-Y. Oувrard, *Probabilités I*
2. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques*
3. M. Cottrell, V. Genon-Catalot, Ch. Duhamel, Th. Meyre, *Exercices de probabilités*
4. F. Rouvière, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*
5. J.-Y. Oувrard, *Probabilités II*
6. G. Zémor, *Cours de cryptographie*
7. A. Szpirglas, *Exercices d'algèbre*
8. B. Perrin-Riou, *Algèbre, arithmétique et Maple*
10. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 1*
11. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 1*
12. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 2*
13. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 2*
14. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 3*
15. H. Krivine, *Exercices de mathématiques pour physiciens*
16. J. Jacod, Ph. Protter, *L'essentiel en théorie des probabilités*
17. M. Willem, *Analyse fonctionnelle élémentaire*
18. É. Amar, É. Matheron, *Analyse complexe*
19. B. Randé, *Problèmes corrigés. Concours 2002 et 2003 (MP)*
20. D. Perrin, *Mathématiques d'école*
21. B. Randé, *Problèmes corrigés. Concours 2004 (MP)*
22. P. Bourgade, *Olympiades internationales de mathématiques 1976-2005*
23. V. Prasolov, *Problèmes et théorèmes d'algèbre linéaire*

JEAN-YVES OUVRARD

Probabilités

TOME I

Licence - CAPES

CASSINI

JEAN-YVES OUVRARD est maître de conférences à l'Université Joseph Fourier de Grenoble. Il est docteur d'état en mathématiques.

Imprimé sur papier permanent.

ISBN 978-2-84225-130-7

© Cassini, Paris, 2007.

(1^{re} édition, 1998. ISBN 2-84225-004-4)

Table des matières

Introduction	1
Chapitre 1. Phénomènes aléatoires et modèles probabilistes	3
1.1. La notion d'expérience aléatoire	3
1.2. L'algèbre des événements	6
1.3. Axiomes des tribus et des probabilités. Premières propriétés	8
1.4. Espaces probabilisés discrets	13
1.5. Variables aléatoires	18
Exercices	21
Chapitre 2. Familles sommables de nombres réels	29
2.1. Somme d'une famille de réels positifs	29
2.2. Arithmétique dans $\overline{\mathbb{R}}$. Somme d'une famille d'éléments de $\overline{\mathbb{R}}$	32
2.3. Somme d'une famille de réels de signe quelconque	38
Exercices	45
Chapitre 3. Indépendance	51
3.1. Introduction	51
3.2. Indépendance d'événements et de variables aléatoires . . .	54
3.3. Loi d'une somme de variables aléatoires indépendantes . .	61
3.4. Indépendance et produits cartésiens : construction d'un modèle	65
3.5. Modèles géométriques et binomiaux	67
Exercices	72
Chapitre 4. Probabilités et lois conditionnelles	87
4.1. Probabilités conditionnelles	87
4.2. Lois conditionnelles	93
4.3. Modélisation d'un phénomène évolutif	96
Exercices	99

Chapitre 5. Moments d'une variable aléatoire discrète	115
5.1. Moyenne, ou espérance mathématique	115
5.2. Moments d'ordre supérieur	122
5.3. Fonctions génératrices	137
Exercices	145
Chapitre 6. Variables aléatoires à densité	175
6.1. Probabilités sur \mathbb{R}^n	175
6.2. Loi d'une variable aléatoire à valeurs dans \mathbb{R}^n	179
6.3. Moyenne et variance d'une variable aléatoire réelle	185
6.4. Loi de Laplace-Gauss à deux dimensions	192
6.5. Indépendance de deux variables aléatoires réelles	196
6.6. Somme de variables aléatoires réelles indépendantes	199
6.7. Densités conditionnelles	200
6.8. Annexe. L'intégrale de Riemann dans \mathbb{R}^n	202
Exercices	210
Chapitre 7. Approximation de lois. Loi faible des grands nombres	219
7.1. Approximation de lois	219
7.2. Loi faible des grands nombres	228
Exercices	230
Tableau des lois de probabilité usuelles	237
Bibliographie	239
Index	241
<i>Liste des chapitres du deuxième tome</i>	
8. Lois et moments de variables aléatoires	
9. Indépendance de tribus, de variables aléatoires	
10. Convergences et lois des grands nombres	
11. Probabilités et espérances conditionnelles	
12. Transformée de Fourier et fonctions caractéristiques	
13. Variables aléatoires gaussiennes	
14. Convergences de mesures et convergence en loi	
15. Processus et martingales discrets	
16. Chaînes de Markov	
Annexe. Résumé de théorie de la mesure	

Introduction

Rien ne m'effraie davantage que les certitudes qui prescrivent le bon droit, la normalité du contrat social, la justice...

Derrière les mots, on oublie ceux qui parlent, leurs regards ou leur absence de regard, leurs sourires ou leurs faux-semblants qui énoncent tout ce qui est tu, sous la parole « bien pensante ».

Yvon Chaix

Théâtre du Rio, Grenoble, 1995.

Objectif et stratégie

L'auteur a longtemps été membre des jurys des concours de l'agrégation de Mathématiques, concours externe (1987 à 1990) et interne (1990 à 1995). Depuis plusieurs années, il s'occupe de la préparation à l'agrégation à l'Université Joseph Fourier (Grenoble). Ainsi est née l'idée d'écrire un ouvrage de probabilités destiné aux candidats aux différents concours de recrutement de l'enseignement secondaire.

Les programmes de Probabilités pour les épreuves du CAPES et du concours interne de l'agrégation sont à quelques nuances près les mêmes, à des niveaux légèrement différents. Ils correspondent à ce qu'on traite d'habitude dans les enseignements de Probabilités des deux premières années d'université. Le point commun essentiel de ces programmes est qu'il n'est pas fait appel à la théorie de la mesure. Il en est tout autrement du programme du concours externe de l'agrégation. Tant pour les épreuves écrites et orales d'analyse et probabilités, communes à tous les candidats, que pour la troisième épreuve orale (épreuve de modélisation), on demande une connaissance solide de la théorie classique du calcul des probabilités, reposant sur la théorie de la mesure.

Cela conduit à la structure « à tiroirs » de cet ouvrage :

Le **premier tome** est destiné à tous les candidats aux concours du CAPES et de l'agrégation, et suffisant pour les candidats au concours du CAPES et au concours interne de l'agrégation ; il peut se lire éventuellement à différents niveaux, en ne prenant que ce qui est nécessaire dans chaque programme.

Il y est essentiellement traité des probabilités *discrètes*, avec un souci permanent de faire prendre conscience de ce qu'est la démarche de *modélisation* probabiliste. Un chapitre traite des variables aléatoires à

densité, *sans entrer dans des considérations de théorie de la mesure.*

Ce premier tome pourra être très utile à tous les étudiants de licence de Mathématiques.

Le **second tome** est destiné aux candidats au concours externe de l'agrégation et développe tous les points du programme. Bien sûr, certains chapitres, en particulier ceux traitant des martingales et des chaînes de Markov, ne concernent, a priori, que les candidats qui choisissent à l'oral l'option probabilités et statistiques. À noter qu'en annexe figure un résumé de théorie de la mesure donnant les résultats utilisés dans l'exposé.

Ce second tome rendra de grands services aux étudiants de master de Mathématiques.

La *pratique* est un élément déterminant pour le candidat, qui trouvera à la fin de chaque chapitre des *exercices corrigés* sur chaque notion. En complément de formation, il est évident que le candidat devra faire un certain nombre de problèmes réels de concours (textes et corrigés, parfois très condensés, se trouvent dans les rapports de jurys de concours publiés par le Centre national de documentation pédagogique).

Caractéristiques générales

Cet ouvrage n'est pas un recueil de leçons toutes faites, car cela conduit souvent à des prestations stéréotypées peu appréciées des interrogateurs. Il parcourt tout le programme en essayant de garder une unité et une progression ; chaque notion est *illustrée* par un *exemple*. Le lecteur trouve là, ainsi que dans les *exercices corrigés* de fin de chapitre, tous les matériaux pour construire *lui-même* une leçon.

Je tiens à remercier les éditions Cassini : en rendant accessible cet ouvrage à un public motivé par la perspective d'un concours, mais aussi curieux d'apprendre et de réfléchir, elles me permettent d'apporter une aide, je l'espère fructueuse, à toute personne qui aura eu la patience de me suivre.

Je remercie tout particulièrement André Bellaïche, avec qui j'ai eu de longues et fructueuses confrontations de points de vue sur cet ouvrage. En particulier, il a eu la gentillesse d'écrire l'annexe sur l'intégrale de Riemann dans \mathbb{R}^n , ce qui permet de donner plus de rigueur au chapitre 6, rigueur difficilement accessible lorsqu'on s'astreint à traiter des variables aléatoires « continues » uniquement avec l'intégrale de Riemann.

Enfin, je remercie les relecteurs de cet ouvrage ; leurs remarques ont contribué au polissage du manuscrit et conduit à la forme définitive de ce livre. Je souhaite que le lecteur trouve ici matière à un travail agréable et enrichissant.

Chapitre 1

Phénomènes aléatoires et modèles probabilistes

*La plupart des hommes ont, comme les plantes,
des propriétés cachées que le hasard fait découvrir.*

La Rochefoucauld (1613-1680)

La théorie des Probabilités est la branche des mathématiques qui étudie les phénomènes aléatoires, c'est-à-dire soumis au hasard. Notons que le mot *aléatoire* vient du latin *alea*, le dé, la chance, et que *hasard* vient de l'arabe *az-zahr*, le dé.

Issue de l'étude des jeux de hasard, qui lui ont fourni ses notions de base (probabilité d'un événement, espérance de gain), la théorie des Probabilités est aujourd'hui une branche des mathématiques au même titre que la Géométrie, l'Algèbre et l'Analyse. Elle entretient naturellement des liens étroits avec les applications : en Sciences physiques, en Biologie (où le hasard est omniprésent, comme source de la diversité biologique des individus – la variabilité des caractères à son tour impose que toute étude dans ce domaine doit être statistique), dans les Sciences économiques et sociales et en technologie (contrôler une situation, c'est maîtriser le hasard).

1.1. La notion d'expérience aléatoire

Pour parler du hasard en termes mathématiques, il est nécessaire d'isoler une situation, du type de celles qui se produisent à intervalles réguliers dans les jeux de hasard, lorsque on lance les dés, ou lorsque on distribue un paquet de cartes. À une telle situation isolée, répétable en principe, on donne le nom d'expérience aléatoire.

D'une façon générale, une **expérience aléatoire** est une expérience renouvelable, en principe sinon en pratique, et qui, renouvelée dans des conditions identiques, ne donne pas à chaque essai le même résultat.

Donnons comme exemples :

a) lancer deux dés ;

- b) distribuer les cartes au bridge, c'est-à dire répartir le paquet de 52 cartes entre les 4 joueurs ;
- c) observer le bris d'un objet fragile, une vitre, par exemple ;
- d) observer la formation d'un caractère génétique d'un individu à partir des caractères correspondants de ses parents ;
- e) observer la désintégration d'un noyau atomique radioactif ;
- f) attendre l'autobus à une station donnée, à partir de 18 heures.

Précisons. Quand nous parlons de conditions identiques, il faut entendre : « identiques, pour autant que l'observateur puisse s'en assurer ».

Dans une première approche, on peut penser que c'est là que se glisse le hasard : entre les conditions idéales et les conditions réelles. Si le résultat d'une expérience comporte une part prévisible et une part imprévisible, on considérera le hasard comme la « cause » de la part imprévisible. C'est la vision courante du hasard.

Le hasard apparaît ainsi souvent comme un nom que nous donnons à notre ignorance de certaines conditions de l'expérience : choisir entre deux boîtes, l'une renfermant un prix et l'autre vide, est aussi une façon de tirer au sort. (On voit ici en passant que ce qui est aléatoire pour un observateur ne l'est pas forcément pour un autre.) Il y a donc, notamment, du hasard quand le nombre de facteurs est si grand qu'il est impossible de tous les contrôler.

Mais il y a beaucoup plus intéressant : le hasard s'introduit presque de force dans certains phénomènes où tout est en principe connu, et qui sont qualifiés de *déterministes*. En mécanique des solides, les équations du mouvement sont connues, et la connaissance des conditions initiales permet en principe de déterminer la position du système à un instant ultérieur donné. Et pourtant, jetons un dé ; qui peut affirmer à l'avance sur quelle face il retombera, même s'il connaît bien les caractéristiques de ce dé, caractéristiques géométriques (forme) et mécaniques (répartition homogène ou non de la matière, position et régularité des incrustations des chiffres, etc.) ? C'est que toute imprécision dans notre connaissance des conditions initiales se trouve amplifiée au cours du mouvement. Dans certains systèmes, cette amplification reste limitée, mais dans d'autres comme celui d'un dé roulant sur un tapis, l'amplification des incertitudes est extraordinairement rapide, et interdit toute prévision. C'est d'une façon analogue à celle-ci que le hasard s'introduit dans la plupart des systèmes déterministes¹. (Il y a toutefois un domaine où le

1. C'est le point de vue de Poincaré, mal compris à son époque, remis en honneur aujourd'hui avec la théorie du chaos. Un joli livre sur ces questions, et sur les rapports

hasard ne provient pas de la dégradation d'une connaissance qui pourrait en théorie être totale, c'est celui des phénomènes physiques à l'échelle atomique ou sub-atomique qu'étudie la Mécanique quantique.)

Voici une liste de phénomènes aléatoires, où le hasard intervient, et qui font l'objet de recherches scientifiques ou technologiques (la discipline concernée est mentionnée entre parenthèses) :

- l'apparition d'objets défectueux dans une chaîne de fabrication (*théorie de la fiabilité, contrôle de qualité*) ;
- les effets d'un engrais chimique sur la croissance de céréales (ces effets sont nécessairement variables et justiciables d'un traitement probabiliste ou statistique : la variabilité est la règle dans le monde vivant) ;
- les appels reçus à un standard téléphonique ou les temps d'occupation de postes d'un réseau d'ordinateurs (*théorie des files d'attente et des réseaux*) ;
- les dates, c'est-à-dire les heures précises, d'arrivée d'un bus à une station (*idem*) ;
- le mouvement brownien (mouvement d'une particule dans un liquide observé par le botaniste écossais Robert Brown en 1827, et appelé depuis mouvement brownien ; son étude mathématique est à l'origine d'une recherche encore très actuelle et vivante et d'applications dans des domaines très divers comme *les télécommunications, la médecine, la théorie des assurances, la théorie des marchés financiers, etc.*) ;
- les incertitudes sur la trajectoire d'un véhicule spatial : l'aléatoire vient de ce que de nombreuses perturbations font que le véhicule ne peut suivre exactement sa trajectoire théorique (*théorie du filtrage, contrôle stochastique*).

Pour étudier un phénomène dépendant du *hasard* (au sens que nous avons essayé de cerner) le praticien essaiera d'isoler une expérience aléatoire, et il en construira un modèle probabiliste lui permettant de faire certaines prévisions ; disons pour être concret qu'il pourra calculer la probabilité de certains résultats ou de certains événements. *Mais les conclusions qu'il pourra tirer n'auront de sens qu'à l'intérieur du modèle !* S'il veut aller plus loin, il lui faudra *valider son modèle* et pour ce, il confrontera à son modèle les données recueillies lors de l'expérience ; c'est l'objet de la *statistique*. Si cette épreuve est passée avec succès, il sera en droit de *faire des prévisions* à l'aide de son modèle ou d'*expliquer*

que la science entretient avec la chance, est celui de Ivar Ekeland : *Au hasard : la chance, la science et le monde* (Éditions du Seuil, 1991).

un phénomène à l'aide de ce modèle.

La première étape de la modélisation mathématique d'une expérience aléatoire consiste à spécifier l'ensemble des *résultats possibles* (on dit aussi *réalisations* ou *issues*) de cette expérience. Il est traditionnel de représenter cet ensemble par la lettre Ω . On emploie la lettre ω pour représenter les points de cet ensemble, c'est-à-dire les résultats possibles de l'expérience en question.

Par exemple, dans l'expérience *a*) citée au début, les résultats possibles sont les couples de deux entiers compris entre 1 et 6; on prend donc $\Omega = \{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\}$. Dans l'expérience *b*), l'ensemble Ω est formé de toutes les répartitions possibles des 52 cartes entre les 4 joueurs. Bien que fini, cet ensemble est très grand : il comporte 53 644 737 765 488 792 839 237 440 000 éléments. Enfin, dans l'exemple de la désintégration radioactive, on peut prendre, si on ne s'intéresse qu'à la durée de vie du noyau, $\Omega =]0, +\infty[$, l'ensemble des nombres réels > 0 .

Une même expérience pourra être décrite de plusieurs façons différentes : si deux joueurs lancent deux dés chacun à leur tour, avec comme objectif d'obtenir le meilleur total, on pourra prendre pour Ω aussi bien l'ensemble $\{1, 2, \dots, 6\}^4$ que l'ensemble $\{2, 3, \dots, 12\} \times \{2, 3, \dots, 12\}$.

Comme on le voit, l'objectif de l'étude peut entrer en ligne de compte pour déterminer ce que l'on entend par réalisation.

En définitive, observons que *toutes les expériences aléatoires se traduisent de la même façon, par le choix « au hasard » d'un point ω dans un ensemble Ω adéquat.*

On désigne l'ensemble Ω choisi pour décrire les résultats possibles sous le nom d'*univers* ou *ensemble des éventualités*, ou *des issues*, ou *espace fondamental* (de l'expérience considérée). Mais le plus souvent on dit simplement « l'espace Ω ».

1.2. L'algèbre des événements

Un **événement aléatoire** est un événement lié à une expérience aléatoire : il peut se réaliser, ou non, et sa réalisation, ou sa non-réalisation, dépend exclusivement du résultat ω de cette expérience. On peut aussi considérer un événement comme une propriété du résultat ω , susceptible d'être vérifiée, ou non, ou encore comme une assertion portant sur ω , qui peut être vraie, ou fausse.

Mathématiquement, un événement aléatoire se représente par la donnée de l'ensemble des points ω de Ω pour lesquels cet événement est

réalisé.

Dans le premier exemple ci-dessus, l'événement « la somme des entiers figurant sur les faces exposées est ≤ 4 » est représenté par le sous-ensemble

$$\{(1, 1), (1, 2), (2, 1), (1, 3), (2, 2), (3, 1)\}$$

de $\Omega = \{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\}$.

C'est l'habitude, en Probabilités, d'identifier un événement et la partie (synonyme de sous-ensemble) de Ω qui le représente. On donne aussi à ce sous-ensemble le nom d'événement, et on le désigne par la même lettre.

Cette identification joue un rôle fondamental en Probabilités : les notions et opérations logiques que l'on définit sur les événements correspondent aux notions et opérations que l'on définit en théorie des ensembles. Cette correspondance est résumée dans le tableau suivant, que nous empruntons à Jacques Neveu.

Terminologie probabiliste	Terminologie ensembliste
résultat possible	ω , point de Ω
événement	A , sous-ensemble de Ω
A est réalisé	$\omega \in A$
A implique B	$A \subset B$
A ou B	$A \cup B$
A et B	$A \cap B$
contraire de A (A ne se produit pas)	A^c (complémentaire de A)
événement impossible	\emptyset
événement certain	Ω
A et B sont incompatibles	$A \cap B = \emptyset$

Par contre, on ne considère pas, en général, qu'un sous-ensemble quelconque de Ω est (ou : est associé à) un événement. Il y a à cela plusieurs raisons qui seront discutées dans le tome II de cet ouvrage. Donnons simplement l'exemple suivant : si dans le lancer de deux dés, on considère les deux dés comme indiscernables, il devient impossible d'observer le résultat $\omega = (1, 2)$. En conséquence, on ne considère pas le sous-ensemble $\{(1, 2)\}$ comme un événement.

Il est évident qu'en général on s'intéresse à toute une liste \mathcal{L} de propriétés du résultat qu'il est raisonnable de supposer stable par négation (c'est-à-dire que si L est dans la liste, son contraire le sera aussi),

conjonction (si L_1 et L_2 sont dans la liste, la propriété L_1 **et** L_2 le sera aussi) et disjonction (c'est-à-dire que si L_1 et L_2 sont dans la liste, la propriété L_1 **ou** L_2 le sera aussi) pour pouvoir dire si respectivement une propriété est satisfaite ou non, si deux propriétés sont satisfaites simultanément ou si l'une des deux est satisfaite. Si \mathcal{A} est la famille des événements (au sens mathématique) correspondant à chacune des propriétés de la liste \mathcal{L} , cela revient à supposer que \mathcal{A} est stable par passage au complémentaire, intersection et réunion. Une famille de parties possédant ces propriétés est appelée une algèbre. On parle donc de **l'algèbre des événements**. Pour des raisons mathématiques que nous verrons plus tard (passages à la limite par exemple), on exige en plus que \mathcal{A} soit stable par union dénombrable.

1.3. Axiomes des tribus et des probabilités. Premières propriétés

Passons maintenant à la théorie mathématique proprement dite. Nous commençons par introduire les notions de tribu et d'espace probabilisable, préalables à la définition des probabilités.

Définition 1.1. Une famille \mathcal{A} de parties d'un ensemble Ω est une tribu ou σ -algèbre sur Ω si elle satisfait aux trois axiomes suivants :

- (i) $\Omega \in \mathcal{A}$.
- (ii) Si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$.
- (iii) Pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , on a $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Exemples 1.1. 1. La famille $\mathcal{P}(\Omega)$ des parties de Ω est une tribu.

2. Si $A \in \mathcal{P}(\Omega)$, la famille $\{A, A^c, \Omega, \emptyset\}$ est une tribu sur Ω dite tribu engendrée par l'événement A .

3. La tribu $\{\Omega, \emptyset\}$ est appelée tribu triviale sur Ω .

Proposition 1.2. Si \mathcal{A} est une tribu sur Ω , on a les propriétés suivantes :

- (i) $\emptyset \in \mathcal{A}$.
- (ii) Pour toute suite finie $(A_i)_{1 \leq i \leq n}$ d'éléments de \mathcal{A} , on a $\bigcup_{1 \leq i \leq n} A_i \in \mathcal{A}$.
- (iii) Pour toute suite finie $(A_i)_{1 \leq i \leq n}$ d'éléments de \mathcal{A} , on a

$$\bigcap_{1 \leq i \leq n} A_i \in \mathcal{A}.$$

(iv) Pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , on a $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

(v) Pour tout A et tout B appartenant à \mathcal{A} , on a $A \setminus B \in \mathcal{A}$.

Démonstration.

(i) Il suffit de remarquer que $\emptyset = \Omega^c$.

(ii) Soit une suite finie $(A_i)_{1 \leq i \leq n}$ d'éléments de \mathcal{A} . On définit la suite $(B_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} en posant $B_i = A_i$ pour $1 \leq i \leq n$, et $B_i = \emptyset$ pour $i = 0$ ou $i \geq n + 1$.

On a : $\bigcup_{1 \leq i \leq n} A_i = \bigcup_{n \in \mathbb{N}} B_n$, ce qui démontre le résultat.

(iii) On passe au complémentaire (c'est un procédé souvent utilisé). Plus précisément, soit $(A_i)_{1 \leq i \leq n}$ une suite finie d'éléments de \mathcal{A} ; on a :

$$\left[\bigcap_{1 \leq i \leq n} A_i \right]^c = \bigcup_{1 \leq i \leq n} (A_i)^c.$$

D'après le deuxième axiome de définition des tribus, pour tout i , A_i^c appartient à \mathcal{A} ; d'après la deuxième propriété, le membre de droite appartient donc aussi à \mathcal{A} et, par passage au complémentaire, $\bigcap_{1 \leq i \leq n} A_i$ aussi.

(iv) On utilise encore un passage au complémentaire (s'en persuader).

(v) Puisque A et B^c appartiennent à \mathcal{A} , l'ensemble $A \cap B^c$ appartient aussi à \mathcal{A} .

Définition 1.3. On appelle **espace probabilisable**³ un couple (Ω, \mathcal{A}) , où Ω est un ensemble et \mathcal{A} une tribu sur l'ensemble Ω .

Quand un espace probabilisable est fixé, on dit que \mathcal{A} est la tribu des événements.

La modélisation d'un phénomène aléatoire et d'une famille de propriétés ou d'événements commence par le choix d'un espace probabilisable qui rend compte de l'ensemble des réalisations envisagées et des propriétés ou événements qui peuvent être sujets de l'étude. Cet espace donne une description qualitative du phénomène et des propriétés liées

2. On rappelle que l'ensemble des points appartenant à A sans appartenir à B est noté $A \setminus B$; c'est l'ensemble $A \cap B^c$. À noter que A et B sont quelconques et ne sont pas forcément inclus l'un dans l'autre!

3. Un espace probabilisable n'est autre que ce qu'on appelle en théorie de la mesure un espace mesurable.

à son étude. Nous allons définir la notion de *probabilité* sur un *espace probabilisable* qui permettra de traduire *quantitativement* (sans que nous cherchions à être plus précis pour l'instant) les idées que l'on a sur le phénomène.

Définition 1.4. Soit (Ω, \mathcal{A}) un espace probabilisable. Une **probabilité** P sur cet espace est une application de \mathcal{A} dans \mathbb{R}^+ satisfaisant aux deux axiomes :

$$(i) \quad P(\Omega) = 1.$$

(ii) La fonction d'ensemble P est **σ -additive**, ce qui signifie que pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments disjoints deux à deux de \mathcal{A} on a :

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{\infty} P(A_n).$$

Le triplet (Ω, \mathcal{A}, P) est appelé **espace probabilisé**⁴ ou **espace de probabilité**.

On emploie souvent, dans certaines situations, l'expression **loi de probabilité** au lieu de probabilité.

La proposition suivante énumère les propriétés élémentaires des espaces probabilisés.

Proposition 1.5. Soit (Ω, \mathcal{A}, P) un espace probabilisé. Les propriétés suivantes sont satisfaites :

$$(i) \quad \text{On a : } P(\emptyset) = 0.$$

(ii) Pour toute suite finie $(A_i)_{1 \leq i \leq n}$ d'éléments disjoints deux à deux de \mathcal{A} , on a⁵ :

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

(iii) Si $A, B \in \mathcal{A}$ et si $A \subset B$, on a :

$$P(A) \leq P(B)$$

et

$$P(B \setminus A) = P(B) - P(A). \quad (1.1)$$

4. Cette axiomatique a été introduite en 1929 par le mathématicien russe Andreï Kolmogorov (1903–1987) afin de rendre rigoureux le calcul des probabilités par l'utilisation de la théorie de la mesure.

5. On dit que P est une fonction d'ensemble **finiment additive**.

(iv) Pour tout $A \in \mathcal{A}$, on a : $P(A) \in [0, 1]$.

(v) Pour tout $A \in \mathcal{A}$, on a :

$$P(A^c) = 1 - P(A). \quad (1.2)$$

(vi) Pour toute suite croissante $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , c'est-à-dire telle que $A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}$, on a⁶ :

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_n \uparrow P(A_n).$$

(vii) Pour toute suite décroissante $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , c'est-à-dire telle que $A_n \supset A_{n+1}$ pour tout $n \in \mathbb{N}$, on a :

$$P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_n \downarrow P(A_n).$$

(viii) Si A et B sont des éléments de \mathcal{A} , on a :

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}. \quad (1.3)$$

Démonstration.

(i) On construit une suite $(B_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} en posant $B_0 = \Omega$ et $B_i = \emptyset$ pour tout entier $i \geq 1$. On a : $\bigcup_{n \in \mathbb{N}} B_n = \Omega$, et les B_n sont disjoints deux à deux. Il suffit alors d'appliquer la σ -additivité de P .

(ii) On définit une suite $(B_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} en posant $B_i = A_i$ si $1 \leq i \leq n$ et $B_i = \emptyset$ si $i = 0$ ou $i \geq n+1$. On a $\bigcup_{1 \leq i \leq n} A_i = \bigcup_{n \in \mathbb{N}} B_n$; il suffit alors d'appliquer la σ -additivité de P et la propriété (i).

(iii) On a⁷ $B = A \uplus (B \setminus A)$ et, par conséquent, $P(B) = P(A) + P(B \setminus A)$; d'où le résultat, puisque P est positive.

(iv) Appliquer la propriété (iii) avec $B = \Omega$.

(v) Idem.

(vi) Soit $A = \bigcup_{n \in \mathbb{N}} A_n$. Pour tout $p \in \mathbb{N}$, on a

$$A = A_p \uplus \left[\biguplus_{j \geq p} (A_{j+1} \setminus A_j) \right].$$

6. On écrit $\lim_n \uparrow u_n$ pour désigner la limite $\lim_n u_n$, tout en indiquant, ou en rappelant, que la suite (u_n) est croissante. De même, *mutatis mutandis*, pour $\lim_n \downarrow u_n$.

7. $A \uplus B$ désigne l'union de A et B et précise en même temps que ces ensembles sont *disjoints*. On emploie une notation analogue pour une suite d'ensembles disjoints deux à deux. Certains auteurs écrivent plutôt $A + B$.

On en déduit, en utilisant la σ -additivité de P , la relation

$$P(A) = P(A_p) + \sum_{j \geq p} P(A_{j+1} \setminus A_j).$$

On observe alors, en faisant $p = 0$ dans la relation ci-dessus, que la série de terme général $P(A_{j+1} \setminus A_j)$ est convergente. Son reste $\sum_{j \geq p} P(A_{j+1} \setminus A_j)$ tend donc vers zéro lorsque p tend vers l'infini. Par conséquent, $P(A_p)$ tend vers $P(A)$ (en croissant) quand p tend vers l'infini.

(vii) Soit une suite *décroissante* $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} . On lui associe la suite *croissante* $(B_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} définie par : $B_n = A_n^c$. On applique alors la propriété précédente et l'égalité 1.2.

(viii) On peut écrire : $A \cup B = A \uplus [B \setminus (A \cap B)]$. Reste à appliquer la propriété d'additivité simple de P et la formule 1.1.

La formule 1.3 se généralise au cas d'une réunion de n sous-ensembles éléments de \mathcal{A} .

Proposition 1.6. (Formule de Poincaré.) *Si $n \geq 2$, pour toute suite finie $(A_i)_{1 \leq i \leq n}$ d'éléments de \mathcal{A} , on a :*

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n).$$

Cette formule peut encore s'écrire de la manière suivante ; si I est un ensemble fini de cardinal ≥ 2 , pour toute famille $(A_i)_{i \in I}$ d'éléments de \mathcal{A} , on a, en notant $|J|$ le cardinal d'un ensemble J :

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{\substack{J \subset I \\ |J|=1}} P\left(\bigcap_{j \in J} A_j\right) - \sum_{\substack{J \subset I \\ |J|=2}} P\left(\bigcap_{j \in J} A_j\right) + \sum_{\substack{J \subset I \\ |J|=3}} P\left(\bigcap_{j \in J} A_j\right) - \dots + (-1)^{|I|-1} P\left(\bigcap_{j \in I} A_j\right).$$

Démonstration. On démontre cette propriété par récurrence. Elle est vraie à l'ordre 2 ; supposons-la vraie à l'ordre n , c'est-à-dire supposons que, pour toute suite $(A_i)_{1 \leq i \leq n}$ d'éléments de \mathcal{A} , la formule annoncée

soit vraie. Soit une suite $(A_i)_{1 \leq i \leq n+1}$ d'éléments de \mathcal{A} . On a : $\bigcup_{i=1}^{n+1} A_i = (\bigcup_{i=1}^n A_i) \cup A_{n+1}$. Chacun des termes est un élément de \mathcal{A} . D'où, en appliquant la formule à l'ordre 2 :

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = P\left(\bigcup_{i=1}^n A_i\right) + P(A_{n+1}) - P\left[\left(\bigcup_{i=1}^n A_i\right) \cap A_{n+1}\right].$$

Reste à appliquer deux fois l'hypothèse de récurrence aux termes extrêmes du membre de droite, après avoir observé que

$$\left(\bigcup_{i=1}^n A_i\right) \cap A_{n+1} = \bigcup_{i=1}^n (A_i \cap A_{n+1})$$

et que chacun des $A_i \cap A_{n+1}$ est un élément de \mathcal{A} ; il suffit alors de regrouper convenablement les termes.

1.4. Espaces probabilisés discrets

1.4.1. Définition

La notion d'ensemble dénombrable joue un grand rôle dans ce cours. Pour la commodité du lecteur, nous rassemblons ici, sans démonstration, quelques définitions et quelques résultats bien connus.

Un ensemble **dénombrable** est, par définition, un ensemble qui peut être mis en bijection avec **une partie** de \mathbb{N} . Au sens où nous employons ici le mot *dénombrable*, un ensemble **fini** est donc dénombrable⁸.

Un sous-ensemble d'un ensemble dénombrable est dénombrable. L'image par une application quelconque d'un ensemble dénombrable est dénombrable. Si A_1, A_2, \dots, A_n sont des ensembles dénombrables, la réunion $A_1 \cup A_2 \cup \dots \cup A_n$ et le produit $A_1 \times A_2 \times \dots \times A_n$ sont dénombrables. Si $(A_n)_{n \in \mathbb{N}}$ est une suite infinie d'ensembles dénombrables, la réunion $\bigcup_{n \in \mathbb{N}} A_n$ est dénombrable.

Les ensembles $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ sont dénombrables. L'ensemble \mathbb{R} , un intervalle de \mathbb{R} non réduit à un point, ne sont pas dénombrables.

Considérons maintenant un espace probabilisable (Ω, \mathcal{A}) , et supposons que Ω soit dénombrable et que les sous-ensembles $\{\omega\}$, où $\omega \in \Omega$, appartiennent à la tribu \mathcal{A} . Tout sous-ensemble A de Ω est alors élément de \mathcal{A} . En effet, on a $A = \biguplus_{\omega \in A} \{\omega\}$, et A apparaît ainsi comme une

8. Dans certains cours, on appelle dénombrable un ensemble qui peut être mis en bijection avec \mathbb{N} . Là où nous parlons d'ensemble dénombrable, on parle alors d'ensemble fini ou dénombrable.

réunion dénombrable d'éléments de la tribu \mathcal{A} . On a par conséquent $\mathcal{A} = \mathcal{P}(\Omega)$. Cette observation conduit à la définition suivante.

Définition 1.7. *On appelle espace probabilisable discret un espace probabilisable (Ω, \mathcal{A}) où Ω est dénombrable, et où $\mathcal{A} = \mathcal{P}(\Omega)$.*

Une probabilité sur un espace probabilisable discret⁹ Ω est définie dès que l'on connaît les probabilités des événements élémentaires $\{\omega\}$. Puisque $\mathcal{A} = \biguplus_{\omega \in \Omega} \{\omega\}$, on a en effet

$$P(\mathcal{A}) = \sum_{\omega \in \Omega} P(\{\omega\}) \quad (1.4)$$

pour tout sous-ensemble¹⁰ A de Ω .

Nous verrons au chapitre suivant quel sens donner précisément à la somme figurant au second membre de 1.4. Poursuivons pour l'instant en raisonnant « formellement », tout ce que nous allons écrire étant trivial dans le cas où Ω est fini.

1.4.2. Germe d'une loi de probabilité sur un espace discret

Posons $g(\omega) = P(\{\omega\})$. Il est immédiat qu'on a $g(\omega) \geq 0$ pour tout $\omega \in \Omega$, et $\sum_{\omega \in \Omega} g(\omega) = 1$.

Inversement, si g est une fonction ≥ 0 définie sur Ω , telle que $\sum_{\omega \in \Omega} g(\omega) = 1$, il existe une et une seule probabilité sur Ω telle que $P(\{\omega\}) = g(\omega)$. C'est ce que montre le lemme suivant.

Lemme 1.8. *Soit Ω un ensemble dénombrable.*

(i) *Soit g une application de Ω dans \mathbb{R}^+ telle que $\sum_{\omega \in \Omega} g(\omega) = 1$. Pour $A \in \mathcal{P}(\Omega)$, on pose*

$$P(A) = \sum_{\omega \in A} g(\omega). \quad (1.5)$$

L'application P ainsi définie est une probabilité sur l'espace probabilisable discret $(\Omega, \mathcal{P}(\Omega))$.

(ii) *Toute probabilité sur $(\Omega, \mathcal{P}(\Omega))$ peut être obtenue de cette façon, à partir de la fonction $g(\omega) = P(\{\omega\})$.*

9. Il arrive souvent, quand on parle d'un espace probabilisable discret $(\Omega, \mathcal{P}(\Omega))$, qu'on omette de mentionner la tribu.

10. Cette propriété n'est pas vraie en général. Nous étudierons plus loin (chap. 6) des lois de probabilité définies sur $\Omega = \mathbb{R}$ telles que $P(\{x\}) = 0$ pour tout réel x (toutes les lois possédant une densité satisfont cette condition). Il est évidemment impossible dans ce cas de déduire la valeur de $P([a, b])$ des valeurs de $P(\{x\})$.

Démonstration.

(i) Évidemment, $P(\emptyset) = 0$. Soit de plus $(A_n)_{n \in \mathbb{N}}$ une suite d'éléments de \mathcal{A} *disjoints deux à deux*; on a, par définition de P :

$$\begin{aligned} P\left(\biguplus_{n \in \mathbb{N}} A_n\right) &= \sum_{\omega \in \biguplus_{n \in \mathbb{N}} A_n} g(\omega) \\ &= \sum_{n \in \mathbb{N}} \left[\sum_{\omega \in A_n} g(\omega) \right] \\ &= \sum_{n \in \mathbb{N}} P(A_n). \end{aligned}$$

(ii) C'est évident à partir de la discussion qui précède.

Remarque. L'égalité $\sum_{n \in \mathbb{N}} \left[\sum_{\omega \in A_n} g(\omega) \right] = \sum_{n \in \mathbb{N}} P(A_n)$ dans la démonstration ci-dessus est une propriété d'« associativité » qui sera établie dans le cadre des familles sommables au chapitre suivant sous le nom de *propriété de sommation par paquets*.

Définition 1.9. Une application g de l'ensemble Ω dans \mathbb{R}^+ telle que $\sum_{\omega \in \Omega} g(\omega) = 1$ est appelée *germe*¹¹ de la probabilité P définie par la relation (1.5).

1.4.3. Exemples de lois de probabilité discrètes

Les probabilités définies sur un espace probabilisable discret sont souvent appelées **lois de probabilités discrètes**.

Nous donnons ci-dessous une série d'exemples classiques d'espaces probabilisés (Ω, \mathcal{A}, P) *discrets* qui sont des *modèles élémentaires*, au sens où ils sont objets d'imitation.

Nous verrons ultérieurement dans quelles circonstances ces différents modèles apparaissent naturellement.

Exemple 1.2. Loi uniforme (ou probabilité uniforme) sur un ensemble fini. La loi uniforme sur un ensemble fini Ω est la probabilité qui attribue la même valeur à la probabilité de chaque événement élémentaire $\{\omega\}$. L'existence d'une telle probabilité P est évidente : elle est associée à un germe g qui, d'une part, est une fonction constante sur Ω , et d'autre part, vérifie la condition :

11. Cette dénomination, bien qu'évocatrice (le germe *engendre* la probabilité), n'est pas d'un usage courant (on la trouve en particulier dans certains ouvrages étrangers); nous l'adopterons systématiquement. Certains auteurs se contentent d'une périphrase pour parler de cette fonction; d'autres emploient le terme « *fonction de probabilité* ».

$$\sum_{\omega \in \Omega} g(\omega) = 1.$$

On doit donc avoir¹²

$$\forall \omega \in \Omega \quad g(\omega) = \frac{1}{|\Omega|}.$$

Pour toute partie A de Ω , on a

$$P(A) = \frac{|A|}{|\Omega|}.$$

Deux observations importantes. 1) Quand on parle de choix « au hasard » sur un ensemble fini, on sous-entend souvent que ce choix est fait au moyen de la probabilité **uniforme**, c'est-à-dire, en donnant à chaque élément de l'ensemble les mêmes chances d'être choisi. 2) Il est impossible de définir par le même procédé que ci-dessus une probabilité uniforme sur un ensemble qui n'est pas fini.

Exemple de modélisation. On lance n fois un dé équilibré ; on cherche la probabilité d'obtenir k fois ($k \leq n$) le chiffre six. Une réalisation est une succession de n entiers compris entre 1 et 6. On prend donc $\Omega = \{1, 2, \dots, 6\}^n$. L'événement « on obtient k fois le chiffre six » est représenté par le sous-ensemble

$$A_k = \{(x_1, x_2, \dots, x_n) \mid x_i = 6 \text{ pour } k \text{ indices exactement}\}$$

de Ω . Le dé étant équilibré, il est naturel¹³ de *modéliser* cette « expérience aléatoire » en munissant l'espace probabilisable *discret* $(\Omega, \mathcal{P}(\Omega))$ de la probabilité *uniforme* P qui donne la même importance à chacune des réalisations. On a : $|A_k| = \binom{n}{k} 5^{n-k}$ (choix des k indices des x_i qui sont égaux à 6, puis affectation d'un entier quelconque compris entre 1 et 5 pour les autres). Puisque $|\Omega| = 6^n$ il vient : $P(A_k) = \binom{n}{k} \frac{5^{n-k}}{6^n}$.

Exemple 1.3. Modèles géométriques sur \mathbb{N} et sur \mathbb{N}^* . Ici, $\Omega = \mathbb{N}$ (resp. \mathbb{N}^*). Le réel $p \in]0, 1[$ étant donné, la **loi géométrique** sur \mathbb{N} (resp. sur \mathbb{N}^*) de paramètre p est la loi de probabilité engendrée par le germe g défini par :

$$\begin{array}{l} \forall n \in \mathbb{N} \quad g(n) = pq^n \\ \text{(resp. } \forall n \in \mathbb{N}^* \quad g(n) = pq^{n-1}), \end{array}$$

où on note classiquement $q = 1 - p$. Ces lois de probabilités sont respectivement notées $\mathcal{G}_{\mathbb{N}}(p)$ et $\mathcal{G}_{\mathbb{N}^*}(p)$.

12. Rappelons que $|A|$ désigne le cardinal (nombre d'éléments) de l'ensemble A, et que $\binom{n}{k}$, aussi noté C_n^k , désigne le nombre de combinaisons de n objets pris k à k — en langage moderne, le nombre de parties à k éléments d'un ensemble à n éléments.

13. Ce choix sera justifié d'une autre manière au chapitre 3.

Exemple 1.4. Modèle binomial. Ici, $\Omega = \{0, 1, \dots, n\}$. Un entier $n \in \mathbb{N}^*$ et un réel $p \in]0, 1[$ étant donnés, la **loi binomiale** (ou probabilité **binomiale**), de paramètres n et p est la loi de probabilité engendrée par le germe g défini par :

$$\forall k \in \{0, 1, \dots, n\} \quad g(k) = \binom{n}{k} p^k q^{n-k},$$

où on note classiquement $q = 1 - p$. Elle est notée $\mathcal{B}(n, p)$.

Remarque. Dans l'exemple de modélisation analysé plus haut (probabilité d'obtenir k fois le chiffre 6 sur n lancers d'un dé), on peut, si on ne s'intéresse qu'au résultat, et pas à la manière dont il est obtenu, adopter un autre modèle : l'ensemble des réalisations est $\Omega' = \{0, 1, \dots, n\}$, et la loi de probabilité P' sur Ω' est définie par la condition $P'(\{k\}) = P(A_k)$. On est alors en présence d'un modèle binomial : la loi P' est la loi binomiale $\mathcal{B}\left(n, \frac{1}{6}\right)$.

Exemple 1.5. Modèle de Poisson¹⁴. Ici, $\Omega = \mathbb{N}$. Un réel λ *strictement positif* étant donné, la **loi de Poisson** de paramètre λ est la loi de probabilité engendrée par le germe g défini par :

$$\forall n \in \mathbb{N} \quad g(n) = \exp(-\lambda) \frac{\lambda^n}{n!}.$$

Elle est notée $\mathcal{P}(\lambda)$.

Exemple 1.6. Modèle hypergéométrique. Soit $U = U_1 \uplus U_2$ un ensemble fini partitionné en deux sous ensembles U_1 et U_2 tels que : $|U| = r$ et $|U_1| = r_1$ (et donc $|U_2| = r - r_1$). Soit un entier n , $1 \leq n < r$. On extrait « au hasard »¹⁵ n éléments de U et l'on cherche la probabilité d'obtenir exactement k_1 éléments de U_1 (et donc $n - k_1$ éléments de U_2).

Une réalisation est une partie de U à n éléments : on prend donc $\Omega = \{A \in \mathcal{P}(U) \mid |A| = n\}$. L'événement étudié est alors représenté par le sous-ensemble $A_{k_1} = \{A \in \Omega \mid |A \cap U_1| = k_1\}$ de Ω . On munit l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$ de la probabilité uniforme P .

On cherche la probabilité $P(A_{k_1})$. L'ensemble A_{k_1} est vide si et seulement si $r_1 < k_1 < n$, ou si $0 < k_1 < n - (r - r_1)$. Sinon, c'est-à-dire si

$$\max(0, n - (r - r_1)) \leq k_1 \leq \min(n, r_1), \quad (1.6)$$

¹⁴ Denis Poisson (1781-1840), a été professeur de Mathématiques à l'École polytechnique (1806), puis à la faculté des sciences de Paris (1809). Son œuvre intéresse l'analyse mathématique, le calcul des probabilités, la mécanique céleste et la physique mathématique.

¹⁵ Rappelons que cette expression consacrée traduit une idée d'uniformité, i.e. de *probabilité uniforme*.

on a : $|A_{k_1}| = \binom{r_1}{k_1} \binom{r-r_1}{n-k_1}$. Puisque $|\Omega| = \binom{r}{n}$, il en résulte que :

$$\boxed{P(A_{k_1}) = \frac{\binom{r_1}{k_1} \binom{r-r_1}{n-k_1}}{\binom{r}{n}}.} \quad (1.7)$$

La **loi hypergéométrique** est la loi de probabilité associée au germe g défini sur l'ensemble des entiers k_1 vérifiant (1.6) par : $g(k_1) = P(A_{k_1})$.

Exemple de modélisation. Un lac contient r poissons, dont r_1 sont malades. On y effectue un prélèvement de n poissons. Quelle est la probabilité d'obtenir exactement k_1 poissons malades ? On adopte le modèle précédent, U représentant la population des poissons et U_1 (respectivement U_2) celle des poissons malades (respectivement sains). La probabilité cherchée est donnée par (1.7).

Remarque. Ce modèle est à la base de la théorie des sondages.

1.5. Variables aléatoires

Quand on étudie un phénomène, on est amené à étudier des grandeurs numériques (ou vectorielles) liées à celui-ci. En termes vagues, une variable aléatoire est un nombre (variable) lié à une expérience aléatoire, dont la valeur dépend exclusivement du résultat ω de cette expérience. Mathématiquement, il s'agit donc tout simplement d'une fonction définie sur l'ensemble Ω . Un exemple permettra de préciser les choses.

Exemple introductif. Un jeu consiste à lancer consécutivement n fois une pièce équilibrée. Le nombre d'obtentions de pile est « fonction » de la suite des lancers. On modélise ce jeu de la façon suivante : on choisit $\Omega = \{0, 1\}^n$, une réalisation étant une suite de 0 (pour face par exemple) et de 1 de longueur n . Si le dé est équilibré et si l'on ne triche pas, il est naturel de mettre la probabilité uniforme P sur l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$. Soit alors X l'application de Ω dans \mathbb{N} :

$$\omega \mapsto \sum_{i=1}^n \omega_i \quad \text{où } \omega = (\omega_1, \omega_2, \dots, \omega_n).$$

On dit que X est une **variable aléatoire** réelle ; elle représente ici le nombre *aléatoire* de « piles » obtenus.

Pour k entier, considérons l'ensemble

$$X^{-1}(\{k\}) = \{\omega \in \Omega \mid X(\omega) = k\},$$

qu'on appelle l'**image inverse** du singleton $\{k\}$. C'est l'événement « au cours des n lancers, on obtient k piles ».

La notation courante pour l'événement $X^{-1}(\{k\})$ est $\{X = k\}$ ou $(X = k)$. Au lieu de $P(\{X = k\})$ on écrit plus simplement $P(X = k)$. On peut noter que

$$P(X = k) = \frac{\binom{n}{k}}{2^n}$$

et que l'application $k \mapsto P(X = k)$ de $\{0, \dots, n\}$ dans $[0, 1]$ est donc le germe de la probabilité binomiale $\mathcal{B}\left(n, \frac{1}{2}\right)$.

Voici maintenant la définition générale d'une variable aléatoire.

Définition 1.10. Soient (Ω, \mathcal{A}) et (E, \mathcal{E}) deux espaces probabilisables. Une application de Ω dans E est dite **variable aléatoire** à valeurs dans E (en abrégé v.a.) si, pour tout $A \in \mathcal{E}$, l'ensemble $X^{-1}(A) \in \mathcal{A}$, où on note

$$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$$

l'image inverse de A par l'application X .

Remarque. Si on a choisi l'espace probabilisable (Ω, \mathcal{A}) pour modéliser une expérience aléatoire, l'image inverse de A par X est l'événement défini par l'ensemble des réalisations ω telles que la valeur $X(\omega)$ soit dans A . Elle est encore notée brièvement $(X \in A)$ et on parle de l'événement « X appartient à A ».

La proposition suivante est préliminaire à la définition des variables aléatoires discrètes (définition 1.12).

Proposition 1.11. Soit X une application de Ω dans E telle que $X(\Omega)$ soit dénombrable. Si la tribu \mathcal{E} sur E contient les points, c'est-à-dire si, pour tout $x \in E$, le singleton $\{x\}$ appartient à \mathcal{E} , pour que X soit une variable aléatoire, il faut et il suffit que, pour tout $x \in E$, on ait $X^{-1}(\{x\}) \in \mathcal{A}$.

Démonstration. La condition nécessaire est triviale. Supposons donc que, pour tout $x \in E$, on ait $X^{-1}(\{x\}) \in \mathcal{A}$. Soit $A \in \mathcal{E}$; on a :

$$X^{-1}(A) = X^{-1}[X(\Omega) \cap A] = \bigcup_{x \in X(\Omega) \cap A} X^{-1}(\{x\}).$$

L'ensemble $X(\Omega) \cap A$ étant dénombrable, il vient : $X^{-1}(A) \in \mathcal{A}$.

Définition 1.12. Soient (Ω, \mathcal{A}) un espace probabilisable, et soit E un ensemble. On dit qu'une application X de Ω dans E est une **variable**

aléatoire discrète (en abrégé v.a.d.) si les deux conditions suivantes sont satisfaites :

- (i) L'ensemble $X(\Omega)$ des valeurs prises par X est dénombrable.
- (ii) Pour tout $x \in E$, on a $X^{-1}(\{x\}) \in \mathcal{A}$.

Si on munit E d'une tribu \mathcal{E} , toute variable aléatoire discrète à valeurs dans E est aussi une variable aléatoire à valeurs dans (E, \mathcal{E}) , au sens de la définition 1.10, pourvu que la tribu \mathcal{E} contienne les points.

Avertissement. Dans un premier temps, nous n'étudierons que les variables aléatoires discrètes et en général on aura $E = \mathbb{N}, \mathbb{Z}$, ou \mathbb{N}^n , ou \mathbb{Z}^n , et $\mathcal{E} = \mathcal{P}(E)$.

Remarque. Si \mathcal{A} est la tribu des parties de Ω , ce qui sera souvent le cas lorsque nous étudierons des variables aléatoires discrètes, toute application de Ω dans un espace E muni de n'importe quelle tribu \mathcal{E} est une variable aléatoire de (Ω, \mathcal{A}) dans (E, \mathcal{E}) . Nous verrons plus tard que si Ω n'est pas dénombrable nous serons contraints de choisir pour \mathcal{A} une tribu plus petite, au sens de l'inclusion, que la tribu des parties de Ω .

Proposition 1.13. Soit X une variable aléatoire définie sur l'espace probabilisable (Ω, \mathcal{A}) , à valeurs dans l'espace probabilisable (E, \mathcal{E}) . L'application : $A \mapsto P[X^{-1}(A)]$ de \mathcal{E} dans $[0, 1]$ est une probabilité sur l'espace probabilisable (E, \mathcal{E}) . Elle est notée P_X . Cette probabilité est appelée **loi (de probabilité) de la variable aléatoire X** .

Démonstration. Il suffit de constater que si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments disjoints deux à deux de \mathcal{E} on a : $X^{-1}(\biguplus_{n \in \mathbb{N}} A_n) = \biguplus_{n \in \mathbb{N}} X^{-1}(A_n)$ et d'appliquer la σ -additivité de P . De plus, puisque $X^{-1}(E) = \Omega$, on a $P_X(E) = 1$.

Remarques. 1. La notion de loi d'une variable aléatoire est fondamentale; elle traduit le comportement probabiliste de la variable aléatoire, et c'est sur elle qu'en pratique le statisticien peut obtenir des renseignements probabilistes. L'espace probabilisé (Ω, \mathcal{A}, P) , bien que fondamental théoriquement pour le probabiliste, n'est en général pas celui que l'on rencontre dans la « nature », et il est souvent difficile à expliciter; l'assurance de son existence suffit en général. Il faut bien noter que la loi d'une variable aléatoire est une probabilité sur l'espace des valeurs prises par celle-ci.

2. Si X est une variable aléatoire discrète à valeurs dans (E, \mathcal{E}) , sa loi est déterminée par le germe de probabilité $x \mapsto P(X = x)$ sur (E, \mathcal{E}) .

En effet, pour tout $A \in \mathcal{E}$, on a :

$$\begin{aligned} P[X^{-1}(A)] &= P\left[\biguplus_{x \in X(\Omega) \cap A} X^{-1}(\{x\})\right] \\ &= \sum_{x \in X(\Omega) \cap A} P[X^{-1}(\{x\})] \\ &= \sum_{x \in A} P(X = x), \end{aligned}$$

la deuxième égalité étant vraie sans problème si $X(\Omega)$ est fini, et vraie avec une justification par les *familles sommables* sinon¹⁶ ; cette dernière notion va être étudiée systématiquement au prochain chapitre.

Exercices

On présente quelques *exercices corrigés* pour illustrer les notions introduites et donner des méthodes de travail.

Dans les premiers exercices, tous classiques, on illustre la construction de modèles simples.

Exercice 1.1. Quelle est la probabilité qu'en jetant six dés équilibrés et discernables (par la couleur par exemple), toutes les faces exhibent un chiffre différent ?

Solution. On prend pour ensemble des réalisations $\Omega = \{1, 2, \dots, 6\}^6$, pour tribu des événements $\mathcal{P}(\Omega)$ et pour probabilité P sur (Ω, \mathcal{A}) , la probabilité uniforme qui correspond au fait que les dés sont équilibrés. L'événement « toutes les faces exhibent un chiffre différent » est représenté par le sous-ensemble

$$A = \{(\sigma(1), \sigma(2), \dots, \sigma(6)) \mid \sigma \in S_6\},$$

où S_6 désigne l'ensemble des permutations de l'ensemble des entiers de 1 à 6.

Alors,

$$|\Omega| = 6^6 \quad \text{et} \quad |A| = 6!$$

ce qui donne :

$$P(A) = \frac{|A|}{|\Omega|} = \frac{6!}{6^6},$$

soit

$$P(A) = \frac{720}{46656} \simeq 0,015.$$

Exercice 1.2. Quelles sont les probabilités que, parmi les familles à n enfants, $n \geq 2$, une famille soit constituée d'enfants des deux sexes (événement

¹⁶. Noter que toutefois $X(\Omega) \cap A$ est dénombrable.

A), puis de garçons et d'au plus une fille (événement B)? Construire un modèle et calculer $P(A \cap B)$; comparer cette quantité au produit $P(A)P(B)$.

Solution. On prend pour ensemble des réalisations $\Omega = \{F, G\}^n$, pour tribu des événements $\mathcal{P}(\Omega)$ et pour probabilité P sur (Ω, \mathcal{A}) la probabilité uniforme. L'événement contraire de l'événement « la famille est constituée d'enfants des deux sexes » est l'événement « la famille n'est constituée que de garçons ou que de filles », représenté¹⁷ par le sous-ensemble

$$A^c = \{F\}^n \cup \{G\}^n.$$

On a alors :

$$|\Omega| = 2^n \quad \text{et} \quad |A^c| = 2,$$

d'où, puisque :

$$P(A) = 1 - P(A^c)$$

$$P(A) = 1 - \frac{1}{2^{n-1}}.$$

Soient B_0 et B_1 les événements « la famille n'a que des garçons » et « la famille a une fille exactement » :

$$B_0 = \{G\}^n, \quad B_1 = \{(G, G, \dots, G, F, G, \dots, G) \mid F \text{ au rang } j, 1 \leq j \leq n\}.$$

On a $B = B_0 \cup B_1$ et donc :

$$P(B) = P(B_0) + P(B_1).$$

Puisque

$$|B_0| = 1 \quad \text{et} \quad |B_1| = n,$$

on a :

$$P(B) = \frac{n+1}{2^n}.$$

Par ailleurs, puisque $A \cap B = B_1$, on a :

$$P(A \cap B) = \frac{n}{2^n}.$$

On constate que

$$P(A \cap B) \neq P(A)P(B),$$

sauf dans le cas $n = 3$.

Exercice 1.3. Problème des anniversaires et des coïncidences.

1. Une urne contient M jetons numérotés de 1 à M . On tire successivement n jetons en remettant chaque fois le jeton tiré et en brassant bien. On cherche la probabilité qu'aucun jeton ne soit tiré plus d'une fois.

2. Des étudiants au nombre de n sont réunis dans un amphithéâtre; quelle est la probabilité qu'au moins deux étudiants aient leur anniversaire le même jour (on suppose qu'aucun n'est né un 29 février et que $n \leq 365$) ?

¹⁷ L'événement A est en fait une réunion d'ensembles non disjoints; il est quelquefois plus facile, comme ici, de décrire le complémentaire de A .

Solution.

1. Une réalisation est un n -uplet d'entiers compris entre 1 et M ; soit :

$$\Omega = \{1, 2, \dots, M\}^n.$$

L'événement étudié s'écrit alors :

$$A = \{(\omega_1, \omega_2, \dots, \omega_n) \mid \omega_i \neq \omega_j \quad \forall (i, j) \text{ tel que } i \neq j\};$$

on a donc¹⁸ :

$$|\Omega| = M^n \quad |A| = A_M^n = M(M-1) \cdots (M-n+1).$$

Il vient alors :

$$P(A) = \left(1 - \frac{1}{M}\right) \left(1 - \frac{2}{M}\right) \cdots \left(1 - \frac{n-1}{M}\right).$$

2. Ici, les jetons sont remplacés par les étudiants et les numéros des jetons par les 365 jours de l'année. On étudie en fait l'événement contraire à l'événement « aucun des étudiants n'a son anniversaire le même jour qu'un autre » ; notons-le B . Il vient :

$$P_n(B) = 1 - \frac{A_{365}^n}{365^n}.$$

Nous donnons quelques résultats numériques suivant les valeurs de n :

n	4	16	22	23	40	64
$P_n(B)$	0,016	0,284	0,476	0,507	0,891	0,997

Exercice 1.4. Problème de la loterie. Des tickets au nombre de M sont édités et numérotés de 1 à M . Pour simplifier (cela ne change rien à la généralité) on suppose que les n ($2n \leq M$) premiers sont gagnants, ce que, bien sûr, les acheteurs ne savent pas ! Quelle est la probabilité qu'un acheteur de n billets achète au moins un billet gagnant ?

Solution. Une réalisation est un n -uplet d'entiers tous différents compris entre 1 et M . On prend donc comme ensemble des réalisations :

$$\Omega = \{A \in \mathcal{P}(\{1, 2, \dots, M\}) \mid |A| = n\}.$$

Soit G l'événement « l'acheteur a reçu au moins un billet gagnant ». On a :

$$G^c = \{A \in \mathcal{P}(\{n+1, \dots, M\}) \mid |A| = n\},$$

si bien que :

$$|\Omega| = \binom{M}{n} \quad \text{et} \quad |G^c| = \binom{M-n}{n}.$$

La probabilité cherchée est donc :

$$P(G) = 1 - P(G^c) = 1 - \frac{\binom{M-n}{n}}{\binom{M}{n}},$$

18. Rappelons que A_p^q désigne le nombre d'arrangements de p objets pris q à q — en langage moderne, le nombre de suites de q objets distincts pris parmi p .

soit :

$$P(G) = 1 - \left(1 - \frac{n}{M}\right) \left(1 - \frac{n}{M-1}\right) \cdots \left(1 - \frac{n}{M-n+1}\right).$$

Exercice 1.5. Une application de la formule de Poincaré. Soient (Ω, \mathcal{A}, P) un espace probabilisé et $\{A_i\}_{1 \leq i \leq n}$ une famille finie d'événements (éléments de \mathcal{A}). Pour toute partie J non vide de l'intervalle d'entiers $I = \{1, 2, \dots, n\}$, on note :

$$\widehat{A}_J = \bigcap_{j \in J} A_j \quad B^J = \widehat{A}_J \cap \left(\bigcap_{j \in I \setminus J} A_j^c \right),$$

puis, si $1 \leq m \leq n$,

$$B_m = \bigoplus_{\substack{J \in \mathcal{P}(I) \\ |J|=m}} B^J.$$

On note enfin

$$B_0 = \bigcap_{j=1}^n A_j^c.$$

On se persuadera que pour deux parties distinctes J et J' non vides de I , les ensembles B^J et $B^{J'}$ sont disjoints. Un point ω appartient à B_m ($1 \leq m \leq n$) si et seulement si il appartient à *exactement* m événements A_i . Un point ω appartient à B_0 si et seulement si il n'appartient à aucun des événements A_i . On note :

$$S_0 = 1 \quad \text{et, pour } 1 \leq r \leq n, \quad S_r = \sum_{\substack{J \in \mathcal{P}(I) \\ |J|=r}} P(\widehat{A}_J).$$

1. Démontrer que :

(a) Pour tout $A \in \mathcal{A}$,

$$\begin{aligned} P(B_0 \cap A) &= P(A) - \sum_{j=1}^n P(A_j \cap A) \\ &+ \sum_{1 \leq j_1 < j_2 \leq n} P(A_{j_1} \cap A_{j_2} \cap A) - \sum_{1 \leq j_1 < j_2 < j_3 \leq n} P(A_{j_1} \cap A_{j_2} \cap A_{j_3} \cap A) \\ &+ \cdots + (-1)^n P(A_1 \cap A_2 \cap \cdots \cap A_n \cap A); \quad (1.8) \end{aligned}$$

(b) si $1 \leq m \leq n$,

$$P(B_m) = S_m - \binom{m+1}{m} S_{m+1} + \cdots + (-1)^{n-m} \binom{n}{m} S_n. \quad (1.9)$$

2. Exemple. Au cours d'une soirée, n personnes déposent dans un chapeau une carte portant leur nom (aucune personne ne portant le même nom qu'une

autre). A la fin de la soirée, on brasse bien les cartes puis chacun tire une carte. Quelles sont les probabilités qu'aucune des personnes ne tire une carte portant son nom, puis qu'exactly m ($1 \leq m \leq n$) personnes tirent une carte portant leur nom ?

Solution.

1. On a successivement,

(a) pour tout $A \in \mathcal{A}$,

$$P(B_0 \cap A) = P(A) - P(B_0^c \cap A) \quad \text{et} \quad P(B_0^c \cap A) = P\left(\bigcup_{i=1}^n (A_i \cap A)\right).$$

Il reste à appliquer la formule de Poincaré.

(b) Soit J une partie non vide de cardinal m de I et i_1, i_2, \dots, i_{n-m} une énumération croissante de $I \setminus J$. On applique alors la formule précédente avec :

$$A = \widehat{A}_J \quad \text{et} \quad B_0 = \bigcap_{h=1}^{n-m} A_{i_h}^c.$$

Il vient :

$$\begin{aligned} P(B^J) &= P(\widehat{A}_J) - \sum_{h=1}^{n-m} P(A_{i_h} \cap \widehat{A}_J) + \sum_{1 \leq j_1 < j_2 \leq n-m} P(A_{i_{j_1}} \cap A_{i_{j_2}} \cap \widehat{A}_J) \\ &\quad + \dots + (-1)^{n-m} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_{n-m}} \cap \widehat{A}_J), \end{aligned}$$

ce qui peut s'écrire :

$$\begin{aligned} P(B^J) &= P(\widehat{A}_J) - \sum_{\substack{J' \supset J \\ |J'|=m+1}} P(\widehat{A}_{J'}) + \sum_{\substack{J' \supset J \\ |J'|=m+2}} P(\widehat{A}_{J'}) \\ &\quad + \dots + (-1)^{n-m} P(\widehat{A}_I). \quad (1.10) \end{aligned}$$

Mais, puisque

$$P(B_m) = \sum_{|J|=m} P(B^J),$$

on a, en utilisant l'égalité (1.10) :

$$\begin{aligned} P(B_m) &= \sum_{|J|=m} P(\widehat{A}_J) - \sum_{|J|=m} \sum_{\substack{J' \supset J \\ |J'|=m+1}} P(\widehat{A}_{J'}) \\ &\quad + \sum_{|J|=m} \sum_{\substack{J' \supset J \\ |J'|=m+2}} P(\widehat{A}_{J'}) + \dots + (-1)^{n-m} \sum_{|J|=m} P(\widehat{A}_I). \end{aligned}$$

Or, par associativité de la somme :

$$\begin{aligned}
 \sum_{|J|=m} \sum_{\substack{J' \supset J \\ |J'|=m+1}} P(\widehat{A}_{J'}) &= \sum_{|J'|=m+1} \sum_{\substack{J \subset J' \\ |J|=m}} P(\widehat{A}_{J'}) \\
 &= \sum_{|J'|=m+1} \binom{m+1}{m} P(\widehat{A}_{J'}) \\
 &= \binom{m+1}{m} S_{m+1}
 \end{aligned}$$

etc., ce qui donne la formule annoncée.

2. Numérotions de 1 à n les noms des personnes. Notons toujours $I = \{1, 2, \dots, n\}$. Nous prendrons pour ensemble des réalisations l'ensemble Ω des bijections de I sur lui-même et nous le munirons de la tribu de ses parties. L'espace probabilisé choisi est alors (Ω, \mathcal{A}, P) où P est la probabilité uniforme. Si $A_i = \{f \in \Omega \mid f(i) = i\}$, ensemble des bijections ayant i pour point fixe, le sous-ensemble

$$B_0 = \bigcap_{i=1}^n A_i^c$$

correspond à l'événement « aucune des personnes ne tire son nom » ; on cherche $P(B_0)$. On utilise l'égalité (1.8) avec $A = \Omega$. Or

$$P(A_i) = \frac{|A_i|}{|\Omega|} = \frac{(n-1)!}{n!},$$

$$\text{et si } i \neq j, \quad P(A_i \cap A_j) = \frac{|A_i \cap A_j|}{|\Omega|} = \frac{(n-2)!}{n!}$$

etc.

En reportant dans l'égalité (1.8) et après simplifications il vient :

$$P(B_0) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots + (-1)^n \frac{1}{n!}.$$

On remarque que lorsque n tend vers l'infini, $P(B_0)$ tend vers $e^{-1} \simeq 0,368$.

L'événement B_m défini dans l'énoncé est l'événement « exactement m personnes tirent une carte portant leur nom ». On cherche $P(B_m)$. On utilise l'égalité (1.9).

Si J est une partie de I de cardinal r , on a :

$$P(\widehat{A}_J) = \frac{(n-r)!}{n!}$$

et donc :

$$S_r = \binom{n}{r} \frac{(n-r)!}{n!} = \frac{1}{r!},$$

d'où, en reportant dans l'égalité (1.9) et après simplifications :

$$P(B_m) = \frac{1}{m!} \left(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots + (-1)^{n-m} \frac{1}{(n-m)!} \right).$$

On remarque que lorsque n tend vers l'infini $P(B_m)$ tend vers $\frac{e^{-1}}{m!}$. Par ailleurs, pour n fixé, le nombre de *bijections* de I sur lui-même ayant exactement m points fixes est $n!P(B_m)$.

Exercice 1.6. Exemple de variable aléatoire : jeu de pile ou face. Un joueur jette N fois une pièce équilibrée ($N \geq 2$) ; on s'intéresse au numéro du jet pour lequel il obtient « pile » pour la première fois. Construire un modèle et étudier la loi de la variable aléatoire correspondante.

Solution. On code « pile » par 1 et « face » par 0. Une réalisation de ce jeu est alors une suite de longueur N à valeurs 0 ou 1. La pièce étant équilibrée, on prend pour espace probabilisé (Ω, \mathcal{A}, P) où $\Omega = \{0, 1\}^N$, \mathcal{A} est la tribu des parties de Ω et P est la probabilité uniforme sur l'espace probabilisable (Ω, \mathcal{A}) . Soit, pour $1 \leq j \leq N$, X_j la j -ième projection sur Ω définie par :

$$\forall \omega = (\omega_1, \omega_2, \dots, \omega_N) \in \Omega \quad X_j(\omega) = \omega_j.$$

On définit alors la variable aléatoire T par :

$$\forall \omega \in \Omega \quad T(\omega) = \inf \{j \geq 1 \mid X_j(\omega) = 1\},$$

où l'on convient *ici* que $\inf \emptyset = N + 1$.

La variable aléatoire T est à valeurs dans l'ensemble des entiers compris entre 1 et $N + 1$; l'entier $T(\omega)$ représente le numéro du jet pour lequel le joueur obtient « pile » pour la première fois lors de la série de jets ω . La loi de T est entièrement déterminée par les probabilités $P(T = j)$ pour $1 \leq j \leq N + 1$.

On a $(T = 1) = (X_1 = 1)$ et :

$$(X_1 = 1) = \{(1, \omega_2, \dots, \omega_N) \mid (\omega_2, \dots, \omega_N) \in \{0, 1\}^{N-1}\}.$$

On a donc : $|(X_1 = 1)| = 2^{N-1}$. Comme $|\Omega| = 2^N$, et que P est la probabilité uniforme,

$$P(T = 1) = \frac{|(X_1 = 1)|}{|\Omega|} = \frac{1}{2}.$$

Si $2 \leq j \leq N$, on a :

$$(T = j) = \left[\bigcap_{i=1}^{j-1} (X_i = 0) \right] \cap (X_j = 1),$$

ensemble qui peut encore se décrire comme

$$(T = j) = \{(0, 0, \dots, 0, 1, \omega_{j+1}, \dots, \omega_N) \mid (\omega_{j+1}, \dots, \omega_N) \in \{0, 1\}^{N-j}\}.$$

On a donc :

$$|(T = j)| = 2^{N-j}$$

et par conséquent

$$\boxed{P(T = j) = \frac{1}{2^j}.$$

Enfin, on a $(T = N + 1) = \{(0, 0, \dots, 0)\}$, et donc $|(T = N + 1)| = 1$, ce qui donne :

$$\boxed{P(T = N + 1) = \frac{1}{2^N} .}$$

Remarque. On vérifie bien que

$$\sum_{j=1}^{N+1} P(T = j) = 1 .$$

Chapitre 2

Familles sommables de nombres réels

Les sommes infinies (sommées d'un « nombre » infini de nombres réels) sont d'un usage constant en Calcul des Probabilités.

La forme la plus maniable de cette notion, comme en témoigne déjà la notation $\sum_{i \in I} x_i$, est celle de *somme d'une famille*, qui fait l'objet de ce chapitre. La notion de somme d'une famille de nombres réels diffère dans sa définition de celle de série convergente, que le lecteur connaît déjà : aucune condition, notamment de dénombrabilité, n'est imposée à l'ensemble des indices I , et la définition de $\sum_{i \in I} x_i$ ne fait pas intervenir d'ordre sur l'ensemble I . Elle s'en distingue aussi dans son emploi : moyennant certaines vérifications, on manie à peu près $\sum_{i \in I} x_i$ comme une somme finie (sommation par paquets, théorème 2.16).

Il ne faut pas croire cependant que le remplacement des séries par les familles sommables constitue une avancée spectaculaire : en pratique, on ne sort pas du dénombrable, et la notion de somme d'une famille est en fait équivalente à celle de somme d'une série absolument convergente.

2.1. Somme d'une famille de réels positifs

2.1.1. La droite achevée $\overline{\mathbb{R}}$

Rappelons que tout sous-ensemble A non vide et majoré (c'est-à-dire possédant un majorant) de \mathbb{R} possède une borne supérieure, notée $\sup A$. Par définition, le nombre réel $\sup A$ est le plus petit des majorants de A . L'existence de $\sup A$ est un théorème. Si le sous-ensemble A n'est pas majoré, on pose $\sup A = +\infty$.

Il est utile dans ce contexte de compléter la droite numérique \mathbb{R} en lui adjoignant les deux éléments $+\infty$ et $-\infty$. L'ensemble

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$$

porte le nom de droite achevée. On prolonge la relation d'ordre usuelle de \mathbb{R} à $\overline{\mathbb{R}}$ en posant, pour tout élément x de $\overline{\mathbb{R}}$,

$$-\infty \leq x \quad \text{et} \quad x \leq +\infty.$$

Certaines choses sont plus simples dans $\overline{\mathbb{R}}$ que dans \mathbb{R} : dans $\overline{\mathbb{R}}$, tout sous-ensemble (même \emptyset) possède une borne supérieure et une borne inférieure. On a par exemple $\sup A = +\infty$ si A contient $+\infty$, ou si A est un sous-ensemble non majoré de \mathbb{R} . En effet, dans ce dernier cas, $+\infty$ est un majorant de A , et aucun élément de $\overline{\mathbb{R}}$ inférieur à $+\infty$ n'est un majorant de A . En introduisant $\overline{\mathbb{R}}$, on a donné une signification précise à une relation comme $\sup A = +\infty$, qui n'était auparavant qu'une convention d'écriture.

On convient parfois, par abus de langage, d'appeler *nombres réels* les éléments de $\overline{\mathbb{R}}$. On appelle alors *nombres réels finis* les éléments de \mathbb{R} . Quand on fait cet abus de langage, on doit naturellement appeler *variables aléatoires réelles* les variables aléatoires à valeurs dans $\overline{\mathbb{R}}$, *fonctions réelles* ou *fonctions numériques* les fonctions à valeurs dans $\overline{\mathbb{R}}$, et on appelle *variables aléatoires réelles finies*, *fonctions numériques finies*, les variables aléatoires et les fonctions à valeurs dans \mathbb{R} . Cet abus de langage sera toujours signalé.

2.1.2. Familles sommables de réels positifs

Nous avons déjà rencontré la notion de famille d'éléments de E , indexée par un ensemble I quelconque. En principe, elle ne se distingue pas de celle d'application de I dans E . La différence réside dans la notation et correspond à une différence de point de vue : en parlant d'une famille $(x_i)_{i \in I}$, on s'intéresse aux éléments x_i qu'on imagine repérés (comme énumérés) par les éléments de I . Une *suite*, par exemple, n'est rien d'autre qu'une famille indexée par \mathbb{N} , ou par un intervalle d'entiers.

La somme d'une famille quelconque de réels positifs (positif signifie ≥ 0) se définit de façon très naturelle à partir de la notion de somme d'une famille finie.

Définition 2.1. Soit $(x_i)_{i \in I}$ une famille de nombres réels positifs. On appelle *somme de la famille* $(x_i)_{i \in I}$ l'élément de $[0, +\infty]^1$ noté

$$\sum_{i \in I} x_i$$

1. On note indifféremment $[0, +\infty]$ ou $\overline{\mathbb{R}}^+$ l'ensemble des éléments positifs de $\overline{\mathbb{R}}$, i.e. l'ensemble $\mathbb{R}^+ \cup \{+\infty\}$.

et défini par la relation

$$\sum_{i \in I} x_i = \sup \left\{ \sum_{i \in F} x_i \mid F \text{ partie finie de } I \right\}.$$

La somme d'une famille de réels positifs est donc toujours définie. C'est un nombre réel fini ou $+\infty$.

Définition 2.2. Si la somme $\sum_{i \in I} x_i$ est finie, on dit que la famille $(x_i)_{i \in I}$ est sommable.

Si $\sum_{i \in I} x_i = +\infty$, la famille $(x_i)_{i \in I}$ n'est pas sommable, bien qu'on ait attribué une valeur à sa somme.

2.1.3. Comparaison avec les séries à termes positifs

Proposition 2.3. Soit $(x_i)_{i \in I}$ une famille sommable de réels positifs. L'ensemble des indices i tels que $x_i > 0$ est dénombrable.

En d'autres termes on a $x_i = 0$ sauf pour un ensemble dénombrable d'indices.

Démonstration. Soit J l'ensemble des $i \in I$ tels que $x_i > 0$. L'ensemble J est réunion des ensembles J_n , $n \geq 1$, où J_n est l'ensemble des $i \in I$ tels que $x_i \geq 1/n$. Montrons que chaque J_n est fini. Si l'ensemble J_n était infini, il posséderait pour tout entier k un sous-ensemble K de cardinal k . On aurait donc

$$\sum_{i \in I} x_i \geq \sum_{i \in K} x_i \geq \frac{k}{n}$$

quel que soit $k \geq 1$, d'où $\sum_{i \in I} x_i = +\infty$. La famille (x_i) ne serait donc pas sommable. L'ensemble J , réunion dénombrable d'ensembles finis, est donc dénombrable.

Proposition 2.4. Soit J un ensemble infini dénombrable et soit $(j_k)_{k \in \mathbb{N}}$ une énumération² de J . Alors on a

$$\sum_{j \in J} x_j = \sum_{k=0}^{\infty} x_{j_k}.$$

Selon l'habitude, la notation $\sum_{k=0}^{\infty} x_{j_k}$ dans le second membre désigne la limite des sommes partielles $\sum_{k=0}^N x_{j_k}$ pour N tendant vers l'infini. Puisque les termes x_{j_k} sont positifs, la suite des sommes partielles est

2. C'est la donnée d'une bijection de \mathbb{N} sur J .

croissante, et cette limite existe toujours. Elle peut être finie (on dit dans ce cas que la série de terme général x_{j_k} converge, ce que l'on écrit encore « la série $\sum x_{j_k}$ est convergente ») ou infinie (on dit alors que la série diverge). La proposition 2.4 a donc pour corollaire l'énoncé suivant : pour que la famille à termes positifs $(x_j)_{j \in J}$ soit sommable, il est nécessaire et suffisant que la série $\sum x_{j_k}$ soit convergente.

Démonstration. Pour tout entier N l'ensemble $J_N = \{j_0, j_1, \dots, j_N\}$ est une partie finie de J , et on a

$$\sum_{k=0}^N x_{j_k} = \sum_{j \in J_N} x_j \leq \sum_{j \in J} x_j.$$

On a donc

$$\sum_{k=0}^{\infty} x_{j_k} \leq \sum_{j \in J} x_j. \quad (2.1)$$

En effet, si $\sum_{j \in J} x_j < +\infty$, alors on a (2.1) par passage à la limite dans \mathbb{R} , et si $\sum_{j \in J} x_j = +\infty$, (2.1) est trivialement vraie.

Inversement, soit F une partie finie de J . Il existe un entier N tel que $F \subset J_N$. On a alors

$$\sum_{j \in F} x_j \leq \sum_{j \in J_N} x_j = \sum_{k=0}^N x_{j_k}$$

d'où, a fortiori,

$$\sum_{j \in F} x_j \leq \sum_{k=0}^{\infty} x_{j_k}. \quad (2.2)$$

Associée à (2.1), cette inégalité fournit le résultat annoncé.

2.2. Arithmétique dans $\overline{\mathbb{R}}$. Somme d'une famille d'éléments de $\overline{\mathbb{R}}$

2.2.1. Arithmétique dans $\overline{\mathbb{R}}$

Pour énoncer commodément les principales propriétés de la notion de somme d'une famille de réels positifs – et notamment la propriété essentielle de sommation par paquets (proposition 2.9 ci-dessous) – il faut admettre la possibilité que les éléments de la famille puissent prendre la valeur $+\infty$, ce qui oblige à définir la somme de $+\infty$ et d'un réel positif.

On peut en fait étendre les opérations de l'arithmétique élémentaire (addition, soustraction, multiplication) au cas où l'un des deux termes,

ou facteurs, est infini, **sauf dans un cas** : la valeur de l'expression $(+\infty) + (-\infty)$ n'est pas définie, de même que celle des expressions qui se ramènent à celle-ci : $(+\infty) - (+\infty)$, $(-\infty) - (-\infty)$.

Dans les autres cas, l'extension des opérations se fait de façon naturelle. On pose

$$x + (+\infty) = +\infty$$

$$x + (-\infty) = -\infty$$

pour x réel fini, et

$$+\infty + (+\infty) = +\infty$$

$$-\infty + (-\infty) = -\infty.$$

La soustraction se définit en convenant que $x - y = x + (-y)$ (et que $-(+\infty) = -\infty$). La multiplication est définie dans tous les cas :

$$x \times (+\infty) = (\text{sign } x) \infty$$

$$x \times (-\infty) = (-\text{sign } x) \infty$$

si $x \neq 0$, et

$$0 \times (\pm\infty) = 0.$$

En notant, comme on le fait souvent, ∞ au lieu de $+\infty$, on retiendra que les seules opérations non définies sont celles qui se ramènent à $\infty - \infty$.

Enfin, on pose, comme il est naturel,

$$|+\infty| = |-\infty| = +\infty.$$

2.2.2. Familles sommables d'éléments positifs de $\overline{\mathbb{R}}$

Définition 2.5. La notion de somme d'une famille de réels positifs éventuellement infinis se définit comme dans le cas des réels positifs finis :

$$\sum_{i \in I} x_i = \sup \left\{ \sum_{i \in F} x_i \mid F \text{ partie finie de } I \right\}.$$

On dit que la famille $(x_i)_{i \in I}$ est sommable si $\sum_{i \in I} x_i < +\infty$.

Le résultat suivant est naturel : si la famille $(x_i)_{i \in I}$ est sommable, alors on a $x_i < +\infty$ pour tout $i \in I$. En effet, supposons que pour un indice $i_0 \in I$ on ait $x_{i_0} = +\infty$ et montrons que $\sum_{i \in I} x_i = +\infty$. On a en effet dans ce cas $\sum_{i \in I} x_i \geq \sum_{i \in \{i_0\}} x_i = x_{i_0} = +\infty$.

2.2.3. Propriétés des familles sommables de réels positifs

Proposition 2.6. Soient $(x_i)_{i \in I}$ et $(y_i)_{i \in I}$ deux familles de nombres réels positifs (éventuellement infinis).

(i) Si $x_i \leq y_i$ pour tout $i \in I$, on a

$$\sum_{i \in I} x_i \leq \sum_{i \in I} y_i ;$$

(ii) si a est un élément positif de $\overline{\mathbb{R}}$, on a

$$a \sum_{i \in I} x_i = \sum_{i \in I} ax_i ;$$

(iii) on a

$$\sum_{i \in I} (x_i + y_i) = \sum_{i \in I} x_i + \sum_{i \in I} y_i.$$

Démonstration. Principe général : pour montrer une inégalité du type $\sup A \leq \beta$ il suffit, puisque $\sup A$ est le plus petit des majorants de A , de montrer que β est un majorant de A , autrement dit qu'on a $\alpha \leq \beta$ pour tout $\alpha \in A$.

En particulier, pour montrer que $\sum_{i \in I} x_i \leq \beta$ (les x_i étant positifs ou égaux à $+\infty$), il suffit de montrer qu'on a $\sum_{i \in F} x_i \leq \beta$ pour toute partie finie F de I .

(i) Pour toute partie finie F de I , on a

$$\sum_{i \in F} x_i \leq \sum_{i \in F} y_i \leq \sum_{i \in I} y_i$$

On conclut à l'aide du principe évoqué plus haut.

(ii) Par définition du produit dans $\overline{\mathbb{R}}^+$, pour toute partie finie F de A , on a

$$\sum_{i \in F} ax_i = a \sum_{i \in F} x_i,$$

d'où

$$\sum_{i \in F} ax_i \leq a \sum_{i \in I} x_i,$$

par définition de la somme d'une famille, et

$$\sum_{i \in I} ax_i \leq a \sum_{i \in I} x_i,$$

en invoquant le principe général. On démontre de la même façon l'inégalité inverse.

(iii) On sait bien que « le sup d'une somme n'est pas toujours la somme des sup ». Nous obtiendrons cette égalité en établissant comme ci-dessus une double inégalité.

Pour toute partie finie F de I, on a

$$\begin{aligned} \sum_{i \in F} (x_i + y_i) &= \sum_{i \in F} x_i + \sum_{i \in F} y_i \\ &\leq \sum_{i \in I} x_i + \sum_{i \in I} y_i \end{aligned}$$

d'où

$$\sum_{i \in I} (x_i + y_i) \leq \sum_{i \in I} x_i + \sum_{i \in I} y_i, \quad (2.3)$$

toujours en application du principe rappelé au début.

En sens inverse, il faut procéder en deux étapes. Si F et G sont deux parties finies de I, on a

$$\sum_{i \in F} x_i + \sum_{i \in G} y_i \leq \sum_{i \in F \cup G} (x_i + y_i) \leq \sum_{i \in I} (x_i + y_i)$$

d'où

$$\sum_{i \in F} x_i + \sum_{i \in G} y_i \leq \sum_{i \in I} (x_i + y_i).$$

En prenant la borne supérieure du membre de gauche de cette inégalité quand G décrit l'ensemble des parties finies de I, on obtient

$$\sum_{i \in F} x_i + \sum_{i \in I} y_i \leq \sum_{i \in I} (x_i + y_i).$$

En procédant de la même façon, on obtient

$$\sum_{i \in I} x_i + \sum_{i \in I} y_i \leq \sum_{i \in I} (x_i + y_i). \quad (2.4)$$

Les inégalités (2.3) et (2.4) fournissent le résultat annoncé.

L'utilisation de la fonction indicatrice $\mathbf{1}_A$ permet de transformer une somme relative à une partie A de I en une somme sur I, ce qui est commode pour certaines démonstrations.

Proposition 2.7. Soit $(x_i)_{i \in I}$ une famille de réels positifs (éventuellement infinis) et soit A une partie quelconque de I. On a

$$\sum_{i \in A} x_i = \sum_{i \in I} \mathbf{1}_A(i) x_i.$$

Démonstration. Soit F une partie finie de A . Puisque $\mathbf{1}_A(i) = 1$ pour tout $i \in F$, on a

$$\sum_{i \in F} x_i = \sum_{i \in F} \mathbf{1}_A(i) x_i$$

et donc, F étant aussi une partie finie de I ,

$$\sum_{i \in F} x_i \leq \sum_{i \in I} \mathbf{1}_A(i) x_i.$$

Il en résulte que

$$\sum_{i \in I} x_i \leq \sum_{i \in I} \mathbf{1}_A(i) x_i.$$

Inversement, soit G une partie finie de I . On a

$$\sum_{i \in G} \mathbf{1}_A(i) x_i = \sum_{i \in G \cap A} x_i$$

et aussi, puisque $G \cap A$ est une partie finie de A ,

$$\sum_{i \in G} \mathbf{1}_A(i) x_i \leq \sum_{i \in A} x_i.$$

L'inégalité précédente étant vérifiée pour toute partie finie G de I , on a

$$\sum_{i \in I} \mathbf{1}_A(i) x_i \leq \sum_{i \in A} x_i,$$

ce qui démontre l'inégalité attendue.

Proposition 2.8. Soit $(x_i)_{i \in I}$ une famille de réels positifs (éventuellement infinis). Soient A et B deux parties de I . Alors

(i) si $A \subset B$, on a

$$\sum_{i \in A} x_i \leq \sum_{i \in B} x_i.$$

(ii) Si A et B sont disjoints, on a

$$\sum_{i \in A \uplus B} x_i = \sum_{i \in A} x_i + \sum_{i \in B} x_i.$$

On énonce la propriété (ii) en disant que l'application $A \mapsto \sum_{i \in A} x_i$ est une « fonction additive d'ensembles ». Évidemment, cette propriété s'étend au cas d'un nombre fini de parties deux à deux disjointes de I .

Démonstration. (i) En observant que toute partie finie F de A est aussi une partie finie de B , on peut écrire

$$\sum_{i \in F} x_i \leq \sum_{i \in B} x_i,$$

d'où le résultat.

(ii) Les parties A et B étant disjointes, on a $\mathbf{1}_{A \uplus B} = \mathbf{1}_A + \mathbf{1}_B$. En utilisant ce qui précède, on obtient alors

$$\sum_{i \in A \uplus B} x_i = \sum_{i \in I} \mathbf{1}_{A \uplus B}(i) x_i = \sum_{i \in I} \mathbf{1}_A(i) x_i + \sum_{i \in I} \mathbf{1}_B(i) x_i = \sum_{i \in A} x_i + \sum_{i \in B} x_i.$$

La proposition suivante est évidente dans le cas où les ensembles considérés sont finis : on peut dans ce cas la considérer comme une généralisation de la propriété d'associativité de la somme. Avec la proposition correspondante dans le cas des réels de signe quelconque, elle joue un rôle essentiel dans la suite de cet ouvrage.

Notons qu'il n'y a pas d'énoncé analogue à la proposition 2.9 comode à formuler pour le cas des séries. C'est là la justification principale de l'emploi de la notion de famille sommable.

Proposition 2.9. (Somme par paquets.) Soit $(x_i)_{i \in I}$ une famille de réels positifs (éventuellement infinis) et soit $(A_j)_{j \in J}$ une partition quelconque de I. Alors on a

$$\boxed{\sum_{i \in I} x_i = \sum_{j \in J} \left(\sum_{i \in A_j} x_i \right).}$$

Démonstration. Pour toute partie finie F de I, on a

$$F = \bigsqcup_{j \in J} (A_j \cap F).$$

L'ensemble F, étant fini, ne rencontre qu'un nombre fini des A_j ; autrement dit, il existe une partie finie J_F de J telle que

$$F = \bigsqcup_{j \in J_F} (A_j \cap F).$$

On a

$$\sum_{i \in F} x_i = \sum_{j \in J_F} \left(\sum_{i \in A_j \cap F} x_i \right)$$

(associativité des sommes finies), d'où

$$\sum_{i \in F} x_i \leq \sum_{j \in J_F} \left(\sum_{i \in A_j} x_i \right)$$

en observant que $(A_j \cap F) \subset A_j$. On peut dans l'inégalité qui précède remplacer d'abord J_F par J, puis F par I (ou, si on préfère, renvoyer à

des propriétés bien connues de la borne supérieure, ou à sa définition). On obtient

$$\sum_{i \in I} x_i \leq \sum_{j \in J} \left(\sum_{i \in A_j} x_i \right).$$

En sens inverse, pour toute partie finie J_0 de J , l'additivité simple (proposition 2.8, (ii)) entraîne

$$\sum_{j \in J_0} \left(\sum_{i \in A_j} x_i \right) = \sum_{i \in \bigcup_{j \in J_0} A_j} x_i \leq \sum_{i \in I} x_i$$

d'où

$$\sum_{j \in J} \left(\sum_{i \in A_j} x_i \right) \leq \sum_{i \in I} x_i.$$

2.3. Somme d'une famille de réels de signe quelconque

On considère maintenant des familles de nombres réels de signe quelconque. On définit très simplement la notion de famille sommable et celle de somme en séparant partie positive et partie négative. Mais il n'est plus possible d'admettre des valeurs infinies, car on risque de rencontrer l'opération interdite $\infty - \infty$.

2.3.1. Familles sommables de réels de signe quelconque

Rappelons pour commencer que, étant donné un réel x , on pose

$$x^+ = \begin{cases} x & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}, \quad x^- = \begin{cases} 0 & \text{si } x \geq 0 \\ |x| & \text{si } x < 0. \end{cases}$$

On a

$$\boxed{x = x^+ - x^-, \quad |x| = x^+ + x^-}.$$

On dit que x^+ est la partie positive de x , et x^- sa partie négative. Ces notions sont souvent utilisées pour écrire une fonction quelconque f comme différence de deux fonctions positives f^+ et f^- : on pose $f^+(x) = (f(x))^+$ et $f^-(x) = (f(x))^-$. Notons que $f^+ = \sup(f, 0)$ et $f^- = \sup(0, -f)$.

Nous allons ici nous servir de ces notions pour écrire une famille quelconque $(x_i)_{i \in I}$ de nombres réels comme différence de deux familles de réels positifs, et étendre ainsi aux familles de réels de signe quelconque la notion de famille sommable.

Définition 2.10. La famille de nombres réels $(x_i)_{i \in I}$ est dite sommable si on a $\sum_{i \in I} x_i^+ < +\infty$ et $\sum_{i \in I} x_i^- < +\infty$. Si $(x_i)_{i \in I}$ est sommable, on définit sa somme comme le nombre réel, noté $\sum_{i \in I} x_i$,

$$\sum_{i \in I} x_i = \sum_{i \in I} x_i^+ - \sum_{i \in I} x_i^-.$$

La somme de la famille $(x_i)_{i \in I}$ n'est donc pas toujours définie, contrairement au cas des familles de réels positifs. Mais si $(x_i)_{i \in I}$ est une famille de réels positifs de somme finie (définition 2.1), alors les deux notions de somme introduites coïncident.

Lemme 2.11. Pour que la famille $(x_i)_{i \in I}$ soit sommable, il faut et il suffit que la famille $(|x_i|)_{i \in I}$ soit sommable, ou, ce qui revient au même, qu'on ait $\sum_{i \in I} |x_i| < +\infty$.

Démonstration. Puisque $|x_i| = x_i^+ + x_i^-$, on a

$$\sum_{i \in I} |x_i| = \sum_{i \in I} x_i^+ + \sum_{i \in I} x_i^-.$$

Si $(x_i)_{i \in I}$ est sommable, on a $\sum_{i \in I} x_i^+ < +\infty$, $\sum_{i \in I} x_i^- < +\infty$, d'où $\sum_{i \in I} |x_i| < +\infty$.

Inversement, si $\sum_{i \in I} |x_i| < +\infty$, on a $\sum_{i \in I} x_i^+ < +\infty$, $\sum_{i \in I} x_i^- < +\infty$ et la famille $(x_i)_{i \in I}$ est sommable.

Il n'y a donc pas de distinction à faire entre « sommable » et « absolument sommable ».

2.3.2. Familles sommables et séries

L'énoncé suivant fait le lien entre la notion de famille sommable et celle de série absolument convergente.

Théorème 2.12. Pour que la famille de réels $(x_i)_{i \in I}$ soit sommable, il est nécessaire que l'ensemble J des indices $i \in I$ tels que $x_i \neq 0$ soit dénombrable.

Supposons cette condition réalisée, et supposons que J soit infini. Soit $(j_n)_{n \in \mathbb{N}}$ une énumération quelconque de J . Alors, pour que la famille $(x_i)_{i \in I}$ soit sommable, il faut et il suffit que la série $\sum_{n=0}^{\infty} x_{j_n}$ soit absolument convergente. Si cette dernière condition est réalisée, on a

$$\sum_{i \in I} x_i = \sum_{n=0}^{\infty} x_{j_n}. \quad (2.5)$$

Démonstration. À l'exception de l'égalité (2.5), les assertions précédentes résultent immédiatement du lemme 2.11 et des propositions 2.3 et 2.4.

Pour démontrer (2.5), observons qu'en vertu de la proposition 2.3 on a

$$\sum_{i \in I} x_i^+ = \sum_{n=0}^{\infty} x_{j_n}^+ \quad \text{et} \quad \sum_{i \in I} x_i^- = \sum_{n=0}^{\infty} x_{j_n}^-.$$

D'autre part, on a par définition

$$\sum_{i \in I} x_i = \sum_{i \in I} x_i^+ - \sum_{i \in I} x_i^-.$$

Il suffit donc de montrer que $\sum_{n=0}^{\infty} x_{j_n} = \sum_{n=0}^{\infty} x_{j_n}^+ - \sum_{n=0}^{\infty} x_{j_n}^-$, ce qui résulte immédiatement de la convergence (déjà démontrée) des séries du second membre et du fait que $x_{j_n} = x_{j_n}^+ - x_{j_n}^-$.

Remarques. On retrouve, comme corollaire immédiat de cet énoncé le fait bien connu que la convergence absolue d'une série $\sum_{n=0}^{\infty} x_n$ (et en cas de convergence absolue, la valeur de la somme) ne dépend pas de l'ordre des termes de la série. Dans la suite, pour désigner la somme d'une série absolument convergente, nous emploierons indifféremment la notation « série » $\sum_{n=0}^{\infty}$ ou la notation « famille sommable » $\sum_{n \in \mathbb{N}}$. De même quand il s'agira d'une série à termes positifs, dont la somme peut être infinie.

La propriété de σ -additivité des probabilités se reformule immédiatement en termes de familles sommables : si I est *dénombrable* et si $(A_i)_{i \in I}$ est une famille d'événements deux à deux disjoints, on a

$$\boxed{P\left(\bigoplus_{i \in I} A_i\right) = \sum_{i \in I} P(A_i).$$

2.3.3. Propriétés des familles sommables

Lemme 2.13. Soit $(x_i)_{i \in I}$ une famille sommable de nombres réels.

(i) Soit A une partie de I , la famille $(x_i)_{i \in A}$ (restriction de la famille au sous-ensemble A) est encore sommable ;

(ii) l'application $A \mapsto \sum_{i \in A} x_i$ est une « fonction additive d'ensembles ».

Démonstration. (i) résulte immédiatement de l'inégalité

$$\sum_{i \in A} |x_i| \leq \sum_{i \in I} |x_i|$$

(proposition 2.8) et du lemme 2.11.

(ii) résulte de l'égalité

$$\sum_{i \in A} x_i = \sum_{i \in A} x_i^+ - \sum_{i \in A} x_i^-$$

et de l'énoncé analogue pour les familles de réels positifs (proposition 2.8, (ii)).

Remarque. L'analogie de l'assertion (i) ci-dessus est faux dans le cas des séries convergentes et non absolument convergentes : penser par exemple à la série $1 - \frac{1}{2} + \frac{1}{3} + \dots + \frac{(-1)^n}{n} + \dots$.

Il n'est pas tout à fait évident, dans le cas de familles de signe quelconque, qu'on ait

$$\sum_{i \in I} (x_i + y_i) = \sum_{i \in I} x_i + \sum_{i \in I} y_i$$

comme c'est le cas pour les familles de réels ≥ 0 . Écrire $x_i = x_i^+ - x_i^-$ revient en fait à découper l'ensemble des indices i en deux : les indices pour lesquels $x_i > 0$ et ceux pour lesquels $x_i < 0$ (on peut négliger les indices pour lesquels $x_i = 0$). Mais il n'y a aucune raison pour que les découpages correspondant aux trois familles $(x_i)_{i \in I}$, $(y_i)_{i \in I}$ et $(x_i + y_i)_{i \in I}$ coïncident.

Proposition 2.14. Soient $(x_i)_{i \in I}$ et $(y_i)_{i \in I}$ deux familles sommables de nombres réels. La famille $(x_i + y_i)_{i \in I}$ est sommable, et on a

$$\sum_{i \in I} (x_i + y_i) = \sum_{i \in I} x_i + \sum_{i \in I} y_i. \quad (2.6)$$

Démonstration. On règle la question de la sommabilité en observant que

$$\sum_{i \in I} |x_i + y_i| \leq \sum_{i \in I} (|x_i| + |y_i|) = \sum_{i \in I} |x_i| + \sum_{i \in I} |y_i|.$$

Pour prouver (2.6) on partage l'ensemble I en six sous-ensembles,

A :	ensemble des $i \in I$ tels que	$x_i \geq 0, y_i \geq 0$
B :	"	$x_i < 0, y_i < 0$
C :	"	$x_i \geq 0, y_i < 0, x_i + y_i \geq 0$
D :	"	$x_i \geq 0, y_i < 0, x_i + y_i < 0$
E :	"	$x_i < 0, y_i \geq 0, x_i + y_i \geq 0$
F :	"	$x_i < 0, y_i \geq 0, x_i + y_i < 0$

Il résulte de la proposition 2.6 (additivité des familles à termes positifs) qu'on a

$$\sum_{i \in A} (x_i + y_i) = \sum_{i \in A} x_i + \sum_{i \in A} y_i.$$

L'égalité analogue où A est remplacé par B s'obtient de la même façon en remplaçant x_i par $-x_i$ et y_i par $-y_i$.

Prouver que

$$\sum_{i \in C} (x_i + y_i) = \sum_{i \in C} x_i + \sum_{i \in C} y_i$$

revient à prouver, puisque la famille $(-y_i)_{i \in C}$ est sommable, que

$$\sum_{i \in C} (x_i + y_i) + \sum_{i \in C} (-y_i) = \sum_{i \in C} x_i.$$

Or, pour $i \in C$, on a $x_i + y_i \geq 0$ et $-y_i \geq 0$. L'égalité résulte donc encore une fois de la proposition 2.6. On prouve de la même façon les inégalités analogues concernant les sous-ensembles D , E , F . On achève la démonstration en appliquant le lemme 2.13.

Corollaire 2.15. Soient $(x_i)_{i \in I}$ et $(y_i)_{i \in I}$ deux familles sommables de nombres réels. Quels que soient les réels λ et μ , la famille $(\lambda x_i + \mu y_i)_{i \in I}$ est sommable et on a

$$\sum_{i \in I} (\lambda x_i + \mu y_i) = \lambda \sum_{i \in I} x_i + \mu \sum_{i \in I} y_i.$$

Démonstration. Cela résulte directement de la proposition précédente.

Il nous reste à examiner ce que devient la *propriété de sommation par paquets* dans le cas de familles de réels de signe quelconque.

Comme nous l'avons dit dans l'introduction de ce chapitre, c'est l'existence de cette propriété qui justifie l'emploi de la notion de famille sommable (plutôt que celui des séries absolument convergentes). C'est elle qui permet de manier les sommes $\sum_{i \in I} x_i$ (et notamment les sommes doubles, triples, etc. dont la définition ne se distingue pas d'ailleurs de celle des sommes « simples ») avec la même facilité que des sommes finies.

Théorème 2.16. (Sommation par paquets.) Soit $(x_i)_{i \in I}$ une famille de nombres réels et soit $(A_j)_{j \in J}$ une partition de I . Alors, pour que la famille $(x_i)_{i \in I}$ soit sommable, il faut et il suffit que chacune des familles $(x_i)_{i \in A_j}$ soit sommable, et que la famille

$$\left(\sum_{i \in A_j} |x_i| \right)_{j \in J}$$

soit sommable.

Si la famille $(x_i)_{i \in I}$ est sommable, on a

$$\boxed{\sum_{i \in I} x_i = \sum_{j \in J} \left(\sum_{i \in A_j} x_i \right)}. \quad (2.7)$$

Remarque. Observer qu'il ne suffit pas, pour que la famille $(x_i)_{i \in I}$ soit sommable, que chacune des familles $(x_i)_{i \in A_j}$ soit sommable, et que la famille des sommes $(\sum_{i \in A_j} x_i)_{j \in J}$ (sans valeurs absolues) soit sommable. On peut facilement construire un contre-exemple en s'inspirant de la série $(1 - 1) + (1 - 1) + (1 - 1) + \dots$.

Démonstration. Il résulte de la proposition 2.9 (sommation par paquets pour les familles à termes positifs) que

$$\sum_{i \in I} |x_i| = \sum_{j \in J} \left(\sum_{i \in A_j} |x_i| \right), \quad (2.8)$$

d'où la première assertion.

Supposons maintenant que les deux membres de (2.8) soient finis. Dans ce cas, on a

$$\begin{aligned} \sum_{i \in I} x_i &= \sum_{i \in I} x_i^+ - \sum_{i \in I} x_i^- \\ &= \sum_{j \in J} \left(\sum_{i \in A_j} x_i^+ \right) - \sum_{j \in J} \left(\sum_{i \in A_j} x_i^- \right) \end{aligned}$$

(proposition 2.9). La proposition 2.9 affirme notamment que les familles

$$\left(\sum_{i \in A_j} x_i^+ \right)_{j \in J} \quad \text{et} \quad \left(\sum_{i \in A_j} x_i^- \right)_{j \in J}$$

sont sommables. En vertu du corollaire 2.15, leur différence l'est aussi, et on peut écrire

$$\sum_{i \in I} x_i = \sum_{j \in J} \left(\sum_{i \in A_j} x_i^+ - \sum_{i \in A_j} x_i^- \right). \quad (2.9)$$

La proposition 2.9 affirme aussi que pour chaque indice j , les familles (x_i^+) et (x_i^-) sont sommables sur A_j . La famille $(x_i)_{i \in A_j}$, qui s'obtient en prenant leur différence l'est donc aussi, et on peut écrire

$$\sum_{i \in A_j} x_i^+ - \sum_{i \in A_j} x_i^- = \sum_{i \in A_j} x_i.$$

En reportant dans (2.9), on obtient (2.7).

Corollaire 2.17. (Théorème de Fubini pour les familles.)

(i) Soit $(x_{ij})_{(i,j) \in I \times J}$ une famille « double » de réels positifs, éventuellement infinis. Alors on a

$$\boxed{\sum_{(i,j) \in I \times J} x_{ij} = \sum_{i \in I} \left(\sum_{j \in J} x_{ij} \right) = \sum_{j \in J} \left(\sum_{i \in I} x_{ij} \right)}. \quad (2.10)$$

(ii) Soit $(x_{ij})_{i \in I, j \in J}$ une famille double de nombres réels (finis). Alors si l'une des sommes

$$\sum_{(i,j) \in I \times J} |x_{ij}|, \quad \sum_{i \in I} \left(\sum_{j \in J} |x_{ij}| \right), \quad \sum_{j \in J} \left(\sum_{i \in I} |x_{ij}| \right)$$

est finie, les deux autres le sont aussi, et on a

$$\boxed{\sum_{(i,j) \in I \times J} x_{ij} = \sum_{i \in I} \left(\sum_{j \in J} x_{ij} \right) = \sum_{j \in J} \left(\sum_{i \in I} x_{ij} \right)}. \quad (2.11)$$

(ce qui sous-entend que toutes les familles considérées dans (2.11) sont sommables).

Démonstration. Il suffit d'appliquer la proposition précédente en observant que

$$I \times J = \bigsqcup_{i \in I} \{i\} \times J = \bigsqcup_{j \in J} I \times \{j\}.$$

Corollaire 2.18. (i) Soient $(x_i)_{i \in I}$ et $(y_j)_{j \in J}$ deux familles de réels positifs, éventuellement infinis. Alors on a

$$\boxed{\sum_{(i,j) \in I \times J} x_i y_j = \left(\sum_{i \in I} x_i \right) \left(\sum_{j \in J} y_j \right)}. \quad (2.12)$$

(ii) Soient $(x_i)_{i \in I}$ et $(y_j)_{j \in J}$ deux familles sommables de réels de signe quelconque. Alors la famille double $(x_i y_j)_{(i,j) \in I \times J}$ est sommable, et l'égalité (2.12) est vérifiée.

Démonstration. Appliquer le corollaire précédent.

Nous utiliserons souvent dans la suite de cet ouvrage une généralisation du corollaire 2.18 au cas d'un nombre fini quelconque de familles :

Corollaire 2.19. (i) Soient I_1, \dots, I_k des ensembles et soient $(x_{i_1}^{(1)})_{i_1 \in I_1}, \dots, (x_{i_k}^{(k)})_{i_k \in I_k}$ des familles de réels positifs indexées par I_1, \dots, I_k respectivement. Alors on a

$$\sum_{(i_1, \dots, i_k) \in I_1 \times \dots \times I_k} x_{i_1}^{(1)} \dots x_{i_k}^{(k)} = \left(\sum_{i_1 \in I_1} x_{i_1}^{(1)} \right) \dots \left(\sum_{i_k \in I_k} x_{i_k}^{(k)} \right). \quad (2.13)$$

(ii) Soient maintenant $(x_{i_1}^{(1)})_{i_1 \in I_1}, \dots, (x_{i_k}^{(k)})_{i_k \in I_k}$ des familles de réels de signe quelconque, supposées sommables. Alors la famille $(x_{i_1}^{(1)} \dots x_{i_k}^{(k)})_{(i_1, \dots, i_k) \in I_1 \times \dots \times I_k}$ est sommable, et l'égalité (2.13) est encore vérifiée.

Démonstration. La démonstration se fait par récurrence à partir du corollaire 2.18.

Exercices

Exercice 2.1. Démontrer que la famille $(x_{mn})_{(m,n) \in \mathbb{N}^{*2}}$ indexée par \mathbb{N}^{*2} et définie par

$$\forall (m, n) \in \mathbb{N}^{*2} \quad x_{mn} = \frac{(-1)^{mn}}{m^2 n^2}$$

est sommable, et calculer sa somme Σ . On rappelle que $S \equiv \sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$, résultat que l'on peut obtenir grâce à la théorie des séries de Fourier.

Solution. On a

$$\sum_{(m,n) \in \mathbb{N}^{*2}} |x_{mn}| = \sum_{(m,n) \in \mathbb{N}^{*2}} \frac{1}{n^2 m^2}.$$

d'où

$$\sum_{(m,n) \in \mathbb{N}^{*2}} |x_{mn}| = \left(\sum_{n \in \mathbb{N}^*} \frac{1}{n^2} \right)^2,$$

en utilisant l'égalité (2.12), cas particulier de la propriété de Fubini. Puisque

$$\sum_{n \in \mathbb{N}^*} \frac{1}{n^2} < +\infty,$$

il vient :

$$\sum_{(m,n) \in \mathbb{N}^{*2}} |x_{mn}| < +\infty.$$

La famille (x_{mn}) est donc sommable sur \mathbb{N}^{*2} . Reste à calculer sa somme Σ ; on a :

$$\Sigma = \sum_{\substack{(m,n) \in \mathbb{N}^{*2} \\ nm \text{ pair}}} \frac{1}{n^2 m^2} - \sum_{\substack{(m,n) \in \mathbb{N}^{*2} \\ nm \text{ impair}}} \frac{1}{n^2 m^2}. \quad (2.14)$$

Mais, en notant P l'ensemble des entiers > 0 pairs, et I l'ensemble des entiers > 0 impairs, on a

$$\left\{ (m, n) \in \mathbb{N}^{*2} \mid nm \text{ pair} \right\} = (P \times P) \uplus (P \times I) \uplus (I \times P)$$

et

$$\left\{ (m, n) \in \mathbb{N}^{*2} \mid nm \text{ impair} \right\} = I \times I.$$

D'où

$$\sum_{\substack{(m,n) \in \mathbb{N}^{*2} \\ nm \text{ pair}}} \frac{1}{n^2 m^2} = \sum_{(m,n) \in P \times P} \frac{1}{n^2 m^2} + \sum_{(m,n) \in P \times I} \frac{1}{n^2 m^2} + \sum_{(m,n) \in I \times P} \frac{1}{n^2 m^2}$$

(additivité par rapport aux ensembles) et

$$\sum_{\substack{(m,n) \in \mathbb{N}^{*2} \\ nm \text{ impair}}} \frac{1}{n^2 m^2} = \sum_{(m,n) \in I \times I} \frac{1}{n^2 m^2}.$$

En posant

$$U = \sum_{n \in P} \frac{1}{n^2} \quad \text{et} \quad V = \sum_{n \in I} \frac{1}{n^2},$$

on obtient, en utilisant la propriété de Fubini (pour les familles positives)

$$\sum_{(m,n) \in P \times P} \frac{1}{n^2 m^2} = \left(\sum_{m \in P} \frac{1}{m^2} \right) \left(\sum_{n \in P} \frac{1}{n^2} \right)$$

et des égalités analogues pour les trois autres termes. D'où

$$\Sigma = U^2 + 2UV - V^2.$$

En observant que

$$S = U + V$$

(additivité) et que

$$U = \sum_{k \in \mathbb{N}^*} \frac{1}{(2k)^2} = \frac{1}{4} \sum_{k \in \mathbb{N}^*} \frac{1}{k^2} = \frac{S}{4};$$

on obtient les valeurs de U et V :

$$U = \frac{S}{4} \quad \text{et} \quad V = \frac{3S}{4}.$$

Il vient alors :

$$\Sigma = S^2 \left[\frac{1}{16} + 2 \times \frac{3}{16} - \frac{9}{16} \right] = -\frac{S^2}{8},$$

soit :

$$\Sigma = -\frac{\pi^4}{288} \simeq -0,338.$$

Exercice 2.2. Familles sommables et germes de probabilité. Soient α et β deux réels strictement compris entre 0 et 1, et g la fonction à valeurs positives définie sur \mathbb{N}^2 par :

$$\forall (i, j) \in \mathbb{N}^2 \quad g(i, j) = \alpha\beta(1-\alpha)^i(1-\beta)^j.$$

1. Vérifier que g est un germe de probabilité sur l'espace probabilisable $(\mathbb{N}^2, \mathcal{P}(\mathbb{N}^2))$. Soit P la probabilité associée à g sur l'espace probabilisable $(\mathbb{N}^2, \mathcal{P}(\mathbb{N}^2))$.

2. Soit X (respectivement Y) la variable aléatoire discrète définie sur l'espace probabilisé $(\mathbb{N}^2, \mathcal{P}(\mathbb{N}^2), P)$ par :

$$\forall (i, j) \in \mathbb{N}^2 \quad X(i, j) = i \quad Y(i, j) = j.$$

Quelles sont les lois des variables aléatoires X et Y ? Les identifier à des lois connues.

3. Calculer :

- (a) $P(X = Y)$,
 (b) $P(X > Y)$.

4. Soit Z la variable aléatoire discrète définie sur l'espace probabilisé $(\mathbb{N}^2, \mathcal{P}(\mathbb{N}^2), P)$ par :

$$\forall (i, j) \in \mathbb{N}^2 \quad Z(i, j) = \begin{cases} 1 & \text{si } i \text{ et } j \text{ sont pairs} \\ -1 & \text{si } i \text{ et } j \text{ sont impairs} \\ 0 & \text{si } i \text{ et } j \text{ sont de parité différente} \end{cases}$$

(a) Démontrer que la famille Zg définie sur \mathbb{N}^2 par :

$$\forall (i, j) \in \mathbb{N}^2 \quad (Zg)(i, j) = Z(i, j)g(i, j)$$

est sommable sur \mathbb{N}^2 et calculer sa somme $\sum_{(i,j) \in \mathbb{N}^2} Zg(i, j)$.

(b) Si D est la diagonale $\{(i, j) \in \mathbb{N}^2 \mid i = j\}$, vérifier que la famille Zg est sommable sur D ; que vaut la somme $\sum_{(i,j) \in D} Zg(i, j)$?

Solution.

1. La fonction g étant *positive* à variables séparées (c'est-à-dire de la forme $g(i, j) = g_1(i)g_2(j)$), on a :

$$\begin{aligned} \sum_{(i,j) \in \mathbb{N}^2} g(i, j) &= \alpha\beta \left(\sum_{i \in \mathbb{N}} (1-\alpha)^i \right) \left(\sum_{j \in \mathbb{N}} (1-\beta)^j \right) \\ &= \alpha\beta \left(\sum_{i=0}^{+\infty} (1-\alpha)^i \right) \left(\sum_{j=0}^{+\infty} (1-\beta)^j \right) \\ &= \alpha\beta \frac{1}{1-(1-\alpha)} \frac{1}{1-(1-\beta)} \\ &= 1, \end{aligned}$$

la deuxième égalité résultant de la convergence des séries géométriques de raison plus petite que 1 en valeur absolue et de la proposition 2.4. La fonction g étant positive et de somme 1 sur \mathbb{N}^2 est bien un germe de probabilité sur l'espace probabilisable $(\mathbb{N}^2, \mathcal{P}(\mathbb{N}^2))$.

2. On peut écrire, pour tout $i \in \mathbb{N}$,

$$(X = i) = \{i\} \times \mathbb{N}.$$

Il résulte de la définition de la probabilité P associée à g et de l'application de la proposition 2.4 que

$$\begin{aligned} P(X = i) &= \sum_{(j,k) \in \{i\} \times \mathbb{N}} g(j, k) \\ &= \alpha(1 - \alpha)^i \sum_{k \in \mathbb{N}} \beta(1 - \beta)^k, \end{aligned}$$

soit :

$$P(X = i) = \alpha(1 - \alpha)^i.$$

Ceci démontre que la loi de la variable aléatoire X est la loi géométrique sur \mathbb{N} de paramètre α . De même, la loi de la variable aléatoire Y est la loi géométrique sur \mathbb{N} de paramètre β .

3. La méthode consiste à partitionner les événements.

(a) On a :

$$(X = Y) = \bigsqcup_{i \in \mathbb{N}} ((X = i) \cap (Y = i)) = \{(i, i) \mid i \in \mathbb{N}\},$$

et donc :

$$\begin{aligned} P(X = Y) &= \sum_{i \in \mathbb{N}} g(i, i) \\ &= \alpha\beta \sum_{i \in \mathbb{N}} (1 - \alpha)^i (1 - \beta)^i. \end{aligned}$$

Puisque $0 < (1 - \alpha)(1 - \beta) < 1$, il vient (somme d'une série géométrique et proposition 2.4) :

$$P(X = Y) = \frac{\alpha\beta}{1 - (1 - \alpha)(1 - \beta)} = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta}.$$

(b) On a

$$(X > Y) = \bigsqcup_{(i,j) \in \mathbb{N}^2} (X = i) \cap (Y = j) = \{(i, j) \in \mathbb{N}^2 \mid i > j\}$$

et par conséquent :

$$P(X > Y) = \sum_{\substack{(i,j) \in \mathbb{N}^2 \\ i > j}} [\alpha\beta(1 - \alpha)^i (1 - \beta)^j].$$

Par application de la propriété de Fubini (possible car g est positive) et de la proposition 2.4, on peut écrire successivement :

$$\begin{aligned}
 P(X > Y) &= \alpha\beta \sum_{\substack{j \in \mathbb{N} \\ +\infty}} \left[(1-\beta)^j \sum_{\substack{i > j \\ +\infty}} (1-\alpha)^i \right] \\
 &= \alpha\beta \sum_{j=0}^{+\infty} \left[(1-\beta)^j \sum_{i=j+1}^{+\infty} (1-\alpha)^i \right] \\
 &= \alpha\beta \sum_{j=0}^{+\infty} \left[(1-\beta)^j \frac{(1-\alpha)^{j+1}}{1-(1-\alpha)} \right] \\
 &= \beta(1-\alpha) \frac{1}{1-(1-\alpha)(1-\beta)}.
 \end{aligned}$$

Il en résulte que :

$$\boxed{P(X > Y) = \frac{\beta(1-\alpha)}{\alpha + \beta - \alpha\beta}}$$

4. On applique ici des propriétés fondamentales des familles sommables.

(a) Puisque $|Z| \leq 1$, on a : $|Zg| \leq g$; la famille g étant sommable sur \mathbb{N}^2 , il en est de même de Zg et on a :

$$\sum_{(i,j) \in \mathbb{N}^2} Zg(i,j) = \sum_{(i,j) \in (2\mathbb{N})^2} g(i,j) - \sum_{(i,j) \in (2\mathbb{N}+1)^2} g(i,j).$$

Or, d'après la propriété de Fubini, on a :

$$\begin{aligned}
 \sum_{(i,j) \in (2\mathbb{N})^2} g(i,j) &= \sum_{(i,j) \in \mathbb{N}^2} [\alpha\beta(1-\alpha)^{2i}(1-\beta)^{2j}] \\
 &= \alpha\beta \left(\sum_{i \in \mathbb{N}} (1-\alpha)^{2i} \right) \left(\sum_{j \in \mathbb{N}} (1-\beta)^{2j} \right) \\
 &= \alpha\beta \frac{1}{1-(1-\alpha)^2} \frac{1}{1-(1-\beta)^2} \\
 &= \frac{1}{(2-\alpha)(2-\beta)}.
 \end{aligned}$$

De même, on a :

$$\begin{aligned}
 \sum_{(i,j) \in (2\mathbb{N}+1)^2} g(i,j) &= \sum_{(i,j) \in \mathbb{N}^2} [\alpha\beta(1-\alpha)^{2i+1}(1-\beta)^{2j+1}] \\
 &= (1-\alpha)(1-\beta) \sum_{(i,j) \in (2\mathbb{N})^2} g(i,j),
 \end{aligned}$$

d'où :

$$\sum_{(i,j) \in \mathbb{N}^2} Zg(i,j) = [1 - (1-\alpha)(1-\beta)] \frac{1}{(2-\alpha)(2-\beta)},$$

soit :

$$\boxed{\sum_{(i,j) \in \mathbb{N}^2} Zg(i,j) = \frac{\alpha + \beta - \alpha\beta}{(2-\alpha)(2-\beta)}}$$

(b) La famille Zg est sommable sur \mathbb{N}^2 , donc aussi sur D , et on a :

$$\begin{aligned}
 \sum_{(i,j) \in D} Zg(i,j) &= \sum_{(i,j) \in D \cap (2\mathbb{N})^2} g(i,j) - \sum_{(i,j) \in D \cap (2\mathbb{N}+1)^2} g(i,j) \\
 &= \sum_{i=0}^{+\infty} [\alpha\beta(1-\alpha)^{2i}(1-\beta)^{2i}] - \sum_{i=0}^{+\infty} [\alpha\beta(1-\alpha)^{2i+1}(1-\beta)^{2i+1}] \\
 &= \alpha\beta[1 - (1-\alpha)(1-\beta)] \frac{1}{1 - (1-\alpha)^2(1-\beta)^2} \\
 &= \frac{\alpha\beta}{1 + (1-\alpha)(1-\beta)},
 \end{aligned}$$

soit :

$$\boxed{\sum_{(i,j) \in D} Zg(i,j) = \frac{\alpha\beta}{2 - (\alpha + \beta) + \alpha\beta}.}$$

Chapitre 3

Indépendance

3.1. Introduction

La modélisation de deux jets consécutifs d'un même dé conduit naturellement, si l'on s'intéresse à des événements relatifs à l'ensemble des deux jets, à prendre comme espace fondamental, ou espace des réalisations, le produit cartésien $\Omega = \Omega_1 \times \Omega_2$ où $\Omega_1 = \Omega_2 = \{1, 2, \dots, 6\}$, et à prendre pour probabilité P sur l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$ la **probabilité uniforme**. Un événement A relatif au premier jet uniquement est représenté par une partie de Ω de la forme $A = A_1 \times \Omega_2$. De même, un événement B relatif au second jet est représenté par une partie de Ω de la forme $B = \Omega_1 \times A_2$. Ces événements sont **indépendants**, au sens intuitif, puisqu'a priori le premier jet n'a aucune influence sur le second. Nous allons calculer la probabilité de réalisation simultanée de ces deux événements. On a

$$A \cap B = (A_1 \times \Omega_2) \cap (\Omega_1 \times A_2) = A_1 \times A_2$$

L'additivité de la probabilité entraîne

$$\begin{aligned} P(A_1 \times A_2) &= \sum_{\omega_1 \in A_1} P(\{\omega_1\} \times A_2) \\ &= \sum_{\omega_1 \in A_1} \frac{|A_2|}{|\Omega_1| |\Omega_2|}, \end{aligned}$$

soit

$$P(A_1 \times A_2) = \frac{|A_1| |A_2|}{|\Omega_1| |\Omega_2|}.$$

En particulier, pour $A_2 = \Omega_2$, on a

$$P(A_1 \times \Omega_2) = \frac{|A_1|}{|\Omega_1|}.$$

On a de même

$$P(\Omega_1 \times A_2) = \frac{|A_2|}{|\Omega_2|}.$$

On vient de démontrer que :

$$P(A \cap B) = P(A)P(B).$$

Nous prendrons cette relation comme définition de l'**indépendance** de deux événements A et B . Dans le cas considéré, elle traduit une conséquence nécessaire de l'indépendance physique des jets de dés.

Une autre façon de décrire la situation précédente consiste à introduire les variables aléatoires X_1 et X_2 définies sur Ω , à valeurs dans Ω_1 et Ω_2 , définies respectivement par

$$\forall (\omega_1, \omega_2) \in \Omega \quad X_1(\omega_1, \omega_2) = \omega_1 \quad \text{et} \quad X_2(\omega_1, \omega_2) = \omega_2.$$

Du point de vue « ensembliste », ce sont les projections du produit cartésien sur chacun des « facteurs ». Du point de vue probabiliste, ou intuitif, X_1 et X_2 représentent respectivement le résultat du premier et le résultat du second jet. L'événement A peut s'écrire $A = (X_1 \in A_1)$. On a de même $B = (X_2 \in A_2)$. Et donc

$$A \cap B = (X_1 \in A_1) \cap (X_2 \in A_2).$$

Pour désigner l'événement figurant au second membre, conjonction ou réalisation simultanée des événements $(X_1 \in A_1)$ et $(X_2 \in A_2)$, on emploie plus couramment la notation $(X_1 \in A_1, X_2 \in A_2)$. En employant cette notation, la formule encadrée ci-dessus s'écrit

$$P(X_1 \in A_1, X_2 \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2). \quad (3.1)$$

Nous prendrons cette relation comme définition de l'indépendance de deux variables aléatoires. Soient deux variables aléatoires X_1 et X_2 , définies sur le même espace probabilisé (Ω, \mathcal{A}, P) , à valeurs respectivement dans les espaces probabilisables $(\Omega_1, \mathcal{A}_1)$ et $(\Omega_2, \mathcal{A}_2)$. On dit que X_1 et X_2 sont **indépendantes** si la relation (3.1) est vérifiée quels que soient $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$.

Il est à noter que (3.1) s'écrit encore, en terme de lois de variables aléatoires :

$$P_{(X_1, X_2)}(A_1 \times A_2) = P_{X_1}(A_1) P_{X_2}(A_2). \quad (3.2)$$

Remarque. Revenant au cas particulier étudié plus haut, on peut observer que

$$P_{X_1}(A_1) = P(X_1 \in A_1) = P(A_1 \times \Omega_1)$$

et donc

$$P_{X_1}(A_1) = \frac{|A_1|}{|\Omega_1|}$$

(voir plus haut). On a une expression analogue pour $P_{X_2}(A_2)$, et on peut écrire

$$\forall A_i \in \mathcal{P}(\Omega_i) \quad P_{X_i}(A_i) = \frac{|A_i|}{|\Omega_i|},$$

ce qui traduit le fait que les variables aléatoires X_1 et X_2 sont de **loi uniforme** sur Ω_1 et Ω_2 respectivement.

Comme le montre l'exemple ci-dessous, dire que la probabilité P est uniforme sur Ω , c'est dire plus (et ce « plus » est, comme nous le verrons (proposition 3.11), une notion d'indépendance) que dire que chacune des variables aléatoires X_i est de **loi uniforme** sur Ω_i . En effet, on va construire une probabilité P sur Ω qui n'est pas uniforme et qui est telle que les variables aléatoires X_i soient de **loi uniforme** sur Ω_i .

Soient Λ_1 et Λ_2 les parties de Ω définies par

$$\Lambda_1 = \{(0, 1), (1, 2), (2, 0)\} \quad \Lambda_2 = \{(0, 2), (2, 1), (1, 0)\}.$$

On se donne un réel ε tel que $0 < \varepsilon < \frac{1}{36}$. Soit g le germe de probabilité défini sur Ω par

$$\forall (i, j) \in \Omega, \quad g(i, j) = \begin{cases} \frac{1}{36} + \varepsilon & \text{si } (i, j) \in \Lambda_1 \\ \frac{1}{36} - \varepsilon & \text{si } (i, j) \in \Lambda_2 \\ \frac{1}{36} & \text{sinon} \end{cases}$$

et P la probabilité (*non uniforme*) sur $(\Omega, \mathcal{P}(\Omega))$ associée à g . Pour tout $i \in \{1, 2, \dots, 6\}$, on a

$$P(X_1 = i) = \bigoplus_{j=1}^6 \{(i, j)\},$$

et donc, comme on le vérifie facilement,

$$P(X_1 = i) = \sum_{j=1}^6 P\{(i, j)\} = \sum_{j=1}^6 g(i, j) = \frac{1}{6}.$$

C'est à dire que la loi de X_1 est uniforme sur Ω_1 ; de même, la loi de X_2 est uniforme sur Ω_2 . Mais les variables aléatoires X_1 et X_2 ne sont pas indépendantes au sens défini par la relation (3.1) puisque, par exemple, on a :

$$P(X_1 = 1, X_2 = 2) = g(1, 2) = \frac{1}{36} + \varepsilon,$$

qui est différent du produit

$$P(X_1 = 1) P(X_2 = 2) = \frac{1}{36}.$$

3.2. Indépendance d'événements et de variables aléatoires

Nous allons étudier la notion d'indépendance de manière générale. Sur certains points, nous devons toutefois nous limiter au cas des variables aléatoires discrètes¹.

3.2.1. Événements indépendants

Définition 3.1. Soit (Ω, \mathcal{A}, P) un espace probabilisé.

(i) Deux événements $A \in \mathcal{A}$ et $B \in \mathcal{A}$ sont **indépendants** si :

$$P(A \cap B) = P(A) P(B).$$

(ii) Soit $(A_i)_{i \in I}$ une famille quelconque d'événements. Ces événements sont dits **indépendants** (on dit parfois² indépendants dans leur ensemble) si, pour toute partie finie non vide J de I , on a :

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

En particulier, des événements A_1, A_2, \dots, A_n sont indépendants si pour toute suite finie $1 \leq j_1 < j_2 < \dots < j_k \leq n$ on a

$$P(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = P(A_{j_1})P(A_{j_2}) \dots P(A_{j_k}).$$

Remarques. 1. Si $I = \{1, 2\}$, on retrouve la définition de l'indépendance de deux événements.

2. L'indépendance des événements A_1, A_2, \dots, A_n représente $2^n - n - 1$ conditions (nombre de parties de cardinal ≥ 2 de $I = \{1, 2, \dots, n\}$).

3. Si n événements sont indépendants **dans leur ensemble**, ils sont indépendants **deux à deux**, mais la réciproque n'est pas nécessairement vraie.

Exemple 3.1. On considère l'espace probabilisé (Ω, \mathcal{A}, P) , où Ω est un ensemble à quatre éléments, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, muni de la tribu de ses parties et où P est la probabilité uniforme sur Ω . On considère les événements

$$A = \{\omega_1, \omega_2\}, \quad B = \{\omega_1, \omega_3\}, \quad C = \{\omega_1, \omega_4\}.$$

1. L'étude générale sera faite ultérieurement (tome II) dans le cadre d'une théorie des probabilités basée sur la théorie de la mesure.

2. On dit aussi parfois que les événements sont **mutuellement** indépendants.

On a :

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$$

et

$$P(A) P(B) = P(A) P(C) = P(B) P(C) = \frac{1}{4}.$$

Les événements A, B et C sont donc indépendants **deux à deux**, mais

$$P(A \cap B \cap C) = \frac{1}{4} \quad \text{et} \quad P(A) P(B) P(C) = \frac{1}{8};$$

les événements A, B et C ne sont donc pas indépendants (dans leur ensemble).

4. On peut avoir :

$$P(A \cap B \cap C) = P(A) P(B) P(C)$$

sans que les événements A, B et C soient indépendants deux à deux.

Exemple 3.2. On considère l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$, où $\Omega = \{1, 2, \dots, 6\}^2$ est muni de la tribu de ses parties et où P est la probabilité uniforme. On considère les trois événements :

$$A = \Omega_1 \times \{1, 2, 5\}, \quad B = \Omega_1 \times \{4, 5, 6\}, \quad C = \{(i, j) \in \Omega \mid i + j = 9\};$$

en fait

$$C = \{(3, 6), (4, 5), (5, 4), (6, 3)\}$$

et on a donc :

$$P(C) = \frac{4}{36}.$$

On a ainsi :

$$\begin{aligned} P(A \cap B) &= \frac{6}{36} \neq \frac{1}{2} \frac{1}{2} = P(A) P(B), \\ P(A \cap C) &= \frac{1}{36} \neq \frac{1}{2} \frac{1}{9} = P(A) P(C), \\ P(B \cap C) &= \frac{3}{36} \neq \frac{1}{2} \frac{1}{9} = P(B) P(C). \end{aligned}$$

Toutefois, on a :

$$P(A \cap B \cap C) = \frac{1}{36} = P(A) P(B) P(C).$$

Les événements A, B et C sont donc **trois à trois** indépendants mais ne sont pas indépendants **dans leur ensemble**.

3.2.2. Événements indépendants et passage au complémentaire

Dire que A et B sont indépendants, c'est dire que la réalisation, ou la non-réalisation, de B n'influe pas sur la probabilité de voir A se réaliser. Il est naturel de penser qu'on peut remplacer dans cette formulation B par l'événement contraire B^c . C'est le contenu de la proposition suivante.

Proposition 3.2. Soit (Ω, \mathcal{A}, P) un espace probabilisé.

(i) Si A et B sont deux événements **indépendants**, les événements A et B^c d'une part, A^c et B^c d'autre part, A^c et B enfin, sont indépendants.

(ii) La propriété ci-dessus se généralise de la manière suivante : soit $(A_i)_{i \in I}$ une famille d'événements **indépendants** indexée par un ensemble **quelconque** I ; on suppose que l'ensemble I est partitionné en deux parties I_1 et I_2 . Soit $(B_i)_{i \in I}$ la famille d'événements définis par :

$$B_i = \begin{cases} A_i & \text{si } i \in I_1 \\ A_i^c & \text{si } i \in I_2. \end{cases}$$

Les événements B_i sont indépendants.

Démonstration.

(i) Nous ne faisons la démonstration que pour A et B^c , les autres résultats s'en déduisant par passage au complémentaire. On a

$$P(A \cap B^c) = P(A) - P(A \cap B)$$

et, par indépendance de A et B :

$$P(A \cap B^c) = P(A) - P(A) P(B),$$

ce qui donne, en factorisant :

$$P(A \cap B^c) = P(A) P(B^c).$$

(ii) Il faut démontrer que, pour toute partie **finie** J non vide de I , on a :

$$P\left(\bigcap_{j \in J} B_j\right) = \prod_{j \in J} P(B_j). \quad (3.3)$$

On note $J_1 = I_1 \cap J$ et $J_2 = I_2 \cap J$. Nous supposons ces deux ensembles non vides (les démonstrations dans les autres cas se traitent de manière analogue). Remarquons que l'on a :

$$\begin{aligned} \bigcap_{j \in J} B_j &= \left(\bigcap_{j \in J_1} A_j\right) \cap \left(\bigcap_{j \in J_2} A_j^c\right) \\ &= \left(\bigcap_{j \in J_1} A_j\right) \cap \left(\bigcup_{j \in J_2} A_j\right)^c; \end{aligned}$$

on obtient alors, en utilisant la formule de Poincaré :

$$\begin{aligned} P\left(\bigcap_{j \in J} B_j\right) &= P\left(\bigcap_{j \in J_1} A_j\right) - \sum_{\substack{K \subset J_2 \\ |K|=1}} P\left(\bigcap_{j \in K \cup J_1} A_j\right) \\ &\quad + \sum_{\substack{K \subset J_2 \\ |K|=2}} P\left(\bigcap_{j \in K \cup J_1} A_j\right) - \cdots + (-1)^{|J_2|} P\left(\bigcap_{j \in J} A_j\right). \end{aligned}$$

On a alors, d'après l'indépendance des A_i ,

$$\begin{aligned} P\left(\bigcap_{j \in J} B_j\right) &= \prod_{j \in J_1} P(A_j) - \sum_{\substack{K \subset J_2 \\ |K|=1}} \prod_{j \in K \cup J_1} P(A_j) \\ &\quad + \sum_{\substack{K \subset J_2 \\ |K|=2}} \prod_{j \in K \cup J_1} P(A_j) - \cdots + (-1)^{|J_2|} \prod_{j \in J} P(A_j), \end{aligned}$$

soit, en factorisant,

$$\begin{aligned} P\left(\bigcap_{j \in J} B_j\right) &= \left[\prod_{j \in J_1} P(A_j) \right] \\ &\quad \times \left[1 - \sum_{j \in J_2} P(A_j) + \sum_{\substack{K \subset J_2 \\ |K|=2}} \prod_{j \in K} P(A_j) - \cdots + (-1)^{|J_2|} \prod_{j \in J_2} P(A_j) \right], \end{aligned}$$

ou encore

$$P\left(\bigcap_{j \in J} B_j\right) = \left[\prod_{j \in J_1} P(A_j) \right] \times \left[\prod_{j \in J_2} (1 - P(A_j)) \right],$$

ce qui donne l'égalité (3.3).

3.2.3. Variables aléatoires indépendantes

On définit maintenant la notion de variables aléatoires indépendantes.

Définition 3.3. *Toutes les variables aléatoires considérées sont définies sur le même espace probabilisé (Ω, \mathcal{A}, P) .*

(i) *Soient deux variables aléatoires X_1 et X_2 à valeurs respectivement dans des espaces probabilisables (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) .*

Les variables aléatoires X_1 et X_2 sont indépendantes si, pour tout $A_1 \in \mathcal{E}_1$ et pour tout $A_2 \in \mathcal{E}_2$, les événements $(X_1 \in A_1)$ et $(X_2 \in A_2)$ sont indépendants.

(ii) *Cette notion se généralise à une famille quelconque de variables aléatoires de la manière suivante.*

Soit une famille $(X_i)_{i \in I}$ de variables aléatoires à valeurs dans des espaces probabilisables respectifs (E_i, \mathcal{E}_i) .

Les variables aléatoires X_i sont **indépendantes** si pour toute famille $(A_i)_{i \in I}$ d'ensembles telle que, pour tout $i \in I$, $A_i \in \mathcal{E}_i$, les événements $(X_i \in A_i)$, $i \in I$, sont indépendants.

Proposition 3.4. Soit I un ensemble fini non vide. Soit $(X_i)_{i \in I}$ une famille de variables aléatoires à valeurs respectivement dans des espaces probabilisables (E_i, \mathcal{E}_i) . Les variables aléatoires X_i sont **indépendantes** si et seulement si pour toute famille $(A_i)_{i \in I}$ d'ensembles telle que, pour tout $i \in I$, $A_i \in \mathcal{E}_i$, on a la relation :

$$P\left[\bigcap_{i \in I} (X_i \in A_i)\right] = \prod_{i \in I} P(X_i \in A_i). \quad (3.4)$$

Démonstration. La condition est évidemment nécessaire.

Pour montrer qu'elle est suffisante, supposons que la relation (3.4) soit vraie. Soit une famille *quelconque* $(B_i)_{i \in I}$ telle que, pour tout $i \in I$, $B_i \in \mathcal{E}_i$ et soit J une partie *quelconque* de I . On définit une nouvelle famille $(A_i)_{i \in I}$ en posant $A_i = B_i$ si $i \in J$ et $A_i = E_i$ sinon. Puisque, pour tout i n'appartenant pas à J , on a

$$(X_i \in A_i) = (X_i \in E_i) = \Omega$$

et donc $P(X_i \in A_i) = 1$, la relation (3.4) permet d'écrire :

$$\begin{aligned} P\left[\bigcap_{j \in J} (X_j \in B_j)\right] \\ = P\left[\bigcap_{i \in I} (X_i \in A_i)\right] = \prod_{i \in I} P(X_i \in A_i) = \prod_{j \in J} P(X_j \in B_j), \end{aligned}$$

ce qui démontre, puisque J est une partie *quelconque* de I , que les événements $(X_i \in B_i)$, $i \in I$, sont indépendants; les variables aléatoires X_i , $i \in I$, sont donc indépendantes.

Dans le cas où les variables aléatoires X_i ($i \in I$) sont discrètes, on peut remplacer dans la proposition 3.4 la considération des événements $\bigcap_{i \in I} (X_i \in A_i)$ par celle des événements $\bigcap_{i \in I} (X_i = x_i)$. C'est ce qu'affirme la proposition 3.5 ci-dessous. Pour simplifier, on énonce la proposition dans le cas où $I = \{1, \dots, n\}$. On emploie aussi la notation

$$(X_1 = x_1, \dots, X_n = x_n)$$

pour représenter l'événement $\bigcap_{i=1}^n (X_i = x_i)$. Nous utiliserons cette notation dans toute la suite.

Proposition 3.5. Soit $(X_i)_{i=1, \dots, n}$ une famille de variables aléatoires discrètes à valeurs respectivement dans des ensembles E_i ($i = 1, \dots, n$). Les variables aléatoires X_i sont indépendantes si et seulement si, pour tout n -uplet $(x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega)$, on a :

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n). \quad (3.5)$$

Démonstration. La condition est nécessaire. Soient $x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)$; les variables aléatoires X_1, \dots, X_n étant indépendantes, la relation (3.4) donne la relation (3.5) en prenant $A_i = \{x_i\}$ pour $i = 1, \dots, n$.

La condition est suffisante. Supposons que pour tout n -uplet $(x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega)$, la relation (3.5) soit vraie. Soit, pour $i = 1, \dots, n$, un ensemble quelconque $A_i \in \mathcal{E}_i$; on a :

$$(X_i \in A_i) = \bigcup_{x_i \in X_i(\Omega) \cap A_i} (X_i = x_i).$$

En utilisant la distributivité des opérations d'union et d'intersection, il vient :

$$P\left[\bigcap_{i \in I} (X_i \in A_i)\right] = P\left[\bigcup_{(x_1, \dots, x_n) \in (X_1(\Omega) \cap A_1) \times \dots \times (X_n(\Omega) \cap A_n)} \left(\bigcap_{i=1}^n (X_i = x_i)\right)\right].$$

Les variables aléatoires X_i étant *discrètes*, les ensembles $X_i(\Omega) \cap A_i$ sont dénombrables. On peut alors écrire, par σ -additivité de P :

$$P\left[\bigcap_{i \in I} (X_i \in A_i)\right] = \sum_{(x_1, \dots, x_n) \in (X_1(\Omega) \cap A_1) \times \dots \times (X_n(\Omega) \cap A_n)} P\left[\bigcap_{i=1}^n (X_i = x_i)\right],$$

soit, par la relation (3.5),

$$P\left[\bigcap_{i \in I} (X_i \in A_i)\right] = \sum_{(x_1, \dots, x_n) \in (X_1(\Omega) \cap A_1) \times \dots \times (X_n(\Omega) \cap A_n)} \prod_{i=1}^n P(X_i = x_i);$$

cette égalité s'écrit encore, en utilisant la propriété de Fubini dans le cas du produit direct de familles :

$$P\left[\bigcap_{i \in I} (X_i \in A_i)\right] = \prod_{i=1}^n \sum_{x_i \in X_i(\Omega) \cap A_i} P(X_i = x_i),$$

ou enfin :

$$P\left[\bigcap_{i \in I} (X_i \in A_i)\right] = \prod_{i=1}^n P(X_i \in A_i).$$

Les variables aléatoires X_1, X_2, \dots, X_n sont donc indépendantes.

Remarque. Par son formalisme, la démonstration donnée de cette proposition peut paraître assez abstraite ; cette proposition est pourtant d'un usage courant. Pour s'assurer que tous les points sont bien compris, on suggère, à titre d'exercice, de refaire cette démonstration dans le cas particulier de **deux** variables aléatoires à valeurs par exemple dans \mathbb{N} .

3.2.4. Fonctions de variables aléatoires (cas discret)

Soit X une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{A}, P) , à valeurs dans un espace probabilisable (E, \mathcal{E}) . Et soit f une variable aléatoire définie sur (E, \mathcal{E}) , à valeurs dans un espace (F, \mathcal{F}) . Alors l'application composée $f \circ X$ est une variable aléatoire définie sur (Ω, \mathcal{A}, P) , à valeurs dans (F, \mathcal{F}) : en effet, pour tout $B \in \mathcal{G}$, l'ensemble

$$(f \circ X)^{-1}(B) = X^{-1}(f^{-1}(B))$$

appartient à \mathcal{A} . Plutôt que $f \circ X$, cette variable aléatoire est classiquement notée $f(X)$. On parle à son propos de **fonction de la variable aléatoire** X .

Si X est discrète, ce qui, rappelons-le, signifie que l'ensemble des valeurs de X est dénombrable, et si, pour tout $x \in E$, l'ensemble $X^{-1}(\{x\})$ appartient à \mathcal{A} , alors $f \circ X$ est une variable aléatoire quelle que soit l'application $f : E \rightarrow F$. Par exemple, si X est une variable aléatoire discrète à valeurs réelles, les fonctions $\sin X, X^2$, définies sur Ω sont encore des variables aléatoires discrètes.

Toujours dans le cas discret³, on peut considérer des fonctions de plusieurs variables aléatoires. Soient X_1, \dots, X_n des variables aléatoires discrètes définies sur (Ω, \mathcal{A}, P) , à valeurs respectivement dans $(E_1, \mathcal{P}(E_1)), \dots, (E_n, \mathcal{P}(E_n))$. Soit F un ensemble quelconque, et soit $f : E_1 \times \dots \times E_n \rightarrow F$ une application quelconque. Alors la composée de l'application $\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$ de Ω dans $E_1 \times \dots \times E_n$ et de l'application f se note $f(X_1, \dots, X_n)$, et c'est une variable aléatoire discrète à valeurs dans F .

Le résultat ci-dessous affirme que « des fonctions de variables aléatoires indépendantes sont indépendantes ». Il est d'usage constant, car il permet souvent de démontrer sans calculs que des variables aléatoires sont indépendantes.

3. Pour éviter un problème technique délicat. Le lecteur intéressé pourra se reporter au tome II, ch 8.

Proposition 3.6. Soit $(X_i)_{i \in I}$ une famille quelconque de variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) à valeurs respectivement dans des espaces probabilisables (E_i, \mathcal{E}_i) . On suppose que ces variables aléatoires sont **indépendantes**. Soit, pour tout $i \in I$, une variable aléatoire f_i définie sur l'espace probabilisable (E_i, \mathcal{E}_i) à valeurs dans un espace probabilisable (F_i, \mathcal{F}_i) .

Alors les variables aléatoires $f_i(X_i)$ sont encore indépendantes.

Il suffit de remarquer que, pour tout $i \in I$ et tout $B_i \in \mathcal{F}_i$, on a

$$(f_i(X_i) \in B_i) = (f_i \circ X_i)^{-1}(B_i) = X_i^{-1}(f_i^{-1}(B_i)) = (X_i \in f_i^{-1}(B_i)),$$

et que $f_i^{-1}(B_i) \in \mathcal{E}_i$. L'indépendance des variables aléatoires X_i assure alors celle des variables aléatoires $f_i(X_i)$.

Corollaire 3.7. Soient X_1, \dots, X_n des variables aléatoires discrètes définies sur un même espace probabilisé (Ω, \mathcal{A}, P) , à valeurs respectivement dans E_1, \dots, E_n , et soient $f_1 : E_1 \times \dots \times E_k \rightarrow F_1$, $f_2 : E_{k+1} \times \dots \times E_n \rightarrow F_2$ des applications quelconques. Si les variables aléatoires X_1, \dots, X_n sont indépendantes, alors les variables aléatoires $f_1(X_1, \dots, X_k)$ et $f_2(X_{k+1}, \dots, X_n)$ sont indépendantes.

Démonstration. Cela résulte aussitôt de la proposition précédente.

Exemple 3.3. On peut affirmer sans calcul que si X_1, X_2, X_3 et X_4 sont quatre variables aléatoires à valeurs dans \mathbb{Z} , les variables aléatoires $X_1 + X_2$ et $X_3 + X_4$ sont indépendantes, de même que les variables aléatoires $X_1 + X_4$ et $X_2 + X_3$ (il suffit d'appliquer le corollaire en « renumérotant » les X_i), ou encore les variables $(-1)^{X_2} \sin \pi X_3$ et $X_1^5 + X_4^{1000}$.

3.3. Loi d'une somme de variables aléatoires indépendantes

On a très souvent besoin de calculer la loi d'une somme de variables aléatoires. Quand les variables aléatoires X_1 et X_2 sont indépendantes, la loi de leur somme peut se calculer à partir des lois de X_1 et de X_2 . Nous donnons la méthode de calcul, dite de **convolution**, dans le cas de variables aléatoires discrètes.

Proposition 3.8. Soient X_1 et X_2 deux variables aléatoires discrètes et indépendantes définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , à valeurs dans le même espace probabilisable (E, \mathcal{E}) , où E est une partie d'un

groupe abélien stable par addition (en-dehors des cas où E est lui-même un groupe comme \mathbb{Z} , \mathbb{Z}^d , \mathbb{R} , etc., les cas les plus typiques sont \mathbb{N} et \mathbb{R}^+).

Alors, la variable aléatoire $X_1 + X_2$ est discrète et sa loi est donnée par l'une quelconque des deux relations :

$$\forall x \in E \quad P(X_1 + X_2 = x) = \sum_{x_1 \in E} P(X_1 = x_1) P(X_2 = x - x_1) \quad (3.6)$$

ou

$$\forall x \in E \quad P(X_1 + X_2 = x) = \sum_{x_2 \in E} P(X_1 = x - x_2) P(X_2 = x_2). \quad (3.7)$$

Démonstration. Nous avons affirmé plus haut (sans le démontrer) qu'une fonction quelconque $f(X_1, \dots, X_n)$ de variables aléatoires discrètes est une variable aléatoire discrète. Voici une démonstration dans le cas de $X_1 + X_2$.

L'ensemble $(X_1 + X_2)(\Omega)$ est dénombrable puisque c'est l'image de l'ensemble $X_1(\Omega) \times X_2(\Omega)$ (dénombrable comme produit de deux ensembles dénombrables) par l'application $(x_1, x_2) \mapsto x_1 + x_2$.

Par ailleurs, pour tout $x \in E$, on a

$$(X_1 + X_2 = x) = \bigcup_{x_1 \in X_1(\Omega)} (X_1 = x_1) \cap (X_2 = x - x_1), \quad (3.8)$$

ce qui montre que l'ensemble $(X_1 + X_2 = x)$ appartient bien à \mathcal{A} .

Ceci montré, on obtient immédiatement la relation (3.6) à partir de (3.8) en utilisant la σ -additivité de P et l'indépendance des variables aléatoires X_1 et X_2 . On obtient la deuxième relation de manière analogue.

Remarque. La relation (3.6) s'écrit aussi, en termes des lois des variables aléatoires,

$$\forall x \in E \quad P_{X_1 + X_2}(\{x\}) = \sum_{x_1 \in E} P_{X_1}(\{x_1\}) P_{X_2}(\{x - x_1\}). \quad (3.9)$$

Définition 3.9. On dit que la loi de probabilité discrète, obtenue à partir de P_{X_1} et P_{X_2} par la relation (3.9), est le produit de convolution, ou simplement la **convolution** des deux probabilités P_{X_1} et P_{X_2} . Elle est notée $P_{X_1} * P_{X_2}$.

Avec cette notation, on a donc $P_{X_1+X_2} = P_{X_1} * P_{X_2}$. La convolution est donc une opération commutative entre lois de probabilités.

Voici deux exemples de calcul de la loi d'une somme de variables aléatoires indépendantes.

Exemple 3.4. Loi triangulaire. Soit un entier $n \geq 1$. Soient deux variables aléatoires X et Y définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , de même loi uniforme sur l'ensemble $\{0, 1, 2, \dots, n\}$, et indépendantes. On étudie la loi de la somme $Z = X + Y$ et on en exhibe une propriété de symétrie.

Solution. Les variables aléatoires X et Y sont considérées comme prenant leurs valeurs dans \mathbb{N} , partie stable par addition du groupe \mathbb{Z} . Alors, pour tout $k \in \mathbb{N}$, on a :

$$P(Z = k) = \sum_{j \in \mathbb{N}} P(X = j) P(Y = k - j).$$

Mais on a

$$P(X = j) = P(Y = j) = \begin{cases} \frac{1}{n+1} & \text{si } j \in \{0, 1, 2, \dots, n\} \\ 0 & \text{sinon,} \end{cases}$$

ce qui donne :

– si $0 \leq k \leq n$,

$$P(Z = k) = \sum_{j=0}^k \left(\frac{1}{n+1}\right)^2 = \frac{k+1}{(n+1)^2},$$

– si $n < k \leq 2n$,

$$P(Z = k) = \sum_{j=k-n}^n \left(\frac{1}{n+1}\right)^2 = \frac{2n-k+1}{(n+1)^2}.$$

On vérifie facilement que

$$\sum_{k=0}^{2n} P(Z = k) = 1,$$

et donc que

$$P_Z(\{0, 1, 2, \dots, 2n\}) = 1.$$

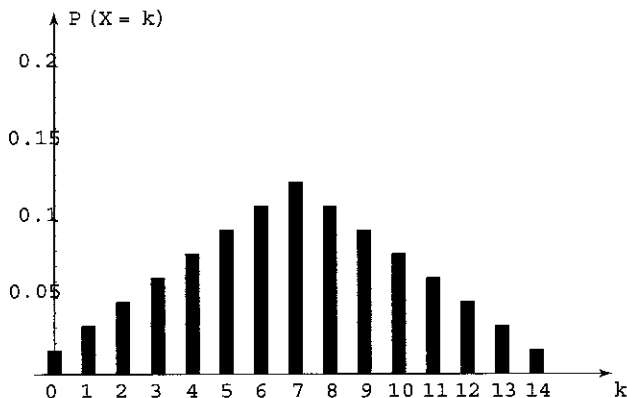
Si k et l sont **symétriques** par rapport à n , c'est-à-dire si $l = 2n - k$, on a :

$$P(Z = l) = \frac{2n - (2n - l) + 1}{(n+1)^2}$$

et donc

$$P(Z = l) = P(Z = k) = \frac{k+1}{(n+1)^2}.$$

On dit que la loi de Z est **symétrique** par rapport à n .

Figure 1. Loi triangulaire ($n = 7$)

Exemple 3.5. Somme de deux variables aléatoires de Poisson indépendantes. Soient deux variables aléatoires X_1 et X_2 définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , indépendantes, de loi de Poisson de paramètres respectifs λ_1 et λ_2 , tous deux strictement positifs. On étudie la loi de la somme $Z = X_1 + X_2$.

Solution. Les variables aléatoires X_1 et X_2 sont à valeurs dans \mathbb{N} ; il en est de même pour la variable aléatoire Z . La relation de convolution s'écrit :

$$\forall k \in \mathbb{N} \quad P(Z = k) = \sum_{j \in \mathbb{N}} P(X_1 = j) P(X_2 = k - j).$$

Puisque, pour $i = 1, 2$, et pour tout $j \in \mathbb{N}$, on a

$$P(X_i = j) = \exp(-\lambda_i) \frac{\lambda_i^j}{j!},$$

il vient, pour tout $k \in \mathbb{N}$,

$$\begin{aligned} P(Z = k) &= \sum_{j=0}^k \left[\left(\exp(-\lambda_1) \frac{\lambda_1^j}{j!} \right) \left(\exp(-\lambda_2) \frac{\lambda_2^{k-j}}{(k-j)!} \right) \right] \\ &= \frac{\exp(-(\lambda_1 + \lambda_2))}{k!} \sum_{j=0}^k \binom{k}{j} \lambda_1^j \lambda_2^{k-j}, \end{aligned}$$

soit encore :

$$P(Z = k) = \frac{\exp(-(\lambda_1 + \lambda_2))}{k!} (\lambda_1 + \lambda_2)^k.$$

On a ainsi démontré la proposition suivante :

Proposition 3.10. *La convolution de deux lois de Poisson $\mathcal{P}(\lambda_1)$ et $\mathcal{P}(\lambda_2)$ de paramètres positifs λ_1 et λ_2 est une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.*

On dit que la famille des lois de Poisson est **stable** par convolution.

3.4. Indépendance et produits cartésiens : construction d'un modèle

Revenons à l'exemple d'introduction, à savoir la **modélisation** de deux jets consécutifs d'un même dé; nous avons été conduits « naturellement » à prendre comme espace fondamental, ou espace des réalisations, le produit cartésien $\Omega = \Omega_1 \times \Omega_2$, où $\Omega_1 = \Omega_2 = \{1, 2, \dots, 6\}$, et à prendre pour probabilité P sur l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$ la probabilité **uniforme**. Nous avons alors montré que les variables aléatoires projections X_1 et X_2 , qui **séparément** traduisent tout ce que l'on peut dire sur chacun des jets, sont des variables aléatoires **indépendantes** de **loi uniforme** sur $\{1, 2, \dots, 6\}$. En fait, si l'on suit le **déroulement** du jeu, ce sont plutôt ces deux dernières propriétés (indépendance, lois uniformes) qui sont « naturelles », et en abordant ainsi la modélisation, on aurait été amené à se poser le problème « inverse » suivant, à savoir : **construire** un espace probabilisé (Ω, \mathcal{A}, P) sur lequel on puisse définir deux variables aléatoires indépendantes de loi uniforme sur $\{1, 2, \dots, 6\}$. Comme nous allons le voir de manière plus générale ci-dessous, **une** réponse à ce problème est l'espace probabilisé que nous avons qualifié de « naturel ».

On se pose le problème général suivant : étant donné une famille (E_i, \mathcal{E}_i) ($i \in I$) d'espaces probabilisables, et une loi de probabilité **donnée** P_i sur chacun de ces espaces, peut-on construire un espace probabilisé (Ω, \mathcal{A}, P) et une famille de variables aléatoires $(X_i)_{i \in I}$ définies sur cet espace, à valeurs respectivement dans (E_i, \mathcal{E}_i) , **indépendantes**, et telles que, pour tout $i \in I$, la variable aléatoire X_i soit de **loi** P_i ?

C'est un problème fondamental pour la construction de modèles. Nous ne l'aborderons ici que dans un cadre restreint, celui où I est **fini** et où les ensembles E_i sont **dénombrables**. Dans ce cadre, une solution est fournie par la proposition suivante.

Proposition 3.11. *Soit $I = \{1, 2, \dots, n\}$; pour tout $i \in I$, soit E_i un ensemble **dénombrable** et P_i une probabilité sur l'espace probabilisable $(E_i, \mathcal{P}(E_i))$ engendrée par le germe de probabilité p_i . On considère le produit cartésien $\Omega = \prod_{i=1}^n E_i$ et la projection X_i de Ω sur le facteur*

E_i . La fonction p définie sur Ω par la relation

$$\forall (\omega_1, \omega_2, \dots, \omega_n) \in \Omega \quad p(\omega_1, \omega_2, \dots, \omega_n) = \prod_{i=1}^n p_i(\omega_i)$$

est le germe d'une probabilité P sur $(\Omega, \mathcal{P}(\Omega))$, appelée **probabilité produit** des probabilités P_i . L'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ est tel que les variables aléatoires X_i sont indépendantes⁴ et de lois respectives P_i .

Démonstration. On a, par la propriété de Fubini,

$$\sum_{(\omega_1, \dots, \omega_n) \in \Omega} p(\omega_1, \dots, \omega_n) = \prod_{i=1}^n \left(\sum_{\omega_i \in E_i} p_i(\omega_i) \right) = 1,$$

ce qui démontre que la fonction *positive* p définit un germe de probabilité.

Par ailleurs, pour tout $i_0 \in I$ et pour tout $\omega_{i_0} \in E_{i_0}$, on a

$$(X_{i_0} = \omega_{i_0}) = \bigsqcup_{\substack{(x_1, \dots, x_n) \in \Omega \\ x_{i_0} = \omega_{i_0}}} \{(x_1, \dots, x_n)\},$$

et donc, toujours par la propriété de Fubini,

$$\begin{aligned} P(X_{i_0} = \omega_{i_0}) &= \sum_{\substack{(x_1, \dots, x_n) \in \Omega \\ x_{i_0} = \omega_{i_0}}} p(x_1, \dots, x_n) \\ &= p_{i_0}(\omega_{i_0}) \prod_{\substack{i \in I \\ i \neq i_0}} \left(\sum_{x_i \in E_i} p_i(x_i) \right); \end{aligned}$$

ceci démontre que, pour tout $\omega_{i_0} \in E_{i_0}$, on a

$$P(X_{i_0} = \omega_{i_0}) = p_{i_0}(\omega_{i_0}),$$

c'est-à-dire que la loi de la variable aléatoire X_{i_0} est P_{i_0} . De plus, d'après la définition de P , on peut écrire

$$\forall (\omega_1, \omega_2, \dots, \omega_n) \in \Omega \quad P \left[\bigcap_{i=1}^n (X_i = \omega_i) \right] = \prod_{i=1}^n P(X_i = \omega_i),$$

ce qui démontre l'indépendance des variables aléatoires X_i .

Si l'on veut modéliser un jeu de pile ou face et étudier des propriétés qui imposent de ne pas choisir le nombre de jets *a priori* (par exemple :

4. On dit aussi que les variables aléatoires sont indépendantes pour la probabilité P .

si on arrête de jouer lorsque l'on a obtenu dix fois « pile », on est conduit à prendre comme ensemble des réalisations l'ensemble $\{0, 1\}^{\mathbb{N}^*}$ des suites infinies à valeurs 0 ou 1 (1 pour « pile », 0 pour « face »). Cet ensemble **n'est pas dénombrable** (il est mis en bijection, à peu de choses près, avec l'intervalle réel $[0, 1]$ au moyen du développement dyadique : $(x_n) \mapsto \sum_{n=1}^{\infty} \frac{x_n}{2^n}$). Nous sortons alors immédiatement du cadre des espaces probabilisés *discrets*. Nous reviendrons plus tard sur cette question⁵; la résolution du problème de modélisation ainsi posé nécessite, comme nous le verrons, l'utilisation de la théorie de la mesure. En attendant **nous admettrons** que la proposition précédente se généralise au cas où $I = \mathbb{N}$ de la manière suivante :

Proposition 3.12. *Soient, pour tout $i \in \mathbb{N}$, E_i un ensemble dénombrable et P_i une probabilité sur l'espace probabilisable $(E_i, \mathcal{P}(E_i))$ engendrée par le germe de probabilité p_i .*

On considère le produit cartésien $\Omega = \prod_{i \in \mathbb{N}} E_i$ et, pour tout $i \in \mathbb{N}$, on note X_i la projection de Ω sur le i -ième facteur E_i . Il existe une tribu \mathcal{A} sur Ω et une probabilité P sur l'espace probabilisable (Ω, \mathcal{A}) telles que les variables aléatoires X_i soient indépendantes et de lois respectives P_i .

3.5. Modèles géométriques et binomiaux

Soit (Ω, \mathcal{A}, P) un espace probabilisé et $(A_n)_{n \in \mathbb{N}}$ une suite d'événements **indépendants de même probabilité** p ($0 < p < 1$). On définit les variables aléatoires N et N' à valeurs dans $\overline{\mathbb{N}}$ par :

$$\forall \omega \in \Omega \quad \begin{cases} N(\omega) = \inf (n \in \mathbb{N} \mid \omega \in A_n) \\ N'(\omega) = \inf (n \in \mathbb{N}^* \mid \omega \in A_n), \end{cases}$$

où on fait la convention : $\inf \emptyset = +\infty$.

Pour tout entier $n \geq 1$, on définit la variable aléatoire S_n à valeurs dans \mathbb{N} par :

$$S_n = \sum_{i=1}^n \mathbf{1}_{A_i}, \quad (3.10)$$

où $\mathbf{1}_{A_i}$ est la **fonction indicatrice** de l'événement A_i .

La donnée de (Ω, \mathcal{A}, P) et de $(A_n)_{n \in \mathbb{N}}$ constitue un modèle de la situation suivante : on répète indéfiniment la « même » expérience aléatoire (on peut lancer une infinité de fois le même dé, par exemple, ou lancer à chaque fois un nouveau dé, possédant des propriétés identiques

5. Voir tome II, ch. 9.

à celles du premier – le modèle est le même), et on s'intéresse à chaque fois à la réalisation du « même » événement.

L'entier $N(\omega)$ représente le **numéro** de l'expérience au cours de laquelle l'événement se réalise pour la **première** fois, si on commence la suite d'expériences avec l'expérience numérotée 0. L'entier $N'(\omega)$ représente le **numéro** de l'expérience au cours de laquelle l'événement se réalise pour la **première** fois, lorsque l'on commence avec l'expérience numérotée 1. L'entier $S_n(\omega)$ représente le **nombre de fois où l'événement considéré se réalise** au cours des expériences numérotées de 1 à n .

Nous allons ci-dessous déterminer les lois des variables aléatoire N , N' et S_n (pour $n \geq 1$). Commençons par une définition.

Définition 3.13. On appelle **variable de Bernoulli**⁶ toute variable aléatoire X à valeurs entières dont la loi est déterminée par

$$P(X = 1) = p \quad \text{et} \quad P(X = 0) = 1 - p,$$

(on suppose généralement que $0 < p < 1$) et on pose classiquement $q = 1 - p$. On parle alors d'une variable de Bernoulli de **paramètre** p .

Remarque. Une variable de Bernoulli, n'est donc rien d'autre qu'une variable ne prenant que les valeurs 0 et 1 (avec probabilité 1). La fonction indicatrice $\mathbf{1}_A$ d'un événement A est une variable de Bernoulli et, inversement, une variable de Bernoulli X ne diffère de la fonction indicatrice d'un événement (à savoir l'événement $(X = 1)$) que sur un ensemble de probabilité nulle.

La variable aléatoire S_n définie en (3.10) est une **somme** de n variables aléatoires de Bernoulli **indépendantes**.

Proposition 3.14. La loi de la variable aléatoire N est la loi géométrique sur \mathbb{N} , notée $\mathcal{G}_{\mathbb{N}}(p)$.

La loi de la variable aléatoire N' est la loi géométrique sur \mathbb{N}^* , notée $\mathcal{G}_{\mathbb{N}^*}(p)$.

De plus, la loi de N' est la même que celle de la variable aléatoire $N + 1$.

6. En mémoire de Jacques Bernoulli (Bâle 1654-1705), premier d'une « dynastie » de mathématiciens et physiciens. Ses travaux portent en particulier sur le calcul infinitésimal, les équations différentielles et le problème des isopérimètres ainsi que sur le calcul des probabilités.

Démonstration. On a

$$(N = 0) = A_0,$$

et

$$\forall k \geq 1 \quad (N = k) = \left(\bigcap_{j=0}^{k-1} A_j^c \right) \cap A_k.$$

Il résulte de l'indépendance des événements A_i , $i \in \mathbb{N}$, que les événements A_j^c , $0 \leq j \leq k-1$, et A_k sont indépendants; on a donc :

$$\forall k \in \mathbb{N} \quad P(N = k) = p(1-p)^k.$$

De plus :

$$P(N = +\infty) = 1 - P(N \in \mathbb{N}) = 1 - \sum_{k=0}^{+\infty} p(1-p)^k,$$

et par conséquent $P(N = +\infty) = 0$. On a bien démontré que $P_{\mathbb{N}} = \mathcal{G}_{\mathbb{N}}(p)$.

On procède de même pour le calcul de la loi de la variable aléatoire N' en remarquant que

$$(N' = 1) = A_1$$

et que

$$\forall k \geq 2 \quad (N = k) = \left(\bigcap_{j=1}^{k-1} A_j^c \right) \cap A_k.$$

On termine par un calcul analogue.

Étudions la loi $P_{\mathbb{N}^*}$ de la variable aléatoire $N'' = N + 1$. On a, pour tout $k \in \mathbb{N}^*$,

$$P(N'' = k) = P(N = k-1) = p(1-p)^{k-1},$$

ce qui montre que $P_{\mathbb{N}''} = \mathcal{G}_{\mathbb{N}^*}(p)$.

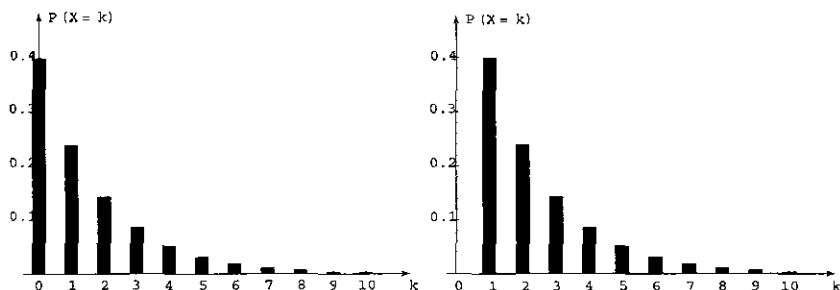


Figure 2. Loïs géométriques sur \mathbb{N} et \mathbb{N}^* ($p = 0,4$)

Proposition 3.15. *La loi de la variable aléatoire S_n est la loi binomiale $\mathcal{B}(n, p)$.*

Démonstration. Remarquons d'abord que $S_n(\Omega) \subset \{0, 1, 2, \dots, n\}$, et que, pour tout $k \in \{0, 1, 2, \dots, n\}$, on a :

$$(S_n = k) = \bigcup_{\substack{I \in \mathcal{P}(\{1, 2, \dots, n\}) \\ |I|=k}} \left[\left(\bigcap_{i \in I} A_i \right) \cap \left(\bigcap_{i \in \{1, 2, \dots, n\} \setminus I} A_i^c \right) \right]$$

(il faut bien remarquer qu'il s'agit d'une union d'ensembles **disjoints**). En tenant compte de l'additivité de P , de l'indépendance des événements A_i (utiliser la proposition 3.2) et du fait que tous les événements A_i ont même probabilité p (et leurs complémentaires $1 - p$), il vient

$$P(S_n = k) = \sum_{\substack{I \in \mathcal{P}(\{1, 2, \dots, n\}) \\ |I|=k}} p^k (1 - p)^{n-k}.$$

On a donc montré que, pour tout $k \in \{0, 1, 2, \dots, n\}$,

$$P(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

ce qui est bien le résultat annoncé.

Proposition 3.16. *La convolution des deux lois binomiales $\mathcal{B}(n_1, p)$ et $\mathcal{B}(n_2, p)$ est la loi binomiale $\mathcal{B}(n_1 + n_2, p)$.*

Remarque. Autrement dit, la famille des lois binomiales de second paramètre p est stable par convolution.

Démonstration. Considérons la variable aléatoire

$$S_{n_1+n_2} = S_{n_1} + \sum_{i=n_1+1}^{n_1+n_2} \mathbf{1}_{A_i}.$$

Les variables aléatoires S_{n_1} et $\sum_{i=n_1+1}^{n_1+n_2} \mathbf{1}_{A_i}$ sont indépendantes et de loi respectives $\mathcal{B}(n_1, p)$ et $\mathcal{B}(n_2, p)$; donc, par définition de la convolution, la convolution des lois $\mathcal{B}(n_1, p)$ et $\mathcal{B}(n_2, p)$ est la loi de $S_{n_1+n_2}$. Mais, puisqu'on a

$$S_{n_1+n_2} = \sum_{i=1}^{n_1+n_2} \mathbf{1}_{A_i},$$

il résulte de la proposition précédente que la loi de $S_{n_1+n_2}$ est la loi binomiale $\mathcal{B}(n_1 + n_2, p)$, ce qui démontre le résultat.

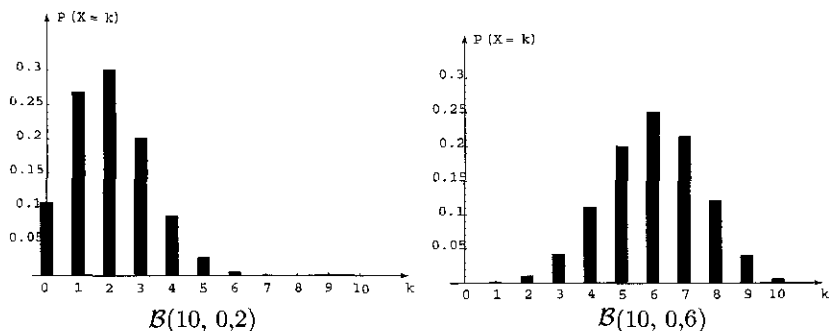


Figure 3. Exemples de lois binomiales

Exemple 3.6. Un jeu. Un joueur jette une pièce a priori non équilibrée (ou note p la probabilité d'obtenir « pile » lors d'un jet de la pièce) jusqu'à ce qu'il obtienne « pile »; si ceci se passe à la suite du k -ième jet, il lance k fois de suite un dé bien équilibré. Il gagne s'il obtient **exactement** un 6. On demande la probabilité pour que le joueur gagne à ce jeu.

Solution. Pour modéliser ce jeu, on considère un espace probabilisé $(\Omega, \mathcal{A}, P_p)$ sur lequel est définie pour tout entier $k \in \mathbb{N}^*$ une variable aléatoire X_k de loi binomiale $\mathcal{B}\left(k, \frac{1}{6}\right)$; celle-ci est destinée à représenter le nombre de 6 obtenus au cours d'une suite de k jets du dé. On considère de plus une variable aléatoire R de loi géométrique de paramètre p sur \mathbb{N}^* ; elle représente le rang du jet de la pièce où « pile » apparaît pour la première fois. On suppose que les variables aléatoires R et X_k , $k \in \mathbb{N}^*$ sont indépendantes (la proposition

3.12 assure l'existence d'un tel modèle). L'événement G « le joueur gagne » s'écrit :

$$G = \bigcup_{k=1}^{+\infty} (R = k, X_k = 1).$$

Alors, par σ -additivité de P_p et indépendance des événements $(R = k)$ et $(X_k = 1)$, on a, en posant $q = 1 - p$:

$$\begin{aligned} P_p(G) &= \sum_{k=1}^{+\infty} P_p(R = k) P_p(X_k = 1) \\ &= \sum_{k=1}^{+\infty} pq^{k-1} \binom{k}{1} \frac{1}{6} \left(\frac{5}{6}\right)^{k-1} \\ &= \frac{p}{6} \sum_{k=1}^{+\infty} k \left(\frac{5q}{6}\right)^{k-1}. \end{aligned}$$

On observe que $0 < \frac{5q}{6} < 1$; les résultats classiques sur la dérivation d'une série entière dans son domaine (réel) de convergence permettent d'écrire que, pour tout x tel que $0 < x < 1$, on a :

$$\begin{aligned} \sum_{k=1}^{+\infty} kx^{k-1} &= \frac{d}{dx} \left(\sum_{k=0}^{+\infty} x^k \right) \\ &= \frac{d}{dx} \left(\frac{1}{1-x} \right) \\ &= \frac{1}{(1-x)^2}. \end{aligned}$$

Un calcul simple conduit alors à :

$$P_p(G) = \frac{6p}{(1+5p)^2}.$$

On remarque que :

$$\max \{P_p(G) \mid p \in]0, 1[\} = 0,3.$$

Ce maximum est atteint pour $p = \frac{1}{5}$; il est strictement plus grand que $\frac{1}{6} \simeq 0,166$.

Exercices

Les probabilités peuvent fournir des méthodes alternatives de démonstration dans différents domaines des mathématiques. En voici un exemple en arithmétique.

Exercice 3.1. Calcul de l'indicateur d'Euler. Soit Ω l'ensemble des entiers $\{1, 2, \dots, n\}$, où n est un entier non premier supérieur ou égal à 2. On considère l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ où P est la probabilité uniforme. Si d divise n on note

$$A_d = \{kd \mid k \in \Omega \text{ et } kd \in \Omega\}.$$

1. Quelle est la probabilité de A_d ?
2. Soit $p_1 < p_2 < \dots < p_r$ la suite des diviseurs premiers de n , rangés par ordre croissant.

(a) Démontrer que $(A_{p_i})_{1 \leq i \leq r}$ est une famille d'événements indépendants.

(b) En déduire le cardinal $\varphi(n)$ de l'ensemble A des nombres inférieurs ou égaux à n et premiers avec n (indicateur d'Euler).

Solution.

1. Soit k_n l'entier tel que : $n = k_n d$. On a $|A_d| = k_n$ et donc, P étant la probabilité uniforme :

$$P(A_d) = \frac{k_n}{n} = \frac{1}{d}.$$

2. On démontre l'indépendance des événements A_{p_i} , puis on utilise le fait que leurs complémentaires sont aussi indépendants.

(a) Les entiers p_i et p_j étant premiers (et distincts) si $i \neq j$, un multiple commun de p_i et p_j est un multiple de $p_i p_j$. Pour $i \neq j$, on a donc

$$A_{p_i} \cap A_{p_j} = \{kp_i p_j \mid k \in \Omega \text{ et } kp_i p_j \in \Omega\} = A_{p_i p_j}.$$

En particulier, l'entier n est un multiple de $p_i p_j$. On a donc :

$$P(A_{p_i} \cap A_{p_j}) = P(A_{p_i p_j}) = \frac{1}{p_i p_j} = P(A_{p_i}) P(A_{p_j}).$$

La même démonstration permet d'établir que, pour une partie non vide quelconque J de $\{1, 2, \dots, r\}$, on a :

$$P\left(\bigcap_{i \in J} A_{p_i}\right) = \prod_{i \in J} P(A_{p_i}),$$

ce qui démontre que les événements A_{p_i} , $1 \leq i \leq r$, sont indépendants.

(b) Les événements $A_{p_i}^c$, $1 \leq i \leq r$, sont donc aussi indépendants. Mais, puisque $\bigcap_{i=1}^r A_i^c$ est l'ensemble A , on a :

$$\frac{\varphi(n)}{n} = \prod_{i=1}^r \left(1 - \frac{1}{p_i}\right),$$

ce qui donne :

$$\varphi(n) = n \prod_{i=1}^r \left(1 - \frac{1}{p_i}\right).$$

Exercice 3.2. Attention aux idées intuitives rapides sur l'indépendance! Soit (Ω, \mathcal{A}, P) un espace probabilisé sur lequel sont définies les variables aléatoires indépendantes U et V , à valeurs dans $\{-1, 1\}$ et de même loi définie par les relations

$$P_U(-1) = \frac{1}{3} \quad \text{et} \quad P_U(1) = \frac{2}{3}.$$

Soient X et Y les variables aléatoires définies par :

$$X = U \quad Y = \text{sign}(U) V.$$

1. Quelle est la loi de la variable aléatoire (X, Y) ? Les variables aléatoires X et Y sont-elles indépendantes?
2. Les variables aléatoires X^2 et Y^2 sont-elles indépendantes?

Solution.

1. Pour tout $x \in \{-1, 1\}$ et tout $y \in \{-1, 1\}$, on a

$$(X = x, Y = y) = (U = x, V = \text{sign}(x) y)$$

et donc, par indépendance des variables aléatoires U et V :

$$P(X = x, Y = y) = P(U = x) P(V = \text{sign}(x) y)$$

soit :

$$\begin{aligned} P(X = 1, Y = 1) &= \frac{4}{9}, & P(X = 1, Y = -1) &= \frac{2}{9}, \\ P(X = -1, Y = 1) &= \frac{1}{9}, & P(X = -1, Y = -1) &= \frac{2}{9}. \end{aligned}$$

Par ailleurs, on a

$$P(X = 1) = \sum_{y \in \{-1, 1\}} P(X = 1, Y = y) = \frac{2}{3}.$$

On a de même

$$P(Y = 1) = \sum_{x \in \{-1, 1\}} P(X = x, Y = 1) = \frac{5}{9}.$$

On constate alors que l'on a

$$P(X = 1) P(Y = 1) = \frac{10}{27} \neq P(X = 1, Y = 1),$$

et donc que les variables aléatoires X et Y ne sont pas indépendantes.

2. Toutefois, on a

$$X^2 = U^2 \quad \text{et} \quad Y^2 = V^2,$$

ce qui implique que les variables aléatoires X^2 et Y^2 sont indépendantes, puisque les variables aléatoires U^2 et V^2 le sont comme fonctions de variables aléatoires indépendantes.

Exercice 3.3. La somme de deux variables aléatoires de loi de Poisson peut être de loi de Poisson sans que ces variables aléatoires soient

indépendantes. On note Ω le produit cartésien $\mathbb{N} \times \mathbb{N}$ et Λ_1 et Λ_2 les parties de Ω définies par :

$$\Lambda_1 = \{(0, 1), (1, 2), (2, 0)\} \quad \Lambda_2 = \{(0, 2), (2, 1), (1, 0)\}.$$

Soient q et r deux réels strictement positifs. On définit, pour tout $i \in \mathbb{N}$, les réels q_i et r_i en posant

$$q_i = \exp(-q) \frac{q^i}{i!} \quad \text{et} \quad r_i = \exp(-r) \frac{r^i}{i!}.$$

On se donne un réel ε tel que :

$$0 < \varepsilon < \min \{q, r_j \mid (i, j) \in \Lambda_1 \cup \Lambda_2\}.$$

Soient μ la fonction définie sur Ω par

$$\forall (i, j) \in \Omega \quad \mu(i, j) = \begin{cases} q_i r_j + \varepsilon & \text{si } (i, j) \in \Lambda_1 \\ q_i r_j - \varepsilon & \text{si } (i, j) \in \Lambda_2 \\ q_i r_j & \text{sinon} \end{cases}$$

et P la fonction définie sur l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$ par :

$$\forall A \in \mathcal{P}(\Omega) \quad P(A) = \sum_{(i, j) \in A} \mu(i, j).$$

1. Démontrer que P est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$.

On note X_1 et X_2 les projections canoniques de Ω sur \mathbb{N} , définies en posant

$$X_1(i, j) = i \quad X_2(i, j) = j$$

pour tout $(i, j) \in \Omega$.

2. Calculer la loi de la somme des variables aléatoires X_1 et X_2 et l'identifier à une loi connue.

3. Quelles sont les lois des variables aléatoires X_1 et X_2 ?

4. Les variables aléatoires X_1 et X_2 sont-elles indépendantes ?

Solution.

1. Par le choix de ε , pour tout $(i, j) \in \Omega$, on a $\mu(i, j) > 0$. Notons brièvement $\mu_{ij} = \mu(i, j)$. Par additivité des familles positives, on a

$$\sum_{(i, j) \in \Omega} \mu_{ij} = \sum_{(i, j) \in \Lambda_1} \mu_{ij} + \sum_{(i, j) \in \Lambda_2} \mu_{ij} + \sum_{(i, j) \in (\Lambda_1 \cup \Lambda_2)^c} \mu_{ij},$$

soit encore

$$\sum_{(i, j) \in \Omega} \mu_{ij} = \left[\sum_{(i, j) \in \Lambda_1} q_i r_j + 3\varepsilon \right] + \left[\sum_{(i, j) \in \Lambda_2} q_i r_j - 3\varepsilon \right] + \sum_{(i, j) \in (\Lambda_1 \cup \Lambda_2)^c} q_i r_j,$$

et donc

$$\sum_{(i, j) \in \Omega} \mu_{ij} = \sum_{(i, j) \in \mathbb{N} \times \mathbb{N}} q_i r_j.$$

Il résulte alors de la propriété de Fubini que l'on a

$$\sum_{(i,j) \in \Omega} \mu_{ij} = \left(\sum_{i \in \mathbb{N}} q_i \right) \left(\sum_{j \in \mathbb{N}} r_j \right).$$

Mais, puisque l'on a

$$\sum_{i \in \mathbb{N}} q_i = 1 \quad \text{et} \quad \sum_{j \in \mathbb{N}} r_j = 1,$$

il vient

$$\sum_{(i,j) \in \Omega} \mu_{ij} = 1.$$

La fonction μ est donc bien un germe de probabilité; elle engendre la probabilité P .

2. Soit $k \in \mathbb{N}$. On a

$$(X_1 + X_2 = k) = \bigsqcup_{i=0}^k (X_1 = i, X_2 = k - i).$$

La variable aléatoire (X_1, X_2) étant l'application identique sur Ω , sa loi est la probabilité P elle-même. Par conséquent, on a

$$\begin{aligned} P(X_1 + X_2 = k) &= \sum_{0 \leq i \leq k} \mu_{i, k-i} \\ &= \sum_{i \in \Lambda_{1,k}} \mu_{i, k-i} + \sum_{i \in \Lambda_{2,k}} \mu_{i, k-i} + \sum_{i \in \Lambda_{3,k}} \mu_{i, k-i}. \end{aligned}$$

où les ensembles $\Lambda_{1,k}$, $\Lambda_{2,k}$ et $\Lambda_{3,k}$ sont définis par

$$\begin{aligned} \Lambda_{1,k} &= \{i \in \mathbb{N} \mid 0 \leq i \leq k \text{ et } (i, k-i) \in \Lambda_1\} \\ \Lambda_{2,k} &= \{i \in \mathbb{N} \mid 0 \leq i \leq k \text{ et } (i, k-i) \in \Lambda_2\} \\ \Lambda_{3,k} &= \{i \in \mathbb{N} \mid 0 \leq i \leq k\} \setminus (\Lambda_{1,k} \cup \Lambda_{2,k}). \end{aligned}$$

En fait, on a

$$\Lambda_{1,k} = \begin{cases} \{0\} & \text{si } k = 1, \\ \{2\} & \text{si } k = 2, \\ \{1\} & \text{si } k = 3, \\ \emptyset & \text{sinon,} \end{cases}$$

et

$$\Lambda_{2,k} = \begin{cases} \{1\} & \text{si } k = 1, \\ \{0\} & \text{si } k = 2, \\ \{2\} & \text{si } k = 3, \\ \emptyset & \text{sinon.} \end{cases}$$

Il en résulte que

$$|\Lambda_{1,k}| = |\Lambda_{2,k}| = \begin{cases} 1 & \text{si } k = 1, 2, \text{ ou } 3, \\ 0 & \text{sinon.} \end{cases}$$

Il vient alors

$$P(X_1 + X_2 = k) = \left[\sum_{i \in \Lambda_{1,k}} q_i r_{k-i} + \varepsilon | \Lambda_{1,k} | \right] + \left[\sum_{i \in \Lambda_{2,k}} q_i r_{k-i} - \varepsilon | \Lambda_{2,k} | \right] \\ + \sum_{i \in \Lambda_{3,k}} q_i r_{k-i},$$

ce qui donne

$$P(X_1 + X_2 = k) = \sum_{0 \leq i \leq k} q_i r_{k-i} \\ = \exp(-(q+r)) \sum_{i=0}^k \frac{q^i r^{k-i}}{i!(k-i)!} \\ = \exp(-(q+r)) \frac{1}{k!} \sum_{i=0}^k \binom{k}{i} q^i r^{k-i},$$

soit encore

$$P(X_1 + X_2 = k) = \exp(-(q+r)) \frac{(q+r)^k}{k!}.$$

La loi de la variable aléatoire $X_1 + X_2$ est donc la loi de Poisson de paramètre $q+r$.

3. Pour tout $k \in \mathbb{N}$, on a

$$(X_1 = k) = \bigsqcup_{j \in \mathbb{N}} \{(k, j)\}$$

et donc

$$P(X_1 = k) = \sum_{j \in \mathbb{N}} \mu_{k,j}. \quad (3.11)$$

On explicite cette formule suivant les valeurs de k :

(a) si $k = 0$, on a

$$P(X_1 = 0) = \mu_{0,0} + \mu_{0,1} + \mu_{0,2} + \sum_{j=3}^{+\infty} \mu_{0,j} \\ = q_0 r_0 + (q_0 r_1 + \varepsilon) + (q_0 r_2 - \varepsilon) + \sum_{j=3}^{+\infty} q_0 r_j = q_0 \left(\sum_{j=0}^{+\infty} r_j \right),$$

soit

$$P(X_1 = 0) = q_0;$$

(b) si $k = 1$, on a

$$P(X_1 = 1) = \mu_{1,0} + \mu_{1,2} + \mu_{1,1} + \sum_{j=3}^{+\infty} \mu_{1,j} \\ = (q_1 r_0 - \varepsilon) + (q_1 r_2 + \varepsilon) + q_1 r_1 + \sum_{j=3}^{+\infty} q_1 r_j$$

$$= q_1 \left(\sum_{j=0}^{+\infty} r_j \right),$$

soit :

$$\boxed{P(X_1 = 1) = q_1 ;}$$

(c) si $k = 2$, on a :

$$\begin{aligned} P(X_1 = 2) &= \mu_{2,0} + \mu_{2,1} + \sum_{j=2}^{+\infty} \mu_{2,j} \\ &= (q_2 r_0 - \varepsilon) + (q_2 r_1 + \varepsilon) + \sum_{j=2}^{+\infty} q_2 r_j \\ &= q_2 \left(\sum_{j=0}^{+\infty} r_j \right), \end{aligned}$$

soit :

$$\boxed{P(X_1 = 2) = q_2 ;}$$

(d) si $k \geq 3$, on a :

$$P(X_1 = k) = \sum_{j=0}^{+\infty} q_k r_j = q_k.$$

En résumé, on a, pour tout $k \in \mathbb{N}$:

$$\boxed{P(X_1 = k) = q_k .}$$

On démontrerait de même que, pour tout $k \in \mathbb{N}$:

$$\boxed{P(X_2 = k) = r_k .}$$

On a ainsi démontré que les variables aléatoires X_1 et X_2 sont de lois de Poisson de paramètres respectifs q et r .

4. On a vu que, pour tout $(i, j) \in \mathbb{N} \times \mathbb{N}$, on a

$$P(X_1 = i, X_2 = j) = \mu_{ij} ;$$

en particulier on a :

$$P(X_1 = 1, X_2 = 0) = \mu_{10} \neq P(X_1 = 1) P(X_2 = 0),$$

ce qui démontre que les variables aléatoires X_1 et X_2 ne sont pas indépendantes !

Exercice 3.4. Loi binomiale négative. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , indépendantes, de même loi géométrique sur \mathbb{N}^* de paramètre p . Pour tout $n \in \mathbb{N}^*$, on définit la variable aléatoire S_n par :

$$S_n = \sum_{j=1}^n X_j .$$

1. Calculer la loi de la variable aléatoire S_2 .
2. Calculer par récurrence la loi de la variable aléatoire S_n .

Solution.

1. Les variables aléatoires X_n sont à valeurs dans \mathbb{N}^* et on a $S_2(\Omega) \subset \mathbb{N}^*$. Si $k \in \mathbb{N}^*$, il résulte de la méthode de convolution que l'on a :

$$P(S_2 = k) = \sum_{j=1}^k P(X_1 = j) P(X_2 = k - j).$$

En particulier, on a :

$$P(S_2 = 1) = 0.$$

Notons $q = 1 - p$; si $k \geq 2$, on a :

$$\begin{aligned} P(S_2 = k) &= \sum_{j=1}^{k-1} pq^{j-1} pq^{k-j-1} \\ &= (k-1) p^2 q^{k-2}, \end{aligned}$$

ce qui s'écrit, si $k \geq 2$:

$$P(S_2 = k) = \binom{k-1}{1} p^2 q^{k-2}.$$

2. On peut faire éventuellement encore le calcul de $P(S_3 = k)$ pour avoir l'intuition d'une formule de récurrence. Supposons que la loi de la variable aléatoire S_n soit donnée par :

$$P(S_n = k) = \begin{cases} 0 & \text{si } 1 \leq k \leq n-1 \\ \binom{k-1}{n-1} p^n q^{k-n} & \text{si } k \geq n. \end{cases} \quad (3.12)$$

Puisque $S_{n+1} = S_n + X_{n+1}$ et que les variables aléatoires S_n et X_{n+1} sont indépendantes, il résulte de la méthode de convolution que :

$$P(S_{n+1} = k) = \sum_{j=1}^k P(S_n = j) P(X_{n+1} = k - j),$$

ce qui donne, d'après l'hypothèse de récurrence :

$$P(S_{n+1} = k) = 0 \quad \text{si } 1 \leq k \leq n,$$

et, si $k \geq n+1$,

$$\begin{aligned} P(S_{n+1} = k) &= \sum_{j=n}^{k-1} \binom{j-1}{n-1} p^n q^{j-n} pq^{k-j-1} \\ &= \left(\sum_{j=n}^{k-1} \binom{j-1}{n-1} \right) p^{n+1} q^{k-(n+1)}. \end{aligned}$$

Reste à calculer le coefficient de $p^{n+1}q^{k-(n+1)}$; pour cela, formons le polynôme $Q(x)$ défini par :

$$Q(x) = \sum_{n=1}^{k-1} \left(\sum_{j=n}^{k-1} \binom{j-1}{n-1} \right) x^n.$$

Par permutation des sommes puis translation d'indices, on obtient successivement

$$\begin{aligned} Q(x) &= \sum_{j=1}^{k-1} \left(\sum_{n=1}^j \binom{j-1}{n-1} \right) x^n \\ &= \sum_{j=1}^{k-1} \left(\sum_{p=0}^{j-1} \binom{j-1}{p} \right) x^{p+1} \\ &= \sum_{j=1}^{k-1} x(1+x)^{j-1} \\ &= x \frac{1 - (1+x)^{k-1}}{1 - (1+x)} \\ &= (1+x)^{k-1} - 1. \end{aligned}$$

Il en résulte que

$$Q(x) = \sum_{j=1}^{k-1} \binom{k-1}{j} x^j.$$

En comparant avec la définition de $Q(x)$, on obtient

$$\sum_{j=n}^{k-1} \binom{j-1}{n-1} = \binom{k-1}{n},$$

ce qui démontre que la loi de S_n est donnée par la relation (3.12). Cette loi est appelée **loi binomiale négative** de paramètres n et p .

Exercice 3.5. Variables aléatoires de loi géométrique ; calculs de lois de maximum et minimum, lois de variables aléatoires bi-dimensionnelles, lois de marginales. Soient X et Y deux variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , indépendantes, de même loi géométrique sur \mathbb{N} de paramètre p .

1. Le calcul des probabilités suivantes est intéressant, par exemple, pour comparer les temps de succès de deux joueurs jouant simultanément et dans les mêmes conditions.

(a) Calculer $P(Y \geq X)$; étudier le cas particulier où $p = \frac{1}{2}$.

(b) Calculer $P(Y = X)$; étudier le cas particulier où $p = \frac{1}{2}$.

(c) Démontrer que $P(Y > X) = P(X > Y)$ et retrouver ainsi la probabilité $P(Y \geq X)$.

2. On définit les variables aléatoires U et V par :

$$U = \max(X, Y) \quad V = \min(X, Y).$$

(a) Calculer, pour tout $(u, v) \in \mathbb{N}^2$, la probabilité $P(U \leq u, V \geq v)$.

(b) En déduire les lois des variables aléatoires U et V .

(c) Identifier la loi de la variable aléatoire V .

3. On définit la variable aléatoire $W = U - V$; déterminer sa loi.

Solution.

1. On partitionne les événements étudiés.

(a) On a

$$(Y \geq X) = \bigcup_{\substack{(i,j) \in \mathbb{N}^2 \\ i \geq j}} (Y = i, X = j)$$

et, par σ -additivité de P :

$$P(Y \geq X) = \sum_{\substack{(i,j) \in \mathbb{N}^2 \\ i \geq j}} P(Y = i, X = j).$$

Les variables aléatoires X et Y étant indépendantes, de loi géométrique sur \mathbb{N} de paramètre p , on peut écrire successivement, en posant comme de coutume $q = 1 - p$:

$$\begin{aligned} P(Y \geq X) &= \sum_{\substack{(i,j) \in \mathbb{N}^2 \\ i \geq j}} pq^i pq^j \\ &= p^2 \sum_{i=0}^{+\infty} \left[q^i \left(\sum_{j=0}^i q^j \right) \right] \\ &= p^2 \sum_{i=0}^{+\infty} \left[q^i \frac{1 - q^{i+1}}{1 - q} \right] \\ &= \sum_{i=0}^{+\infty} pq^i - pq \sum_{i=0}^{+\infty} q^{2i} \\ &= 1 - \frac{pq}{1 - q^2}, \end{aligned}$$

soit encore :

$$\boxed{P(Y \geq X) = \frac{1}{1 + q}}.$$

En particulier, si $p = \frac{1}{2}$, on trouve $P(Y \geq X) = \frac{2}{3}$.

(b) On a

$$(Y = X) = \bigcup_{i \in \mathbb{N}} (Y = i, X = i)$$

et, par σ -additivité de P :

$$P(Y = X) = \sum_{i \in \mathbb{N}} P(Y = i, X = i).$$

Les variables aléatoires X et Y étant indépendantes, de même loi géométrique sur \mathbb{N} de paramètre p , on peut écrire :

$$\begin{aligned} P(Y = X) &= \sum_{i \in \mathbb{N}} p^2 q^{2i} \\ &= \frac{p^2}{1 - q^2}, \end{aligned}$$

soit encore :

$$\boxed{P(Y = X) = \frac{p}{1 + q}}.$$

En particulier, si $p = \frac{1}{2}$, on trouve $P(Y = X) = \frac{1}{3}$.

(c) On a

$$P(Y > X) = \sum_{\substack{(i,j) \in \mathbb{N}^2 \\ i > j}} P(Y = i) P(X = j)$$

et, les variables aléatoires X et Y ayant même loi :

$$\begin{aligned} P(Y > X) &= \sum_{\substack{(i,j) \in \mathbb{N}^2 \\ i > j}} P(X = i) P(Y = j) \\ &= P(X > Y). \end{aligned}$$

Il en résulte que

$$2 P(Y > X) + P(Y = X) = 1,$$

soit encore :

$$P(Y > X) = \frac{1}{2} [1 - P(Y = X)];$$

on obtient :

$$\boxed{P(Y > X) = \frac{q}{1 + q}}.$$

En particulier, si $p = \frac{1}{2}$, on a :

$$\boxed{P(Y > X) = P(Y < X) = P(Y = X) = \frac{1}{3}}.$$

On retrouve la probabilité $P(Y \geq X)$ de la manière suivante :

$$\begin{aligned} P(Y \geq X) &= P(Y > X) + P(Y = X) \\ &= \frac{q}{1 + q} + \frac{p}{1 + q} \\ &= \frac{1}{1 + q}. \end{aligned}$$

2. (a) \diamond Si $v > u \geq 0$, on a

$$(U \leq u) \cap (V \geq v) = \emptyset,$$

et donc :

$$P(U \leq u, V \geq v) = 0.$$

\diamond Si $0 \leq v \leq u$, on a

$$(U \leq u, V \geq v) = (v \leq X \leq u, v \leq Y \leq u).$$

Les variables aléatoires X et Y étant indépendantes, de même loi géométrique sur \mathbb{N} de paramètre p , on peut écrire

$$\begin{aligned} P(U \leq u, V \geq v) &= [P(v \leq X \leq u)]^2 \\ &= \left(\sum_{j=v}^u pq^j \right)^2, \end{aligned}$$

soit, après simplifications :

$$P(U \leq u, V \geq v) = \begin{cases} [q^v - q^{u+1}]^2 & \text{si } 0 \leq v \leq u \\ 0 & \text{sinon.} \end{cases} \quad (3.13)$$

(b) On a $(V \geq 0) = \Omega$; il en résulte que :

$$P(U \leq u) = \begin{cases} (1 - q^{u+1})^2 & \text{si } u \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

Mais, puisque l'on a

$$P(U = u) = P(U \leq u) - P(U \leq u - 1),$$

il vient, après calculs,

$$P(U = u) = \begin{cases} p [2q^u - q^{2u} (1 + q)] & \text{si } u \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

(c) La suite d'ensembles $(U \leq u)_{u \in \mathbb{N}}$ est **croissante** (pour l'inclusion) et de réunion Ω ; on peut donc écrire que

$$P(V \geq v) = \lim_{u \rightarrow +\infty} P[(U \leq u) \cap (V \geq v)],$$

soit :

$$P(V \geq v) = \begin{cases} q^{2v} & \text{si } v \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

Mais, puisque l'on a

$$P(V = v) = P(V \geq v) - P(V \geq v + 1),$$

il vient, après calculs,

$$P(V = v) = \begin{cases} q^{2v} [1 - q^2] & \text{si } v \geq 0 \\ 0 & \text{sinon,} \end{cases}$$

ce qui démontre que **la loi de V est la loi géométrique sur \mathbb{N} de paramètre $1 - q^2$.**

3. On peut écrire

$$P(U \leq u, V = v) = P(U \leq u, V \geq v) - P(U \leq u, V \geq v + 1),$$

soit, d'après (3.13), et après simplifications :

◇ si $u \geq v + 1$ et $v \geq 0$,

$$P(U \leq u, V = v) = pq^v [q^v (1 + q) - 2q^{u+1}], \quad (3.14)$$

◇ si $u = v \geq 0$,

$$P(U \leq u, V = u) = P(V = U = u) = [q^u - q^{u+1}]^2 = p^2 q^{2u}.$$

On remarque que dans ce dernier cas l'égalité (3.14) est encore vraie.

Alors, puisque pour tout $w \in \mathbb{N}$ on a

$$(W \leq w) = \bigcup_{v \in \mathbb{N}} (U \leq w + v, V = v),$$

il vient en utilisant l'égalité (3.14) et après calculs :

$$\forall w \in \mathbb{N} \quad P(W \leq w) = 1 - \frac{2}{1+q} q^{w+1}.$$

On en déduit la probabilité $P(W = w)$:

◇ si $w \geq 1$, on a

$$P(W = w) = P(W \leq w) - P(W \leq w - 1).$$

soit :

$$P(W = w) = \frac{2p}{1+q} q^w ;$$

◇ si $w = 0$, on a

$$P(W = 0) = P(X = Y),$$

soit :

$$P(W = 0) = \frac{p}{1+q}.$$

Exercice 3.6. Loi binomiale, loi de Poisson et gestion de stock. Le nombre de clients arrivant dans un magasin pendant une journée de vente est supposé suivre une loi de Poisson⁷ de paramètre λ . Un client achète un article A avec la probabilité p (il en achète au plus un). Le **stock** d'articles A à l'ouverture du magasin est de s ($s \geq 1$) articles. On veut calculer la loi du nombre total d'articles demandés en une journée et la probabilité qu'il n'y ait pas **rupture de stock** de l'article A durant cette journée.

On modélise la situation de la manière suivante : on se donne une variable aléatoire N , représentant le nombre de clients entrant dans le magasin durant la journée, et une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$, toutes définies sur le même espace de probabilité (Ω, \mathcal{A}, P) , représentant la décision d'achat du

7. Le choix de cette hypothèse peut être justifié par le théorème de Poisson (théorème 7.1).

n -ième client : $X_n = 1$ s'il achète, $X_n = 0$ sinon. On suppose toutes ces variables aléatoires **indépendantes**. On suppose que la loi de N est la **loi de Poisson** $\mathcal{P}(\lambda)$ et que, pour tout $n \in \mathbb{N}^*$, la variable aléatoire X_n suit une **loi de Bernoulli** de paramètre p . Le nombre total d'articles demandés en une journée est la variable aléatoire T définie par :

$$T = \mathbf{1}_{(N \geq 1)} \sum_{j=1}^N X_j.$$

On demande la loi de T et la probabilité $P(T \leq s)$.

Solution. On a

$$(T = 0) = (N = 0) \bigcup_{n \geq 1} \left[\bigcup_{j=1}^n (N = n, \sum_{j=1}^n X_j = 0) \right].$$

En notant $q = 1 - p$, par σ -additivité de P et indépendance des variables aléatoires N et X_n , il vient

$$P(T = 0) = P(N = 0) + \sum_{n \geq 1} \exp(-\lambda) \frac{\lambda^n}{n!} q^n,$$

soit encore

$$P(T = 0) = \exp(-\lambda) + \sum_{n \geq 1} \exp(-\lambda) \frac{\lambda^n}{n!} q^n,$$

et, après calculs :

$$\boxed{P(T = 0) = \exp(-\lambda p)}.$$

De plus, pour tout $k \geq 1$, on peut écrire :

$$(T = k) = \bigcup_{n \geq 1} \left(N = n, \sum_{j=1}^n X_j = k \right).$$

La loi de $\sum_{j=1}^n X_j$ étant la loi binomiale $\mathcal{B}(n, p)$, il vient alors

$$P(T = k) = \sum_{n \geq k} \exp(-\lambda) \frac{\lambda^n}{n!} \binom{n}{k} p^k q^{n-k},$$

soit encore

$$P(T = k) = \exp(-\lambda) \frac{(p\lambda)^k}{k!} \sum_{n \geq k} \frac{(\lambda q)^{n-k}}{(n-k)!},$$

et donc :

$$\boxed{P(T = k) = \exp(-\lambda p) \frac{(p\lambda)^k}{k!}}.$$

Nous venons de démontrer que la loi de T est la **loi de Poisson** de paramètre λp .

La probabilité qu'il n'y ait pas rupture de stock est alors :

$$P(T \leq s) = \exp(-\lambda p) \sum_{k=0}^s \frac{(p\lambda)^k}{k!} .$$

Chapitre 4

Probabilités et lois conditionnelles

Nous avons étudié au chapitre précédent la notion d'événements indépendants. Dans ce chapitre, nous allons voir comment traduire l'influence d'un événement sur un autre ; la notion de probabilité conditionnelle va permettre, dans une certaine mesure, de rendre compte de l'information apportée par la réalisation d'un événement sur la réalisation (éventuelle) d'un autre événement.

4.1. Probabilités conditionnelles

Dans la suite, P est une probabilité quelconque sur l'espace probabilisable (Ω, \mathcal{A}) .

4.1.1. Définitions. Formule des probabilités totales

Définition 4.1. Soit $B \in \mathcal{A}$ un événement tel que $P(B) > 0$. Soit $A \in \mathcal{A}$ un événement ; on appelle **probabilité de A conditionnée par B** (ou sachant B) le nombre réel, noté $P(A | B)$ ou $P^B(A)$, défini par :

$$P(A | B) \equiv P^B(A) = \frac{P(A \cap B)}{P(B)}. \quad (4.1)$$

Proposition 4.2. Soit $B \in \mathcal{A}$ un événement tel que $P(B) > 0$. L'application P^B de \mathcal{A} dans \mathbb{R}^+ qui à tout $A \in \mathcal{A}$ fait correspondre $P^B(A)$ est une probabilité sur l'espace probabilisable (Ω, \mathcal{A}) . Elle est appelée **probabilité conditionnée¹ à B** .

Démonstration. On a d'abord

$$P^B(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = 1;$$

L'application P^B est σ -additive ; en effet, pour toute suite $(A_i)_{i \in \mathbb{N}}$ d'événements disjoints deux à deux on a

1. On parle souvent brièvement de la **probabilité conditionnelle** P^B .

$$\left(\bigcup_{i \in \mathbb{N}} A_i \right) \cap B = \bigcup_{i \in \mathbb{N}} (A_i \cap B),$$

et donc :

$$P^B \left(\bigcup_{i \in \mathbb{N}} A_i \right) = \frac{\sum_{i \in \mathbb{N}} P(A_i \cap B)}{P(B)} = \sum_{i \in \mathbb{N}} P^B(A_i).$$

Remarques. 1. Cette proposition est très importante puisqu'elle dit que toutes les propriétés établies jusqu'à maintenant (et ultérieurement) pour une probabilité quelconque sont vraies pour la **probabilité conditionnelle** P^B .

2. Soient un événement $B \in \mathcal{A}$ tel que $P(B) > 0$ et un événement A quelconque. Pour que les événements A et B soient indépendants, il faut et il suffit que :

$$P^B(A) = P(A).$$

L'exemple ci-dessous montre comment une probabilité conditionnelle rend compte de l'information apportée par la réalisation d'un événement sur la réalisation d'un autre événement.

Exemple 4.1. On jette deux fois un dé. Soit A l'événement « on obtient un 6 au premier jet », et soient B_k , $2 \leq k \leq 12$, les événements « la somme des deux entiers obtenus est k ». On modélise les deux lancers de dé par l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ où $\Omega = \{1, 2, \dots, 6\}^2$ et P est la probabilité uniforme. On a :

$$A = \{6\} \times \{1, 2, \dots, 6\}$$

et

$$B_k = \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 = k\}.$$

En particulier, on a :

$$B_{12} \subset A \quad \text{et} \quad B_{11} = \{(5, 6), (6, 5)\}.$$

Il en résulte que :

$$P(A) = \frac{1}{6}, \quad P(A \mid B_{12}) = 1, \quad P(A \mid B_{11}) = \frac{1}{2}.$$

La proposition suivante résulte immédiatement de la définition.

Proposition 4.3. Si $P(B) > 0$, on a

$$\boxed{P(A \cap B) = P(A \mid B) P(B).}$$

C'est très souvent sous la forme de la relation de la proposition 4.3 qu'on utilise la définition de $P(A | B)$. En effet, dans beaucoup de problèmes conduisant au calcul d'une probabilité, les données fournies par l'expérience ou tirées d'une première analyse du problème s'interprètent de façon très naturelle comme des probabilités conditionnelles.

La proposition 4.3 se généralise de la façon suivante.

Proposition 4.4. (Probabilités conditionnelles en cascade.) Soit $(A_i)_{1 \leq i \leq n}$ une suite finie d'événements tels que $P\left(\bigcap_{1 \leq i \leq n-1} A_i\right) > 0$. On a :

$$P\left(\bigcap_{1 \leq i \leq n} A_i\right) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \dots \\ \dots P(A_n | A_1 \cap \dots \cap A_{n-1}).$$

Démonstration. On remarque que toutes les probabilités conditionnelles introduites sont bien définies puisque, pour tout $j \in \{1, 2, \dots, n-1\}$, on a :

$$P\left(\bigcap_{1 \leq i \leq j} A_i\right) \geq P\left(\bigcap_{1 \leq i \leq n-1} A_i\right) > 0;$$

la formule résulte alors d'une simple récurrence.

Exemple 4.2. On tire successivement 4 cartes d'un jeu de 52 cartes. Quelle est la probabilité de tirer les 4 as ?

Solution. Notons A_i ($i = 1, \dots, 4$) l'événement « la i -ième carte tirée est un as ». On demande de calculer $P(A_1 \cap A_2 \cap A_3 \cap A_4)$. Les probabilités immédiatement accessibles sont, en-dehors de $P(A_1)$, les probabilités conditionnelles

$$P(A_2 | A_1) = \frac{3}{51}, \quad P(A_3 | A_1 \cap A_2) = \frac{2}{50}, \quad P(A_4 | A_1 \cap A_2 \cap A_3) = \frac{1}{49}.$$

On a donc

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) \\ = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2)P(A_4 | A_1 \cap A_2 \cap A_3) \\ = \frac{4 \cdot 3 \cdot 2 \cdot 1}{52 \cdot 51 \cdot 50 \cdot 49} \simeq 0,00000369.$$

On observera que nous ne nous sommes pas préoccupés dans cet exemple de la construction d'un espace Ω dont les A_i pourraient être considérés comme des sous-ensembles. Une justification de cette attitude sera donnée plus loin (à la sous-section 4.3 ; c'est en particulier l'objet de la proposition 4.9).

Définition 4.5. Soit $(A_i)_{i \in I}$ une famille dénombrable d'événements disjoints deux à deux² telle que :

$$P\left(\bigcup_{i \in I} A_i\right) = 1. \quad (4.2)$$

Une telle famille est appelée **système complet de constituants**.

Soit N le complémentaire de $\bigcup_{i \in I} A_i$ dans Ω . On a $P(N) = 0$. Les événements $(A_i)_{i \in I}$ et N forment une *partition* de Ω . On traduit cette situation en termes probabilistes en disant : avec probabilité 1, l'un des événements A_i – et un seul – se réalise.

Exemple 4.3. On lance une pièce de monnaie jusqu'à ce qu'on obtienne un « pile ». Soit A_i l'événement « pile apparaît pour la première fois au i -ième lancer ». Les A_i ($i = 1, 2, \dots$) sont évidemment deux à deux incompatibles. D'autre part, il est possible qu'aucun des A_i ne se réalise mais la probabilité de cet événement est nulle.

Les événements A_i ($i = 1, 2, \dots$) forment donc un système complet de constituants.

Théorème 4.6. (Formule des probabilités totales.) Soit $(A_i)_{i \in I}$ un système complet de constituants tel que $P(A_i) > 0$ pour tout $i \in I$. On a alors, pour tout événement $A \in \mathcal{A}$:

$$P(A) = \sum_{i \in I} P(A | A_i) P(A_i). \quad (4.3)$$

Démonstration. Soit N le complémentaire de $\bigcup_{i \in I} A_i$.

$$A = \left[\bigcup_{i \in I} (A \cap A_i) \right] \cup (A \cap N).$$

Comme on a $P(A \cap N) \leq P(N) = 0$, on obtient immédiatement, en utilisant la σ -additivité de P :

$$P(A) = P\left[\bigcup_{i \in I} (A \cap A_i)\right] = \sum_{i \in I} P[(A \cap A_i)].$$

Il ne reste plus qu'à utiliser la définition des probabilités conditionnelles.

2. On dit souvent de manière équivalente que ces événements sont **incompatibles**.

Un cas particulier de système de constituants est le système (B, B^c) , où B est un événement tel que $0 < P(B) < 1$. La formule des probabilités totales devient dans ce cas

$$P(A) = P(A | B) P(B) + P(A | B^c) P(B^c).$$

Exemple 4.4. Une urne U_1 (respectivement U_2) contient b_1 (respectivement b_2) boules blanches et n_1 (respectivement n_2) boules noires. On choisit « au hasard » une urne et on tire ensuite une boule dans cette urne. On cherche la probabilité de tirer une boule noire.

Solution. Nous reviendrons par la suite (voir l'exemple 4.9) sur la construction d'un modèle associé à cette expérience. Supposons donc construit un espace probabilisé (Ω, \mathcal{A}, P) tel que si N , U_1 , U_2 sont les événements « on tire une boule noire », « le tirage a lieu dans l'urne U_1 », « le tirage a lieu dans l'urne U_2 », on ait, pour $i = 1, 2$:

$$P(U_i) = \frac{1}{2} \quad \text{et} \quad P(N | U_i) = \frac{n_i}{n_i + b_i}.$$

Les événements U_1 et U_2 forment un système complet de constituants et la probabilité $P(N)$ de tirer une boule noire est donnée par :

$$P(N) = P(U_1) P(N | U_1) + P(U_2) P(N | U_2),$$

soit :

$$P(N) = \frac{1}{2} \left(\frac{n_1}{n_1 + b_1} + \frac{n_2}{n_2 + b_2} \right).$$

Exemple 4.5. Le second problème du chevalier de Méré. On date la naissance du calcul des probabilités de la résolution par Pascal³, en 1654, de deux problèmes que lui avait posé le chevalier de Méré. L'un de ces problèmes est celui-ci : deux joueurs jouent à un jeu de hasard en plusieurs parties ; celui qui, le premier, a gagné trois parties, emporte la totalité de l'enjeu. Si les joueurs doivent arrêter le jeu alors qu'il ne manque au premier joueur, pour l'emporter, qu'une partie, et au second que deux parties, comment répartir équitablement l'enjeu ?

Pour calculer la fraction de l'enjeu qui doit être attribuée à chaque joueur (dont il explique qu'elle doit être égale à ce que nous appellerions aujourd'hui la probabilité qu'à ce joueur de l'emporter), Pascal emploie une démarche très proche de celle que nous adopterions, en utilisant les probabilités conditionnelles et la formule des probabilités totales. Voici, en substance, son raisonnement.

3. Blaise Pascal (1623-1662), savant, penseur et écrivain. Avant 1654, ses œuvres mathématiques sont essentiellement d'ordre géométrique. En 1654, le chevalier de Méré le met en contact avec Fermat. Les relations épistolaires entre Pascal et Fermat sont en partie à l'origine du calcul des probabilités.

Si on joue une partie de plus, le premier joueur a une chance sur deux de l'emporter. Donnons-lui la moitié de l'enjeu. S'il perd cette partie, les deux joueurs se retrouvent à égalité, et ont désormais des chances égales de l'emporter. Partageons également entre eux le reste de l'enjeu. Le premier joueur reçoit donc les trois quarts de l'enjeu, et le second le quart restant.

4.1.2. « Probabilités des causes » et formule de Bayes

Dans une expérience effectuée en deux temps, comme celle de l'exemple ci-dessus, la probabilité conditionnelle possède une interprétation très intuitive : a priori, c'est-à-dire avant que ne soit effectué le choix de l'urne, la probabilité de tirer une boule noire est $P(N)$. Réalisons le premier temps de l'expérience, le choix au hasard d'une urne, et supposons que l'urne choisie soit la première. Tout le monde s'accordera pour dire que « maintenant » la probabilité de tirer une boule noire est $\frac{n_1}{n_1 + b_1} = P(N | U_1)$.

En fait, le temps n'a rien à voir là-dedans. Supposons que l'expérience ait été réalisée complètement, et qu'on ait seulement pu observer que l'urne choisie était l'urne n°1. Autrement dit, on dispose sur le résultat de l'expérience d'une information partielle. On considérera encore que, compte tenu de cette information, la probabilité de tirer une boule noire est $P(N | U_1)$. Pour toute décision à prendre (pari, etc.) dépendant de la probabilité d'apparition d'une boule noire, il convient d'utiliser le nombre $P(N | U_1)$ à la place du nombre $P(N)$.

Supposons au contraire qu'on connaisse la couleur de la boule tirée – disons une boule noire –, autre information partielle, mais qu'on ignore dans quelle urne elle a été tirée. Le nombre $P(U_1 | N)$ représente la probabilité d'avoir tiré dans l'urne 1 sachant que c'est une boule noire qui a été tirée. On parle à ce propos de *probabilité a posteriori*, ou même de *probabilité des causes* : si, pour fixer les idées, l'urne n°1 contient beaucoup de boules noires, et l'urne n°2 contient peu de boules noires, le fait que ce soit une boule noire qui ait été tirée laisse soupçonner que c'est dans l'urne n°1 que le tirage a été effectué. Le nombre $P(U_1 | N)$ mesure la valeur de la vraisemblance que l'on doit accorder à cette hypothèse. Mathématiquement, toutefois, il n'y a pas de différence de nature entre une probabilité conditionnelle comme $P(N | U_1)$ et une probabilité conditionnelle comme $P(U_1 | N)$.

Proposition 4.7. Soient deux événements $A_1, A_2 \in \mathcal{A}$ de probabilités non nulles. Alors :

$$P(A_2 | A_1) = P(A_1 | A_2) \frac{P(A_2)}{P(A_1)}.$$

Démonstration. Par définition d'une probabilité conditionnelle on a :

$$P(A_2 | A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} = \frac{P(A_1 | A_2) P(A_2)}{P(A_1)}.$$

Théorème 4.8. (Théorème de Bayes.) Soit $(A_i)_{i \in I}$ un système complet de constituants tel que $P(A_i) > 0$ pour tout $i \in I$. On a alors, pour tout événement $A \in \mathcal{A}$ de probabilité non nulle et pour tout $i \in I$:

$$P(A_i | A) = \frac{P(A | A_i) P(A_i)}{\sum_{j \in I} P(A | A_j) P(A_j)} \quad (4.4)$$

(formule de Bayes).

Démonstration. Il résulte de la proposition 4.7 que, pour tout $i \in I$:

$$P(A_i | A) = \frac{P(A | A_i) P(A_i)}{P(A)};$$

il suffit alors d'appliquer le théorème 4.6.

Exemple 4.6. On reprend l'exemple 4.4. On cherche la probabilité **a posteriori** d'avoir tiré dans l'urne U_i sachant qu'une boule noire a été tirée.

Solution. On applique le théorème de Bayes avec le système complet de constituants formé des événements U_1 et U_2 . Il vient, pour $i = 1, 2$,

$$P(U_i | N) = \frac{P(U_i) P(N | U_i)}{P(U_1) P(N | U_1) + P(U_2) P(N | U_2)},$$

soit encore :

$$P(U_i | N) = \frac{\frac{n_i}{n_i + b_i}}{\frac{n_1}{n_1 + b_1} + \frac{n_2}{n_2 + b_2}}.$$

4.2. Lois conditionnelles

Lors de l'étude d'un phénomène aléatoire, une information partielle est souvent apportée par une variable aléatoire X à valeurs dans un espace probabilisable (E, \mathcal{E}) . Supposons que X soit une variable aléatoire discrète, à valeurs dans un ensemble E , et soit x tel que $P(X = x) > 0$. Soit

Y une autre variable aléatoire, à valeurs dans un espace probabilisable (F, \mathcal{F}) . Par définition, la loi de Y conditionnelle⁴ à l'événement $(X = x)$ est la probabilité sur (F, \mathcal{F}) définie par l'application

$$B \mapsto P^{(X=x)}(Y \in B) \equiv P(Y \in B \mid X = x);$$

elle est notée $P_Y^{(X=x)}$.

Si Y est aussi une variable aléatoire discrète, la famille des probabilités $P_Y^{(X=x)}$, où x décrit l'ensemble des éléments de E tels que $P(X = x) > 0$, et la loi P_X de X déterminent entièrement la loi du couple aléatoire (X, Y) .

En effet, la variable aléatoire (X, Y) est aussi discrète. Sa loi est donc entièrement déterminée par la donnée des probabilités $P(X = x, Y = y)$ pour $x \in E, y \in F$. Or si $x \in E$ et $y \in F$,

– soit $P(X = x) = 0$, auquel cas on a

$$P(X = x, Y = y) \leq P(X = x) = 0,$$

et donc :

$$P(X = x, Y = y) = 0;$$

– soit $P(X = x) > 0$, auquel cas on a

$$\begin{aligned} P(X = x, Y = y) &= P(Y = y \mid X = x) P(X = x) \\ &= P^{(X=x)}(Y = y) P(X = x). \end{aligned}$$

On sait donc toujours bien calculer $P(X = x, Y = y)$.

Exemple 4.7. Soient X et Y deux variables aléatoires discrètes à valeurs dans \mathbb{N} définies sur l'espace probabilisé (Ω, \mathcal{A}, P) . On suppose que la loi de X est la loi de Poisson $\mathcal{P}(\lambda)$, où $\lambda > 0$. On suppose de plus que, pour tout entier $n > 0$, la loi conditionnelle de Y sachant $(X = n)$ est la loi binomiale $\mathcal{B}(n, p)$, et que si $n = 0$, la variable aléatoire Y prend la valeur 0 avec probabilité 1.

On cherche la loi de Y , ainsi que, pour $k \in \mathbb{N}$, la loi de X conditionnelle à l'événement $(Y = k)$.

Exemple de situation modélisée comme ci-dessus. On se place à un embranchement routier; le nombre de véhicules arrivant pendant un intervalle de temps d'une heure est une variable aléatoire X , suivant une loi de Poisson $\mathcal{P}(\lambda)$ (c'est une hypothèse courante dans ce genre de situation⁵). Les véhicules ne peuvent prendre que l'une des directions A ou B, et la variable aléatoire Y représente le nombre de véhicules empruntant la direction A pendant cet intervalle de temps. Chaque véhicule prend la direction A avec la probabilité

4. On dit aussi la « loi de Y sachant que $X = x$. »

5. Voir la note de l'exercice 3.6.

p , et les choix sont faits de manière indépendante. C'est pourquoi on suppose que si n véhicules arrivent à l'embranchement pendant une heure donnée, la loi (conditionnelle) de Y est la loi binomiale $\mathcal{B}(n, p)$.

La probabilité conditionnelle $P^{(Y=k)}(X = n)$ est la probabilité que n véhicules soient arrivés à l'embranchement sachant que k véhicules ont emprunté la direction A. Le calcul de cette probabilité est donc un exemple de « probabilité des causes ».

Solution.

– On note $q = 1 - p$. Pour tout $(n, k) \in \mathbb{N}^2$, on a

$$P(X = n, Y = k) = P^{(X=n)}(Y = k) P(X = n),$$

soit encore (il faut traiter séparément le cas $n = 0$) :

$$P(X = n, Y = k) = \begin{cases} 0 & \text{si } 0 \leq n < k \\ \exp(-\lambda) \frac{\lambda^n}{n!} \binom{n}{k} p^k q^{n-k} & \text{si } 0 \leq k \leq n. \end{cases}$$

Puisque $\bigsqcup_{n \in \mathbb{N}} (X = n) = \Omega$, on a

$$P(Y = k) = \sum_{n \in \mathbb{N}} P(X = n, Y = k).$$

On a donc

◇ si $k < 0$,

$$P(Y = k) = 0;$$

◇ si $0 \leq k$,

$$P(Y = k) = \sum_{n \geq k} \exp(-\lambda) \frac{\lambda^n}{n!} \binom{n}{k} p^k q^{n-k},$$

soit

$$P(Y = k) = \exp(-\lambda) \frac{(\lambda p)^k}{k!} \sum_{n \geq k} \frac{1}{(n-k)!} (\lambda q)^{n-k},$$

ou encore :

$$P(Y = k) = \exp(-\lambda p) \frac{(\lambda p)^k}{k!}.$$

La loi de Y est donc la loi de Poisson de paramètre λp .

– Alors, pour tout $n \in \mathbb{N}$ et tout $k \in \mathbb{N}$, on a

$$P^{(Y=k)}(X = n) = \frac{P(X = n, Y = k)}{P(Y = k)};$$

on a donc :

◇ si $n < k$,

$$P^{(Y=k)}(X = n) = 0;$$

◇ si $0 \leq k \leq n$,

$$P^{(Y=k)}(X = n) = \frac{\exp(-\lambda) \frac{\lambda^n}{n!} \binom{n}{k} p^k q^{n-k}}{\exp(-\lambda p) \frac{(\lambda p)^k}{k!}} = \exp(-\lambda q) \frac{(\lambda q)^{n-k}}{(n-k)!}.$$

Pour tout entier k fixé, la loi conditionnelle de X sachant $(Y = k)$ est donc la **loi de Poisson** de paramètre λq **translatée** de k .

4.3. Modélisation d'un phénomène évolutif

La construction d'un modèle probabiliste pour une expérience aléatoire « concrète » commence en principe par la définition d'un espace probabilisé (Ω, \mathcal{A}, P) . On définit ensuite les événements auxquels on s'intéresse (comme des parties de Ω) et certaines variables aléatoires (comme des fonctions définies sur Ω) représentant des grandeurs liées à l'expérience. Le plus souvent, cependant, la connaissance explicite de Ω n'est pas utilisée. On a seulement besoin de savoir que tel et tel événement sont indépendants, que telle variable aléatoire X suit telle loi, etc. L'espace Ω passe au second plan.

Dans un problème courant de modélisation, l'espace Ω est d'ailleurs très rarement fourni par la situation considérée – sauf dans quelques situations tirées des jeux de hasard. Ce qui est immédiatement accessible, ce sont des conditions imposées à certaines données : les événements A_1, A_2, \dots, A_n sont équiprobables – cela provient souvent de considérations de symétrie –, les variables aléatoires X et Y sont indépendantes, la variable aléatoire X suit telle loi. Et c'est au mathématicien qu'il revient de montrer l'existence d'un espace Ω , d'événements A_1, A_2, \dots, A_n définis comme parties de cet espace, de variables aléatoires définies comme fonctions sur Ω , possédant les propriétés demandées.

C'est ce que nous avons fait au chapitre 3 lorsque nous avons montré qu'étant données des lois de probabilité discrètes P_1, \dots, P_n sur des ensembles dénombrables E_1, \dots, E_n , il existe un espace probabilisé (Ω, \mathcal{A}, P) et des variables aléatoires X_1, \dots, X_n définies sur (Ω, \mathcal{A}, P) à valeurs dans E_1, \dots, E_n respectivement et telles que

- (a) X_i suit la loi P_i ($i = 1, \dots, n$);
- (b) X_1, \dots, X_n sont indépendantes.

(Rappelons qu'on prend $\Omega = E_1 \times \dots \times E_n$ et pour X_i la i -ième projection $\text{pr}_i : \Omega \rightarrow E_i$.)

Dans de nombreuses expériences aléatoires constituées d'étapes successives comme celles qui nous ont servi d'exemple dans ce chapitre, c'est en fait à des **probabilités conditionnelles** que l'on a immédiatement

accès.

Proposition 4.9. Soient E_1, \dots, E_n des ensembles dénombrables. Supposons que soient donnés

- un germe de loi de probabilité sur E_1 , noté p_1 ;
- pour chaque $x_1 \in E_1$, un germe de loi de probabilité sur E_2 , noté $p_2^{x_1}$;
- ...
- pour chaque $(x_1, \dots, x_{n-1}) \in E_1 \times \dots \times E_{n-1}$, un germe de loi de probabilité sur E_n , noté $p_n^{x_1, \dots, x_{n-1}}$.

Alors, il existe un espace probabilisé (Ω, \mathcal{A}, P) et des variables aléatoires X_1, \dots, X_n définies sur cet espace, à valeurs respectivement dans E_1, \dots, E_n , telles que

$$\begin{cases} P(X_1 = x_1) = p_1(x_1) \\ P(X_2 = x_2 \mid X_1 = x_1) = p_2^{x_1}(x_2) \\ \dots \\ P(X_n = x_n \mid X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = p_n^{x_1, \dots, x_{n-1}}(x_n) \end{cases} \quad (4.5)$$

quels que soient $x_1 \in E_1, \dots, x_n \in E_n$.

On peut par exemple prendre $\Omega = E_1 \times \dots \times E_n$, $\mathcal{A} = \mathcal{P}(\Omega)$, définir P à partir de son germe p en posant pour $\omega = (x_1, \dots, x_n)$,

$$p(\omega) = p_1(x_1)p_2^{x_1}(x_2)\dots p_n^{x_1, \dots, x_{n-1}}(x_n) \quad (4.6)$$

et prendre pour X_i ($1 \leq i \leq n$) la i -ième projection définie par

$$X_i(x_1, \dots, x_n) = x_i.$$

Démonstration. Il suffit de vérifier que la solution proposée dans le dernier alinéa de la proposition convient.

La propriété de Fubini pour les familles à termes positifs permet d'écrire

$$\sum_{\omega \in \Omega} p(\omega) = \sum_{x_1 \in E_1} p_1(x_1) \left(\sum_{x_2 \in E_2} p_2^{x_1}(x_2) \dots \left(\sum_{x_n \in E_n} p_n^{x_1, \dots, x_{n-1}}(x_n) \right) \right).$$

En effectuant successivement les sommations de droite à gauche, on vérifie que

$$\sum_{\omega \in \Omega} p(\omega) = 1$$

et que p est donc bien un germe de probabilité. Les relations (4.5) résultent de la formule des probabilités conditionnelles en cascade.

Remarques. 1. Si $p_i^{x_1, \dots, x_{i-1}}(x_i)$ ne dépend que de x_i , autrement dit, si, pour $i = 1, \dots, n$, et pour tout $(x_1, \dots, x_{i-1}) \in E_1 \times \dots \times E_{i-1}$, le germe $p_i^{x_1, \dots, x_{i-1}}$ est un germe g_i sur E_i indépendant de x_1, \dots, x_{i-1} , alors la probabilité P est la **probabilité produit** des probabilités associées aux germes g_i .

Dans ce cas, des variables aléatoires X_1, \dots, X_n satisfaisant aux relations (4.5) sont nécessairement indépendantes, puisque des relations

$$P(X_i = x_i \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) = P(X_i = x_i)$$

on tire, au moyen d'un raisonnement par récurrence élémentaire, les relations

$$\begin{aligned} P(X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x_i) \\ = P(X_1 = x_1) \dots P(X_{i-1} = x_{i-1}) P(X_i = x_i). \end{aligned}$$

On retrouve ainsi la construction de la section 3.4.

2. Si $p_i^{x_1, \dots, x_{i-1}}(x_i)$ ne dépend que de x_i et x_{i-1} , autrement dit, si, pour $i = 2, \dots, n$, le germe $p_i^{x_1, \dots, x_{i-1}}$ est un germe sur E_i ne dépendant que de x_{i-1} , alors on a

$$P(X_i = x_i \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) = P(X_i = x_i \mid X_{i-1} = x_{i-1}). \quad (4.7)$$

Une suite de variables aléatoire $(X_i)_{1 \leq i \leq n}$ satisfaisant la condition (4.7) est dite **markovienne**⁶.

Exemple 4.8. Soit $(U_i)_{1 \leq i \leq n}$ une suite de variables aléatoires indépendantes à valeurs dans \mathbb{Z} . Soit, pour $i = 1, \dots, n$, $X_i = U_1 + \dots + U_i$; la suite $(X_i)_{1 \leq i \leq n}$ est *markovienne* à valeurs dans \mathbb{Z} .

En effet, si $i \geq 2$, $X_i = X_{i-1} + U_i$, si bien qu'avec les notations ci-dessus, on a

$$\begin{aligned} P(X_i = x_i \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \\ = P(U_i = x_i - x_{i-1} \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \end{aligned}$$

quels que soient $x_1, \dots, x_i \in \mathbb{Z}$; la variable aléatoire (X_1, \dots, X_{i-1}) étant fonction de U_1, \dots, U_{i-1} est indépendante de U_i , d'où il résulte que

$$P(X_i = x_i \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) = P(U_i = x_i - x_{i-1}).$$

De plus, le même raisonnement conduit à l'égalité

$$P(X_i = x_i \mid X_{i-1} = x_{i-1}) = P(U_i = x_i - x_{i-1}),$$

ce qui démontre le résultat.

6. En référence au mathématicien russe Andreï Markov (1856-1922), élève de P. Tchebichev, auteur de nombreux travaux sur les probabilités.

Exemple 4.9. Reprenons l'exemple 4.4 pour construire, comme annoncé, l'espace probabilisé associé. Reformulons d'abord le problème de modélisation posé en termes de variables aléatoires.

Représentons l'ensemble des choix d'urne par $E_1 = \{1, 2\}$, et celui des couleurs par $E_2 = \{n, b\}$. On souhaite en fait construire un modèle constitué d'une part d'un espace probabilisable (Ω, \mathcal{A}, P) , et d'autre part de deux variables aléatoires $X_1 : \Omega \rightarrow E_1$ et $X_2 : \Omega \rightarrow E_2$ satisfaisant aux relations

$$\begin{cases} P(X_1 = i) = \frac{1}{2}, \\ P(X_2 = n \mid X_1 = i) = \frac{n_i}{n_i + b_i}, \quad P(X_2 = b \mid X_1 = i) = \frac{b_i}{n_i + b_i} \end{cases}$$

($i = 1, 2$).

La proposition 4.9 indique qu'on peut prendre $\Omega = E_1 \times E_2$, $\mathcal{A} = \mathcal{P}(\Omega)$, et définir P en prenant

$$\begin{cases} P(\{(i, n)\}) = \frac{1}{2} \frac{n_i}{n_i + b_i}, \\ P(\{(i, b)\}) = \frac{1}{2} \frac{b_i}{n_i + b_i} \end{cases}$$

($i = 1, 2$). Les variables X_1 et X_2 seront bien sûr les projections du produit cartésien $\Omega = E_1 \times E_2$ sur les espaces facteurs.

Exercices

Exercice 4.1. Formules de Bayes et des probabilités totales; jeu de pile ou face. Un joueur joue à pile ou face avec deux pièces équilibrées de la manière suivante : il lance simultanément ces deux pièces ; s'il n'obtient aucun pile, son gain est nul et la partie s'arrête. S'il obtient au moins un pile, il relance simultanément les deux pièces autant de fois qu'il a obtenu pile à la première phase du jeu et gagne autant d'unités que le nombre de piles obtenu lors de cette deuxième série de jets. On cherche la probabilité que son gain soit non nul, puis la probabilité d'avoir obtenu deux piles au premier jet sachant qu'il a obtenu un seul pile à la seconde étape.

Solution. On suppose construit un espace probabilisé (Ω, \mathcal{A}, P) tel que, X_1 désignant la variable aléatoire donnant le nombre de piles obtenu à la première étape, sa loi soit donnée par :

$$P(X_1 = 0) = \frac{1}{4}, \quad P(X_1 = 1) = \frac{1}{2}, \quad P(X_1 = 2) = \frac{1}{4},$$

et, X_2 désignant la variable aléatoire gain du joueur, on ait :

$$(X_1 = 0) \subset (X_2 = 0).$$

On suppose de plus que la loi conditionnelle de X_2 à X_1 soit (partiellement) donnée par :

$$P(X_2 = 0 | X_1 = 1) = \frac{1}{4}, \quad P(X_2 = 1 | X_1 = 1) = \frac{1}{2}, \quad P(X_2 = 2 | X_1 = 1) = \frac{1}{4},$$

$$P(X_2 = 0 | X_1 = 2) = \frac{1}{16}, \quad P(X_2 = 1 | X_1 = 2) = \frac{1}{4}.$$

On a

$$(X_2 = 0) = (X_1 = 0) \cup [(X_1 > 0) \cap (X_2 = 0)],$$

et donc :

$$P(X_2 = 0) = P(X_1 = 0) +$$

$$P(X_1 = 1) P(X_2 = 0 | X_1 = 1) + P(X_1 = 2) P(X_2 = 0 | X_1 = 2),$$

soit :

$$P(X_2 = 0) = \frac{25}{64}.$$

On cherche la probabilité d'avoir un gain non nul; elle est donnée par :

$$P(X_2 > 0) = 1 - P(X_2 = 0) = 1 - \frac{25}{64},$$

soit :

$$P(X_2 > 0) = \frac{39}{64}.$$

La probabilité d'avoir obtenu deux piles au premier jet sachant qu'il a obtenu un seul pile à la seconde étape est $P(X_1 = 2 | X_2 = 1)$. On a, d'après le théorème de Bayes :

$$P(X_1 = 2 | X_2 = 1) = \frac{P(X_1 = 2) P(X_2 = 1 | X_1 = 2)}{\sum_{j=0}^2 P(X_1 = j) P(X_2 = 1 | X_1 = j)} = \frac{\frac{1}{4} \frac{1}{4}}{\frac{1}{2} \frac{1}{2} + \frac{1}{4} \frac{1}{4}},$$

soit :

$$P(X_1 = 2 | X_2 = 1) = \frac{1}{5}.$$

Remarque. Pour compléter cet exercice on pourrait essayer de construire un espace probabilisé (Ω, \mathcal{A}, P) avec la méthode étudiée dans la section 4.3. On serait amené à définir complètement le système de loi conditionnelle de X_2 sachant X_1 , que nous n'avons utilisé que partiellement ici.

Exercice 4.2. Modèle d'urnes. Deux urnes A et B contiennent respectivement 6 boules blanches et 5 noires d'une part, 4 blanches et 8 noires d'autre part. On transfère « au hasard » deux boules de l'urne B dans l'urne A puis on tire ensuite « au hasard » une boule dans l'urne A. On cherche

1. la probabilité que la boule tirée soit blanche,
2. la probabilité que l'une au moins des boules transférées soit blanche sachant que la boule tirée est blanche.

Solution. On suppose construit un espace probabilisé (Ω, \mathcal{A}, P) permettant de traduire les tirages uniformes successifs.

1. Si X_1 désigne la variable aléatoire donnant le nombre de boules blanches transférées et si B est l'événement « une boule blanche est tirée dans l'urne A », on peut écrire :

$$P(X_1 = 2) = \frac{\binom{4}{2}}{\binom{12}{2}}, \quad P(X_1 = 1) = \frac{\binom{8}{1}\binom{4}{1}}{\binom{12}{2}}, \quad P(X_1 = 0) = \frac{\binom{8}{2}}{\binom{12}{2}},$$

soit :

$$P(X_1 = 2) = \frac{1}{11}, \quad P(X_1 = 1) = \frac{16}{33}, \quad P(X_1 = 0) = \frac{14}{33}.$$

Par ailleurs, les tirages dans l'urne A étant uniformes, on a :

$$P(B | X_1 = 2) = \frac{8}{13}, \quad P(B | X_1 = 1) = \frac{7}{13}, \quad P(B | X_1 = 0) = \frac{6}{13}.$$

On a alors, par la formule des probabilités totales :

$$P(B) = \sum_{j=0}^2 P(B | X_1 = j) P(X_1 = j),$$

soit, tous calculs faits :

$$P(B) = \frac{220}{429}.$$

2. On doit calculer $P(X_1 \geq 1 | B)$. On a :

$$P(X_1 = 2 | B) = \frac{P(B | X_1 = 2) P(X_1 = 2)}{P(B)} = \frac{\frac{1}{11} \frac{8}{13}}{\frac{220}{429}},$$

soit :

$$P(X_1 = 2 | B) = \frac{6}{55},$$

et

$$P(X_1 = 1 | B) = \frac{P(B | X_1 = 1) P(X_1 = 1)}{P(B)} = \frac{\frac{16}{33} \frac{7}{13}}{\frac{220}{429}},$$

soit :

$$P(X_1 = 1 | B) = \frac{28}{55}.$$

Par additivité d'une probabilité conditionnelle, il vient alors :

$$P(X_1 \geq 1 | B) = P(X_1 = 1 | B) + P(X_1 = 2 | B),$$

soit :

$$P(X_1 \geq 1 | B) = \frac{34}{55}.$$

Exercice 4.3. Un modèle démographique simple. Dans une population donnée, on suppose que la probabilité p_k pour qu'une famille ait k enfants est définie par :

$$p_0 = p_1 = a \quad \text{et, pour } k \geq 2, \quad p_k = (1 - 2a) 2^{-(k-1)},$$

où a est un réel strictement compris entre 0 et 1. On suppose de plus que la probabilité d'avoir un garçon ou une fille lors d'une naissance est la même.

1. Quelle est la probabilité pour qu'une famille ayant deux garçons ait deux enfants seulement ?
2. Quelle est la probabilité pour qu'une famille ait deux filles sachant qu'elle a deux garçons ?

Solution. On suppose construit un espace probabilisé (Ω, \mathcal{A}, P) qui soit tel qu'en notant, pour tout entier n , respectivement E_n , G_n et F_n les événements « la famille se compose de n enfants », « la famille a n garçons » et « la famille a n filles », on ait, pour tout $n \in \mathbb{N}$:

$$P(E_n) = p_n.$$

1. On doit calculer la probabilité conditionnelle $P(E_2 | G_2)$. On a

$$P(E_2 | G_2) = \frac{P(E_2) P(G_2 | E_2)}{P(G_2)}, \quad (4.8)$$

et :

$$P(G_2 | E_n) = \binom{n}{2} \frac{1}{2^n}.$$

Le théorème des probabilités totales donne ici :

$$P(G_2) = \sum_{n=2}^{+\infty} P(E_n) P(G_2 | E_n);$$

évidemment, on a

$$P(G_2 | E_0) = P(G_2 | E_1) = 0.$$

On a donc :

$$P(G_2) = \sum_{n=2}^{+\infty} p_n \binom{n}{2} \frac{1}{2^n} = \frac{(1-2a)}{4^2} \sum_{n=2}^{+\infty} n(n-1) \frac{1}{4^{n-2}}.$$

Reste à calculer cette somme ; or si $x \in [0, 1[$, le théorème de dérivation terme à terme des séries entières donne :

$$\sum_{n=2}^{+\infty} n(n-1) x^{n-2} = \frac{d^2}{dx^2} \left(\sum_{n=0}^{+\infty} x^n \right) = \frac{d^2}{dx^2} \left(\frac{1}{1-x} \right) = \frac{2}{(1-x)^3}.$$

On a donc, après simplifications :

$$P(G_2) = \frac{8(1-2a)}{27}.$$

On a alors, en reportant dans l'égalité (4.8) :

$$P(E_2 | G_2) = \frac{p_2 \binom{2}{2} \frac{1}{2^2}}{\frac{8(1-2a)}{27}},$$

soit, après simplification :

$$P(E_2 | G_2) = \frac{27}{64} \simeq 0,421.$$

2. On calcule maintenant la probabilité conditionnelle $P(F_2 | G_2)$. Toujours d'après le théorème des probabilités totales, on a

$$P(F_2 \cap G_2) = \sum_{n=0}^{+\infty} P(E_n) P(F_2 \cap G_2 | E_n),$$

soit, puisque toutes les probabilités conditionnelles sont nulles sauf celle pour $n = 4$:

$$P(F_2 \cap G_2) = P(E_4) P(F_2 \cap G_2 | E_4).$$

On obtient donc numériquement

$$\begin{aligned} P(F_2 \cap G_2) &= (1-2a) \frac{1}{2^3} \binom{4}{2} \frac{1}{2^4} \\ &= \frac{3(1-2a)}{64}. \end{aligned}$$

Il reste à utiliser la définition de la probabilité conditionnelle

$$P(F_2 | G_2) = \frac{P(F_2 \cap G_2)}{P(G_2)}$$

pour obtenir :

$$P(F_2 | G_2) = \frac{81}{512} \simeq 0,158.$$

Exercice 4.4. Probabilités conditionnelles en cascade. Dans une urne se trouvent n boules rouges et n boules blanches. On tire deux par deux, sans remise, les boules jusqu'à vider l'urne. Quelle est la probabilité que l'on tire une boule de chaque couleur à chaque tirage ?

Solution. On suppose construit un espace probabilisé (Ω, \mathcal{A}, P) qui soit tel qu'à chaque étape les tirages soient uniformes. Pour $j = 1, 2, \dots, n$ appelons E_j l'événement « au j -ième tirage, on obtient une boule de chaque couleur ». On a, pour tout $j = 1, 2, \dots, n-1$:

$$P\left(E_{j+1} \mid \bigcap_{i=1}^j E_i\right) = \frac{(n-j)^2}{\binom{2(n-j)}{2}}.$$

De plus, on a

$$P(E_1) = \frac{n^2}{\binom{2n}{2}}.$$

La formule des probabilités totales en cascade donne ici

$$P\left(\bigcap_{i=1}^n E_j\right) = P(E_1) \prod_{j=1}^{n-1} P\left(E_{j+1} \mid \bigcap_{i=1}^j E_i\right).$$

Il vient donc, après simplifications :

$$P\left(\bigcap_{i=1}^n E_j\right) = \frac{2^n (n!)^2}{(2n)!}.$$

Exercice 4.5. Probabilités conditionnelles en cascade et temps de succès ; loi binomiale négative. Soit (Ω, \mathcal{A}, P) un espace probabilisé et $(A_n)_{n \in \mathbb{N}^*}$ une suite d'événements indépendants de même probabilité p (on note $q = 1 - p$). Avec la convention $\inf \emptyset = +\infty$, on définit la suite de variables aléatoires $(T_n)_{n \in \mathbb{N}^*}$ en posant, pour tout $\omega \in \Omega$:

$$\begin{aligned} T_1(\omega) &= \inf \{j \geq 1 \mid \omega \in A_j\}, \\ T_2(\omega) &= \inf \{j > T_1(\omega) \mid \omega \in A_j\}, \\ &\dots \\ T_{n+1}(\omega) &= \inf \{j > T_n(\omega) \mid \omega \in A_j\}. \end{aligned}$$

1. (a) Calculer, pour tout $k \in \mathbb{N}^*$, la loi de la variable aléatoire T_k .
- (b) Pour toute suite strictement croissante d'entiers n_1, n_2, \dots, n_{k+1} , calculer la probabilité conditionnelle :

$$P(T_{k+1} - T_k = n_{k+1} - n_k \mid T_1 = n_1, \dots, T_k = n_k).$$

- (c) En déduire que, si $n_1 < n_2 < \dots < n_{k+1}$, on a :

$$P(T_{k+1} = n_{k+1} \mid T_1 = n_1, \dots, T_k = n_k) = P(T_{k+1} = n_{k+1} \mid T_k = n_k).$$

- (d) Que vaut la probabilité $P(T_1 = n_1, \dots, T_k = n_k)$?

2. Démontrer que les variables aléatoires $T_1, T_2 - T_1, \dots, T_k - T_{k-1}$ sont indépendantes, déterminer leur loi et retrouver ainsi le fait que la k -ième convolution de la loi géométrique $\mathcal{G}_{\mathbb{N}^*}(p)$ (c'est-à-dire la loi de la somme de k variables indépendantes de même loi $\mathcal{G}_{\mathbb{N}^*}(p)$) est la loi binomiale négative $\mathcal{B}^-(k, p)$.

3. Calculer la probabilité $P(T_1 = 3, T_5 = 9, T_7 = 17)$. En donner une valeur numérique avec trois chiffres significatifs pour les valeurs de p suivantes : 0,3 ; 0,5 ; 0,8.

Exemple de situation modélisée dans ce cadre : jeu de pile ou face. Si A_n est l'événement « on obtient un pile au n -ième jet », la variable

aléatoire T_k représente le numéro du jet qui fait apparaître pour la k -ième fois un pile lors d'une succession de jets de la pièce.

Plus généralement, cela s'applique à toute suite d'expériences indépendantes à l'issue desquelles une propriété peut être réalisée ou non et au cours desquelles la probabilité p de réalisation de la propriété est indépendante du numéro de l'expérience. Dans ce cas A_n est l'événement « la propriété étudiée est réalisée à la n -ième expérience » ; la variable aléatoire T_k représente le numéro de l'expérience qui fait apparaître pour la k -ième fois la propriété étudiée.

Solution. Commençons par observer qu'on peut écrire

$$T_{n+1} = \inf (j > T_n \mid \mathbf{1}_{A_j} = 1).$$

1. (a) On sait déjà que la loi de T_1 est la loi géométrique sur \mathbb{N}^* de paramètre p . Si $l \geq k \geq 2$, on a :

$$(T_k = l) = \left(\sum_{j=1}^{l-1} \mathbf{1}_{A_j} = k-1 \right) \cap (\mathbf{1}_{A_l} = 1).$$

Les événements A_n étant indépendants et de même probabilité p , la loi de la variable aléatoire $\sum_{j=1}^{l-1} \mathbf{1}_{A_j}$ est la loi binomiale $\mathcal{B}(l-1, p)$ et les événements

$\left(\sum_{j=1}^{l-1} \mathbf{1}_{A_j} = k-1 \right)$ et $(\mathbf{1}_{A_l} = 1)$ sont indépendants. Il en résulte que l'on a

$$\boxed{P(T_k = l) = \binom{l-1}{k-1} p^k q^{l-k} .}$$

La loi de la variable aléatoire T_k est la **loi binomiale négative** de paramètres k et p , notée $\mathcal{B}^-(k, p)$.

(b) Avec les notations de l'énoncé, on peut écrire :

$$\left[\bigcap_{j=1}^k (T_j = n_j) \right] \cap (T_{k+1} - T_k = n_{k+1} - n_k) = \\ \left[\bigcap_{j=1}^k (T_j = n_j) \right] \cap \left(\sum_{j=n_k+1}^{n_{k+1}-1} \mathbf{1}_{A_j} = 0 \right) \cap (\mathbf{1}_{A_{n_{k+1}}} = 1).$$

Toujours par indépendance des événements figurant au second membre de cette dernière égalité, on a

$$P(T_{k+1} - T_k = n_{k+1} - n_k \mid T_1 = n_1, \dots, T_k = n_k) = \\ P \left[\left(\sum_{j=n_k+1}^{n_{k+1}-1} \mathbf{1}_{A_j} = 0 \right) \cap (\mathbf{1}_{A_{n_{k+1}}} = 1) \right],$$

soit :

$$\boxed{P(T_{k+1} - T_k = n_{k+1} - n_k \mid T_1 = n_1, \dots, T_k = n_k) = pq^{n_{k+1} - n_k - 1}} \quad (4.9)$$

(c) Un calcul analogue démontre que :

$$P(T_{k+1} - T_k = n_{k+1} - n_k \mid T_k = n_k) = pq^{n_{k+1} - n_k - 1}.$$

Il en résulte, puisque l'on a

$$\begin{aligned} & P(T_{k+1} = n_{k+1} \mid T_1 = n_1, \dots, T_k = n_k) = \\ & P(T_{k+1} - T_k = n_{k+1} - n_k \mid T_1 = n_1, \dots, T_k = n_k) \end{aligned}$$

et

$$P(T_{k+1} = n_{k+1} \mid T_k = n_k) = P(T_{k+1} - T_k = n_{k+1} - n_k \mid T_k = n_k),$$

que :

$$\boxed{P(T_{k+1} = n_{k+1} \mid T_1 = n_1, \dots, T_k = n_k) = P(T_{k+1} = n_{k+1} \mid T_k = n_k)}.$$

Remarque. Cette relation exprime le caractère **markovien** de la suite des variables aléatoires T_k , c'est-à-dire que l'évolution probabiliste de la suite à l'instant $k+1$ ne dépend que de sa valeur prise à l'instant k , et non de toutes les valeurs prises aux instants précédents. La valeur commune de ces probabilités conditionnelles est donnée par :

$$\boxed{P(T_{k+1} = n_{k+1} \mid T_k = n_k) = pq^{n_{k+1} - n_k - 1}}.$$

(d) Si $n_1 < n_2 < \dots < n_k$, il résulte de la formule des probabilités en cascade que l'on a

$$\begin{aligned} & P(T_1 = n_1, \dots, T_k = n_k) = \\ & P(T_1 = n_1) P(T_2 = n_2 \mid T_1 = n_1) \dots P\left[(T_k = n_k) \mid \bigcap_{j=1}^{k-1} (T_j = n_j)\right], \end{aligned}$$

soit, d'après la question précédente,

$$P(T_1 = n_1, \dots, T_k = n_k) = pq^{n_1-1} pq^{n_2-n_1-1} \dots pq^{n_k-n_{k-1}-1},$$

ou encore, après simplifications :

$$\boxed{P(T_1 = n_1, \dots, T_k = n_k) = p^k q^{n_k - k}}.$$

2. De l'égalité (4.9) résulte que, pour tout $m \geq 1$, on a :

$$P(T_{k+1} - T_k = m \mid T_1 = n_1, \dots, T_k = n_k) = pq^{m-1}.$$

Donc, pour tout $k \geq 1$, les variables aléatoires $T_{k+1} - T_k$ et (T_1, T_2, \dots, T_k) sont indépendantes. Les variables aléatoires $T_1, T_2 - T_1, \dots, T_k - T_{k-1}$ sont alors aussi indépendantes; elles sont de plus de loi géométrique sur \mathbb{N}^* , notée $\mathcal{G}_{\mathbb{N}^*}(p)$. En remarquant que l'on a

$$T_k = T_1 + \sum_{j=1}^{k-1} (T_{j+1} - T_j),$$

il résulte de la définition de la convolution de lois (loi de somme de variables aléatoires indépendantes) que la k -ième convolution de la loi géométrique sur \mathbb{N}^* est la loi binomiale négative $\mathcal{B}^-(k, p)$.

3. Par indépendance des variables aléatoires $T_1, T_5 - T_1, T_7 - T_5$, il vient alors :

$$\begin{aligned} P(T_1 = 3, T_5 = 9, T_7 = 17) \\ &= P(T_1 = 3, T_5 - T_1 = 6, T_7 - T_5 = 8) \\ &= P(T_1 = 3) P(T_5 - T_1 = 6) P(T_7 - T_5 = 8). \end{aligned}$$

Puisque l'on a

$$T_5 - T_1 = \sum_{j=2}^5 (T_j - T_{j-1}),$$

la loi de $T_5 - T_1$ est la loi binomiale négative $\mathcal{B}^-(4, p)$. De même, puisque l'on a

$$T_7 - T_5 = \sum_{j=6}^7 (T_j - T_{j-1}),$$

la loi de $T_7 - T_5$ est la loi binomiale négative $\mathcal{B}^-(2, p)$. Il vient alors

$$P(T_1 = 3, T_5 = 9, T_7 = 17) = pq^{3-1} \binom{6-1}{4-1} p^4 q^2 \left(\binom{8-1}{2-1} p^2 q^6 \right),$$

soit encore :

$$P(T_1 = 3, T_5 = 9, T_7 = 17) = 70p^7 q^{10}.$$

Les valeurs approchées de ces probabilités pour les valeurs de p indiquées sont données dans le tableau ci-dessous :

p	0,3	0,5	0,8
$P(T_1 = 3, T_5 = 9, T_7 = 17)$	$4,32 \cdot 10^{-4}$	$5,34 \cdot 10^{-4}$	$1,5 \cdot 10^{-6}$

Exercice 4.6. Construction d'une probabilité sur un espace produit et formule des probabilités totales. On dispose de trois cartes ; la première a ses deux faces rouges, la deuxième ses deux faces blanches, la troisième une face blanche et l'autre rouge. Un distributeur tire « au hasard » une carte puis choisit « au hasard » d'exposer une face au \hat{i} qui doit parier sur la couleur de la face cachée. On demande quelle est la meilleure stratégie du \hat{i} .

Solution. On modélise cette expérience à deux étapes de la façon suivante : pour modéliser la première étape, c'est-à-dire le choix de la carte, on choisit l'espace probabilisé $(\Omega_1, \mathcal{P}(\Omega_1), P_1)$ où

$$\Omega_1 = \{RR, BB, RB\},$$

et P_1 est la probabilité uniforme. Ensuite, on définit l'ensemble $\Omega_2 = \{R, B\}$ permettant de donner la couleur de la face exposée. On définit les germes de probabilité suivants : le germe p_1 qui engendre la probabilité uniforme sur

Ω_1 et qui est défini, pour tout $\omega_1 \in \Omega_1$, par $p_1(\omega_1) = \frac{1}{3}$, puis la famille de germes sur Ω_2 définis par :

$$\begin{aligned} p_2^{\text{RR}}(\text{R}) &= 1 & \text{et} & & p_2^{\text{RR}}(\text{B}) &= 0, \\ p_2^{\text{BB}}(\text{R}) &= 0 & \text{et} & & p_2^{\text{BB}}(\text{B}) &= 1, \\ p_2^{\text{RB}}(\text{R}) &= \frac{1}{2} & \text{et} & & p_2^{\text{RB}}(\text{B}) &= \frac{1}{2}. \end{aligned}$$

Cette famille de germes de probabilité définit une probabilité unique P sur l'espace probabilisable $(\Omega_1 \times \Omega_2, \mathcal{P}(\Omega_1 \times \Omega_2))$ par la relation :

$$\forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 \quad P(\{\omega_1, \omega_2\}) = p_1(\omega_1) p_2^{\omega_1}(\omega_2).$$

Soit C_R (respectivement C_B) l'événement « la face cachée est rouge » (respectivement « la face cachée est blanche »); notons X_1 et X_2 les projections canoniques de $\Omega_1 \times \Omega_2$ sur Ω_1 et Ω_2 . On cherche d'abord la probabilité conditionnelle $P(C_R \mid X_2 = \text{R})$.

L'événement C_R se partitionne de la manière suivante

$$C_R = (X_1 = \text{RR}) \bigsqcup [(X_1 = \text{RB}) \cap (X_2 = \text{B})],$$

ce qui permet d'écrire, par additivité d'une probabilité conditionnelle :

$$\begin{aligned} P(C_R \mid X_2 = \text{R}) \\ = P((X_1 = \text{RR}) \mid X_2 = \text{R}) + P((X_1 = \text{RB}) \cap (X_2 = \text{B}) \mid X_2 = \text{R}). \end{aligned}$$

Mais on a

$$P[(X_1 = \text{RB}) \cap (X_2 = \text{B}) \mid X_2 = \text{R}] = 0$$

et

$$P(X_1 = \text{RR} \mid X_2 = \text{R}) = P(X_2 = \text{R} \mid X_1 = \text{RR}) \frac{P(X_1 = \text{RR})}{P(X_2 = \text{R})}.$$

Il en résulte que :

$$P(C_R \mid X_2 = \text{R}) = \frac{1}{3}.$$

De plus, d'après le théorème des probabilités totales, on a

$$\begin{aligned} P(X_2 = \text{R}) &= P(X_2 = \text{R} \mid X_1 = \text{RR}) P(X_1 = \text{RR}) \\ &\quad + P(X_2 = \text{R} \mid X_1 = \text{BB}) P(X_1 = \text{BB}) \\ &\quad + P(X_2 = \text{R} \mid X_1 = \text{RB}) P(X_1 = \text{RB}), \end{aligned}$$

soit encore :

$$P(X_2 = \text{R}) = \frac{1}{3} + \left(\frac{1}{2} \times \frac{1}{3}\right) = \frac{1}{2}.$$

On obtient donc

$$\boxed{P(C_R \mid X_2 = \text{R}) = \frac{2}{3}},$$

et par conséquent :

$$P(C_B \mid X_2 = \text{R}) = \frac{1}{3}.$$

Par échange des rôles de R et B on obtient de même :

$$P(C_B | X_2 = B) = \frac{2}{3} \quad \text{et} \quad P(C_R | X_2 = B) = \frac{1}{3}.$$

La meilleure stratégie du $\hat{1}$ est donc de parier sur la couleur de la face exposée.

Exercice 4.7. Variables aléatoires de loi géométriques, conditionnement et indépendance.

Toutes les variables aléatoires introduites sont définies sur le même espace probabilisable (Ω, \mathcal{A}) . On se donne une probabilité P sur cet espace.

Soient q et r deux réels strictement compris entre 0 et 1. On considère deux variables aléatoires discrètes U et V , indépendantes, de lois géométriques sur \mathbb{N}^* de paramètres respectifs $1 - q$ et $1 - r$; ces lois sont notées respectivement $\mathcal{G}_{\mathbb{N}^*}(1 - q)$ et $\mathcal{G}_{\mathbb{N}^*}(1 - r)$.

1. Le calcul des probabilités suivantes est intéressant, par exemple, pour comparer les temps de succès de deux joueurs jouant simultanément dans des conditions qui peuvent être différentes.

(a) Calculer $P(U < V)$.

(b) Que vaut cette probabilité dans le cas où $q = r$, puis dans celui où $q = r = \frac{1}{2}$?

2. On étudie la loi de U sous la probabilité conditionnelle $P^{(U < V)}$.

(a) Calculer, pour tout $k \in \mathbb{N}^*$, la probabilité conditionnelle :

$$P(U = k | U < V).$$

Identifier la loi conditionnelle de U sachant que « $U < V$ », c'est-à-dire la probabilité sur \mathbb{N}^* engendrée par le germe : $k \mapsto P(U = k | U < V)$.

(b) Déterminer, pour tout $k \in \mathbb{N}$, la probabilité conditionnelle

$$P(U > k | U < V).$$

3. Soient X_1 et X_2 deux variables aléatoires discrètes indépendantes de même loi géométrique sur \mathbb{N}^* de paramètre q . On définit les ensembles

$$A_1 = \{\omega \in \Omega \mid X_1(\omega) < X_2(\omega)\} \quad A_2 = \{\omega \in \Omega \mid X_2(\omega) < X_1(\omega)\}$$

et

$$H = \{\omega \in \Omega \mid X_1(\omega) \neq X_2(\omega)\}.$$

On considère les variables aléatoires

$$X = \min(X_1, X_2) \quad \text{et} \quad J = \mathbf{1}_{A_1} + 2\mathbf{1}_{A_2}$$

(c'est-à-dire que $J(\omega)$ vaut 1 si $\omega \in A_1$, 2 si $\omega \in A_2$ et 0 sinon).

On note P^H la probabilité (conditionnelle) sur l'espace probabilisable (Ω, \mathcal{A}) définie, pour tout $A \in \mathcal{A}$, par $P^H(A) = P(A | H)$.

(a) Calculer, pour $k \in \mathbb{N}^*$, la probabilité $P[H \cap (J = 1) \cap (X > k)]$.

(b) Justifier l'égalité $P(A_1) = P(A_2)$ et en déduire, pour tout $k \in \mathbb{N}$, la valeur de la probabilité $P^H[(J = 1) \cap (X = k)]$.

(c) Que vaut alors $P^H[(J = 1)]$?

4. On étudie une propriété d'indépendance des variables aléatoires X et J sous la probabilité P^H .

(a) Remarquer que l'on a l'égalité

$$P[H \cap (J = 1) \cap (X > k)] = P[H \cap (J = 2) \cap (X > k)]$$

et déduire de ce qui précède la valeur de la probabilité $P^H(X > k)$.

(b) Montrer que les variables aléatoires X et J sont indépendantes lorsqu'on les considère comme définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P^H)$.

Exemple de situation modélisée dans ce cadre : jeu de pile ou face à deux joueurs.

Deux joueurs jouent à pile ou face avec une pièce identique, et lancent alternativement leur pièce. La variable aléatoire X_1 (respectivement X_2) est le numéro du jet à l'issue duquel le premier (respectivement le deuxième) joueur obtient pile pour la première fois (les jets sont comptés indépendamment pour chacun des joueurs). La variable aléatoire X est le numéro du jet à l'issue duquel l'un des deux joueurs obtient pile pour la première fois ; la variable aléatoire J vaut 1 (respectivement 2) si le premier (respectivement le deuxième) joueur obtient pile en premier, vaut 0 si les deux joueurs obtiennent pile en un même nombre de coups. On démontre que **conditionnellement** au fait qu'ils n'obtiennent pas pile en un même nombre de coups, les variables aléatoires X et J sont **indépendantes**.

Solution.

1. On partitionne l'événement étudié.

(a) On a

$$(U < V) = \bigsqcup_{\substack{1 \leq u < v \\ (u, v) \in \mathbb{N}^* \times \mathbb{N}^*}} [(U = u) \cap (V = v)],$$

ce qui donne, par σ -additivité de P et indépendance des variables aléatoires U et V :

$$P(U < V) = \sum_{\substack{1 \leq u < v \\ (u, v) \in \mathbb{N}^* \times \mathbb{N}^*}} (1 - q) q^{u-1} (1 - r) r^{v-1}.$$

Puisque la famille sous le signe somme est positive, la propriété de Fubini permet d'écrire :

$$\begin{aligned} P(U < V) &= (1-q)(1-r) \sum_{u=1}^{+\infty} \left[q^{u-1} \left(\sum_{v=u+1}^{+\infty} r^{v-1} \right) \right] \\ &= (1-q)(1-r) \sum_{u=1}^{+\infty} \frac{r^u}{1-r} q^{u-1}, \end{aligned}$$

soit encore :

$$P(U < V) = \frac{(1-q)r}{1-rq}.$$

(b) On trouve en particulier :

$$\text{si } q = r, \quad P(U < V) = \frac{q}{1+q},$$

$$\text{si } q = r = \frac{1}{2}, \quad P(U < V) = \frac{1}{3}.$$

2. L'événement $(U < V)$ est de probabilité non nulle.

(a) On a

$$P(U = k \mid U < V) = \frac{P[(U = k) \cap (k < V)]}{P(U < V)},$$

soit, par indépendance de U et V :

$$P(U = k \mid U < V) = \frac{P(U = k) P(k < V)}{P(U < V)}.$$

Mais on a

$$P(k < V) = \sum_{v=k+1}^{+\infty} (1-r)r^{v-1} = r^k \quad (4.10)$$

et donc, en utilisant la première question :

$$P(U = k \mid U < V) = \frac{(1-q) q^{k-1} r^k}{\frac{(1-q)r}{1-rq}}.$$

On a établi que, pour tout $k \in \mathbb{N}^*$, on a :

$$P(U = k \mid U < V) = (1-rq)(rq)^{k-1}.$$

Ceci démontre que la **loi conditionnelle** engendrée par le germe $P(U = \cdot \mid U < V)$ est la **loi géométrique** sur \mathbb{N}^* de paramètre $1-rq$, notée $\mathcal{G}_{\mathbb{N}^*}(1-rq)$.

(b) Il résulte alors de l'égalité (4.10) que l'on a, pour tout $k \in \mathbb{N}$:

$$P(U > k \mid U < V) = (rq)^k.$$

3. Cette question fait apparaître certaines propriétés des variables aléatoires J et X sous le conditionnement par l'événement H .

(a) Puisque $(J = 1) \subset H$, et que, sur l'ensemble $(J = 1)$, on a $X = X_1$, il vient

$$P[H \cap (J = 1) \cap (X > k)] = P[(J = 1) \cap (X_1 > k)],$$

et donc :

$$P[H \cap (J = 1) \cap (X > k)] = P(X_1 > k \mid X_1 < X_2) P(X_1 < X_2). \quad (4.11)$$

En transposant les résultats des deux premières questions à ce cas particulier, il vient

$$P(X_1 > k \mid X_1 < X_2) = (1 - q)^{2k} \quad \text{et} \quad P(X_1 < X_2) = \frac{1 - q}{2 - q},$$

et donc :

$$P[H \cap (J = 1) \cap (X > k)] = \frac{(1 - q)^{2k+1}}{2 - q}.$$

(b) Les variables aléatoires X_1 et X_2 sont indépendantes et de même loi ; on peut alors écrire :

$$\begin{aligned} P(A_1) &= \sum_{x_1 < x_2} P_{X_1}(\{x_1\}) P_{X_2}(\{x_2\}) \\ &= \sum_{x_1 < x_2} P_{X_2}(\{x_1\}) P_{X_1}(\{x_2\}), \end{aligned}$$

ce qui démontre l'égalité :

$$P(A_1) = P(A_2).$$

On en déduit que l'on a

$$P(H) = 2P(A_1),$$

et donc que, d'après l'égalité (4.11) :

$$P^H[(J = 1) \cap (X > k)] = \frac{1}{2} P(X_1 > k \mid X_1 < X_2).$$

Il en résulte que, pour tout $k \in \mathbb{N}$:

$$P^H[(J = 1) \cap (X > k)] = \frac{1}{2} (1 - q)^{2k}.$$

(c) La définition d'une probabilité conditionnelle donne

$$P^H(J = 1) = \frac{P[(J = 1) \cap H]}{P(H)} = \frac{P(A_1)}{P(H)},$$

soit :

$$P^H(J = 1) = \frac{1}{2}. \quad (4.12)$$

4. On étudie les lois des variables aléatoires X et J sous la probabilité P^H et on conclut à l'indépendance sous P^H de ces variables aléatoires.

(a) On a les égalités d'ensembles :

$$H \cap (J = 1) \cap (X > k) = (X_1 > k) \cap (X_1 < X_2)$$

et

$$H \cap (J = 2) \cap (X > k) = (X_2 > k) \cap (X_2 < X_1).$$

Les variables aléatoires X_1 et X_2 étant indépendantes et de même loi, un calcul analogue à celui de la question 3b assure que ces ensembles ont même probabilité. De plus, on remarque que

$$H \cap (J = 0) = \emptyset.$$

On a donc

$$\begin{aligned} P^H(X > k) &= P^H[(J = 1) \cap (X > k)] + P^H[(J = 2) \cap (X > k)] \\ &= 2 P^H[(J = 1) \cap (X > k)], \end{aligned}$$

soit encore :

$$P^H(X > k) = (1 - q)^{2k}.$$

(b) Puisque $P^H(J = 0) = 0$, il résulte de l'égalité (4.12) que :

$$P^H(J = 2) = \frac{1}{2}.$$

En rassemblant les résultats des questions précédentes, on a alors démontré que, pour $j = 0, 1, 2$ et pour tout $k \in \mathbb{N}$, on a

$$P^H[(J = j) \cap (X > k)] = P^H(J = j) P^H(X > k),$$

et donc que, **sous la probabilité P^H , les variables aléatoires X et J sont indépendantes.**

Chapitre 5

Moments d'une variable aléatoire discrète

Les moments d'une variable aléatoire sont des paramètres numériques qui donnent des renseignements sur la loi de cette variable aléatoire (sans toutefois, en général, la déterminer complètement). Les plus couramment utilisés sont la moyenne, ou espérance mathématique, et la variance.

Dans ce chapitre, nous n'introduisons les moments que pour des variables aléatoires réelles discrètes, le procédé de sommation étant la sommation des familles de nombres réels. Dans le chapitre sur les variables aléatoires à densité, nous définirons ces mêmes notions; formellement, il suffira de changer le procédé de sommation, la somme Σ étant remplacée par l'intégrale \int . L'étude générale qui permet d'unifier ces deux cas relève de la théorie de la mesure et de l'intégration; elle sera abordée dans le tome II.

Dans ce chapitre, sauf mention du contraire, toutes les variables aléatoires sont définies sur un espace probabilisé (Ω, \mathcal{A}, P) et sont **discrètes**. Elles sont de plus à valeurs réelles, finies ou infinies. (Dans la pratique, les variables considérées seront toujours finies avec probabilité 1.)

5.1. Moyenne, ou espérance mathématique

5.1.1. Définition

Définition 5.1. Soit X une variable aléatoire réelle discrète. Si la famille de nombres réels $(x P(X = x))_{x \in \mathbb{R}}$ est sommable, autrement dit si la somme $\sum_{x \in \mathbb{R}} |x| P(X = x)$ est finie, on dit que la variable aléatoire X possède une moyenne, et le nombre

$$\sum_{x \in \mathbb{R}} x P(X = x)$$

est appelé **moyenne, ou espérance mathématique, de la variable aléatoire X** . Ce nombre est noté, suivant les besoins typographiques, EX ou $E(X)$.

Remarques. 1. La notion de moyenne est définie ici sans faire référence à l'espace probabilisé (Ω, \mathcal{A}, P) et apparaît comme une notion relative à la loi de la variable aléatoire.

2. Puisque X est supposée discrète, l'ensemble de ses valeurs $X(\Omega)$ est un ensemble dénombrable; l'ensemble $\text{val}(X)$ de ses valeurs prises avec une probabilité non nulle est donc aussi dénombrable.

On peut bien sûr dans la définition de EX restreindre la sommation à l'ensemble $\text{val}(X)$ constitué des réels x tels que $P(X = x) > 0$. On écrit donc

$$EX = \sum_{x \in \text{val}(X)} x P(X = x).$$

C'est cette formule qui sert pour le calcul de la moyenne à partir de la loi de X . Le calcul explicite se fait en remarquant que, selon que l'ensemble des valeurs de X est fini ou non, la définition de EX prend l'une des formes suivantes :

◇ si $\text{val}(X)$ est fini et s'énumère sous la forme $\{x_1, x_2, \dots, x_n\}$, alors X a une moyenne et on a

$$EX = \sum_{i=1}^n x_i P(X = x_i); \quad (5.1)$$

◇ si $\text{val}(X)$ est infini et énuméré sous la forme $\{x_n \mid n \in \mathbb{N}\}$, il faut et il suffit, pour que X admette une moyenne, que l'on ait $\sum_{i=0}^{+\infty} |x_i| P(X = x_i) < +\infty$ (autrement dit, que la série soit absolument convergente), et on a dans ce cas :

$$EX = \sum_{i=0}^{+\infty} x_i P(X = x_i). \quad (5.2)$$

L'existence de la moyenne et sa valeur ne dépendent pas de l'ordre d'énumération choisie pour les valeurs de X . (Voir pour tout ceci le théorème 2.5 et les commentaires qui l'accompagnent.)

3. On peut enfin observer que pour une variable aléatoire $X \geq 0$, la somme $\sum_{x \in \text{val}(X)} xP(X = x)$ a toujours un sens, même si elle peut prendre la valeur $+\infty$. On écrit dans ce cas $EX = +\infty$. Avec cette convention, la condition pour que X possède une moyenne s'écrit $E|X| < +\infty$.

En langage courant, la moyenne EX n'est autre que la **moyenne pondérée** des valeurs x prises par la variable aléatoire X , valeurs

pondérées par la probabilité que X prenne la valeur x .

5.1.2. Premières propriétés de l'espérance

Les propriétés suivantes de l'espérance sont utilisées en permanence.

1. Soit un réel a ; si la variable aléatoire est telle que $P(X = a) = 1$ (on dit dans ce cas que X est constante avec probabilité 1), elle admet une moyenne égale à a . En particulier, on a :

$$E(a) = a.$$

2. Toute variable aléatoire X discrète **bornée** (au sens où il existe un réel positif M tel que l'on ait, pour tout $\omega \in \Omega$, $|X(\omega)| \leq M$) admet une moyenne.

3. Si $X \geq 0$, alors $EX \geq 0$. Si $X \geq 0$ et $EX = 0$, alors $X = 0$ avec probabilité 1. En effet, on a $\sum_{x>0} x P(X = x) = 0$, donc $x P(X = x) = 0$, et, par suite, $P(X = x) = 0$ pour tout réel $x > 0$. Puisque X est discrète, on a donc $P(X > 0) = \sum_{x>0} P(X = x) = 0$.

4. Si X et Y possèdent une moyenne et vérifient $X \leq Y$, on a : $EX \leq EY$. Enfin, on a l'inégalité :

$$|EX| \leq E|X|.$$

Ces propriétés résultent des propriétés analogues pour les familles sommables.

Exemple 5.1. Espérance d'une variable aléatoire de loi uniforme. Si X est une variable aléatoire réelle discrète, de loi la probabilité uniforme¹ sur l'ensemble $\{x_1, x_2, \dots, x_n\}$, sa moyenne EX est la moyenne arithmétique des valeurs x prises par X :

$$EX = \frac{1}{n} \sum_{i=1}^n x_i.$$

Exemple 5.2. Espérance d'une fonction indicatrice. Espérance d'une loi de Bernoulli. Soit un événement $A \in \mathcal{A}$; sa fonction indicatrice $\mathbf{1}_A$ est une variable aléatoire réelle discrète dont la loi est déterminée par les relations

$$P(\mathbf{1}_A = 1) = P(A), \quad P(\mathbf{1}_A = 0) = 1 - P(A);$$

on a immédiatement

$$E \mathbf{1}_A = P(A).$$

1. On dit que la variable aléatoire X est **uniformément répartie** sur l'ensemble $\{x_1, x_2, \dots, x_n\}$.

Une autre façon de présenter le même résultat consiste à dire que si X est une variable de Bernoulli de paramètre p , on a

$$EX = p.$$

5.1.3. Espérance d'une fonction de variable aléatoire

Théorème 5.2. (Théorème de transfert.) Soit X une variable aléatoire discrète prenant ses valeurs dans l'ensemble E . Soit f une application de E dans \mathbb{R} . L'application composée $Y = f \circ X$ (couramment notée $f(X)$) est une variable aléatoire réelle discrète. Pour que Y admette une moyenne il faut et il suffit que la somme

$$E(f(X)) = \sum_{x \in E} |f(x)| P(X = x)$$

soit finie. Dans ce cas, on a :

$$\boxed{E(f(X)) = \sum_{x \in E} f(x) P(X = x).} \quad (5.3)$$

Grâce à cet énoncé, on peut, dans le calcul de $E(f(X))$, effectuer la sommation sur l'ensemble des valeurs de X , au lieu de l'effectuer sur l'ensemble des valeurs de $f(X)$. D'où la dénomination de *théorème de transfert*. Il suffit donc de connaître la loi de X pour s'assurer de l'existence de la moyenne de $f(X)$ et calculer sa valeur.

Démonstration. Soit $y \in \mathbb{R}$. L'événement $\{f(X) = y\}$ est la réunion disjointe des événements $\{X = x\}$ où x décrit $f^{-1}(y)$. Comme l'ensemble des valeurs prises par X est dénombrable, les antécédents x de y pour lesquels l'événement $\{X = x\}$ n'est pas vide forment un sous-ensemble dénombrable de $f^{-1}(y)$, qu'on notera G_y . En vertu de la σ -additivité de P , on a donc

$$P(f(X) = y) = \sum_{x \in G_y} P(X = x).$$

Puisque $P(X = x) = 0$ si $x \in f^{-1}(y) \setminus G_y$, on a aussi

$$P(f(X) = y) = \sum_{x \in f^{-1}(y)} P(X = x).$$

On peut alors écrire

$$|y| P(f(X) = y) = |y| \sum_{x \in f^{-1}(y)} P(X = x) = \sum_{x \in f^{-1}(y)} |f(x)| P(X = x)$$

Comme on a évidemment $E = \bigsqcup_{y \in \mathbb{R}} f^{-1}(y)$, on a, d'après la propriété de sommation par paquets,

$$\begin{aligned} \sum_{x \in E} |f(x)| P(X = x) &= \sum_{y \in \mathbb{R}} \left(\sum_{x \in f^{-1}(y)} |f(x)| P(X = x) \right) \\ &= \sum_{y \in \mathbb{R}} |y| P(f(X) = y). \end{aligned}$$

La finitude de la somme

$$\sum_{x \in E} |f(x)| P(X = x)$$

équivalent donc à celle de la somme

$$\sum_{y \in \mathbb{R}} |y| P(f(X) = y)$$

c'est-à-dire à l'existence de la moyenne de $f(X)$. Lorsque cette condition est réalisée, la sommation par paquets est licite, et on a alors

$$\begin{aligned} E(f(X)) &= \sum_{y \in \mathbb{R}} y P(f(X) = y) \\ &= \sum_{y \in \mathbb{R}} \left(\sum_{x \in f^{-1}(y)} f(x) P(X = x) \right) \\ &= \sum_{x \in E} f(x) P(X = x) \end{aligned}$$

Dans le cas où Ω est lui-même **dénombrable**, en prenant $E = \Omega$, et en prenant pour X l'application identique de Ω sur lui-même, l'application du théorème 5.2 conduit à l'énoncé suivant :

Corollaire 5.3. *Soit Y une variable aléatoire réelle discrète. Pour que Y ait une moyenne, il faut et il suffit que la famille $(Y(\omega) P(\{\omega\}))_{\omega \in \Omega}$ soit sommable et dans ce cas, on a :*

$$EY = \sum_{\omega \in \Omega} Y(\omega) P(\{\omega\}).$$

Notation. On note $\mathcal{L}_d^1(\Omega, \mathcal{A}, P)$ l'ensemble des variables aléatoires réelles discrètes définies sur l'espace probabilisé (Ω, \mathcal{A}, P) et qui admettent une moyenne, c'est-à-dire qui sont telles que $E|X| < +\infty$.

2. Il faut savoir que cette notation n'est pas « officielle ». De plus, on attire l'attention du lecteur du tome II sur le fait que, contrairement aux espaces \mathcal{L}^1 classiques qui y seront définis, ces ensembles n'ont aucune propriété **topologique** intéressante ; en particulier, ils ne sont pas complets.

Proposition 5.4. *L'ensemble $\mathcal{L}_d^1(\Omega, \mathcal{A}, P)$ est un espace vectoriel et l'application $: X \mapsto EX$ est une forme linéaire sur $\mathcal{L}_d^1(\Omega, \mathcal{A}, P)$.*

Démonstration. Soient X et Y deux éléments de $\mathcal{L}_d^1(\Omega, \mathcal{A}, P)$ et a et b deux réels quelconques. L'application $aX + bY$ est une variable aléatoire réelle discrète ; la variable aléatoire $Z = (X, Y)$ est discrète et à valeurs dans $E = X(\Omega) \times Y(\Omega)$. Soient π_1 et π_2 les projections canoniques, c'est-à-dire les applications définies sur E par : pour tout $(x, y) \in E$, $\pi_1(x, y) = x$ et $\pi_2(x, y) = y$. La variable aléatoire $aX + bY$ est fonction de la variable aléatoire Z ; elle s'écrit :

$$aX + bY = a\pi_1(Z) + b\pi_2(Z).$$

Il résulte alors du théorème 5.2 que les familles $\{\pi_i(z) P(Z = z)\}_{z \in E}$, $i = 1, 2$, sont sommables ; par l'inégalité triangulaire, il en est de même de la famille $\{(a\pi_1(z) + b\pi_2(z)) P(Z = z)\}_{z \in E}$, ce qui démontre que la variable aléatoire $aX + bY$ admet une moyenne.

Toujours d'après le théorème 5.2, on a :

$$\begin{aligned} E(aX + bY) &= \sum_{z \in E} (a\pi_1(z) + b\pi_2(z)) P(Z = z) \\ &= a \sum_{z \in E} \pi_1(z) P(Z = z) + b \sum_{z \in E} \pi_2(z) P(Z = z), \end{aligned}$$

ce qui démontre que

$$E(aX + bY) = aEX + bEY.$$

Remarque. Le théorème 5.2 s'applique sans changement, avec la même démonstration, au cas où f est une fonction à valeurs dans $\overline{\mathbb{R}}$, et où $Y = f(X)$ peut prendre des valeurs infinies (il résulte de toutes façons des hypothèses que si $Y = \pm\infty$, c'est avec une probabilité nulle).

La proposition 5.4 s'étend de la même façon, en utilisant la convention habituelle $0 \times \infty = 0$, à l'exception du cas où le calcul de $aX(\omega) + bY(\omega)$ conduit à l'expression $\infty - \infty$. Comme ci-dessus, si cela se produit, c'est avec une probabilité nulle.

5.1.4. Moyennes des lois discrètes classiques

On se réfère à un même espace probabilisé (Ω, \mathcal{A}, P) .

Exemple 5.3. Loi binomiale. Soit $(A_i)_{1 \leq i \leq n}$ une famille de n événements indépendants de même probabilité p ($0 < p < 1$). On rappelle que la loi de la

variable aléatoire $S_n = \sum_{i=1}^n \mathbf{1}_{A_i}$, est la loi binomiale $\mathcal{B}(n, p)$. On a, d'après la linéarité de l'espérance,

$$ES_n = \sum_{i=1}^n E(\mathbf{1}_{A_i}) = \sum_{i=1}^n P(A_i),$$

soit :

$$\boxed{ES_n = np.}$$

Exemple 5.4. Lois géométriques. Soit $(A_i)_{i \in \mathbb{N}}$ une suite d'événements indépendants de même probabilité p ($0 < p < 1$).

◇ On rappelle que la variable aléatoire N définie par, pour tout $\omega \in \Omega$,

$$N(\omega) = \inf(n \in \mathbb{N} \mid \omega \in A_n) \quad (\text{avec la convention } \inf \emptyset = +\infty),$$

est de loi géométrique $\mathcal{G}_N(p)$. On a, d'après l'égalité (5.2) et en notant $q = 1 - p$,

$$EN = \sum_{n=0}^{+\infty} n P(N = n) = \sum_{n=0}^{+\infty} n pq^n,$$

soit encore :

$$\boxed{EN = \frac{q}{p}.}$$

◇ On rappelle que la variable aléatoire N' définie par, pour tout $\omega \in \Omega$,

$$N'(\omega) = \inf(n \in \mathbb{N}^* \mid \omega \in A_n) \quad (\text{avec la convention } \inf \emptyset = +\infty),$$

est de loi géométrique $\mathcal{G}_{N^*}(p)$ et que, si $N'' = 1 + N$, la loi de N'' est la loi de N' . On en déduit que :

$$EN' = EN'' = 1 + EN,$$

soit :

$$\boxed{EN' = \frac{1}{p}.}$$

Exemple 5.5. Loi de Poisson. Si X est une variable aléatoire de loi de Poisson $\mathcal{P}(\lambda)$, où $\lambda > 0$, on a, d'après l'égalité (5.2) :

$$EX = \sum_{n=0}^{+\infty} nP(X = n) = \sum_{n=0}^{+\infty} n \exp(-\lambda) \frac{\lambda^n}{n!},$$

soit encore :

$$\boxed{EX = \lambda.}$$

Remarque. Si on sait qu'une variable aléatoire suit une loi géométrique (respectivement une loi de Poisson), pour identifier entièrement sa loi, il suffit de connaître sa moyenne. Ce n'est pas le cas pour la loi binomiale! Nous en reparlerons.

5.2. Moments d'ordre supérieur

5.2.1. Moments d'ordre quelconque

Avertissement. Le lecteur préparant un concours et qui ne s'intéresse qu'aux moments d'ordre un et deux (voir les programmes) peut se contenter de lire la définition 5.5 et passer à la sous-section 5.2.2.

Définition 5.5. Soit p un réel supérieur ou égal à 1. On note $\mathcal{L}_d^p(\Omega, \mathcal{A}, P)$ l'ensemble des variables aléatoires discrètes réelles X telles que $E(|X|^p) < \infty$.

Si p est un entier ≥ 1 et si $X \in \mathcal{L}_d^p(\Omega, \mathcal{A}, P)$ le nombre réel $E(X^p)$ est appelé **moment d'ordre p** de X .

Nous établissons les inégalités de **Hölder** et de **Minkowski** dans le contexte des familles sommables et nous en déduisons les propriétés de l'ensemble $\mathcal{L}_d^p(\Omega, \mathcal{A}, P)$.

Définition 5.6. Deux réels p et q strictement positifs sont dits **conjugués** s'ils satisfont à l'égalité :

$$\frac{1}{p} + \frac{1}{q} = 1;$$

dans ce cas, on a bien sûr $p > 1$ et $q > 1$.

Lemme 5.7. Soient p et q deux réels conjugués. Pour a et $b \geq 0$, on a

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \quad (5.4)$$

Démonstration. La fonction $x \mapsto -\ln x$ étant **convexe** sur $]0, +\infty[$, on a, quels que soient $x, y > 0$,

$$-\ln \left(\frac{x}{p} + \frac{y}{q} \right) \leq -\frac{1}{p} \ln x - \frac{1}{q} \ln y,$$

soit

$$\ln x^{\frac{1}{p}} + \ln y^{\frac{1}{q}} \leq \ln \left(\frac{x}{p} + \frac{y}{q} \right),$$

et donc :

$$x^{\frac{1}{p}} y^{\frac{1}{q}} \leq \frac{x}{p} + \frac{y}{q}.$$

Il suffit alors de remplacer x par a^p et y par b^q pour obtenir l'inégalité (5.4) lorsque $a > 0$ et $b > 0$. Si $a = 0$ ou $b = 0$, (5.4) est évidente (à condition de donner un sens à 0^p et 0^q).

On déduit de ce lemme l'inégalité de Hölder.

Proposition 5.8. (Inégalité de Hölder.) Soient p et q deux réels conjugués. Soient $(x_i)_{i \in I}$, $(y_i)_{i \in I}$, $(\rho_i)_{i \in I}$ trois familles de nombres réels positifs indexées par un même ensemble I . On a :

$$\boxed{\sum_{i \in I} x_i y_i \rho_i \leq \left(\sum_{i \in I} x_i^p \rho_i \right)^{\frac{1}{p}} \left(\sum_{i \in I} y_i^q \rho_i \right)^{\frac{1}{q}}.} \quad (5.5)$$

Démonstration. Il suffit de démontrer l'inégalité (5.5) lorsque les deux termes du membre de droite sont finis. Si l'un de ceux-ci est nul, par exemple le premier, alors, pour tout $i \in I$, on a $x_i^p \rho_i = 0$, et donc aussi $x_i y_i \rho_i = 0$. On a alors $\sum_{i \in I} x_i y_i \rho_i = 0$.

Si ces deux termes sont non nuls, il résulte du lemme 5.7 que l'on a

$$\frac{x_i}{\left(\sum_{i \in I} x_i^p \rho_i \right)^{\frac{1}{p}}} \frac{y_i}{\left(\sum_{i \in I} y_i^q \rho_i \right)^{\frac{1}{q}}} \leq \frac{1}{p} \frac{x_i^p}{\left(\sum_{i \in I} x_i^p \rho_i \right)} + \frac{1}{q} \frac{y_i^q}{\left(\sum_{i \in I} y_i^q \rho_i \right)};$$

après multiplication par ρ_i qui est positif, et sommation sur I , on obtient

$$\frac{\sum_{i \in I} x_i y_i \rho_i}{\left(\sum_{i \in I} x_i^p \rho_i \right)^{\frac{1}{p}} \left(\sum_{i \in I} y_i^q \rho_i \right)^{\frac{1}{q}}} \leq \frac{1}{p} + \frac{1}{q} = 1,$$

ce qui démontre l'inégalité 5.5.

On en déduit l'inégalité de Minkowski.

Proposition 5.9. (Inégalité de Minkowski.) Sous les mêmes hypothèses que dans la proposition 5.8 on a :

$$\boxed{\left(\sum_{i \in I} (x_i + y_i)^p \rho_i \right)^{\frac{1}{p}} \leq \left(\sum_{i \in I} x_i^p \rho_i \right)^{\frac{1}{p}} + \left(\sum_{i \in I} y_i^p \rho_i \right)^{\frac{1}{p}}.} \quad (5.6)$$

Démonstration. Puisque $p > 1$, l'additivité de la somme entraîne

$$\sum_{i \in I} (x_i + y_i)^p \rho_i = \sum_{i \in I} \left[(x_i + y_i)^{p-1} x_i \right] \rho_i + \sum_{i \in I} \left[(x_i + y_i)^{p-1} y_i \right] \rho_i.$$

En appliquant l'inégalité (5.5) à chacun des facteurs du membre de droite, on obtient :

$$\begin{aligned} \sum_{i \in I} (x_i + y_i)^p \rho_i &\leq \left(\sum_{i \in I} (x_i + y_i)^{q(p-1)} \rho_i \right)^{\frac{1}{q}} \left(\sum_{i \in I} x_i^p \rho_i \right)^{\frac{1}{p}} \\ &\quad + \left(\sum_{i \in I} (x_i + y_i)^{q(p-1)} \rho_i \right)^{\frac{1}{q}} \left(\sum_{i \in I} y_i^p \rho_i \right)^{\frac{1}{p}}. \end{aligned}$$

Il suffit alors de remarquer que : $q(p-1) = p$.

Proposition 5.10. (i) Si $p \geq 1$, $\mathcal{L}_d^p(\Omega, \mathcal{A}, P)$ est un espace vectoriel.

(ii) Soient p et q deux réels **conjugués**. Soient $X \in \mathcal{L}_d^p(\Omega, \mathcal{A}, P)$ et $Y \in \mathcal{L}_d^q(\Omega, \mathcal{A}, P)$. Alors $XY \in \mathcal{L}_d^1(\Omega, \mathcal{A}, P)$ et :

$$\mathbb{E}|XY| \leq (\mathbb{E}|X^p|)^{\frac{1}{p}} (\mathbb{E}|Y^q|)^{\frac{1}{q}}. \quad (5.7)$$

En prenant $p = q = 2$ dans (5.7) et en observant que $|\mathbb{E}(XY)| \leq \mathbb{E}|XY|$, on obtient l'inégalité

$$|\mathbb{E}(XY)| \leq (\mathbb{E}|X^2|)^{\frac{1}{2}} (\mathbb{E}|Y^2|)^{\frac{1}{2}} \quad (5.8)$$

(inégalité de **Schwarz**).

(iii) Soient α et β deux entiers tels que : $1 \leq \alpha \leq \beta$; alors, on a l'inclusion ensembliste $\mathcal{L}_d^\beta(\Omega, \mathcal{A}, P) \subset \mathcal{L}_d^\alpha(\Omega, \mathcal{A}, P)$ et

$$(\mathbb{E}|X^\alpha|)^{\frac{1}{\alpha}} \leq (\mathbb{E}|X^\beta|)^{\frac{1}{\beta}}. \quad (5.9)$$

En particulier, une variable aléatoire réelle discrète ayant un moment d'ordre $p \geq 1$ admet un moment de tout ordre p' tel que $1 \leq p' \leq p$.

Démonstration.

(i) Soient X et Y deux éléments de $\mathcal{L}_d^p(\Omega, \mathcal{A}, P)$ et a et b deux réels quelconques. La fonction $aX + bY$ est une variable aléatoire réelle discrète. La variable aléatoire $Z = (X, Y)$ est discrète et à valeurs dans $E = \text{val}(X) \times \text{val}(Y)$. Soient π_1 et π_2 les projections canoniques du produit $\text{val}(X) \times \text{val}(Y)$, c'est-à-dire les applications définies sur E par les relations $\pi_1(x, y) = x$ et $\pi_2(x, y) = y$ ($(x, y) \in E$). La variable aléatoire $aX + bY$ est **fonction** de la variable aléatoire Z ; elle s'écrit :

$$aX + bY = a\pi_1(Z) + b\pi_2(Z).$$

Il résulte du théorème 5.2 que les familles $(|\pi_i(z)|^p P(Z = z))_{z \in E}$, $i = 1, 2$, sont sommables. L'inégalité de Minkowski (5.6) permet alors d'écrire :

$$\begin{aligned} & \left[\sum_{z \in E} |a\pi_1(z) + b\pi_2(z)|^p P(Z = z) \right]^{\frac{1}{p}} \\ & \leq |a| \left[\sum_{z \in E} |\pi_1(z)|^p P(Z = z) \right]^{\frac{1}{p}} + |b| \left[\sum_{z \in E} |\pi_2(z)|^p P(Z = z) \right]^{\frac{1}{p}}; \end{aligned}$$

le second membre étant fini, la famille

$$\{|a\pi_1(z) + b\pi_2(z)|^p P(Z = z)\}_{z \in E}$$

est alors sommable. Il en résulte, toujours d'après le théorème 5.2, que la variable aléatoire $aX + bY$ admet un moment d'ordre p . Ceci prouve que $\mathcal{L}_d^p(\Omega, \mathcal{A}, P)$ est un espace vectoriel.

(ii) Soient X et Y deux éléments de $\mathcal{L}_d^p(\Omega, \mathcal{A}, P)$. Toujours avec les mêmes notations, il résulte de l'inégalité de Hölder (5.5) que :

$$\begin{aligned} \sum_{z \in E} |\pi_1(z) \pi_2(z)| P(Z = z) \\ \leq \left[\sum_{z \in E} |\pi_1(z)|^p P(Z = z) \right]^{\frac{1}{p}} \left[\sum_{z \in E} |\pi_2(z)|^q P(Z = z) \right]^{\frac{1}{q}}. \end{aligned}$$

Le second membre est fini par hypothèse, et donc aussi le premier, ce qui, d'après le théorème 5.2, démontre que $|XY| \equiv |\pi_1(Z) \pi_2(Z)|$ admet une moyenne et que l'inégalité (5.7) est vraie.

(iii) Soit $X \in \mathcal{L}_d^\beta(\Omega, \mathcal{A}, P)$. On peut supposer $\beta > \alpha$. Choisissons γ tel que :

$$\frac{1}{\beta} + \frac{1}{\gamma} = \frac{1}{\alpha},$$

c'est-à-dire $\gamma = \frac{\beta}{\beta - \alpha}$. L'inégalité de Hölder (5.5) permet d'écrire

$$\sum_{x \in \mathbb{R}} [|x|^\alpha \cdot 1] P(X=x) \leq \left[\sum_{x \in \mathbb{R}} (|x|^\alpha)^\beta P(X=x) \right]^{\frac{\alpha}{\beta}} \left[\sum_{x \in \mathbb{R}} 1^\gamma P(X=x) \right]^{\frac{1}{\gamma}},$$

soit encore, puisque $\sum_{x \in \mathbb{R}} P(X = x) = 1$ et que $X \in \mathcal{L}_d^\beta(\Omega, \mathcal{A}, P)$:

$$\sum_{x \in \mathbb{R}} |x|^\alpha P(X = x) \leq \left[\sum_{x \in \mathbb{R}} |x|^\beta P(X = x) \right]^{\frac{\alpha}{\beta}} < +\infty.$$

Toujours d'après le théorème 5.2, il en résulte que $X \in \mathcal{L}_d^\alpha(\Omega, \mathcal{A}, P)$ et que l'inégalité (5.9) sur les moments est vérifiée.

5.2.2. Moments d'ordre deux

Nous faisons une étude particulière des moments d'ordre deux qui jouent un rôle primordial dans beaucoup de théorèmes (et de problèmes) en Probabilités, ainsi qu'en Statistique.

Nous redonnons de manière indépendante de la sous-section précédente une proposition liant les moments d'ordre un et deux d'une variable aléatoire numérique discrète (voir la définition 5.5).

Proposition 5.11. Si $X \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$, alors $X \in \mathcal{L}_d^1(\Omega, \mathcal{A}, P)$ et

$$\boxed{E|X| \leq (EX^2)^{\frac{1}{2}}.} \quad (5.10)$$

Par ailleurs, si X et Y sont des variables aléatoires réelles discrètes admettant un moment d'ordre deux (ce sont donc des éléments de $\mathcal{L}_d^2(\Omega, \mathcal{A}, P)$), la variable aléatoire XY admet une moyenne (c'est donc un élément de $\mathcal{L}_d^1(\Omega, \mathcal{A}, P)$) et on a l'inégalité de Schwarz :

$$\boxed{|E(XY)| \leq (E|X^2|)^{\frac{1}{2}} (E|Y|^2)^{\frac{1}{2}}.} \quad (5.11)$$

Démonstration. On vérifie facilement que, pour tout $x \in \overline{\mathbb{R}}$, on a $|x| \leq 1 + x^2$, et donc :

$$|X| \leq 1 + X^2.$$

La variable aléatoire X^2 appartient, par hypothèse, à $\mathcal{L}_d^1(\Omega, \mathcal{A}, P)$; il en est de même pour X . Définissons le polynôme Q du second degré à coefficients réels en posant

$$Q(\lambda) = E\left[(|X| - \lambda E(|X|))^2 \right],$$

soit

$$Q(\lambda) = \lambda^2 (E|X|)^2 - 2\lambda (E|X|)^2 + E(X^2).$$

Puisque $Q(\lambda) \geq 0$ pour tout $\lambda \in \mathbb{R}$, le discriminant réduit Δ' est négatif ou nul, ce qui s'écrit

$$\Delta' = (E|X|)^2 \left[(E|X|)^2 - E(X^2) \right] \leq 0.$$

Ceci démontre l'inégalité (5.10).

Le même raisonnement appliqué au polynôme S défini par :

$$S(\lambda) = E\left[(X + \lambda Y)^2 \right]$$

permet de démontrer l'inégalité (5.11), après avoir remarqué que, puisque X et Y sont des variables aléatoires réelles discrètes, $X + \lambda Y$ est bien définie.

Remarque. Sous les mêmes hypothèses, par application de l'inégalité (5.11) aux variables aléatoires $|X|$ et $|Y|$, on obtient l'inégalité :

$$\boxed{E(|XY|) \leq (E|X^2|)^{\frac{1}{2}} (E|Y|^2)^{\frac{1}{2}}.} \quad (5.12)$$

Définition 5.12. (i) Si $X \in \mathcal{L}_d^1(\Omega, \mathcal{A}, P)$, la variable aléatoire numérique discrète $\overset{\circ}{X} = X - EX$ est appelée **variable aléatoire centrée** associée à X . On a $E(\overset{\circ}{X}) = 0$.

(ii) Si $X \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$, la variable aléatoire X admet une moyenne; le réel $E[X - EX]^2$ est appelé **variance**³ de la variable aléatoire X et noté $\text{var}(X)$. Sa racine carrée positive est appelée **écart-type** de X et notée σ_X . On a donc $\text{var}(X) = \sigma_X^2$.

(iii) Si $X \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$, et si $\sigma_X \neq 0$, la variable aléatoire $\tilde{X} = \frac{X - EX}{\sigma_X}$ est appelée **variable aléatoire centrée réduite** associée à X . On a $E(\tilde{X}) = 0$ et $\sigma_{\tilde{X}} = 1$.

Remarques. D'après le théorème de transfert 5.2, on a si $X \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$:

$$\sigma_X^2 = \sum_{x \in \text{val}(X)} (x - EX)^2 P(X = x). \quad (5.13)$$

Cette formule sert pour le **calcul** de la variance à **partir de la loi** de X . La variable aléatoire étant **discrète**, l'ensemble $X(\Omega)$ est **dénombrable** et, suivant qu'il est fini ou non, cette dernière formule est souvent explicitée sous l'une des formes suivantes :

◇ si $X(\Omega)$ est **fini** et s'énumère sous la forme $\{x_1, x_2, \dots, x_n\}$, on a

$$\sigma_X^2 = \sum_{i=1}^n (x_i - EX)^2 P(X = x_i); \quad (5.14)$$

◇ si $X(\Omega)$ est **infini** et énuméré sous la forme $\{x_n \mid n \in \mathbb{N}\}$, on a :

$$\sigma_X^2 = \sum_{i=0}^{+\infty} (x_i - EX)^2 P(X = x_i). \quad (5.15)$$

La proposition suivante permet en général un calcul plus simple de la variance.

3. La variance de X est donc le moment d'ordre deux de la variable centrée associée à X , ou le « moment centré d'ordre deux de X ». Cette terminologie est peu usitée.

Proposition 5.13. Si $X \in \mathcal{L}_d^2(\Omega, \mathcal{A}, \mathbb{P})$, on a

(i)

$$\sigma_X^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2. \quad (5.16)$$

(ii) Pour tous réels a et b , on a

$$\sigma_{aX+b}^2 = a^2 \sigma_X^2. \quad (5.17)$$

Démonstration.

(i) On applique la linéarité de l'espérance après avoir développé le carré.

(ii) On a

$$\mathbb{E}(aX + b) = a\mathbb{E}X + b,$$

et donc :

$$\sigma_{aX+b}^2 = \mathbb{E}\left[\left((aX + b) - (a\mathbb{E}X + b)\right)^2\right] = a^2 \mathbb{E}\left[(X - \mathbb{E}X)^2\right].$$

Remarque. La formule 5.16 est parfois citée sous le nom de **formule de Leibniz** et elle est d'un usage courant pour calculer une variance. La formule 5.17 traduit l'**invariance par translation** de l'écart-type et son **homogénéité**. En résumé, changements d'origine et d'échelle se traduisent par la formule (5.17).

5.2.3. Covariance de deux variables aléatoires

Définition 5.14. Si X et $Y \in \mathcal{L}_d^2(\Omega, \mathcal{A}, \mathbb{P})$, il résulte de l'inégalité de Schwarz que la variable aléatoire $(X - \mathbb{E}X)(Y - \mathbb{E}Y)$ admet une moyenne ; celle-ci est appelée **covariance** de X et Y et notée $\text{cov}(X, Y)$:

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)].$$

La proposition suivante permet en général un calcul plus simple de la covariance de deux variables aléatoires. Elle permet aussi de calculer la **variance d'une somme** de variables aléatoires.

Proposition 5.15. (i) Si X et $Y \in \mathcal{L}_d^2(\Omega, \mathcal{A}, \mathbb{P})$, on a :

$$\text{cov}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y). \quad (5.18)$$

(ii) Si X et $Y \in \mathcal{L}_d^2(\Omega, \mathcal{A}, \mathbb{P})$, on a :

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2 \text{cov}(X, Y). \quad (5.19)$$

Démonstration.

(i) On applique la linéarité de l'espérance après avoir développé le produit.

(ii) On commence par remarquer que la variable aléatoire $X + Y$ est bien définie. De plus, on a

$$\begin{aligned}\sigma_{X+Y}^2 &= E\left([(X-EX) + (Y-EY)]^2\right) \\ &= E\left([X-EX]^2\right) + E\left([Y-EY]^2\right) + 2 E[(X-EX)(Y-EY)] \\ &= \sigma_X^2 + \sigma_Y^2 + 2 \operatorname{cov}(X, Y).\end{aligned}$$

La proposition suivante est souvent utilisée.

Proposition 5.16. Soient deux variables aléatoires indépendantes X et $Y \in \mathcal{L}_d^1(\Omega, \mathcal{A}, P)$, la variable aléatoire discrète XY admet une moyenne et on a :

$$E(XY) = (EX)(EY). \quad (5.20)$$

Démonstration. Le produit XY est une variable aléatoire numérique discrète, $Z = (X, Y)$ est aussi une variable aléatoire discrète et, en utilisant les mêmes notations qu'à la proposition 5.4, on a : $XY = \pi_1(Z)\pi_2(Z)$. Les variables aléatoires X et Y étant indépendantes, on a : pour tout $z = (x, y) \in X(\Omega) \times Y(\Omega)$,

$$P(Z = z) = P(X = x)P(Y = y). \quad (5.21)$$

Alors, d'après la propriété de Fubini pour les familles positives :

$$\begin{aligned}\sum_{z \in X(\Omega) \times Y(\Omega)} |\pi_1(z)\pi_2(z)| P(Z = z) \\ &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} |xy| P(X = x)P(Y = y) \\ &= \left[\sum_{x \in X(\Omega)} |x| P(X = x) \right] \left[\sum_{y \in Y(\Omega)} |y| P(Y = y) \right] \\ &= (E|X|)(E|Y|) < +\infty.\end{aligned}$$

Il en résulte que la variable aléatoire $|XY|$ admet une moyenne. Par un calcul similaire, mais sans les valeurs absolues, calcul maintenant licite, la formule (5.20) est vérifiée.

Corollaire 5.17. Si X et $Y \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$ et sont indépendantes, on a :

$$\boxed{\operatorname{cov}(X, Y) = 0.}$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2.$$

Démonstration. La proposition précédente assure que $\text{cov}(X, Y) = 0$; pour obtenir la seconde égalité, il suffit de reporter dans (5.19).

Remarques. 1. Ce corollaire est d'un usage très courant. Il permet notamment de montrer que si X_1, \dots, X_n sont des variables aléatoires indépendantes et de même loi (de variance finie), on a

$$\sigma_{X_1 + \dots + X_n} = \sigma_{X_1} \sqrt{n}.$$

2. Comme le montre l'exemple ci-dessous, l'égalité $\text{cov}(X, Y) = 0$ n'entraîne pas que X et Y sont indépendantes.

Exemple 5.6. Soient U et $V \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$ et indépendantes. Soient a, b, c, d des réels non nuls; on définit les variables aléatoires réelles discrètes X et Y par :

$$\begin{cases} X = aU + bV \\ Y = cU + dV. \end{cases}$$

Alors, puisque U et V sont indépendantes, on a :

$$\begin{aligned} \text{cov}(X, Y) &= E \left[\left(a\overset{\circ}{U} + b\overset{\circ}{V} \right) \left(c\overset{\circ}{U} + d\overset{\circ}{V} \right) \right] \\ &= ac \sigma_U^2 + (ad + bc) \text{cov}(U, V) + bd \sigma_V^2 \\ &= ac \sigma_U^2 + bd \sigma_V^2. \end{aligned}$$

Supposons de plus que U et V aient même loi déterminée par :

$$P(U = 1) = p, \quad P(U = 0) = 1 - p.$$

On a alors $\sigma_U^2 = \sigma_V^2$ (puisque U et V ont même loi); si on suppose enfin que $ac + bd = 0$, il vient : $\text{cov}(X, Y) = 0$, et pourtant, si $a + b \neq 0$ et $c + d \neq 0$, les variables aléatoires X et Y ne sont pas indépendantes. En effet, on a les égalités entre événements

$$(X = 0) = (U = 0, V = 0) = (Y = 0).$$

On a donc, d'une part

$$P(X = 0) = P(Y = 0) = P(U = 0, V = 0) = P(U = 0)P(V = 0) = (1 - p)^2,$$

et d'autre part,

$$P(X = 0, Y = 0) = P(U = 0, V = 0) = (1 - p)^2.$$

Si $0 < p < 1$, on a donc

$$P(X = 0)P(Y = 0) \neq P(X = 0, Y = 0),$$

et les variables aléatoires X et Y ne sont pas indépendantes!

5.2.4. Variances des lois discrètes classiques

Nous gardons les notations des exemples 5.2, 5.3, 5.4 et 5.5.

Loi de Bernoulli

On a

$$\sigma_{\mathbf{1}_A}^2 = E(\mathbf{1}_A^2) - [E(\mathbf{1}_A)]^2 = p(1-p),$$

soit

$$\boxed{\sigma_{\mathbf{1}_A}^2 = p(1-p)}.$$

Loi binomiale

La famille de variables aléatoires $(\mathbf{1}_{A_i})_{1 \leq i \leq n}$ est une famille de variables aléatoires indépendantes, de même loi et donc de même variance. Il résulte du corollaire 5.17 que :

$$\sigma_{S_n}^2 = \sum_{i=1}^n \sigma_{\mathbf{1}_{A_i}}^2 = n \sigma_{\mathbf{1}_{A_1}}^2.$$

On a donc :

$$\boxed{\sigma_{S_n}^2 = np(1-p)}.$$

Lois géométriques sur \mathbb{N} et sur \mathbb{N}^*

(a) La variable aléatoire N suit la loi $\mathcal{G}_{\mathbb{N}}(p)$. On a

$$EN^2 = E[N(N-1)] + EN,$$

et, par le théorème de transfert :

$$E[N(N-1)] = \sum_{n=0}^{+\infty} n(n-1)pq^n.$$

Mais, d'après le théorème de dérivation des séries entières (argument déjà rencontré), on a

$$\sum_{n=2}^{+\infty} n(n-1)x^{n-2} = \frac{d^2}{dx^2} \left(\sum_{n=0}^{+\infty} x^n \right);$$

pour tout x tel que $0 \leq x < 1$; il en résulte que l'on a

$$EN^2 = \frac{2pq^2}{(1-q)^3} + \frac{q}{p} = \frac{q}{p} \left(2\frac{q}{p} + 1 \right).$$

On a alors, en utilisant l'égalité (5.16) et en simplifiant :

$$\sigma_N^2 = \frac{q}{p^2}.$$

(b) Les variables aléatoires N' et $1+N$ ont même loi géométrique $\mathcal{G}_{N^*}(p)$ et donc même variance. La variance étant invariante par translation, il vient :

$$\sigma_{N'}^2 = \frac{q}{p^2}.$$

Loi de Poisson

Soit X une variable aléatoire de loi de Poisson $\mathcal{P}(\lambda)$. On a

$$E(X^2) = E[X(X-1)] + EX,$$

et, par le théorème de transfert :

$$E[X(X-1)] = \sum_{n=0}^{+\infty} n(n-1) \exp(-\lambda) \frac{\lambda^n}{n!}.$$

Mais, par translation d'indice, on a :

$$\sum_{n=2}^{+\infty} n(n-1) \frac{\lambda^n}{n!} = \lambda^2 \exp(-\lambda).$$

Il en résulte que l'on a

$$E(X^2) = \lambda^2 + \lambda,$$

et, d'après l'égalité (5.16) :

$$\sigma_X^2 = \lambda.$$

Remarque. Pour une variable aléatoire de Poisson, le paramètre représente à la fois la moyenne et la variance.

5.2.5. Inégalités de Markov et de Tchebichev

L'inégalité classique de Tchebichev majore la probabilité pour une variable aléatoire de « trop » s'écarter de sa moyenne. On la qualifie de grossière par comparaison avec des estimations plus précises qui peuvent être données dans des cas particuliers (voir chapitre 7, exemple 7.1 et exercice 7.2). Elle a surtout une importance théorique et intervient notamment dans la démonstration de la loi (faible) des grands nombres.

Proposition 5.18. (Inégalité de Markov.) Si $X \in \mathcal{L}_d^1(\Omega, \mathcal{A}, P)$, on a, pour tout $\varepsilon > 0$:

$$\boxed{P(|X| > \varepsilon) \leq \frac{E|X|}{\varepsilon}.} \quad (5.22)$$

Démonstration. Soit l'ensemble $D = \{x \in X(\Omega) \mid |x| > \varepsilon\}$. On a les minoration successives suivantes :

$$E|X| = \sum_{x \in \mathbb{R}} |x| P(X = x) \geq \sum_{x \in D} |x| P(X = x) \geq \varepsilon \sum_{x \in D} P(X = x).$$

Puisque $X(\Omega)$ est dénombrable, il en résulte que l'on a :

$$E|X| \geq \varepsilon P(X \in D) = \varepsilon P[X^{-1}(D)],$$

ce qui démontre l'inégalité.

On en déduit l'inégalité de Tchebichev.

Proposition 5.19. (Inégalité de Tchebichev.) Si $X \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$, on a : pour tout $\varepsilon > 0$

$$\boxed{P(|X - EX| > \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2}.} \quad (5.23)$$

L'inégalité de Tchebichev s'utilise parfois sous la forme

$$P(|X - EX| > \varepsilon \sigma_X) \leq 1/\varepsilon^2.$$

Démonstration. Il suffit d'appliquer l'inégalité de Markov à la variable aléatoire $(X - EX)^2$ et de remarquer que

$$\{(X - EX)^2 > \varepsilon^2\} = \{|X - EX| > \varepsilon\}.$$

La définition de la variance permet de conclure.

Remarques. 1. On peut bien sûr faire une démonstration directe de cette inégalité de la manière suivante. Soit l'ensemble

$$D = \{x \in X(\Omega) \mid |x - EX| > \varepsilon\}.$$

On a les minoration successives :

$$\begin{aligned} \sigma_X^2 &= \sum_{x \in \mathbb{R}} (x - EX)^2 P(X = x) \\ &\geq \sum_{x \in D} (x - EX)^2 P(X = x) \geq \varepsilon^2 \sum_{x \in D} P(X = x). \end{aligned}$$

Puisque $X(\Omega)$ est dénombrable, on a :

$$\sigma_X^2 \geq \varepsilon^2 P(X \in D) = \varepsilon^2 P[X^{-1}(D)],$$

ce qui donne l'inégalité.

2. Plus l'écart-type σ_X est petit, plus la variable aléatoire X est concentrée autour de sa moyenne. Le cas extrême de cette concentration est atteint quand $\sigma_X = 0$. Il est clair, en effet, que si la variable aléatoire X est égale à une constante avec probabilité 1, sa moyenne est égale à cette constante, et on $\sigma_X = 0$. L'inégalité de Tchebichev fournit la réciproque : si $\sigma_X = 0$, on a pour tout n entier, $P(|X - EX| \geq 1/n) = 0$, d'où

$$P(|X - EX| > 0) = \lim_n \downarrow P(|X - EX| \geq 1/n) = 0$$

et $P(X = EX) = 1$. (On peut aussi faire une démonstration directe.)

5.2.6. Coefficient de corrélation. Droite de régression

Comme on va le voir, le **coefficient de corrélation** de deux variables aléatoires permet de « mesurer » une certaine liaison entre ces variables aléatoires.

Définition 5.20. Soient X et $Y \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$ telles que $\sigma_X \neq 0$ et $\sigma_Y \neq 0$. On appelle **coefficient de corrélation** de X et Y le réel :

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Proposition 5.21. Soient X et $Y \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$ de variances non nulles et de coefficient de corrélation $\rho_{X,Y}$.

(i) On a :

$$|\rho_{X,Y}| \leq 1.$$

(ii) Pour que $|\rho_{X,Y}| = 1$ il faut et il suffit qu'il existe trois réels a , b , c non tous nuls tels que

$$P(aX + bY + c = 0) = 1.$$

Démonstration.

(i) L'inégalité de Schwarz permet d'écrire :

$$|\text{cov}(X, Y)| = \left| E\left(\overset{\circ}{X}\overset{\circ}{Y}\right) \right| \leq \left(E\overset{\circ}{X}^2 \right)^{\frac{1}{2}} \left(E\overset{\circ}{Y}^2 \right)^{\frac{1}{2}} = \sigma_X \sigma_Y,$$

ce qui démontre l'inégalité annoncée.

(ii)

(a) Si $|\rho_{X,Y}| = 1$, le polynôme du second degré en λ : $E \left[\overset{\circ}{X} + \lambda \overset{\circ}{Y} \right]^2$ admet une racine double λ_0 . Son discriminant est alors nul, c'est-à-dire que l'on a $E \left[\overset{\circ}{X} + \lambda_0 \overset{\circ}{Y} \right]^2 = 0$. Il en résulte que, d'après la remarque suivant la proposition 5.19, on a $P \left(\overset{\circ}{X} + \lambda_0 \overset{\circ}{Y} = 0 \right) = 1$.

(b) Réciproquement, supposons qu'il existe trois réels a, b, c non tous nuls tels que :

$$P(aX + bY + c = 0) = 1. \quad (5.24)$$

Si $c \neq 0$, les nombres a et b sont différents de 0. En effet, si par exemple $a = 0$, on a alors $P(bY + c = 0) = 1$, et donc $\sigma_{bY+c}^2 = \sigma_0^2 = 0$, soit encore $b^2 \sigma_Y^2 = 0$ et donc $b = 0$, ce qui est impossible d'après (5.24). Dans ce cas on a

$$P(X = \alpha Y + \beta) = 1, \quad (5.25)$$

où $\alpha \neq 0$.

Si $c = 0$, on a $a \neq 0$ ou $b \neq 0$. Si par exemple $a \neq 0$, l'égalité (5.25) est encore satisfaite avec $\beta = 0$ (si on avait $b \neq 0$, on ferait un calcul analogue).

Dans ces deux cas on a donc

$$\text{cov}(X, Y) = E \left[\left(\alpha \overset{\circ}{Y} \right) \overset{\circ}{Y} \right] = \alpha \sigma_Y^2$$

et

$$\sigma_X^2 = \sigma_{\alpha Y + \beta}^2 = \alpha^2 \sigma_Y^2,$$

d'où :

$$\rho_{X,Y} = \frac{\alpha \sigma_Y^2}{|\alpha| \sigma_Y^2}.$$

Il en résulte que : $|\rho_{X,Y}| = 1$.

Remarque. On a utilisé le résultat suivant : si $X \in \mathcal{L}_d^1(\Omega, \mathcal{A}, P)$, si Y est une variable aléatoire discrète telle que $P(X = Y) = 1$, alors $Y \in \mathcal{L}_d^1(\Omega, \mathcal{A}, P)$ et $EX = EY$.

En effet, soit $D = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$. Si $Z = (X, Y)$, on a $P_Z(D) = 1$, et toujours avec les notations de la proposition 5.4, on a :

$$E|X| = E|\pi_1(Z)| = \sum_{z \in \mathbb{R}^2} |\pi_1(z)| P(Z = z).$$

Puisque $P(Z \in D) = 1$ et que $\pi_1(z) = \pi_2(z)$ sur D , il vient :

$$\begin{aligned} E|X| &= \sum_{z \in D} |\pi_1(z)| P(Z = z) = \sum_{z \in D} |\pi_2(z)| P(Z = z) \\ &= \sum_{z \in \mathbb{R}^2} |\pi_2(z)| P(Z = z), \end{aligned}$$

ce qui prouve que $Y \in \mathcal{L}_d^1(\Omega, \mathcal{A}, P)$. Le même calcul sans les modules est alors licite et conduit à l'égalité : $EX = EY$.

La proposition 5.21 affirme que si $\rho_{X,Y} = \pm 1$, alors l'une des deux variables X et Y est fonction linéaire (affine) de l'autre. Si X et Y sont indépendantes, on a $\rho_{X,Y} = 0$ (la réciproque est fautive). Par suite, si $\rho_{X,Y} \neq 0$, on peut affirmer que X et Y ne sont pas indépendantes. Le coefficient de corrélation est surtout utilisé en Statistique, pour des séries d'observations empiriques — qu'on peut considérer comme des variables aléatoires prenant un nombre fini de valeurs et de loi uniforme sur cet ensemble de valeurs — et on considère en général que Y , disons, est d'autant plus près d'être une fonction affine de X que $|\rho_{X,Y}|$ est proche de 1.

Le problème de la régression linéaire

Les variables aléatoires X et $Y \in \mathcal{L}_d^2(\Omega, \mathcal{A}, P)$ étant données, on cherche une « meilleure » approximation de Y comme fonction affine de X au sens des moindres carrés, à savoir une solution en le couple $(a, b) \in \mathbb{R}^2$ du problème de minimisation :

$$\inf (\Phi(a, b) \mid (a, b) \in \mathbb{R}^2),$$

où

$$\Phi(a, b) = E[Y - (aX + b)]^2.$$

Ce problème est appelé, improprement d'ailleurs, **problème de régression linéaire**.

On a :

$$\begin{aligned} \Phi(a, b) &= E \left[\overset{\circ}{Y} - a\overset{\circ}{X} + (EY - aEX - b) \right]^2 \\ &= E \left[\overset{\circ}{Y} - a\overset{\circ}{X} \right]^2 + [EY - aEX - b]^2. \end{aligned}$$

Pour tout a fixé, cette quantité est minimum pour $\hat{b}_a = EY - aEX$. Reste à minimiser en a le polynôme

$$\Phi(a, \hat{b}_a) \equiv \sigma_Y^2 - 2a \operatorname{cov}(X, Y) + a^2 \sigma_X^2.$$

La valeur \hat{a} minimisant ce polynôme du second degré est solution unique de l'équation

$$-\text{cov}(X, Y) + a\sigma_X^2 = 0.$$

La solution du problème de régression linéaire est donc le couple (\hat{a}, \hat{b}_a) donné par :

$$\begin{cases} \hat{a} = \frac{\text{cov}(X, Y)}{\sigma_X^2} = \rho_{X, Y} \frac{\sigma_Y}{\sigma_X} \\ \hat{b}_a = EY - \hat{a}EX = EY - EX \cdot \rho_{X, Y} \frac{\sigma_Y}{\sigma_X}. \end{cases}$$

La droite D d'équation :

$$(y - EY) - \rho_{X, Y} \frac{\sigma_Y}{\sigma_X} (x - EX) = 0$$

est appelée **droite de régression linéaire** de Y en X. La « meilleure » approximation de Y comme fonction affine de X au sens des moindres carrés est

$$EY + \rho_{X, Y} \frac{\sigma_Y}{\sigma_X} (X - EX),$$

et on a $P[(X, Y) \in D] = 1$ si et seulement si $\Phi(\hat{a}, \hat{b}_a) = 0$.

Cas particulier. Si la variable aléatoire est de loi uniforme sur l'ensemble des n points du plan : $\{(x_i, y_i)\}_{1 \leq i \leq n}$, alors $\Phi(a, b) = \frac{1}{n} \sum_{i=1}^n [y_i - (ax_i + b)]^2$. On retrouve la droite d'approximation des moindres carrés des physiciens (exercice : déterminer l'équation de la droite D).

5.3. Fonctions génératrices

La fonction génératrice d'une variable aléatoire à valeurs dans \mathbb{N} caractérise la loi de cette variable aléatoire. Les fonctions génératrices sont un outil de calcul commode. Elles se prêtent bien au calcul de loi de sommes de variables aléatoires indépendantes, ainsi qu'à l'étude de la convergence en loi, comme nous le verrons dans le deuxième tome.

Dans cette section, sauf mention du contraire, les variables aléatoires seront définies sur l'espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{N} .

5.3.1. Définition

Lemme 5.22. Soit X une variable aléatoire définie sur l'espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{N} . Pour tout $s \in [-1, 1]$, la variable aléatoire s^X appartient à $\mathcal{L}_d^1(\Omega, \mathcal{A}, P)$.

Démonstration. On a $|s^X| \leq 1$. La variable aléatoire constante égale à 1 admet une moyenne, donc aussi s^X .

On note, quand cette quantité existe :

$$G_X(s) = E s^X.$$

Définition 5.23. La fonction G_X est appelée **fonction génératrice** de la variable aléatoire X .

Proposition 5.24. Soit X une variable aléatoire définie sur l'espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{N} , de fonction génératrice G_X , et de loi déterminée par, pour tout $n \in \mathbb{N}$,

$$P(X = n) = p_n.$$

(i) Le domaine de définition de G_X contient l'intervalle $[-1, 1]$.
On a :

$$\forall s \in [-1, 1], \quad |G_X(s)| \leq 1 \quad \text{et} \quad G_X(1) = 1.$$

(ii) Pour tout $s \in [-1, 1]$, on a :

$$G_X(s) = \sum_{n=0}^{+\infty} p_n s^n.$$

(iii) La fonction G_X est continue sur $[-1, 1]$, et C^∞ sur $] -1, 1[$.

(iv) La fonction génératrice G_X (et même sa restriction à l'intervalle $[-1, 1]$) caractérise la loi de X ; on a, pour tout $n \in \mathbb{N}$:

$$P(X = n) = \frac{G_X^{(n)}(0)}{n!}.$$

Démonstration.

(i) Le lemme précédent affirme que $G_X(s) = E(s^X)$ est défini pour $s \in [-1, 1]$. Pour établir l'inégalité, on constate que, pour tout $s \in [-1, 1]$,

$$|G_X(s)| \leq E|s|^X \leq E1 = 1.$$

(ii) Cela résulte du théorème de transfert.

(iii) Il suffit de remarquer que G_X est une série entière de rayon de convergence supérieur ou égal à 1 (cela vaut aussi pour le dernier point).

5.3.2. Fonctions génératrices des lois classiques à valeurs dans \mathbb{N}

Nous conservons les notations déjà employées. Soit $s \in [-1, 1]$ quelconque.

- Si $P_X = \mathcal{B}(n, p)$, on a

$$\boxed{G_X(s) = (ps + q)^n ;}$$

en effet,

$$G_X(s) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} s^k.$$

- Si $P_X = \mathcal{P}(\lambda)$, loi de Poisson de paramètre $\lambda > 0$, on a

$$\boxed{G_X(s) = \exp[\lambda(s - 1)] ;}$$

en effet,

$$G_X(s) = \sum_{k=0}^{+\infty} \exp(-\lambda) \frac{\lambda^k}{k!} s^k.$$

- Si $P_X = \mathcal{G}_{\mathbb{N}}(p)$, loi géométrique de paramètre $p \in]0, 1[$, on a

$$\boxed{G_X(s) = \frac{p}{1 - qs} ;}$$

en effet,

$$G_X(s) = \sum_{k=0}^{+\infty} pq^k s^k$$

- Si $P_X = \mathcal{G}_{\mathbb{N}^*}(p)$, loi géométrique de paramètre $p \in]0, 1[$:

$$\boxed{G_X(s) = \frac{ps}{1 - qs} ;}$$

en effet,

$$G_X(s) = \sum_{k=1}^{+\infty} pq^{k-1} s^k.$$

Proposition 5.25. Soient X et Y deux variables aléatoires indépendantes définies sur l'espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{N} , de fonctions génératrices G_X et G_Y ; la fonction génératrice de $X + Y$ est donnée, pour tout $s \in [-1, 1]$, par :

$$\boxed{G_{X+Y}(s) = G_X(s) G_Y(s) .}$$

Démonstration. On a $G_{X+Y}(s) = E(s^X s^Y)$ et les variables aléatoires s^X et s^Y sont indépendantes.

Corollaire 5.26. Soit X une variable aléatoire définie sur l'espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{N} , de loi **binomiale négative** $\mathcal{B}^-(n, p)$, sa fonction génératrice est donnée pour tout $s \in [-1, 1]$ par :

$$G_X(s) = \left(\frac{ps}{1-qs} \right)^n.$$

Démonstration. Si $(X_i)_{1 \leq i \leq n}$ est une famille de variables aléatoires indépendantes de même loi géométrique $\mathcal{G}_{N^*}(p)$, on a vu que la loi de la variable aléatoire $S_n = \sum_{i=1}^n X_i$ est la loi binomiale négative $\mathcal{B}^-(n, p)$. Reste à appliquer la proposition précédente, généralisée au cas de n variables aléatoires indépendantes.

5.3.3. Fonction génératrice et moments

La fonction génératrice d'une variable aléatoire déterminant sa loi, il est naturel qu'elle donne ses **moments**, lorsque ceux-ci existent.

Nous noterons, pour $r \in \mathbb{N}^*$, $G_X^{(r)}(1^-)$ la r -ième dérivée à gauche de G_X en 1 lorsqu'elle existe.

Proposition 5.27. Soit X une variable aléatoire définie sur l'espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{N} .

Pour que X admette un moment d'ordre $r \in \mathbb{N}^*$, il faut et il suffit que sa fonction génératrice G_X soit r fois dérivable à gauche en 1 et, dans ce cas, on a

$$G_X^{(r)}(1^-) = \sum_{k=r}^{+\infty} k(k-1) \cdots (k-r+1) p_k, \quad (5.26)$$

ce qui s'écrit encore :

$$E[X(X-1) \cdots (X-r+1)] = G_X^{(r)}(1^-). \quad (5.27)$$

En particulier (cas $r = 1$), on a

$$EX = G'_X(1^-). \quad (5.28)$$

Démonstration. Nous ne ferons la démonstration que pour $r = 1$, laissant le cas général en exercice. Pour tout $s \in [0, 1[$ et tout $n \in \mathbb{N}^*$, on a :

$$G_X(1) - G_X(s) = \sum_{n=1}^{+\infty} p_n (1 - s^n)$$

et, puisque

$$1 - s^n = (1 - s) \sum_{j=0}^{n-1} s^j,$$

on a

$$\frac{G_X(1) - G_X(s)}{1 - s} = \sum_{n=1}^{+\infty} p_n \left(\sum_{j=0}^{n-1} s^j \right), \quad (5.29)$$

avec, pour tout $n \in \mathbb{N}^*$ et tout $s \in [0, 1]$:

$$0 \leq p_n \left(\sum_{j=0}^{n-1} s^j \right) \leq np_n.$$

- Si X admet une moyenne, la série de fonctions de terme général $p_n \left(\sum_{j=0}^{n-1} s^j \right)$ est alors normalement donc uniformément convergente sur $[0, 1]$; il en résulte que, d'après (5.29), la limite à gauche en 1 de $\frac{G_X(1) - G_X(s)}{1 - s}$ existe et vaut $\sum_{n=1}^{+\infty} np_n \equiv EX$.
- Réciproquement, si la limite à gauche en 1 de $\frac{G_X(1) - G_X(s)}{1 - s}$ existe, il résulte de l'égalité (5.29) que l'on a, pour tout $N \in \mathbb{N}^*$:

$$\begin{aligned} 0 \leq \sum_{n=1}^N np_n &= \lim_{s \rightarrow 1^-} \sum_{n=1}^N p_n \left(\sum_{j=0}^{n-1} s^j \right) \\ &\leq \lim_{s \rightarrow 1^-} \sum_{n=1}^{+\infty} p_n \left(\sum_{j=0}^{n-1} s^j \right) = G'_X(1^-); \end{aligned}$$

la série de terme général *positif* np_n est donc convergente et la moyenne de X existe.

5.3.4. Somme d'un nombre aléatoire de variables aléatoires

Nous avons déjà rencontré dans des exemples des situations où nous avons à étudier la loi d'une **somme** de variables aléatoires indépendantes de **longueur aléatoire**. Les fonctions génératrices peuvent être un outil précieux pour ce faire.

Proposition 5.28. Soient $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires de même loi non dégénérée, à valeurs dans \mathbb{N} , et T une variable aléatoire à valeurs dans \mathbb{N}^* telle que $P(T = 1) \neq 1$. On suppose que les variables aléatoires $T, X_n, n \in \mathbb{N}^*$, sont indépendantes. On définit, pour tout

$n \in \mathbb{N}^*$, la variable aléatoire

$$S_n = \sum_{j=1}^n X_j.$$

Soit S la variable aléatoire définie par : pour tout $\omega \in \Omega$,

$$S(\omega) = S_{T(\omega)}(\omega).$$

Si G_T et G_{X_1} désignent respectivement les fonctions génératrices de T et X_1 , la fonction génératrice de S est donnée par :

$$G_S = G_T \circ G_{X_1}.$$

Démonstration. Puisque T prend ses valeurs dans \mathbb{N}^* , on a, pour tout $s \in [-1, 1]$:

$$\begin{aligned} G_S(s) &= \sum_{k=0}^{+\infty} s^k P(S = k) \\ &= \sum_{k=0}^{+\infty} s^k P \left[(S = k) \cap \left(\bigcup_{n \in \mathbb{N}^*} (T = n) \right) \right] \\ &= \sum_{k \in \mathbb{N}} s^k \left[\sum_{n \in \mathbb{N}^*} P(S = k, T = n) \right]. \end{aligned}$$

Mais on a

$$\sum_{k \in \mathbb{N}} |s|^k \left[\sum_{n \in \mathbb{N}^*} P(S = k, T = n) \right] \leq \sum_{k \in \mathbb{N}} \left[\sum_{n \in \mathbb{N}^*} P(S = k, T = n) \right] = 1.$$

La famille $\{s^k P(S = k, T = n)\}_{(k,n) \in \mathbb{N} \times \mathbb{N}^*}$ est donc sommable et, d'après la propriété de Fubini, puis l'indépendance des variables aléatoires S_n et T , on peut écrire :

$$\begin{aligned} G_S(s) &= \sum_{n \in \mathbb{N}^*} \left[\sum_{k \in \mathbb{N}} s^k P(S_n = k, T = n) \right] \\ &= \sum_{n \in \mathbb{N}^*} \left[\sum_{k \in \mathbb{N}} s^k P(S_n = k) \right] P(T = n) \\ &= \sum_{n \in \mathbb{N}^*} G_{S_n}(s) P(T = n). \end{aligned}$$

Mais les variables aléatoires X_n sont indépendantes et de même loi (donc de même fonction génératrice). Il résulte de la proposition 5.25 que :

$$G_S(s) = \sum_{n=1}^{+\infty} [G_{X_1}(s)]^n P(T = n) = G_T[G_{X_1}(s)].$$

C'est le résultat annoncé.

Nous donnons dans ce même contexte des relations sur les **moyenne** et **variance** de S .

Corollaire 5.29. (Identités de Wald.) *Sous les mêmes hypothèses que dans la proposition 5.28 :*

(i) *Si X_1 et T admettent une moyenne, S admet une moyenne donnée par :*

$$\boxed{ES = (EX_1) (ET)}. \quad (5.30)$$

(ii) *Si X_1 et T admettent un moment d'ordre deux, il en est de même pour S et sa variance est donnée par :*

$$\boxed{\sigma_S^2 = (\sigma_{X_1}^2) (ET) + (EX_1)^2 \sigma_T^2}. \quad (5.31)$$

Démonstration.

(i) Il faut d'abord remarquer qu'il ne peut pas exister de théorème général sur la dérivation à gauche (ou à droite) de fonctions composées (réfléchir à des contre-exemples). On ne peut donc pas obtenir (5.30) en observant que les fonctions génératrices G_T et G_{X_1} admettent une dérivée à gauche en 1 (proposition 5.27) et en écrivant directement

$$G'_S(1^-) = G'_T(1^-) G'_{X_1}(1^-).$$

Toutefois, cette écriture peut être justifiée en observant que puisque $G_{X_1}(1) = 1$ et $G'_{X_1}(1^-) > 0$ ($G'_{X_1}(1^-) = E(X_1)$ et X_1 est non dégénérée), il existe $s_0 \in [0, 1[$ tel que, pour tout $s \in [s_0, 1[$, on ait : $G_{X_1}(s) < 1$. On a alors, pour tout $s \in [s_0, 1[$,

$$\frac{G_S(1) - G_S(s)}{1 - s} = \frac{1 - G_T[G_{X_1}(s)]}{1 - s} = \frac{1 - G_T[G_{X_1}(s)]}{1 - G_{X_1}(s)} \times \frac{1 - G_{X_1}(s)}{1 - s}.$$

Puisque

$$\lim_{s \rightarrow 1} G_{X_1}(s) = 1,$$

il en résulte que la limite de $\frac{G_S(1) - G_S(s)}{1 - s}$ quand s tend vers 1, autrement dit la dérivée à gauche en 1 de G_S , existe, et est donnée par :

$$G'_S(1^-) \equiv \lim_{s \rightarrow 1} \frac{G_S(1) - G_S(s)}{1 - s} = G'_T(1^-) G'_{X_1}(1^-);$$

la proposition 5.27 assure alors que S admet une moyenne donnée par l'égalité (5.30).

(ii) On a, pour tout $s \in]-1, 1[$:

$$G_S''(s) = G_T''(G_{X_1}(s)) [G_{X_1}''(s)]^2 + G_T'(G_{X_1}(s)) G_{X_1}''(s).$$

Il résulte de la proposition 5.27 que, puisque $\lim_{s \rightarrow 1} G_{X_1}(s) = 1$, la limite à gauche de $G_S''(s)$ quand s tend vers 1 existe et est donnée par :

$$\lim_{s \rightarrow 1^-} G_S''(s) = E[T(T-1)] (EX_1)^2 + (ET) (E[X(X-1)]). \quad (5.32)$$

Démontrons alors que $E[S(S-1)]$ existe. D'après la proposition 5.27, pour tout $s \in]-1, 1[$:

$$G_S''(s) = \sum_{n=2}^{+\infty} n(n-1) s^{n-2} P(S=n).$$

Alors, G_S'' étant bornée au voisinage de 1, il existe $c > 0$ et $s_1 \in [0, 1[$ tels que :

$$\forall s \in [s_1, 1[, \quad \forall N > 2, \quad 0 \leq \sum_{n=2}^N n(n-1) s^{n-2} P(S=n) \leq c.$$

Par continuité, on a donc, pour tout $N > 2$,

$$0 \leq \sum_{n=2}^N n(n-1) P(S=n) \leq c,$$

ce qui démontre que $E[S(S-1)]$ existe et vaut $\lim_{s \rightarrow 1^-} G_S''(s)$. On a donc, d'après l'égalité (5.32) :

$$E[S(S-1)] = E[T(T-1)] (EX_1)^2 + (ET) (E[X(X-1)]).$$

Un calcul algébrique simple, tenant compte de la relation

$$\sigma_S^2 = E[S(S-1)] + ES - (ES)^2$$

conduit à l'égalité (5.31).

Remarque. On peut bien sûr établir les identités de Wald en raisonnant directement sur les moments, sans passer par les fonctions génératrices (ce qui revient à étudier la loi de S) ; ce peut être un exercice !

Exercices

Exercice 5.1. Calcul de moyenne. Un joueur joue à **pile ou face** contre une banque avec une pièce non nécessairement équilibrée (p est la probabilité d'obtenir pile en un coup). Sa mise initiale est $a (\in \mathbb{N}^*)$. Tant qu'il obtient face, il perd ce qu'il a misé et décide de miser au prochain coup $k (> 1)$ fois la mise précédente. Quand il obtient pile, il gagne k fois sa dernière mise et s'arrête de jouer. On demande le gain moyen du joueur.

Solution. Soit sur l'espace probabilisé (Ω, \mathcal{A}, P) une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes, de même loi de Bernoulli $\mathcal{B}(1, p)$, l'événement $(X_n = 1)$ étant associé au fait que le joueur obtienne pile au n -ième coup. Soit T la variable aléatoire donnant le numéro du jet qui sort pile pour la première fois :

$$T = \inf \{n \geq 1 \mid X_n = 1\},$$

avec la convention usuelle que $\inf \emptyset = +\infty$. La variable aléatoire G , gain algébrique du joueur, est alors définie par,

$$\text{sur } (T = 1), \quad G = -a + ka,$$

et, pour tout $n \geq 2$,

$$\text{sur } (T = n), \quad G = -a(1 + k + \dots + k^{n-1}) + k^n a,$$

soit encore, après réduction :

$$\text{sur } (T = n), \quad G = k^n a \left(\frac{2-k}{1-k} \right) - \frac{a}{1-k}.$$

On constate que cette formule est valide dans le cas $n = 1$ et que $P(T < +\infty) = 1$.

On définit, pour tout $n \in \mathbb{N}^*$,

$$u_n = k^n a \left(\frac{2-k}{1-k} \right) - \frac{a}{1-k}.$$

Soit Φ l'application de $\mathbb{N}^* \cup \{+\infty\}$ dans \mathbb{R} définie par : pour tout $t \in \mathbb{N}^* \cup \{+\infty\}$,

$$\Phi(t) = \sum_{n \in \mathbb{N}^*} \mathbf{1}_{\{n\}}(t) u_n.$$

Le gain peut s'écrire

$$G = \Phi(T);$$

le théorème de transfert donne alors :

$$EG = \sum_{t \in \mathbb{N}^*} \Phi(t) P(T = t),$$

soit, puisque $P(T = +\infty) = 0$:

$$EG = \sum_{n \in \mathbb{N}^*} u_n P(T = n).$$

Mais T suit une loi géométrique sur \mathbb{N}^* de paramètre p et donc, en notant $q = 1 - p$:

$$EG = \sum_{n \in \mathbb{N}^*} u_n p q^{n-1}.$$

Il en résulte que l'on a :

- si $qk \geq 1$ et $k \neq 2$, $EG = +\infty$,
- si $k = 2$, $EG = a$ (puisque alors $G = a$),
- si $qk < 1$, en reconnaissant la somme d'une série géométrique,

$$EG = a \left(\frac{2-k}{1-k} \right) \sum_{n=1}^{+\infty} k^n p q^{n-1} - \frac{a}{1-k} = a \left(\frac{2-k}{1-k} \right) \left(\frac{pk}{1-qk} \right) - \frac{a}{1-k},$$

soit, après réduction :

$$EG = a \frac{pk - 1}{1 - qk}.$$

Exercice 5.2. Moyenne et moments d'ordre supérieur. Soient X et Y deux variables aléatoires discrètes définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , indépendantes, de même loi uniforme sur l'ensemble $\{0, 1, 2, \dots, n\}$. Soient Z et T les variables aléatoires

$$Z = |X - Y| \quad \text{et} \quad T = \min(X, Y).$$

1. Cette question est relative au théorème de transfert et à la propriété de linéarité de l'espérance.

(a) Justifier l'existence des moments de tout ordre de Z et T .

(b) Calculer EZ (sans calculer la loi de Z) et en donner un équivalent lorsque n tend vers l'infini.

(c) En déduire ET et en donner un équivalent lorsque n tend vers l'infini. On rappelle l'égalité, pour tout $(a, b) \in \mathbb{R}^2$:

$$|a - b| = a + b - 2 \min(a, b).$$

2. Soit U une variable aléatoire à valeurs dans \mathbb{N} , telle qu'existe $K \in \mathbb{N}^*$ vérifiant :

$$0 \leq U \leq K.$$

(a) Exprimer $\sum_{j=1}^K P(U \geq j)$ en fonction de EU .

(b) Calculer de même $\sum_{j=1}^K j^2 P(U \geq j)$ en fonction de EU , EU^2 et EU^3 .

3. Pour étudier la loi d'un minimum de variables aléatoires, il est agréable de calculer la probabilité que celui-ci soit supérieur ou égal à un nombre quelconque.

(a) Calculer, pour tout $j \in \mathbb{N}$, la probabilité $P(T \geq j)$.

(b) En utilisant la question 2a, retrouver la valeur de ET .

4. Il s'agit de tirer parti au maximum de la propriété de linéarité de l'espérance.

(a) Calculer EZ^2 en fonction de la variance σ_X^2 de la variable aléatoire X .

(b) Que vaut la variance σ_Z^2 de la variable aléatoire Z ?

Solution.

1. Toute variable aléatoire **positive dominée** par une variable aléatoire qui admet une moyenne en admet aussi une.

(a) On a, X et Y étant à valeurs positives :

$$0 \leq Z \leq X + Y \quad \text{et} \quad 0 \leq T \leq X.$$

Les variables aléatoires X et Y admettent une moyenne, il en est alors de même de Z et T .

Remarque. On peut aussi dire, qu'avec probabilité 1, ces variables aléatoires sont **bornées** et admettent donc une moyenne.

(b) Utilisant successivement le théorème de transfert puis l'indépendance de X et Y , il vient, en notant E l'ensemble $\{0, 1, 2, \dots, n\}$,

$$\begin{aligned} EZ &= \sum_{(x,y) \in E^2} |x - y| P(X = x, Y = y) \\ &= \sum_{(x,y) \in E^2} |x - y| P(X = x) P(Y = y), \end{aligned}$$

soit :

$$\begin{aligned} EZ &= \frac{1}{(n+1)^2} \sum_{(x,y) \in E^2} |x - y| \\ &= \frac{2}{(n+1)^2} \sum_{i=1}^n \left(\sum_{j=0}^{i-1} (i - j) \right) \\ &= \frac{2}{(n+1)^2} \sum_{i=1}^n \left[i^2 - \sum_{j=0}^{i-1} j \right] \\ &= \frac{2}{(n+1)^2} \sum_{i=1}^n \left[i^2 - \frac{(i-1)i}{2} \right] \\ &= \frac{2}{(n+1)^2} \sum_{i=1}^n \left[\frac{i^2 + i}{2} \right]. \end{aligned}$$

En se rappelant que

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6},$$

et en simplifiant, il vient :

$$EZ = \frac{n(n+2)}{3(n+1)}.$$

Quand n tend vers l'infini, on a $EZ \sim \frac{n}{3}$.

(c) On a

$$T = \frac{1}{2} [X + Y - |X - Y|],$$

et donc :

$$ET = \frac{1}{2} [EX + EY - EZ].$$

Mais, puisque

$$EX = EY = \sum_{j=0}^n \frac{j}{n+1} = \frac{n}{2},$$

il vient :

$$ET = \frac{n(2n+1)}{6(n+1)}.$$

Quand n tend vers l'infini, $ET \sim \frac{n}{3}$.

2. (a) On a

$$\sum_{j=1}^K P(U \geq j) = \sum_{j=1}^K E[\mathbf{1}_{(U \geq j)}] = E\left[\sum_{j=1}^K \mathbf{1}_{(U \geq j)}\right] = E\left[\sum_{j=1}^{\min(K,U)} 1\right],$$

et, puisque $0 \leq U \leq K$:

$$EU = \sum_{j=1}^K P(U \geq j).$$

Cette formule liant la moyenne et les probabilités $P(U \geq j)$ est en fait très générale et d'un usage très fructueux.

(b) De même, on a

$$\sum_{j=1}^K j^2 P(U \geq j) = \sum_{j=1}^K j^2 E[\mathbf{1}_{(U \geq j)}] = E\left[\sum_{j=1}^K j^2 \mathbf{1}_{(U \geq j)}\right] = E\left[\sum_{j=1}^{\min(K,U)} j^2\right],$$

soit

$$\sum_{j=1}^K j^2 P(U \geq j) = E\left[\sum_{j=1}^U j^2\right] = E\left[\frac{U(U+1)(2U+1)}{6}\right],$$

ou encore :

$$\sum_{j=1}^K j^2 P(U \geq j) = \frac{1}{3} EU^3 + \frac{1}{2} EU^2 + \frac{1}{6} EU.$$

3. Dire que le minimum de deux nombres est supérieur ou égal à un troisième est équivalent à dire que ces deux nombres sont supérieurs ou égaux à ce troisième.

(a) Il résulte de l'indépendance de X et Y que, pour tout $j \in \mathbb{N}$, on a

$$P(T \geq j) = P(X \geq j, Y \geq j) = P(X \geq j) P(Y \geq j),$$

et donc

$$P(T \geq j) = \begin{cases} 0 & \text{si } j > n \\ \left(\sum_{k=j}^n P(X = k) \right)^2 & \text{si } 0 \leq j \leq n, \end{cases}$$

soit :

$$P(T \geq j) = \begin{cases} \left(\frac{n-j+1}{n+1} \right)^2 & \text{si } 0 \leq j \leq n \\ 0 & \text{si } j > n. \end{cases}$$

(b) Puisque $0 \leq T \leq n$, on a, d'après les résultats précédents :

$$ET = \sum_{j=1}^n P(T \geq j) = \sum_{j=1}^n \left(\frac{n-j+1}{n+1} \right)^2,$$

soit, après simplifications :

$$ET = \frac{n(2n+1)}{6(n+1)}.$$

4. Il s'agit d'appliquer au mieux la formule de Leibniz.

(a) On a

$$EZ^2 = E(X - Y)^2 = EX^2 + EY^2 - 2E(XY),$$

et, puisque X et Y ont même loi (donc en particulier même moment) et sont indépendantes :

$$EZ^2 = 2EX^2 - 2(EX)^2,$$

ce qui conduit à :

$$EZ^2 = 2\sigma_X^2.$$

(b) La variance de Z est donnée par :

$$\sigma_Z^2 = EZ^2 - (EZ)^2 = 2\sigma_X^2 - (EZ)^2;$$

or on a

$$EX^2 = \sum_{j=0}^n \frac{j^2}{n+1} = \frac{n(2n+1)}{6}$$

et

$$\sigma_X^2 = EX^2 - (EX)^2 = \frac{n(2n+1)}{6} - \frac{n^2}{4},$$

soit

$$\sigma_X^2 = \frac{n(n+2)}{12},$$

ce qui donne :

$$\sigma_Z^2 = \frac{n(n+2)}{6} - \frac{n^2(n+2)^2}{9(n+1)^2}.$$

Après simplifications, on obtient :

$$\sigma_Z^2 = \frac{n(n+2)(n^2+2n+3)}{18(n+1)^2}.$$

On remarque que, lorsque n tend vers l'infini, $\sigma_X^2 \sim \frac{n^2}{12}$ et $\sigma_Z^2 \sim \frac{n^2}{18}$.

Exercice 5.3. Loi d'un maximum, moyenne et moments d'ordre supérieur. Toutes les variables aléatoires sont définies sur l'espace probabilisé (Ω, \mathcal{A}, P) .

1. Soit U une variable aléatoire discrète à valeurs dans \mathbb{N} admettant un moment d'ordre deux.

(a) Démontrer la formule : $EU = \sum_{j \geq 1} P(U \geq j)$.

(b) Exprimer de même la somme $\sum_{j \geq 1} jP(U \geq j)$ en fonction de $E(U^2)$ et $E(U)$.

2. Soient $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires discrètes à valeurs dans \mathbb{N} , indépendantes, de même loi μ . On note, pour tout $k \in \mathbb{N}$, $p_k = P(X_n = k)$ et $F_k = \sum_{j=0}^k p_j$. Soit, pour tout $n \in \mathbb{N}^*$, la variable aléatoire discrète $M_n = \max_{1 \leq i \leq n} (X_i)$.

Calculer, pour tout $k \in \mathbb{N}$, la probabilité $P(M_n \leq k)$ en fonction de F_k et n .

3. On suppose que μ est la loi uniforme sur l'ensemble des entiers $\{1, 2, \dots, K\}$, où K est un entier strictement supérieur à 1.

(a) Calculer, pour tout $k \in \{1, 2, \dots, K\}$, la probabilité $P(M_n = k)$.

(b) On jette trois dés équilibrés; quelle est la probabilité que le maximum des chiffres obtenus soit 4?

4. On suppose maintenant que μ est la loi géométrique sur \mathbb{N}^* de paramètre p ($0 < p < 1$) et on note $q = 1 - p$.

(a) Calculer la moyenne $E(M_n)$.

(b) Trois joueurs jouent à pile ou face avec une pièce équilibrée et s'arrêtent dès qu'ils ont obtenu un pile. La variable aléatoire M_3 est alors le nombre de jets effectués par le ou les joueurs ayant obtenu pile en dernier. Calculer $E(M_3)$. On donnera une valeur sous forme de fraction irréductible, puis décimale au centième près.

Solution.

1. La question est analogue à celle abordée à l'exercice précédent; ici, on partitionne l'ensemble $(U \geq j)$ en une suite infinie d'ensembles et on utilise la propriété de Fubini pour les familles positives.

(a) Puisque l'on a

$$(U \geq j) = \bigcup_{k \geq j} (U = k),$$

il résulte de la σ -additivité de P , de la positivité des termes et de la propriété de Fubini que l'on a successivement

$$\begin{aligned} \sum_{j \geq 1} P(U \geq j) &= \sum_{j \geq 1} P\left(\bigcup_{k \geq j} (U = k)\right) \\ &= \sum_{j \geq 1} \left[\sum_{k \geq j} P(U = k) \right] \\ &= \sum_{k \geq 1} \left[\sum_{j=1}^k P(U = k) \right] \\ &= \sum_{k \geq 1} k P(U = k), \end{aligned}$$

soit :

$$\boxed{\sum_{j \geq 1} P(U \geq j) = EU.}$$

(b) Pour les mêmes raisons, on a

$$\begin{aligned} \sum_{j \geq 1} j P(U \geq j) &= \sum_{j \geq 1} j P\left(\bigcup_{k \geq j} (U = k)\right) \\ &= \sum_{j \geq 1} \left[\sum_{k \geq j} j P(U = k) \right] \\ &= \sum_{k \geq 1} \left[\sum_{j=1}^k j P(U = k) \right] \\ &= \sum_{k \geq 1} \frac{k(k+1)}{2} P(U = k), \end{aligned}$$

soit :

$$\boxed{\sum_{j \geq 1} j P(U \geq j) = \frac{1}{2} [EU^2 + EU].}$$

2. Maximum et fonction de répartition sont des notions relatives à la structure d'ordre; il est naturel d'accéder à la loi d'un maximum de variables aléatoires en termes de fonctions de répartition. On a :

$$(M_n \leq k) = \bigcap_{i=1}^n (X_i \leq k).$$

Les variables aléatoires X_i étant indépendantes de même loi, on a, pour tout $k \in \mathbb{N}$:

$$P(M_n \leq k) = \prod_{i=1}^n P(X_i \leq k) = (F_k)^n.$$

3. On illustre les résultats de la question précédente.

(a) Si μ est la loi uniforme sur $\{1, 2, \dots, K\}$, pour tout $k \in \{1, 2, \dots, K\}$, on a $p_k = \frac{1}{K}$; de plus $F_0 = 0$. Pour tout $k \in \{1, 2, \dots, K\}$, il vient alors

$$F_k = \frac{k}{K} \text{ et}$$

$$P(M_n = k) = P(M_n \leq k) - P(M_n \leq k-1) = (F_k)^n - (F_{k-1})^n,$$

soit :

$$P(M_n = k) = \frac{1}{K^n} [k^n - (k-1)^n].$$

(b) Les dés étant équilibrés, μ est la loi uniforme sur $\{1, 2, \dots, 6\}$. On cherche $P(M_3 = 4)$, soit

$$P(M_3 = 4) = \frac{1}{6^3} (4^3 - 3^3) = \frac{37}{216} \simeq 0,171.$$

4. Il s'agit de calculer la moyenne d'un maximum de variables aléatoires indépendantes de loi géométrique et de l'illustrer par un jeu à trois joueurs.

(a) Si μ est la loi géométrique sur \mathbb{N}^* de paramètre p , on a

$$F_k = \sum_{j=1}^k pq^{j-1} = 1 - q^k,$$

et donc

$$P(M_n \leq k) = (1 - q^k)^n.$$

On a alors

$$P(M_n \geq 1) = 1,$$

et, pour $k \geq 2$:

$$\begin{aligned} P(M_n \geq k) &= 1 - P(M_n \leq k-1) \\ &= 1 - (1 - q^{k-1})^n = \sum_{j=1}^n \binom{n}{j} (-1)^{j+1} q^{j(k-1)}. \end{aligned}$$

Il résulte alors de la première question que

$$\begin{aligned} EM_n &= 1 + \sum_{k \geq 2} \left(\sum_{j=1}^n \binom{n}{j} (-1)^{j+1} q^{j(k-1)} \right) \\ &= 1 + \sum_{j=1}^n \left[\binom{n}{j} (-1)^{j+1} \left(\sum_{k \geq 2} q^{j(k-1)} \right) \right], \end{aligned}$$

soit :

$$EM_n = 1 + \sum_{j=1}^n \binom{n}{j} (-1)^{j+1} \frac{q^j}{1 - q^j}.$$

(b) Dans ce cas, on a $p = q = \frac{1}{2}$ et $n = 3$, ce qui donne :

$$EM_3 = 1 + \binom{3}{1} \frac{\frac{1}{2}}{1 - \frac{1}{2}} - \binom{3}{2} \frac{\frac{1}{4}}{1 - \frac{1}{4}} + \binom{3}{3} \frac{\frac{1}{8}}{1 - \frac{1}{8}},$$

ou

$$EM_3 = \frac{22}{7} \approx 3,14.$$

Exercice 5.4. Moyenne et indépendance. Soient X et Y deux variables aléatoires discrètes définies sur le même espace probabilisé (Ω, \mathcal{A}, P) , indépendantes et de même loi géométrique sur \mathbb{N}^* de paramètre p ($0 < p < 1$) ; on note $q = 1 - p$. On définit les variables aléatoires T , Z et G par :

$$T = \min(X, Y) \quad Z = |X - Y| \quad G = \frac{Z}{T}.$$

1. Il s'agit d'étudier la loi d'un **minimum** de variables aléatoires.

(a) Calculer, pour tout $x \in \mathbb{N}^*$, la probabilité $P(X \geq x)$.

(b) Calculer, pour tout $t \in \mathbb{N}^*$, la probabilité $P(T \geq t)$ et identifier la loi de T .

2. On fait des calculs de moyennes et on fait l'étude de la loi du couple (T, Z) .

(a) Calculer les moyennes $E(X)$ et $E\left(\frac{1}{X}\right)$.

(b) Calculer, pour $(t, z) \in \mathbb{N}^* \times \mathbb{N}$, la probabilité $P(T \geq t, Z = z)$; étudier séparément le cas $z = 0$.

(c) En déduire la loi de Z .

3. Démontrer alors que les variables aléatoires T et Z sont indépendantes, c'est-à-dire que, pour tout $(t, z) \in \mathbb{N}^* \times \mathbb{N}$, on a

$$P(T = t, Z = z) = P(T = t) P(Z = z).$$

On remarquera que $P(T \geq t, Z \geq z)$ s'écrit sous la forme $f(t)g(z)$, où f et g sont des fonctions définies respectivement sur \mathbb{N}^* et \mathbb{N} .

4. Que vaut la moyenne $E(G)$?

Solution.

1. On a déjà remarqué que la loi d'une variable aléatoire discrète X à valeurs dans \mathbb{N} est caractérisée par la donnée, pour tout $x \in \mathbb{N}$, des probabilités

$P(X \geq x)$.

(a) Pour tout $x \in \mathbb{N}^*$, on a

$$P(X \geq x) = \sum_{k=x}^{+\infty} pq^{k-1} = p \frac{q^{x-1}}{1-q},$$

soit :

$$P(X \geq x) = q^{x-1}.$$

(b) Les variables aléatoires X et Y sont indépendantes ; on a donc, pour tout $x \in \mathbb{N}^*$,

$$P(T \geq t) = P(X \geq t, Y \geq t) = q^{2(t-1)}.$$

Puisque l'on a

$$P(T = t) = P(T \geq t) - P(T \geq t+1),$$

il vient :

$$P(T = t) = q^{2(t-1)} (1 - q^2),$$

c'est-à-dire que la loi de T est la loi géométrique sur \mathbb{N}^* de paramètre $1 - q^2$.

2. Il s'agit d'appliquer le théorème de transfert pour calculer une moyenne.

(a) On rappelle que :

$$E(X) = \frac{1}{p}.$$

D'après le théorème de transfert, on a

$$E\left(\frac{1}{X}\right) = \sum_{k=1}^{+\infty} \frac{1}{k} P(X = k),$$

soit, après réduction :

$$E\left(\frac{1}{X}\right) = p \sum_{k=0}^{+\infty} \frac{1}{k+1} q^k.$$

Mais, si $0 < x < 1$, on a

$$\sum_{k=0}^{+\infty} \frac{x^{k+1}}{k+1} = -\ln(1-x);$$

il vient alors :

$$E\left(\frac{1}{X}\right) = -\frac{p}{q} \ln p.$$

– Pour tout $(t, z) \in \mathbb{N}^* \times \mathbb{N}$, on a

$$P(T \geq t, Z = z) = P(X \geq t, Y \geq t, |X - Y| = z),$$

soit, en notant

$$D_{t,z} = \{(x, y) \in \mathbb{N}^* \times \mathbb{N}^* \mid x \geq t, y \geq t, |x - y| = z\},$$

$$P(T \geq t, Z = z) = \sum_{(x,y) \in D_{t,z}} p^2 q^{x+y-2}.$$

On a

$$D_{t,0} = \{(x, y) \in \mathbb{N}^* \times \mathbb{N}^* \mid x \geq t, x = y\},$$

ce qui donne

$$P(T \geq t, Z = 0) = \sum_{x \geq t} p^2 q^{2(x-1)},$$

soit, après calcul, pour tout $t \in \mathbb{N}^*$:

$$P(T \geq t, Z = 0) = \frac{p}{1+q} q^{2(t-1)}.$$

- Pour tout $z \in \mathbb{N}^*$, on a

$$D_{t,z} = \{(x, y) \in \mathbb{N}^* \times \mathbb{N}^* \mid x \geq t, y = x + z\}$$

$$\cup \{(x, y) \in \mathbb{N}^* \times \mathbb{N}^* \mid y \geq t, x = y + z\},$$

ce qui permet d'écrire

$$P(T \geq t, Z = z) = 2p^2 \sum_{y \geq t} q^{2(y-1)+z} = 2p^2 \frac{q^{2(t-1)+z}}{1-q^2}.$$

On a donc, pour tout $(t, z) \in \mathbb{N}^* \times \mathbb{N}$:

$$P(T \geq t, Z = z) = \frac{2p}{1+q} q^{2(t-1)+z}.$$

(b) On a, pour tout $z \in \mathbb{N}^*$,

$$P(Z = z) = P(T \geq 1, Z = z) = \frac{2p}{1+q} q^z,$$

et

$$P(Z = 0) = \frac{p}{1+q}.$$

3. Il résulte de la question précédente qu'il existe des fonctions positives f et g telles que l'on ait, pour tout $(t, z) \in \mathbb{N}^* \times \mathbb{N}$:

$$P(T \geq t, Z = z) = f(t) g(z).$$

Il en résulte que

$$P(T \geq t) = \sum_{z \in \mathbb{N}} P(T \geq t, Z = z) = f(t) \sum_{z \in \mathbb{N}} g(z),$$

soit, si $K = \sum_{z \in \mathbb{N}} g(z)$, pour tout $t \in \mathbb{N}^*$,

$$P(T \geq t) = K f(t).$$

En prenant $t = 1$, il vient $K = \frac{1}{f(1)}$. De plus, pour tout $z \in \mathbb{N}$, on a

$$P(Z = z) = P(T \geq 1, Z = z) = f(1) g(z);$$

en combinant ces résultats, on obtient donc, pour tout $(t, z) \in \mathbb{N}^* \times \mathbb{N}$:

$$P(T \geq t) P(Z = z) = f(t) g(z).$$

Comme par ailleurs pour tout $(t, z) \in \mathbb{N}^* \times \mathbb{N}$

$$\begin{aligned} P(T = t, Z = z) &= P(T \geq t, Z = z) - P(T \geq t + 1, Z = z) \\ &= f(t) g(z) - f(t + 1) g(z) \\ &= [f(t) - f(t + 1)] g(z) \\ &= \frac{1}{K} [P(T \geq t) - P(T \geq t + 1)] g(z), \end{aligned}$$

il vient, pour tout $(t, z) \in \mathbb{N}^* \times \mathbb{N}$

$$P(T = t, Z = z) = P(T = t) P(Z = z),$$

ce qui démontre que **les variables aléatoires T et Z sont indépendantes.**

4. Il en résulte que les variables aléatoires Z et $\frac{1}{T}$ sont indépendantes, et donc que l'on a :

$$EG = (EZ) \left[E\left(\frac{1}{T}\right) \right].$$

La loi de la variable aléatoire T étant la loi géométrique sur \mathbb{N}^* de paramètre $1 - q^2$, il résulte de la deuxième question que :

$$E\left(\frac{1}{T}\right) = -\frac{1 - q^2}{q^2} \ln(1 - q^2) = -\frac{(1 + q)p}{q^2} \ln[p(1 + q)];$$

de plus, d'après la question précédente, on a

$$EZ = \frac{2p}{1 + q} \sum_{z \geq 1} zq^z = \frac{2q}{1 + q} \sum_{z \geq 1} zpq^{z-1};$$

en identifiant la somme figurant au membre de droite à la moyenne d'une variable aléatoire de loi géométrique sur \mathbb{N}^* de paramètre p , on obtient :

$$EZ = \frac{2q}{(1 + q)p},$$

ce qui conduit à

$$EG = -\frac{2}{q} \ln[p(1 + q)].$$

Exercice 5.5. Moments et indépendance. Soient U et V deux variables aléatoires discrètes définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{Z} , indépendantes, admettant un moment d'ordre deux. On suppose U **centrée**. On définit les deux variables aléatoires discrètes X et Y par :

$$X = (-1)^V U \quad Y = V.$$

1. On étudie quelques questions d'indépendance et de corrélation.

- (a) Justifier l'existence de la moyenne de X et la calculer.
 (b) Justifier l'existence de la moyenne de XY et la calculer.
 (c) Que vaut la covariance de X et Y ?
 (d) Les variables aléatoires X^2 et Y^2 sont-elles indépendantes ?

2. Dans cette question, on suppose que la loi de U est donnée par :

$$P(U = -2) = \frac{1}{3} \quad \text{et} \quad P(U = 1) = \frac{2}{3},$$

et celle de V par :

$$P(V = 1) = \frac{1}{2} \quad \text{et} \quad P(V = 2) = \frac{1}{2}.$$

- (a) Calculer $E(X^3)$ et $E(U^3)$.
 (b) Calculer $E[\mathbf{1}_{(V=1)} X^3]$.
 (c) Les variables aléatoires X et Y sont-elles indépendantes ?

3. On suppose ici que U a une **loi symétrique** (c'est-à-dire que U et $-U$ ont même loi) et que $P(U = 0) = 0$. Soient f et g deux fonctions réelles **bornées** quelconques définies sur \mathbb{Z} .

- (a) Justifier l'existence de la moyenne de la variable aléatoire discrète $f(X)g(Y)$.
 (b) Démontrer que :

$$E[f(X)g(Y)] = \begin{cases} E[f(U)] E[g(V)] & \text{si } f \text{ paire} \\ 0 & \text{si } f \text{ impaire.} \end{cases}$$

(c) En déduire que, pour toutes fonctions f et g réelles **bornées** définies sur \mathbb{Z} , on a :

$$E[f(X)g(Y)] = E[f(X)] E[g(Y)].$$

(d) Que dire de l'indépendance des variables aléatoires X et Y ?

Solution.

1. Il est évident qu'en général les variables aléatoires X et Y ne sont pas indépendantes.

(a) On a $|X| \leq |U|$; la variable aléatoire U admet une moyenne donc X aussi; les variables aléatoires $(-1)^V$ et U sont indépendantes, donc :

$$EX \equiv E[(-1)^V U] = E[(-1)^V] EU,$$

et, puisque U est centrée, on a :

$$EX = 0.$$

(b) On a : $|XY| \leq |U||V|$. Les variables aléatoires U et V admettant un moment d'ordre 2, le produit UV admet une moyenne; il en est donc de même de la variable aléatoire XY . Les variables aléatoires $(-1)^V V$ et U étant indépendantes, on a

$$E(XY) = E[(-1)^V V] EU,$$

et donc :

$$E(XY) = 0.$$

(c) Il en résulte que :

$$\text{cov}(X, Y) \equiv E(XY) - (EX)(EY) = 0.$$

(d) On a $X^2 = U^2$ et $Y^2 = V^2$; les variables aléatoires U et V étant indépendantes, il en est de même de U^2 et V^2 , ce qui démontre que les variables aléatoires X^2 et Y^2 sont indépendantes.

2. Il s'agit de démontrer que, dans ce cas, les variables aléatoires X et Y ne sont pas indépendantes.

(a) Remarquons que, dans ce cas, U est bien centrée. On a

$$\begin{aligned} EX^3 &= E[\mathbf{1}_{(V=1)}X^3] + E[\mathbf{1}_{(V=2)}X^3] \\ &= E[\mathbf{1}_{(V=1)}(-1)^3 U^3] + E[\mathbf{1}_{(V=2)}(-1)^6 U^3], \end{aligned}$$

soit, par indépendance des variables aléatoires U et V :

$$\begin{aligned} EX^3 &= -E[\mathbf{1}_{(V=1)}] E[U^3] + E[\mathbf{1}_{(V=2)}] E[U^3] \\ &= E[U^3] [P(V=2) - P(V=1)]. \end{aligned}$$

On a donc :

$$EX^3 = 0.$$

Par ailleurs, on a

$$E[U^3] = (-2)^3 \frac{1}{3} + (1)^3 \frac{2}{3},$$

soit :

$$EU^3 = -2.$$

(b) On a :

$$E[\mathbf{1}_{(V=1)}X^3] = E[\mathbf{1}_{(V=1)}(-1)^3 U^3],$$

soit, par indépendance des variables aléatoires U et V :

$$E[\mathbf{1}_{(V=1)}X^3] = -E[\mathbf{1}_{(V=1)}] E[U^3] = -P(V=1) E[U^3].$$

On a donc :

$$E[\mathbf{1}_{(V=1)}X^3] = 1.$$

(c) On vient de démontrer que :

$$E[\mathbf{1}_{(Y=1)} X^3] \neq E[\mathbf{1}_{(Y=1)}] E[X^3].$$

Les variables aléatoires X et Y ne sont donc pas indépendantes (bien que X² et Y² le soient).

3. Il s'agit de démontrer que, dans ce cas, les variables aléatoires X et Y sont indépendantes.

(a) Notant $\|f\|_\infty = \sup |f|$, on a l'inégalité :

$$|f(X) g(Y)| \leq \|f\|_\infty \|g\|_\infty;$$

la variable aléatoire $f(X) g(Y)$ est **bornée** et admet donc une moyenne.

– Par le théorème de transfert, on a

$$\begin{aligned} E[f(X) g(Y)] &= E[f((-1)^V U) g(V)] \\ &= \sum_{(u,v) \in \mathbb{Z}^2} f((-1)^v u) g(v) P(U = u, V = v), \end{aligned}$$

et, les variables aléatoires U et V étant indépendantes :

$$E[f(X) g(Y)] = \sum_{(u,v) \in \mathbb{Z}^2} f((-1)^v u) g(v) P(U = u) P(V = v).$$

Mais, les fonctions f et g étant bornées, on a

$$\sum_{(u,v) \in \mathbb{Z}^2} |f((-1)^v u) g(v)| P(U = u) P(V = v) < +\infty,$$

et, par la propriété de Fubini :

$$E[f(X) g(Y)] = \sum_{v \in \mathbb{Z}} g(v) \left[\sum_{u \in \mathbb{Z}} f((-1)^v u) P(U = u) \right] P(V = v). \quad (5.33)$$

Si f est **paire**, il vient, par le théorème de transfert,

$$E[f(X) g(Y)] = \sum_{v \in \mathbb{Z}} g(v) [E f(U)] P(V = v),$$

soit, encore par le théorème de transfert :

$$\boxed{E[f(X) g(Y)] = [E f(U)] [E g(V)].}$$

– Si f est **impaire**, il résulte de l'égalité (5.33) que :

$$\begin{aligned} E[f(X) g(Y)] &= \sum_{v \in 2\mathbb{Z}} g(v) [E f(U)] P(V = v) - \\ &\quad \sum_{v \in 2\mathbb{Z}+1} g(v) [E f(U)] P(V = v). \end{aligned}$$

Mais la famille $\{f(u) P(U = u)\}_{u \in \mathbb{Z}}$ étant sommable, il résulte du théorème de transfert que l'on a :

$$\begin{aligned} E f(U) &= \sum_{u \in \mathbb{N}} f(u) P(U = u) + \sum_{u \in -\mathbb{N}^*} f(u) P(U = u) \\ &= \sum_{u \in \mathbb{N}} f(u) P(U = u) + \sum_{u \in \mathbb{N}^*} f(-u) P(U = -u), \end{aligned}$$

soit, en tenant compte de la symétrie de la loi P_U et du fait que f est impaire :

$$E f(U) = \sum_{u \in \mathbb{N}} f(u) P(U = u) - \sum_{u \in \mathbb{N}^*} f(u) P(U = u) = f(0) P(U = 0).$$

Puisque $P(U = 0) = 0$, il vient :

$$\boxed{E[f(X)g(Y)] = 0.}$$

(b) Si la fonction f est **quelconque**, on la décompose en sa partie paire f_P et sa partie impaire f_I . Il résulte de la question précédente que :

$$E[f(X)g(Y)] = [E f_P(U)] [E g(V)].$$

En prenant en particulier $g = 1$, il vient

$$E[f(X)] = E[f_P(U)],$$

et donc :

$$\boxed{E[f(X)g(Y)] = [E f(X)] [E g(Y)].}$$

(c) En prenant $f = \mathbf{1}_A$ et $g = \mathbf{1}_B$ où A et B sont des parties quelconques de \mathbb{Z} , on obtient l'indépendance de X et Y .

Exercice 5.6. Inégalité de Tchebichev et polynômes de Bernstein.

Soit f une fonction réelle continue sur l'intervalle fermé $[0, 1]$. Pour $n \in \mathbb{N}^*$, on note B_n le polynôme de Bernstein défini par

$$B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k},$$

où on fait la convention d'écriture : $0^0 = 1$.

Soit, sur l'espace probabilisé (Ω, \mathcal{A}, P) , pour tout $x \in]0, 1[$, une suite de variables aléatoires indépendantes, de même **loi de Bernoulli** de paramètre x . On note

$$S_n = \sum_{k=1}^n X_k.$$

1. Déterminer la moyenne $E\left[f\left(\frac{S_n}{n}\right)\right]$.

2. Soit, pour tout $\varepsilon > 0$, le réel $\delta(\varepsilon)$ défini par :

$$\delta(\varepsilon) = \sup \{|f(x) - f(y)| \mid x, y \in [0, 1] \text{ et } |x - y| \leq \varepsilon\}.$$

(a) Démontrer que $\delta(\varepsilon)$ tend vers 0 avec ε .

(b) Démontrer que :

$$\sup_{x \in [0,1]} |B_n(x) - f(x)| \leq \delta(\varepsilon) + \frac{2 \|f\|_\infty}{n\varepsilon^2}.$$

En déduire que la suite des polynômes B_n converge vers f uniformément sur $[0, 1]$.

Solution.

1. On sait que la loi de S_n est la loi binomiale $\mathcal{B}(n, x)$. Il résulte du théorème de transfert que l'on a :

$$\mathbb{E} \left[f \left(\frac{S_n}{n} \right) \right] = \sum_{k=0}^n \left[f \left(\frac{k}{n} \right) \binom{n}{k} x^k (1-x)^{n-k} \right].$$

2. Il s'agit d'établir une démonstration constructive du **théorème d'approximation de Weierstrass** dans le cas où les fonctions sont définies sur l'intervalle $[0, 1]$; cette méthode d'approximation conduit aux **polynômes de Bezier** et à la procédure Bezier utilisée pour la modélisation linéaire et la génération de courbes de « forme arbitraire » et passant par des points donnés.

(a) La fonction f est continue sur le compact $[0, 1]$; il en résulte qu'elle est uniformément continue, ce qui peut s'exprimer par le fait que $\delta(\varepsilon)$ tend vers 0 avec ε .

(b) Pour tout $x \in]0, 1[$, on a

$$|B_n(x) - f(x)| = \left| \mathbb{E} \left[f \left(\frac{S_n}{n} \right) \right] - f(x) \right|,$$

et donc

$$|B_n(x) - f(x)| \leq \mathbb{E} \left[\mathbf{1}_{\left(\left| \frac{S_n}{n} - x \right| \leq \varepsilon \right)} \left| f \left(\frac{S_n}{n} \right) - f(x) \right| \right] + \mathbb{E} \left[\mathbf{1}_{\left(\left| \frac{S_n}{n} - x \right| > \varepsilon \right)} \left| f \left(\frac{S_n}{n} \right) - f(x) \right| \right],$$

soit

$$|B_n(x) - f(x)| \leq \delta(\varepsilon) + 2 \|f\|_\infty \mathbb{P} \left(\left| \frac{S_n}{n} - x \right| > \varepsilon \right).$$

Or, d'après l'inégalité de Tchebichev, on a :

$$\mathbb{P} \left(\left| \frac{S_n}{n} - \mathbb{E} \left(\frac{S_n}{n} \right) \right| > \varepsilon \right) \leq \frac{1}{\varepsilon^2} \sigma_{\frac{S_n}{n}}^2.$$

D'une part, on a

$$\mathbb{E} \left(\frac{S_n}{n} \right) = x,$$

et d'autre part, puisque les variables aléatoires X_k sont indépendantes et de même loi, (donc de même variance) :

$$\sigma_{\frac{S_n}{n}}^2 = \frac{1}{n^2} \sigma_{S_n}^2 = \frac{1}{n} \sigma_{X_1}^2.$$

Puisque $\sigma_{X_1}^2 = x(1-x) \leq 1$, il en résulte que, pour tout $x \in]0, 1[$, on a

$$|B_n(x) - f(x)| \leq \delta(\varepsilon) + \frac{2\|f\|_\infty}{n\varepsilon^2}.$$

La fonction $B_n - f$ étant continue sur $[0, 1]$, on a alors :

$$\sup_{x \in [0,1]} |B_n(x) - f(x)| \leq \delta(\varepsilon) + \frac{2\|f\|_\infty}{n\varepsilon^2}.$$

Il s'ensuit que, pour tout $\varepsilon > 0$, on a :

$$0 \leq \limsup_n \sup_{x \in [0,1]} |B_n(x) - f(x)| \leq \delta(\varepsilon),$$

ce qui, compte tenu de la question (a), démontre que

$$\limsup_n \sup_{x \in [0,1]} |B_n(x) - f(x)| = 0.$$

Il en résulte que :

$$\boxed{\lim_{n \rightarrow +\infty} \sup_{x \in [0,1]} |B_n(x) - f(x)| = 0,}$$

ce qui démontre que **la suite des polynômes de Bernstein converge donc uniformément sur $[0, 1]$ vers la fonction f .**

Exercice 5.7. Moyenne, fonction génératrice, inégalité de Tchebichev.

Soit, sur l'espace probabilisé (Ω, \mathcal{A}, P) , une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes, de même loi géométrique sur \mathbb{N}^* de paramètre $p \in]0, 1[$. On note pour $n \in \mathbb{N}^*$:

$$S_n = \sum_{k=1}^n X_k.$$

1. Calculer la moyenne $E\left(\frac{S_n}{n}\right)$ et la variance $\sigma_{\frac{S_n}{n}}^2$ de la variable aléatoire $\frac{S_n}{n}$.

2. Calculer la fonction génératrice G_{S_n} de la variable aléatoire S_n en tout $t \in [-1, 1]$.

3. Justifier le fait que G_{S_n} est développable en série entière dans $]-1, 1[$ et en déduire, pour tout $k \in \mathbb{N}^*$, la probabilité $P(S_n = k)$.

Soit maintenant f une fonction réelle, uniformément continue et bornée sur la demi-droite $[1, +\infty[$. Pour $n \in \mathbb{N}^*$, on note B_n la fonction définie sur $]0, 1[$ par : pour tout $x \in]0, 1[$,

$$B_n(x) = x^n \sum_{k=0}^{+\infty} f\left(1 + \frac{k}{n}\right) \binom{k+n-1}{n-1} (1-x)^k.$$

4. Calculer la moyenne $E \left[f \left(\frac{S_n}{n} \right) \right]$ en fonction de $B_n(p)$, dont on justifiera a posteriori l'existence.

5. Soit, pour tout $\varepsilon > 0$, le réel $\delta(\varepsilon)$ défini par :

$$\delta(\varepsilon) = \sup \{ |f(x) - f(y)| \mid |x| \geq 1 \quad |y| \geq 1 \quad |x - y| \leq \varepsilon \}$$

(a) Démontrer que :

$$\left| E \left[f \left(\frac{S_n}{n} \right) \right] - f \left(\frac{1}{p} \right) \right| \leq \delta(\varepsilon) + 2 \|f\|_\infty P \left(\left| \frac{S_n}{n} - \frac{1}{p} \right| > \varepsilon \right).$$

(b) En déduire que, pour tout $a > 1$, on a :

$$\lim_{n \rightarrow +\infty} \sup_{x \in [1, a]} \left| B_n \left(\frac{1}{x} \right) - f(x) \right| = 0.$$

Solution.

1. Par linéarité de l'espérance, on a :

$$E \left(\frac{S_n}{n} \right) = \frac{1}{n} \sum_{j=1}^n EX_j,$$

soit, puisque les variables aléatoires X_j sont de même loi géométrique sur \mathbb{N}^* de paramètre p (donc de même moyenne) :

$$E \left(\frac{S_n}{n} \right) = \frac{1}{p}.$$

De plus, puisque ces variables aléatoires sont indépendantes et de même loi (donc de même variance), on a

$$\sigma_{\frac{S_n}{n}}^2 = \frac{1}{n^2} \sigma_{S_n}^2 = \frac{1}{n} \sigma_{X_1}^2,$$

soit :

$$\sigma_{\frac{S_n}{n}}^2 = \frac{1-p}{np^2}.$$

2. On rappelle que la fonction génératrice de X_1 est donnée par : pour tout $t \in [-1, 1]$,

$$G_{X_1}(t) = \frac{pt}{1-t(1-p)}.$$

Les variables aléatoires X_j sont indépendantes et de même loi (donc de même fonction génératrice); il en résulte que, pour tout $t \in [-1, 1]$:

$$G_{S_n}(t) = \prod_{j=1}^n G_{X_j}(t) = \left[\frac{pt}{1-t(1-p)} \right]^n. \quad (5.34)$$

Il résulte par ailleurs du théorème de transfert que, pour tout $t \in [-1, 1]$:

$$G_{S_n}(t) = \sum_{k=1}^{+\infty} t^k P(S_n = k).$$

Puisque $\sum_{k=1}^{+\infty} P(S_n = k) = 1$, la fonction G_{S_n} est développable en série entière dans $] -1, 1[$, ce que l'on peut aussi voir d'après l'égalité (5.34); cette même égalité permet d'écrire, pour tout $t \in] -1, 1[$,

$$G_{S_n}(t) = (pt)^n \left[1 + \sum_{k=1}^{+\infty} \frac{(-n)(-n-1) \cdots (-n-k+1)}{k!} [t(1-p)]^k \right],$$

soit, après réduction :

$$G_{S_n}(t) = p^n \sum_{k=0}^{+\infty} \binom{n+k-1}{n-1} (1-p)^k t^{k+n},$$

ou encore, après le changement d'indices $l = k + n$:

$$G_{S_n}(t) = \sum_{l=n}^{+\infty} \binom{l-1}{n-1} p^n (1-p)^{l-n} t^l.$$

3. L'unicité du développement en série entière donne donc :

$$P(S_n = k) = \begin{cases} \binom{k-1}{n-1} p^n (1-p)^{k-n} & \text{si } k \geq n \\ 0 & \text{sinon.} \end{cases}$$

La loi de S_n est donc la **loi binomiale négative** $\mathcal{B}^-(n, p)$.

4. La variable aléatoire $f\left(\frac{S_n}{n}\right)$ est bornée et admet donc une moyenne; le théorème de transfert donne

$$E f\left(\frac{S_n}{n}\right) = \sum_{k=1}^{+\infty} f\left(\frac{k}{n}\right) P(S_n = k) = \sum_{k=n}^{+\infty} f\left(\frac{k}{n}\right) \binom{k-1}{n-1} p^n (1-p)^{k-n},$$

soit, après changement d'indices $l = k - n$:

$$E f\left(\frac{S_n}{n}\right) = p^n \sum_{l=0}^{+\infty} f\left(1 + \frac{l}{n}\right) \binom{l+n-1}{n-1} (1-p)^l.$$

On obtient ainsi l'existence de $B_n(p)$ pour tout $p \in]0, 1[$; de plus $B_n(1) = 0$ et :

$$E f\left(\frac{S_n}{n}\right) = B_n(p).$$

5. (a) On a :

$$E \left| f\left(\frac{S_n}{n}\right) - f\left(\frac{1}{p}\right) \right| \leq E \left[\mathbf{1}_{\left(\left|\frac{S_n}{n} - \frac{1}{p}\right| \leq \varepsilon\right)} \left| f\left(\frac{S_n}{n}\right) - f\left(\frac{1}{p}\right) \right| \right] + E \left[\mathbf{1}_{\left(\left|\frac{S_n}{n} - \frac{1}{p}\right| > \varepsilon\right)} \left| f\left(\frac{S_n}{n}\right) - f\left(\frac{1}{p}\right) \right| \right];$$

en majorant chacun des termes (observer que $\frac{S_n}{n} \geq 1$ et $\frac{1}{p} > 1$), on obtient l'inégalité :

$$\left| \mathbb{E} \left[f \left(\frac{S_n}{n} \right) \right] - f \left(\frac{1}{p} \right) \right| \leq \delta(\varepsilon) + 2 \|f\|_\infty \mathbb{P} \left(\left| \frac{S_n}{n} - \frac{1}{p} \right| > \varepsilon \right). \quad (5.35)$$

(b) D'après l'inégalité de Tchebichev et la première question, on a :

$$\mathbb{P} \left(\left| \frac{S_n}{n} - \frac{1}{p} \right| > \varepsilon \right) \leq \frac{\sigma_{\frac{S_n}{n}}^2}{\varepsilon^2} = \frac{1-p}{np^2\varepsilon^2}.$$

Alors, d'après l'égalité (5.35) et la troisième question, on a, pour tout $a > 1$:

$$\sup_{x \in [1, a]} \left| B_n \left(\frac{1}{x} \right) - f(x) \right| = \sup_{p \in [1/a, 1]} \left| B_n(p) - f \left(\frac{1}{p} \right) \right| \leq \delta(\varepsilon) + 2 \|f\|_\infty \frac{a^2}{n\varepsilon^2}.$$

L'uniforme continuité de f se traduit par : $\lim_{\varepsilon \rightarrow 0} \delta(\varepsilon) = 0$. On pourrait conclure comme à l'exercice 5.6. on donne ici un autre argument pour conclure. Pour

tout $\eta > 0$, on choisit $\varepsilon > 0$ tel que $\delta(\varepsilon) \leq \frac{\eta}{2}$, puis N tel que $2 \|f\|_\infty \frac{a^2}{N\varepsilon^2} \leq \frac{\eta}{2}$.

Il vient alors, pour tout $n \geq N$:

$$\sup_{x \in [1, a]} \left| B_n \left(\frac{1}{x} \right) - f(x) \right| \leq \eta,$$

ce qui démontre que :

$$\lim_{n \rightarrow +\infty} \sup_{x \in [1, a]} \left| B_n \left(\frac{1}{x} \right) - f(x) \right| = 0.$$

Contrairement au résultat sur les polynômes de Bernstein de l'exercice précédent, ce résultat d'approximation n'a pas a priori d'application pratique.

Exercice 5.8. Fonction génératrice et moments. Le nombre N de clients entrant pendant une journée dans un grand magasin est une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$. Les probabilités respectives pour chaque client d'acheter zéro, un ou deux articles de marque A sont $\frac{1}{6}$, $\frac{1}{2}$ et $\frac{1}{3}$. Le nombre d'articles en question achetés pendant une journée est une variable aléatoire S . On étudie la loi de S grâce à sa fonction génératrice.

On modélise le problème de façon plus précise de la manière suivante : soient, sur un espace probabilisé (Ω, \mathcal{A}, P) , une variable aléatoire N de loi de Poisson de paramètre $\lambda > 0$ et une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ indépendantes, de même loi (X_n représente le nombre d'articles achetés par le n -ième client) définie par :

$$P(X_1 = 0) = \frac{1}{6}, \quad P(X_1 = 1) = \frac{1}{2}, \quad P(X_1 = 2) = \frac{1}{3}.$$

On suppose de plus que N et les X_n , $n \in \mathbb{N}^*$, constituent une famille de variables aléatoires indépendantes. On définit enfin la variable aléatoire S par

$$S = \mathbf{1}_{(N \geq 1)} \sum_{j=1}^N X_j,$$

où on fait la convention d'écriture : $\sum_{j=1}^0 = 0$.

1. Calculer la fonction génératrice G_S de la variable aléatoire S , en tout $t \in [-1, 1]$.

2. En déduire la probabilité $P(S = 3)$ et la calculer numériquement dans le cas $\lambda = 6$.

3. Justifier l'existence de la moyenne ES et de la variance σ_S^2 de la variable aléatoire S et donner leur valeur. Les calculer numériquement dans le cas $\lambda = 6$.

Solution.

1. Puisque S prend ses valeurs dans \mathbb{N} , le théorème de transfert permet d'écrire, pour tout $t \in [-1, 1]$,

$$G_S(t) = \sum_{k=0}^{+\infty} t^k P(S = k),$$

ce qui s'écrit, en faisant apparaître le système complet de constituants formé des ensembles $(N = n)$:

$$\begin{aligned} G_S(t) &= \sum_{k=0}^{+\infty} t^k P \left[(S = k) \cap \left(\bigcup_{n \in \mathbb{N}} (N = n) \right) \right] \\ &= \sum_{k \in \mathbb{N}} t^k \left[\sum_{n \in \mathbb{N}} P(S = k, N = n) \right]. \end{aligned}$$

Mais

$$\sum_{k \in \mathbb{N}} |t|^k \left[\sum_{n \in \mathbb{N}} P(S = k, N = n) \right] \leq \sum_{k \in \mathbb{N}} \left[\sum_{n \in \mathbb{N}} P(S = k, N = n) \right] = 1.$$

La famille $\{t^k P(S = k, N = n)\}_{(k,n) \in \mathbb{N} \times \mathbb{N}}$ est donc sommable, et d'après la propriété de Fubini, on a :

$$G_S(t) = \sum_{n \in \mathbb{N}} \left[\sum_{k \in \mathbb{N}} t^k P(S = k, N = n) \right].$$

Mais on a l'inclusion $(N = 0) \subset (S = 0)$, et, par conséquent :

$$G_S(t) = P(N = 0) + \sum_{n \in \mathbb{N}^*} \left[\sum_{k \in \mathbb{N}} t^k P(S = k, N = n) \right].$$

Si, pour $n \geq 1$, on note $S_n = \sum_{j=1}^n X_j$, on a $(S = k) \cap (N = n) = (S_n = k) \cap (N = n)$ et donc :

$$G_S(t) = P(N = 0) + \sum_{n \in \mathbb{N}^*} \left[\sum_{k \in \mathbb{N}} t^k P(S_n = k, N = n) \right].$$

En tenant compte de l'indépendance de S_n et N , il vient :

$$\begin{aligned} G_S(t) &= P(N = 0) + \sum_{n \in \mathbb{N}^*} \left[\sum_{k \in \mathbb{N}} t^k P(S_n = k) \right] P(N = n) \\ &= P(N = 0) + \sum_{n \in \mathbb{N}^*} G_{S_n}(t) P(N = n). \end{aligned}$$

Mais les variables aléatoires X_n sont indépendantes et de même loi, et donc de même fonction génératrice ; il résulte de la proposition 5.25 que :

$$G_S(t) = P(N = 0) + \sum_{n=1}^{+\infty} [G_{X_1}(t)]^n P(N = n),$$

soit :

$$\boxed{G_S(t) = G_N[G_{X_1}(t)]}.$$

On rappelle que :

$$G_N(t) = \exp(\lambda(t-1));$$

par ailleurs, on a

$$G_{X_1}(t) = E(t^{X_1}) = \frac{1}{6} + \frac{1}{2}t + \frac{1}{3}t^2,$$

et donc, pour tout $t \in [-1, 1]$:

$$\boxed{G_S(t) = \exp \left[\lambda \left(-\frac{5}{6} + \frac{1}{2}t + \frac{1}{3}t^2 \right) \right]}.$$

2. De l'unicité du développement en série entière dans $]-1, 1[$ de la fonction G_S résulte que l'on a :

$$P(S = 3) = \frac{G_S'''(0)}{3!}.$$

Si on note

$$f(t) = \lambda \left(-\frac{5}{6} + \frac{1}{2}t + \frac{1}{3}t^2 \right),$$

un calcul simple conduit à

$$G_S'''(t) = \left[f'''(t) + 3f'(t)f''(t) + (f'(t))^3 \right] \exp[f(t)]$$

et

$$f'(t) = \lambda \left(\frac{1}{2} + \frac{2}{3}t \right), \quad f''(t) = \frac{2}{3}\lambda, \quad f'''(t) = 0,$$

ce qui conduit à :

$$\boxed{P(S = 3) = \frac{\lambda^2}{6} \left[1 + \frac{\lambda}{8} \right] \exp \left(-\frac{5\lambda}{6} \right)}.$$

Dans le cas $\lambda = 6$, on obtient :

$$P(S = 3) = \frac{21}{2} \exp(-5) \simeq 0,07.$$

3. On a

$$0 \leq S \leq 2N$$

et la variable aléatoire N admet des moments de tout ordre donc S aussi. La formule de Wald est visiblement encore valable dans ce contexte puisque la fonction génératrice de S est obtenue par composition de celles de N et X_1 et donc :

$$ES = G'_S(1) = (EN)(EX_1).$$

On a

$$EX_1 = \frac{1}{2} + 2 \times \frac{1}{3} = \frac{7}{6},$$

soit :

$$ES = \frac{7\lambda}{6}.$$

Dans le cas $\lambda = 6$, on obtient : $ES = 7$.

De même, on a

$$\sigma_S^2 = (\sigma_{X_1}^2)(EN) + (EX_1)^2 \sigma_N^2.$$

On a

$$EX_1^2 = \frac{1}{2} + 4 \times \frac{1}{3} = \frac{11}{6} \quad \text{et} \quad \sigma_{X_1}^2 = \frac{11}{6} - \left(\frac{7}{6}\right)^2 = \frac{17}{36}$$

et donc, puisque $\sigma_N^2 = \lambda$,

$$\sigma_S^2 = \frac{11}{6} \lambda.$$

Dans le cas $\lambda = 6$, on obtient : $\sigma_S^2 = 11$.

Exercice 5.9. Loi trinomiale, loi binomiale, fonction génératrice et indépendance ; caractérisation de la loi de Poisson. Toutes les variables aléatoires introduites sont définies sur le même espace probabilisé (Ω, \mathcal{A}, P) . Soient p, q, r trois réels strictement positifs tels que $p + q + r = 1$. On considère, pour tout entier $n \geq 1$, la variable aléatoire $Y_n = (U_n, V_n)$ à valeurs dans \mathbb{N}^2 , de loi **trinomiale** déterminée par : pour tout $(k, l) \in \mathbb{N}^2$ tel que $0 \leq k + l \leq n$,

$$P(Y_n = (k, l)) = \frac{n!}{k! l! [n - (k + l)]!} p^k q^l r^{n - (k + l)}.$$

On pose $Y_0 = (0, 0)$ (avec la convention $0! = 1$, on remarquera que Y_0 satisfait encore à l'égalité ci-dessus).

1. Démontrer que les variables aléatoires à valeurs entières U_n et V_n sont de loi **binomiale**. Déterminer leur paramètres.
2. Les variables aléatoires U_n et V_n sont-elles indépendantes ?

3. Démontrer la formule :

$$\forall x, y \in \mathbb{R} \quad \sum_{k=1}^n k \binom{n}{k} x^{k-1} y^{n-k} = n(x+y)^{n-1}, \quad (5.36)$$

et en déduire la moyenne $E(U_n V_n)$.

4. Calculer la covariance des variables aléatoires U_n et V_n et la variance de la variable aléatoire $U_n + V_n$.

Soit N une variable aléatoire à valeurs dans \mathbb{N} ; on note $a_n = P(N = n)$, pour tout $n \in \mathbb{N}$. On suppose que les variables aléatoires de la famille $\{Y_n, n \in \mathbb{N}; N\}$ sont **indépendantes**. On définit les variables aléatoires U, V par : pour tout $\omega \in \Omega$,

$$U(\omega) = U_{N(\omega)}(\omega) \quad \text{et} \quad V(\omega) = V_{N(\omega)}(\omega);$$

on pose $Y = (U, V)$.

5. On suppose que N suit la **loi de Poisson** de paramètre $\lambda > 0$. Calculer, pour tout $(k, l) \in \mathbb{N}^2$, la probabilité $P[Y = (k, l)]$ (on introduira le système complet de constituants $\{(N = n)\}_{n \in \mathbb{N}}$). En déduire que les variables aléatoires U et V sont indépendantes et identifier leur loi.

6. On suppose toujours que N suit la **loi de Poisson** de paramètre $\lambda > 0$ et on veut retrouver l'indépendance des variables aléatoires U et V en utilisant la méthode des fonctions génératrices. La **fonction génératrice** de la variable aléatoire Y est définie sur $[0, 1]^2$ par : pour tout $(a, b) \in [0, 1]^2$,

$$G_Y(a, b) = E(a^U b^V).$$

Calculer la fonction génératrice de Y et en déduire celles des variables aléatoires U et V ; retrouver ainsi leur loi. Déduire de ces résultats que ces variables aléatoires sont indépendantes.

7. On suppose maintenant que N a une loi quelconque, mais que les variables aléatoires U et V sont indépendantes. Démontrer, à l'aide de la fonction génératrice de Y que la variable aléatoire N suit une **loi de Poisson**.

Solution.

1. La famille des événements $(V_n = l)$, $l \in \mathbb{N}$, est un système complet de constituants; on a donc :

$$P(U_n = k) = \sum_{l \in \mathbb{N}} P(U_n = k, V_n = l),$$

soit

$$P(U_n = k) = \sum_{l=0}^{n-k} \frac{n!}{k! l! [n - (k+l)]!} p^k q^l r^{n-(k+l)},$$

ou encore

$$P(U_n = k) = \frac{p^k n!}{k! (n-k)!} \sum_{l=0}^{n-k} \binom{n-k}{l} q^l r^{(n-k)-l} = \binom{n}{k} p^k (q+r)^{n-k};$$

puisque $q + r = 1 - p$, il vient

$$P(U_n = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

c'est-à-dire que **la loi de U_n est la loi binomiale $\mathcal{B}(n, p)$** . De même, **la loi de V_n est la loi binomiale $\mathcal{B}(n, q)$** .

2. On a

$$P(U_n = 0, V_n = 0) = r^n \neq (1-p)^n (1-q)^n = P(U_n = 0) P(V_n = 0),$$

ce qui démontre que **les variables aléatoires U_n et V_n ne sont pas indépendantes**.

3. En dérivant par rapport à x les deux membres de l'égalité du binôme

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k},$$

on obtient de suite l'égalité (5.36). Le théorème de transfert permet d'écrire :

$$\begin{aligned} E(U_n V_n) &= \sum_{(k,l) \in \mathbb{N}^2} kl P(U_n = k, V_n = l) \\ &= \sum_{\substack{0 \leq k+l \leq n \\ k, l \geq 1}} kl \frac{n!}{k! l! [n - (k+l)]!} p^k q^l r^{n-(k+l)}, \end{aligned}$$

soit encore

$$E(U_n V_n) = \sum_{k=1}^n k p^k \frac{n!}{k! [n-k]!} \left[\sum_{l=1}^{n-k} l \binom{n-k}{l} q^l r^{(n-k)-l} \right].$$

En utilisant l'égalité (5.36), il vient :

$$\begin{aligned} E(U_n V_n) &= \sum_{k=1}^n k \binom{n}{k} p^k [(n-k) q (q+r)^{(n-k)-1}] \\ &= q \sum_{k=1}^n k (n-k) \binom{n}{k} p^k (1-p)^{(n-k)-1}. \end{aligned}$$

Un simple calcul montre que

$$k(n-k) \binom{n}{k} = nk \binom{n-1}{k},$$

ce qui donne

$$E(U_n V_n) = qn \sum_{k=1}^{n-1} k \binom{n-1}{k} p^k (1-p)^{(n-1)-k},$$

soit, en utilisant à nouveau l'égalité (5.36),

$$E(U_n V_n) = pqn(n-1).$$

4. Les moyennes de U_n et V_n sont respectivement np et nq ; la covariance de U_n et V_n est donc :

$$\text{cov}(U_n, V_n) = E(U_n V_n) - E(U_n) E(V_n) = pqn(n-1) - n^2 pq,$$

soit

$$\boxed{\text{cov}(U_n, V_n) = -npq.}$$

La variance de $U_n + V_n$ est alors donnée par :

$$\sigma_{U_n+V_n}^2 = \sigma_{U_n}^2 + \sigma_{V_n}^2 + 2 \text{cov}(U_n, V_n) = np(1-p) + nq(1-q) - 2npq,$$

soit, après réduction :

$$\boxed{\sigma_{U_n+V_n}^2 = nr(1-r).}$$

5. La famille des événements $(N = n)$, $n \in \mathbb{N}$, est un système complet de constituants; on a donc

$$P[Y = (k, l)] = \sum_{n \in \mathbb{N}} P[Y = (k, l), N = n] = \sum_{n \in \mathbb{N}} P[Y_n = (k, l), N = n],$$

soit, en tenant compte de l'indépendance des variables aléatoires Y_n et N :

$$P[Y = (k, l)] = \sum_{n \in \mathbb{N}} P[Y_n = (k, l)] P(N = n).$$

En remplaçant les probabilités par leurs valeurs, on obtient

$$P[Y = (k, l)] = \sum_{n=l+k}^{+\infty} \frac{n!}{k! l! [n - (k+l)]!} p^k q^l r^{n-(k+l)} \exp(-\lambda) \frac{\lambda^n}{n!},$$

soit

$$\begin{aligned} P[Y = (k, l)] &= \frac{(\lambda p)^k}{k!} \frac{(\lambda q)^l}{l!} \exp(-\lambda) \sum_{n=l+k}^{+\infty} \frac{(r\lambda)^{n-(k+l)}}{[n - (k+l)]!} & (5.37) \\ &= \frac{(\lambda p)^k}{k!} \frac{(\lambda q)^l}{l!} \exp[-\lambda(1-r)], \end{aligned}$$

c'est-à-dire

$$\boxed{P[Y = (k, l)] = \left[\exp(-\lambda p) \frac{(\lambda p)^k}{k!} \right] \left[\exp(-\lambda q) \frac{(\lambda q)^l}{l!} \right].} \quad (5.38)$$

On obtient alors les lois des marginales U et V ; on a pour tout $k \in \mathbb{N}$,

$$P(U = k) = \sum_{l \in \mathbb{N}} P[Y = (k, l)] = \exp(-\lambda p) \frac{(\lambda p)^k}{k!},$$

et pour tout $l \in \mathbb{N}$,

$$P(V = l) = \sum_{k \in \mathbb{N}} P[Y = (k, l)] = \exp(-\lambda q) \frac{(\lambda q)^l}{l!}.$$

Ainsi, les variables aléatoires U et V sont de loi de Poisson de paramètres respectifs λp et λq . L'égalité (5.38) démontre alors l'indépendance des variables aléatoires U et V .

6. Le théorème de transfert permet d'écrire la fonction génératrice de Y sous la forme : pour tout $(a, b) \in [0, 1]^2$,

$$G_Y(a, b) = E(a^U b^V) = \sum_{(k, l) \in \mathbb{N}^2} a^k b^l P[Y = (k, l)].$$

La famille des événements $(N = n)$, $n \in \mathbb{N}$, est un système complet de constituants ; on a donc

$$\begin{aligned} G_Y(a, b) &= \sum_{(k, l) \in \mathbb{N}^2} a^k b^l \left[\sum_{n \in \mathbb{N}} P(Y = (k, l), N = n) \right] \\ &= \sum_{(k, l) \in \mathbb{N}^2} a^k b^l \left[\sum_{n \in \mathbb{N}} P(Y_n = (k, l), N = n) \right] \end{aligned}$$

soit, en tenant compte de l'indépendance des variables aléatoires Y_n et N :

$$G_Y(a, b) = \sum_{(k, l) \in \mathbb{N}^2} a^k b^l \left[\sum_{n \in \mathbb{N}} P[Y_n = (k, l)] P(N = n) \right].$$

Les termes étant positifs, on a encore

$$G_Y(a, b) = \sum_{n \in \mathbb{N}} \left[\sum_{(k, l) \in \mathbb{N}^2} a^k b^l P[Y_n = (k, l)] \right] P(N = n),$$

c'est-à-dire

$$G_Y(a, b) = \sum_{n \in \mathbb{N}} G_{Y_n}(a, b) P(N = n), \quad (5.39)$$

où $G_{Y_n}(a, b)$ est la fonction génératrice de Y_n .

Le théorème de transfert donne :

$$\begin{aligned} G_{Y_n}(a, b) &= \sum_{(k, l) \in \mathbb{N}^2} a^k b^l P[Y_n = (k, l)] \\ &= \sum_{\substack{0 \leq k+l \leq n \\ k, l \geq 1}} a^k b^l \frac{n!}{k! l! [n - (k+l)]!} p^k q^l r^{n-(k+l)}, \end{aligned}$$

soit, par la formule du trinôme et en tenant compte de ce que $p + q + r = 1$:

$$G_{Y_n}(a, b) = (pa + qb + r)^n. \quad (5.40)$$

Si N suit la loi de Poisson $\mathcal{P}(\lambda)$, il vient :

$$G_Y(a, b) = \sum_{n=0}^{+\infty} \left[(pa + qb + r)^n \exp(-\lambda) \frac{\lambda^n}{n!} \right] = \exp[\lambda(pa + qb + r - 1)],$$

ce qui s'écrit encore :

$$G_Y(a, b) = \exp[\lambda p(a - 1)] \exp[\lambda q(b - 1)].$$

Il faut remarquer que, de manière générale, on obtient les fonctions génératrices des marginales U et V par les relations : pour tout $a \in [0, 1]$,

$$G_U(a) = E(a^U) = G_Y(a, 1),$$

et pour tout $b \in [0, 1]$,

$$G_V(b) = E(b^V) = G_Y(1, b),$$

soit ici pour tout $a \in [0, 1]$,

$$G_U(a) = \exp[\lambda p(a-1)],$$

et pour tout $b \in [0, 1]$,

$$G_V(b) = \exp[\lambda q(b-1)].$$

On vient de démontrer que, pour tout $(a, b) \in [0, 1]^2$,

$$\boxed{G_Y(a, b) = G_U(a) G_V(b).} \quad (5.41)$$

Le théorème de transfert permet d'écrire alors que, pour tout $(a, b) \in [0, 1]^2$, on a :

$$\sum_{(k,l) \in \mathbb{N}^2} a^k b^l P[Y = (k, l)] = \sum_{k \in \mathbb{N}} a^k P[U = k] \sum_{l \in \mathbb{N}} b^l P[V = l],$$

soit, puisque les termes sont positifs, par la propriété de Fubini :

$$\sum_{(k,l) \in \mathbb{N}^2} a^k b^l P[Y = (k, l)] = \sum_{(k,l) \in \mathbb{N}^2} a^k b^l P[U = k] P[V = l] < +\infty.$$

Il en résulte que, pour tout $(k, l) \in \mathbb{N}^2$, on a

$$P[Y = (k, l)] = P[U = k, V = l] = P[U = k] P[V = l],$$

ce qui démontre à nouveau l'**indépendance** des variables aléatoires U et V.

Remarque. La méthode suivie est générale ; l'égalité (5.41) est une condition nécessaire et suffisante d'indépendance des marginales U et V de la variable aléatoire Y.

7. Les égalités (5.39) et (5.40) donnent ici, pour tout $(a, b) \in [0, 1]^2$,

$$G_Y(a, b) = \sum_{n \in \mathbb{N}} G_{Y_n}(a, b) a_n = \sum_{n \in \mathbb{N}} (pa + qb + r)^n a_n = G_N(pa + qb + r),$$

où G_N est la fonction génératrice de N. L'hypothèse d'indépendance de U et V entraîne, d'après la méthode de calcul des fonctions génératrices des marginales vue à la question précédente, que

$$G_Y(a, b) = G_U(a, 1) G_V(1, b),$$

ce qui donne la relation : pour tout $(a, b) \in [0, 1]^2$,

$$G_N[p(a-1) + q(b-1) + 1] = G_N[p(a-1) + 1] G_N[q(b-1) + 1].$$

En posant, pour tout $a \in [0, 1]$, $g(a) = G_N(a + 1)$, on a donc, pour tout $(a, b) \in [0, 1]^2$:

$$g(a + b) = g(a)g(b);$$

la fonction g étant de plus continue est une exponentielle. Puisque g est supérieure ou égale à 1, il existe donc $\lambda > 0$ tel que, pour tout $a \in [0, 1]$, on ait $g(a) = \exp(\lambda a)$, ce qui donne :

$$G_N(a) = \exp[\lambda(a - 1)].$$

La fonction génératrice déterminant la loi d'une variable aléatoire, il s'ensuit que **la loi de N est la loi de Poisson $\mathcal{P}(\lambda)$** .

Remarque. En résumé, pour que la variable aléatoire N suive la loi de Poisson, il faut et il suffit que les variables aléatoires U et V soient indépendantes.

Chapitre 6

Variables aléatoires à densité

L'extension à des espaces Ω non dénombrables de la théorie développée dans les chapitres précédents se heurte à des difficultés mathématiques sérieuses.

Par exemple, si on prend $\Omega = [0, 1] \times [0, 1]$, il n'existe pas sur $(\Omega, \mathcal{P}(\Omega))$ de probabilité P qui, d'une part, attribue la **même probabilité** à deux parties de $[0, 1] \times [0, 1]$ translatées l'une de l'autre et qui, d'autre part, soit telle que $P([a, b] \times [c, d])$ soit égale à l'aire du rectangle $[a, b] \times [c, d]$. Ce sont pourtant des propriétés que l'on attend pour une probabilité uniforme.

Par contre, une probabilité P possédant ces propriétés peut être définie sur une tribu plus petite que $\mathcal{P}(\Omega)$, qu'on appelle la tribu des ensembles boréliens. La notion d'ensemble borélien appartient à la théorie de la mesure, qui est l'outil adéquat pour l'étude approfondie du calcul des probabilités, mais à laquelle on s'est abstenu de faire appel dans ce volume.

L'objet de ce chapitre est d'introduire la notion de variable aléatoire à densité de manière élémentaire, sans se soucier en particulier des problèmes d'existence; le point de vue est en quelque sorte descriptif; il répond ainsi à la demande des programmes des concours internes et externes du CAPES ainsi que du concours interne de l'Agrégation.

6.1. Probabilités sur \mathbb{R}^n

6.1.1. Densité d'une probabilité sur \mathbb{R}^n

Nous affirmons, sans démonstration, l'existence de la tribu borélienne.

Proposition 6.1. *Il existe une plus petite tribu (plus petite au sens de l'inclusion) sur \mathbb{R}^n contenant la famille des pavés ouverts $\prod_{i=1}^n]a_i, b_i[$. Cette tribu est appelée **tribu borélienne** de \mathbb{R}^n et notée $\mathcal{B}_{\mathbb{R}^n}$. Ses éléments sont appelés **sous-ensembles boréliens** de \mathbb{R}^n . Tout ouvert est borélien, tout fermé est borélien. Tout pavé $\prod_{i=1}^n]a_i, b_i]$ est borélien.*

Soit f une fonction de \mathbb{R}^n dans \mathbb{R}^+ intégrable au sens de Riemann sur \mathbb{R}^n et telle que $\int_{\mathbb{R}^n} f(x) dx = 1$. On démontre, et nous l'admettrons, qu'il existe une probabilité unique P sur l'espace probabilisable $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ telle que, pour tout pavé A de \mathbb{R}^n de la forme $\prod_{i=1}^n]a_i, b_i]$, on ait :

$$P(A) = \int_A f(x) dx. \quad (6.1)$$

La fonction f est appelée **densité** de la probabilité P ; on dit aussi que la probabilité P est de densité f .

Cas particulier. Si $n = 1$, et si f est une fonction définie sur \mathbb{R} , à valeurs réelles positives, continue sauf peut-être en un nombre fini de points, et telle que $\int_{-\infty}^{+\infty} f(x) dx = 1$, il existe une probabilité unique P sur l'espace probabilisable $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ telle que, pour tout réel a et tout réel b tels que $a < b$ on ait :

$$P(]a, b]) = \int_a^b f(x) dx.$$

Nous admettrons que l'égalité (6.1) est alors encore vérifiée si A est une réunion finie d'intervalles de \mathbb{R} , fermés, ouverts ou semi-ouverts, bornés ou non.

6.1.2. Exemples classiques de lois de probabilités sur \mathbb{R}

Voici des exemples classiques de lois de probabilité définies sur l'espace probabilisable $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ par une densité f :

(1) **Loi uniforme sur l'intervalle $[a, b]$.** Elle est notée $\mathcal{U}([a, b])$. La densité f est définie par : pour tout $x \in \mathbb{R}$,

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$$

(a et b sont deux réels tels que $a < b$).

Elle donne la même probabilité à deux sous-intervalles de même longueur de l'intervalle $[a, b]$.

(2) **Loi exponentielle de paramètre $p > 0$.** Elle est notée $\exp(p)$. La densité f est définie par : pour tout $x \in \mathbb{R}$,

$$f(x) = \mathbf{1}_{\mathbb{R}^+}(x) p \exp(-px).$$

Remarquer que cette probabilité donne une probabilité nulle à tout intervalle contenu dans \mathbb{R}^- . Elle intervient dans la modélisation des **temps d'attente**.

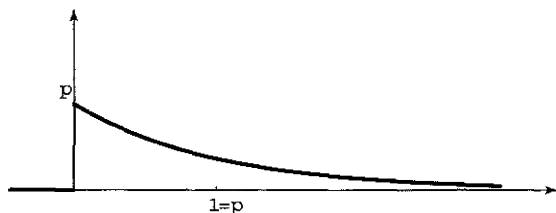


Figure 4. La loi exponentielle

(3) **Loi de Cauchy.** La densité f est définie par : pour tout $x \in \mathbb{R}$,

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

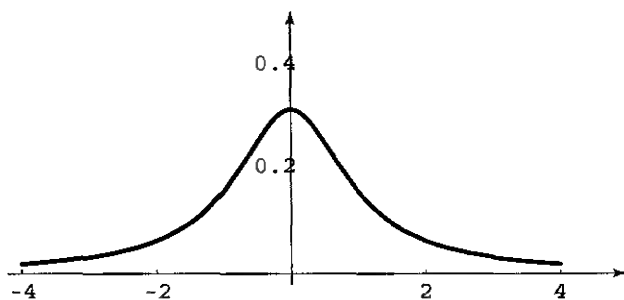


Figure 5. La loi de Cauchy

(4) **Loi de Laplace-Gauss, ou normale, de paramètres $m \in \mathbb{R}$ et $\sigma^2 > 0$.** Elle est notée $\mathcal{N}(m, \sigma^2)$. La densité¹ f est définie par : pour tout $x \in \mathbb{R}$,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Son graphe est la fameuse **courbe en cloche** de Gauss. Elle admet deux points d'inflexion d'abscisses $x_1 = m - \sigma$ et $x_2 = m + \sigma$ et on

1. La formule classique $\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$ permet, par changement de variable, de s'assurer que f est bien une densité.

a $f(x_1) = f(x_2) = \frac{1}{\sigma\sqrt{2\pi e}}$, ce qui montre que le pic est d'autant plus élevé et resserré que σ est petit. Cette loi apparaît très souvent dans les modélisations, en raison du **théorème limite central** que nous énoncerons par la suite.

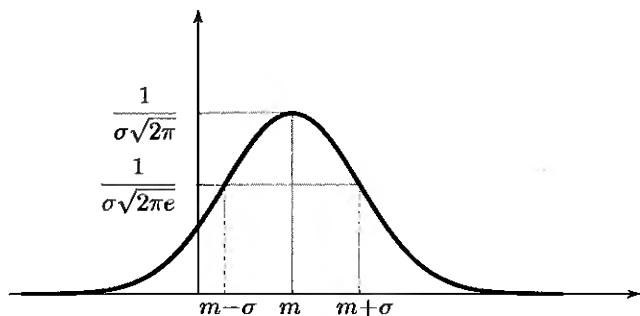


Figure 6. La loi normale

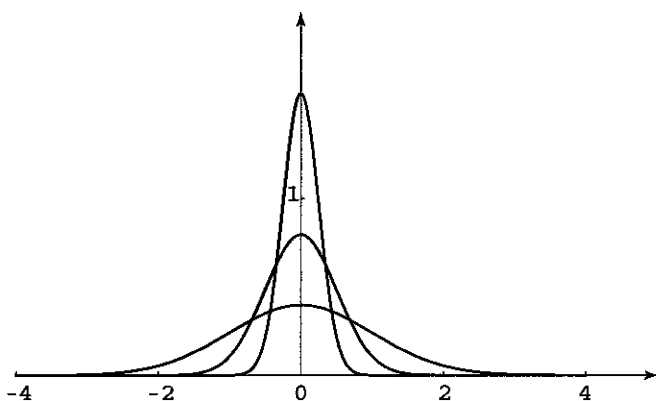


Figure 7. Différentes valeurs de σ

(5) **Loi du chi-deux à n degrés de liberté.** Elle est notée χ_n^2 . (La lettre grecque χ se transcrit *chi*, mais se prononce *ki*.) La densité f est définie par : pour tout $x \in \mathbb{R}$,

$$f(x) = \mathbf{1}_{\mathbb{R}^+}(x) \frac{1}{K_n} \exp\left(-\frac{x}{2}\right) x^{\frac{n}{2}-1},$$

où, pour tout entier $p \geq 1$, on note :

$$K_{2p} = 2^p (p-1)! \quad \text{et} \quad K_{2p+1} = (2p-1)(2p-3)\cdots 3 \cdot 1 \cdot \sqrt{2\pi}.$$

On peut montrer que c'est la loi d'une variable aléatoire de la forme $X_1^2 + \dots + X_n^2$, où X_1, \dots, X_n sont des variables indépendantes de même loi $\mathcal{N}(0, 1)$, ce qui explique l'expression « n degrés de liberté ». L'exercice 6.5 donne une démonstration² de ce fait dans les cas $n = 1$ et $n = 2$. La loi du chi-deux joue un grand rôle en Statistique.

6.1.3. Exemples classiques de lois de probabilités sur \mathbb{R}^2

(1) **Loi uniforme sur le rectangle** $[a, b] \times [c, d]$. Elle est notée $\mathcal{U}([a, b] \times [c, d])$. La densité f est définie par : pour tout $x \in \mathbb{R}^2$,

$$f(x) = \frac{1}{(b-a)(d-c)} \mathbf{1}_{[a,b] \times [c,d]}(x),$$

où a, b, c et d sont des réels tels que $a < b$ et $c < d$.

(2) **Loi uniforme sur le disque** $D(0, r)$. La densité f est définie par : pour tout $x \in \mathbb{R}^2$,

$$f(x) = \frac{1}{\pi r^2} \mathbf{1}_{D(0,r)}(x),$$

où r est un réel strictement positif.

Plus généralement, la loi uniforme sur une partie A du plan \mathbb{R}^2 dont on peut définir l'aire (et dont on suppose qu'elle est non nulle) est définie par la densité

$$f(x) = \frac{1}{\text{aire}(A)} \mathbf{1}_A(x).$$

(3) **Loi gaussienne, ou normale, centrée réduite sur \mathbb{R}^2** . La densité f est définie par : pour tout $x = (x_1, x_2) \in \mathbb{R}^2$,

$$f(x) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right).$$

6.2. Loi d'une variable aléatoire à valeurs dans \mathbb{R}^n

6.2.1. Densité d'une variable aléatoire, fonction de répartition

Dans ce qui suit, n est un entier supérieur ou égal à 1 et X est une variable aléatoire définie sur l'espace probabilisé (Ω, \mathcal{A}, P) et à valeurs

² Une fois lu ce chapitre, ce pourra être un exercice de retrouver directement la formule donnée pour la densité de la loi χ_n^2 dans les cas $n = 2, n = 3$. Pour $n = 2$, il faut savoir intégrer en coordonnées polaires, et pour $n = 3$, il faut utiliser les coordonnées sphériques.

dans $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$. On rappelle que sa loi P_X est une probabilité définie sur l'espace probabilisable $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ par : pour tout $A \in \mathcal{B}_{\mathbb{R}^n}$,

$$P_X(A) = P(X \in A).$$

Définition 6.2. S'il existe une fonction f_X définie sur \mathbb{R}^n , à valeurs ≥ 0 , Riemann-intégrable, telle que

$$\int_{\mathbb{R}^n} f_X(x) dx = 1,$$

et que, pour tout pavé A de \mathbb{R}^n on ait :

$$P_X(A) = \int_A f_X(x) dx,$$

cette fonction f_X est appelée **densité de la variable aléatoire X** (elle détermine entièrement la loi de X et elle n'est autre que la densité de la probabilité P_X).

Définition 6.3. Si $n = 1$, l'application F_X de \mathbb{R} dans $[0, 1]$ définie par : pour tout $x \in \mathbb{R}$

$$F_X(x) = P(X \leq x)$$

est appelée **fonction de répartition de la variable aléatoire X** .

Remarque. Si $n = 1$, et si la variable aléatoire X admet une densité f_X , il résulte des définitions que pour tout réel a et tout réel b tels que $a < b$, on a

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx. \quad (6.2)$$

Intuitivement, on peut écrire $P(x < X \leq x + dx) = f_X(x) dx$ où dx est considéré comme « infiniment petit ». Cette formulation très parlante, bien que non rigoureuse, est souvent employée par les physiciens et les ingénieurs. Elle revient, pour une fonction f_X régulière, à faire un développement de Taylor d'ordre 1 de F_X , dont on néglige le reste.

Voici les principales propriétés des fonctions de répartition.

Proposition 6.4. Soit X une variable aléatoire réelle de fonction de répartition F_X .

- (i) La fonction de répartition F_X détermine entièrement la loi de X .
- (ii) Pour tout réel a et tout réel b tels que $a < b$, on a

$$\boxed{F_X(b) - F_X(a) = P(a < X \leq b),} \quad (6.3)$$

et, pour tout $x \in \mathbb{R}$,

$$\boxed{\lim_{x \rightarrow +\infty} F_X(x) = 1 \quad \text{et} \quad \lim_{x \rightarrow -\infty} F_X(x) = 0.} \quad (6.4)$$

(iii) La fonction de répartition F_X est une fonction croissante, continue à droite, et admet une limite à gauche en tout point. De plus, pour tout $x \in \mathbb{R}$, on a :

$$\boxed{P(X = x) = F_X(x) - \lim_{y \nearrow x} F_X(y),} \quad (6.5)$$

c'est-à-dire que $P(X = x)$ est égal au saut de F_X en x .

(iv) Si la variable aléatoire X admet une densité f_X , la fonction de répartition F_X est continue en tout point et on a, pour tout $x \in \mathbb{R}$:

$$P(X = x) = 0.$$

On a de plus, pour tout $x \in \mathbb{R}$,

$$\boxed{P(X \leq x) \equiv F_X(x) = \int_{-\infty}^x f_X(u) du,} \quad (6.6)$$

et la fonction de répartition F_X est dérivable en tout point de continuité de la densité f_X .

Démonstration.

(i) Ce point ressort de la théorie de la mesure et il est donc admis.

(ii) Si $a < b$, on a, en raison de l'inclusion des ensembles $(X \leq a) \subset (X \leq b)$:

$$F_X(b) - F_X(a) = P(X \leq b) - P(X \leq a) = P(a < X \leq b).$$

De plus, pour toute suite $(x_n)_{n \in \mathbb{N}}$ de réels croissante et convergeant vers $+\infty$, on a :

$$\bigcup_{n \in \mathbb{N}} (X \leq x_n) = \Omega ;$$

la suite des ensembles $(X \leq x_n)$ étant croissante, on a donc :

$$\lim_{n \rightarrow +\infty} P(X \leq x_n) = P(\Omega) = 1.$$

De même, pour toute suite $(x_n)_{n \in \mathbb{N}}$ décroissante et convergeant vers $-\infty$, on a :

$$\bigcap_{n \in \mathbb{N}} (X \leq x_n) = \emptyset ;$$

la suite des ensembles $(X \leq x_n)$ étant décroissante, on a :

$$\lim_{n \rightarrow +\infty} P(X \leq x_n) = P(\emptyset) = 0.$$

Les relations (6.4) sont ainsi démontrées.

(iii) Il résulte de (6.3) que la fonction F_X est croissante. De plus, pour toute suite $(b_n)_{n \in \mathbb{N}}$ décroissante et convergente vers b , on a :

$$F_X(b_n) - F_X(b) = P(X \in]b, b_n]);$$

or, la suite des ensembles $(X \in]b, b_n])$ est décroissante et d'intersection vide. Il en résulte que :

$$\lim_{n \rightarrow +\infty} F_X(b_n) = F_X(b),$$

ce qui démontre la **continuité à droite**. La fonction F_X étant croissante et bornée admet une limite à gauche en tout point³. Enfin, puisque, pour toute suite $(x_n)_{n \in \mathbb{N}}$ strictement croissante et convergente vers x , on a

$$(X = x) = \bigcap_{n \in \mathbb{N}} (x_n < X \leq x),$$

on obtient :

$$P(X = x) = \lim_{n \rightarrow +\infty} P(x_n < X \leq x) = \lim_{n \rightarrow +\infty} [F_X(x) - F_X(x_n)],$$

ce qui démontre la relation (6.5).

(iv) Si X admet une densité f_X , il résulte de la relation (6.5) que, pour tout $x \in \mathbb{R}$, on a :

$$\begin{aligned} P(X = x) &= \lim_{n \rightarrow +\infty} \left[F_X(x) - F_X\left(x - \frac{1}{n}\right) \right] \\ &= \lim_{n \rightarrow +\infty} \int_{x - \frac{1}{n}}^x f_X(u) du = 0. \end{aligned}$$

De plus, la relation (6.6) s'obtient à partir de (6.3) et (6.4). La propriété de dérivabilité de la fonction de répartition en résulte.

Remarques. Une variable aléatoire X qui vérifie, pour tout $x \in \mathbb{R}$, la condition $P(X = x) = 0$, est dite de **loi diffuse**. Cette propriété n'entraîne pas que la variable aléatoire admette une densité.

Dans le contexte de la remarque suivant la définition 6.3, il résulte de cette proposition que la formule 6.2 conduit à l'égalité

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx.$$

3. L'ensemble de ses points de discontinuité est au plus dénombrable.

6.2.2. Marginales d'une variable aléatoire à valeurs dans \mathbb{R}^2

Définition 6.5. Si $X = (X_1, X_2)$ est une variable aléatoire à valeurs dans \mathbb{R}^2 , les variables aléatoires X_1 et X_2 s'appellent les **marginales** de la variable aléatoire X .

Remarque. Évidemment cette définition se généralise à \mathbb{R}^n et à toute projection de la variable aléatoire X sur un sous-espace engendré par des vecteurs de la base canonique.

Pour simplifier, nous ne décrivons la **loi d'une marginale** que dans le cas où la variable aléatoire est à valeurs dans \mathbb{R}^2 , la généralisation ne soulevant qu'un petit problème d'écriture.

Par la suite, on dira qu'une fonction définie sur \mathbb{R}^2 est **régulière** si on peut lui appliquer le théorème de Fubini⁴ sur \mathbb{R}^2 .

Proposition 6.6. Si la variable aléatoire $X = (X_1, X_2)$ à valeurs dans \mathbb{R}^2 admet une densité f_X régulière, les marginales X_1 et X_2 admettent des densités respectives f_{X_1} et f_{X_2} . Elles sont données par :

– pour tout $x_1 \in \mathbb{R}$,

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_X(x_1, x_2) dx_2,$$

– pour tout $x_2 \in \mathbb{R}$,

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} f_X(x_1, x_2) dx_1.$$

Démonstration. Puisque l'on a, pour tout $x_1 \in \mathbb{R}$,

$$(X_1 \leq x_1) = \bigcup_{n \in \mathbb{N}} [(X_1 \leq x_1) \cap (X_2 \leq n)]$$

et que la suite des ensembles entre crochets est croissante, on a

$$\begin{aligned} P(X_1 \leq x_1) &= \lim_{n \rightarrow +\infty} P[(X_1 \leq x_1) \cap (X_2 \leq n)] \\ &= \lim_{n \rightarrow +\infty} \int_{]-\infty, x_1] \times]-\infty, n]} f_X(u_1, u_2) du_1 du_2 \\ &= \int_{]-\infty, x_1] \times]-\infty, +\infty]} f_X(u_1, u_2) du_1 du_2, \end{aligned}$$

4. Des conditions suffisantes de **régularité** sont données en annexe (cf 6.8.2. Le théorème de Fubini). Cette terminologie n'est pas officielle et n'a de portée que dans le cadre d'application des résultats de ce chapitre.

et, par le théorème de Fubini⁵

$$P(X_1 \leq x_1) = \int_{-\infty}^{x_1} \left[\int_{-\infty}^{+\infty} f_X(u_1, u_2) du_2 \right] du_1.$$

Ceci démontre l'existence de la densité f_{X_1} et donne son expression.

Exemple 6.1. Soit une variable aléatoire X de loi normale centrée réduite sur \mathbb{R}^2 . Déterminons les lois des marginales X_1 et X_2 . On a, pour tout $x_1 \in \mathbb{R}$,

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) dx_2;$$

de plus on a

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_2^2}{2}\right) dx_2 = 1,$$

Il en résulte que, pour tout $x_1 \in \mathbb{R}$:

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right).$$

Les marginales X_1 et X_2 suivent donc la loi **normale de paramètres 0 et 1**, notée $\mathcal{N}(0, 1)$.

6.2.3. Loi d'une fonction de variable aléatoire

Il est utile de savoir étudier la **loi d'une fonction d'une variable aléatoire**. Toujours pour les mêmes raisons, dans ce chapitre, nous n'abordons ce problème que dans le cas où cette fonction est **monotone**.

Proposition 6.7. Soit X une variable aléatoire réelle admettant une densité f_X et soit g une fonction réelle définie sur \mathbb{R} **strictement monotone et dérivable**. La variable aléatoire réelle $Y = g(X)$ admet une densité f_Y donnée par :

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) |(g^{-1})'(y)| & \text{si } y \in g(\mathbb{R}) \\ 0 & \text{sinon.} \end{cases}$$

Démonstration. La fonction g est bijective de \mathbb{R} sur l'intervalle $g(\mathbb{R})$. Supposons-la *strictement décroissante*; sa fonction dérivée est alors strictement négative et ne s'annule en aucun point. Alors, pour tout réel $y \in g(\mathbb{R})$ (attention à l'inversion de l'inégalité) :

$$P(Y \leq y) = P(X \geq g^{-1}(y)) = P(X > g^{-1}(y)) = \int_{g^{-1}(y)}^{+\infty} f_X(x) dx,$$

5. Voir à ce sujet l'annexe 6.8. sur l'intégrale de Riemann dans \mathbb{R}^n .

la deuxième égalité étant vraie parce que X admet une densité; reste à effectuer le changement de variable défini par

$$v = g(x)$$

pour obtenir

$$\begin{aligned} P(Y \leq y) &= \int_y^{-\infty} f_X(g^{-1}(v)) (g^{-1})'(v) dv \\ &= \int_{-\infty}^y f_X(g^{-1}(v)) |(g^{-1})'(v)| dv. \end{aligned}$$

Si de plus $g(\mathbb{R}) =]\alpha, \beta[$, on a, pour tout $y < \alpha$,

$$P(Y \leq y) = P(g(X) \leq y) = 0,$$

et, pour tout $y > \beta$,

$$P(Y \leq y) = P(g(X) \leq y) = 1.$$

La proposition en résulte dans le cas d'une fonction strictement décroissante. Si g est *strictement croissante*, la démonstration est analogue et laissée au lecteur à titre d'exercice.

6.3. Moyenne et variance d'une variable aléatoire réelle

Définition 6.8. Soit X une variable aléatoire définie sur l'espace probabilisé (Ω, \mathcal{A}, P) , à valeurs réelles, possédant une densité f_X . Si la fonction $x \mapsto |x|f_X(x)$ est Riemann-intégrable sur \mathbb{R} , on dit que X admet une moyenne. La **moyenne** ou **espérance mathématique** de X est alors notée EX , ou $E(X)$, et définie par :

$$E(X) = \int_{\mathbb{R}} x f_X(x) dx.$$

On observera l'analogie entre cette définition et celle de l'espérance d'une variable aléatoire discrète.

Si maintenant Φ est une fonction définie sur \mathbb{R} , à valeurs réelles, telle que $Y = \Phi(X)$ soit une variable aléatoire, on peut espérer calculer l'espérance de $\Phi(X)$ au moyen d'une formule analogue à la formule discrète (5.3), donnée par le théorème de transfert, à savoir

$$E[\Phi(X)] = \int_{\mathbb{R}} \Phi(x) f_X(x) dx, \quad (6.7)$$

à condition bien sûr que la fonction $|\Phi|f_X$ soit Riemann-intégrable.

Malheureusement, dans ce contexte, une telle formule est difficile, non seulement à démontrer, mais même à énoncer rigoureusement. Son premier membre n'a de sens que si $Y = \Phi(X)$ possède une densité (ce qui est le cas si Φ est une fonction dérivable et strictement monotone⁶), ou est une variable aléatoire discrète (ce qui se produit par exemple si Φ est une constante, ou si $\Phi(x) = [x]$ (partie entière de x)). Mais il y a d'autres possibilités : si $\Phi(x) = 0$ pour $x < 0$, $\Phi(x) = 1 + x$ pour $x \geq 0$, la loi de $\Phi(X)$ peut comporter une partie discrète placée en 0 et une partie continue, et on peut construire des exemples beaucoup plus compliqués. Nous admettrons que, même si $\Phi(X)$ n'est pas discrète ou n'a pas de densité, on peut donner un sens à l'expression $E[\Phi(X)]$ (si la fonction $|\Phi|f_X$ est intégrable) et que la formule (6.7) permet de calculer sa valeur dans tous les cas.

Nous énonçons (sans démonstration) le théorème ci-dessous pour une variable aléatoire $X = (X_1, \dots, X_n)$ à valeurs dans \mathbb{R}^n .

Théorème 6.9. *Soit X une variable aléatoire définie sur l'espace probabilisé (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{R}^n , de densité f_X . Soit Φ une fonction définie sur \mathbb{R}^n , à valeurs réelles. Si la fonction $|\Phi|f_X$ est Riemann-intégrable sur \mathbb{R}^n , alors la variable aléatoire $\Phi(X)$ possède une moyenne, donnée par la formule :*

$$E[\Phi(X)] = \int_{\mathbb{R}^n} \Phi(x) f_X(x) dx.$$

Proposition 6.10. (i) *Soient X_1 et X_2 deux variables aléatoires réelles telles que la variable aléatoire (X_1, X_2) admette une densité régulière⁷. Si X_1 et X_2 possèdent une moyenne, il en est de même de la variable aléatoire $\lambda_1 X_1 + \lambda_2 X_2$, quels que soient λ_1 et λ_2 , et l'on a :*

$$E(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 E(X_1) + \lambda_2 E(X_2).$$

(ii) *Soit X une variable aléatoire réelle de densité f_X et admettant une moyenne. Pour tous réels a et b , la variable aléatoire $aX + b$ admet une moyenne et l'on a :*

$$E(aX + b) = aE(X) + b. \tag{6.8}$$

6. Plus généralement, $Y = \Phi(X)$ admet une densité si on peut décomposer \mathbb{R} en une réunion d'un nombre fini d'intervalles sur lesquels Φ est dérivable et strictement monotone. Voir une démonstration pour $\Phi(X) = X^2$ dans l'exercice 6.5.

7. Voir note de la proposition 6.6.

Démonstration.

(i) Prenons pour Φ la fonction définie par : pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$\Phi(x_1, x_2) = |\lambda_1 x_1 + \lambda_2 x_2|.$$

On a, par l'inégalité triangulaire, pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$0 \leq \Phi(x_1, x_2) f_{(X_1, X_2)}(x_1, x_2) \leq (|\lambda_1| |x_1| + |\lambda_2| |x_2|) f_{(X_1, X_2)}(x_1, x_2).$$

Montrons que le membre de droite est régulier. D'après la proposition 6.6, la fonction $x_2 \mapsto |x_1| f_{(X_1, X_2)}(x_1, x_2)$ est Riemann-intégrable, d'intégrale $|x_1| f_{X_1}(x_1)$; de plus, puisque X_1 admet une moyenne, la fonction $x_1 \mapsto |x_1| f_{X_1}(x_1)$ est encore Riemann-intégrable. En échangeant les rôles de x_1 et x_2 , on obtient la régularité annoncée. Le théorème de Fubini permet alors d'écrire :

$$\begin{aligned} \int_{\mathbb{R}^2} (|\lambda_1| |x_1| + |\lambda_2| |x_2|) f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 \\ = |\lambda_1| E|X_1| + |\lambda_2| E|X_2| < +\infty. \end{aligned}$$

Ceci démontre que la variable aléatoire $\lambda_1 X_1 + \lambda_2 X_2$ admet une moyenne. On peut alors écrire, par le théorème de Fubini :

$$\begin{aligned} E(\lambda_1 X_1 + \lambda_2 X_2) &= \int_{\mathbb{R}^2} (\lambda_1 x_1 + \lambda_2 x_2) f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 \\ &= \lambda_1 \int_{-\infty}^{+\infty} x_1 \left(\int_{-\infty}^{+\infty} f_{(X_1, X_2)}(x_1, x_2) dx_2 \right) dx_1 \\ &\quad + \lambda_2 \int_{-\infty}^{+\infty} x_2 \left(\int_{-\infty}^{+\infty} f_{(X_1, X_2)}(x_1, x_2) dx_1 \right) dx_2, \end{aligned}$$

ce qui assure le résultat.

(ii) Prenant pour Φ l'application définie par : pour tout $x \in \mathbb{R}$,

$$\Phi(x) = ax + b,$$

on a :

$$|\Phi(x)| f_X(x) \leq (|a| |x| + |b|) f_X(x),$$

ce qui démontre, puisque les applications f_X et $x \mapsto |x| f_X(x)$ sont Riemann-intégrables, que la fonction $|\Phi| f_X$ est Riemann-intégrable, et donc que la variable aléatoire $aX + b$ admet une moyenne; on a de plus :

$$E(ax + b) = \int_{-\infty}^{+\infty} (ax + b) f_X(x) dx,$$

ce qui donne le résultat, par linéarité de l'intégrale et par le fait que $\int_{-\infty}^{+\infty} f_X(x) dx = 1$.

Voici un exemple de calcul de la moyenne d'une variable aléatoire.

Exemple 6.2. Si la variable aléatoire X suit la loi normale $\mathcal{N}(m, \sigma^2)$, on a :

$$\boxed{EX = m.}$$

En effet, on a

$$EX = \int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx,$$

soit, par le changement de variable $y = \frac{x-m}{\sigma}$:

$$EX = \int_{-\infty}^{+\infty} (\sigma y + m) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy,$$

ou encore

$$EX = \sigma \int_{-\infty}^{+\infty} y \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy + m \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy.$$

La première intégrale est nulle (intégrande paire) et la deuxième vaut 1 (intégrale d'une densité). Il en résulte que $EX = m$.

6.3.1. Moments d'ordre 2. Variance

Proposition 6.11. Si X est une variable aléatoire réelle de densité f_X et si X^2 admet une moyenne, alors X admet elle-même une moyenne.

Démonstration. Il suffit de constater que, pour tout $x \in \mathbb{R}$,

$$|x| \leq x^2 + 1;$$

la fonction $x \mapsto |x| f_X(x)$ est alors Riemann-intégrable sur \mathbb{R} .

Définition 6.12. Si X est une variable aléatoire réelle de densité f_X telle que X^2 admet une moyenne, le réel positif $E[X^2]$ est appelé **moment d'ordre deux** de X , et le réel positif $E[(X - EX)^2]$ est appelé **variance** de X et noté σ_X^2 .

Proposition 6.13. Si X est une variable aléatoire réelle de densité f_X telle que X^2 admet une moyenne, la variance de X vérifie :

(i)

$$\boxed{\sigma_X^2 = \int_{-\infty}^{+\infty} (x - EX)^2 f_X(x) dx,} \quad (6.9)$$

(ii) et

$$\boxed{\sigma_X^2 = E(X^2) - (EX)^2.}$$

(iii) Pour tous réels a et b , on a :

$$\sigma_{aX+b}^2 = a^2 \sigma_X^2.$$

Démonstration.

(i) Il suffit d'appliquer le théorème 6.9 avec comme fonction Φ la fonction définie par : pour tout $x \in \mathbb{R}$, $\Phi(x) = (x - EX)^2$.

(ii) En développant le carré et en utilisant la linéarité de l'intégrale dans l'égalité (6.9), il vient successivement :

$$\begin{aligned} \sigma_X^2 &= \int_{-\infty}^{+\infty} [x^2 - 2xEX + (EX)^2] f_X(x) dx \\ &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx - 2EX \int_{-\infty}^{+\infty} x f_X(x) dx + (EX)^2 \int_{-\infty}^{+\infty} f_X(x) dx \\ &= E(X^2) - 2(EX)^2 + (EX)^2 \\ &= E(X^2) - (EX)^2. \end{aligned}$$

(iii) Il résulte de l'égalité

$$(aX + b) - E(aX + b) = a(X - EX)$$

que l'on a :

$$\sigma_{aX+b}^2 = E[a^2(X - EX)^2] = a^2 E[(X - EX)^2] = a^2 \sigma_X^2.$$

Voici un exemple de calcul de la variance d'une variable aléatoire.

Exemple 6.3. Variance d'une loi normale Si la variable aléatoire X a la loi normale $\mathcal{N}(m, \sigma^2)$, sa variance σ_X^2 est σ^2 . En effet, puisque sa moyenne est $EX = m$, on a

$$\sigma_X^2 = E[X - m]^2,$$

soit, d'après l'égalité (6.9) :

$$\sigma_X^2 = \int_{-\infty}^{+\infty} [x - m]^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right) dx.$$

Le changement de variable $y = \frac{x - m}{\sigma}$ donne alors :

$$\begin{aligned} \sigma_X^2 &= \int_{-\infty}^{+\infty} (\sigma y)^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \\ &= \sigma^2 \int_{-\infty}^{+\infty} y^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \\ &= \sigma^2, \end{aligned}$$

puisque, par une intégration par parties (justifiée par la convergence des deux intégrales), on a :

$$\int_{-\infty}^{+\infty} y^2 \exp\left(-\frac{y^2}{2}\right) dy = -y \exp\left(-\frac{y^2}{2}\right) \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \exp\left(-\frac{y^2}{2}\right) dy = \sqrt{2\pi}.$$

Définition 6.14. Si X est une variable aléatoire réelle de densité f_X admettant un moment d'ordre deux, la variable aléatoire $X - EX$ est appelée variable **centrée** (sa moyenne est nulle), la variable aléatoire $\frac{X - EX}{\sigma_X}$ est appelée variable **centrée réduite** associée à X (sa moyenne est nulle et son écart-type est égal à 1).

Nous renvoyons à l'exercice 6.7 pour le calcul des moyenne et variance de variables aléatoires de lois classiques.

Nous donnons l'**inégalité de Bienaymé-Tchebichev** pour une variable aléatoire à densité. Bien que grossière, elle donne une idée de la répartition de X autour de sa moyenne ; elle sert à démontrer des convergences en probabilité.

Proposition 6.15. (Inégalité de Bienaymé-Tchebichev.) Soit X une variable aléatoire réelle de densité f_X , admettant un moment d'ordre deux. Pour tout réel $\varepsilon > 0$, on a

$$\boxed{P(|X - EX| > \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2} .}$$

Démonstration. Pour tout $\varepsilon > 0$, on a, en notant $m = EX$,

$$\begin{aligned} \sigma_X^2 &= \int_{-\infty}^{+\infty} [x - m]^2 f_X(x) dx \\ &\geq \int_{-\infty}^{m-\varepsilon} [x - m]^2 f_X(x) dx + \int_{m+\varepsilon}^{+\infty} [x - m]^2 f_X(x) dx \\ &\geq \varepsilon^2 \left(\int_{-\infty}^{m-\varepsilon} f_X(x) dx + \int_{m+\varepsilon}^{+\infty} f_X(x) dx \right) \\ &= \varepsilon^2 P(|X - EX| > \varepsilon), \end{aligned}$$

ce qui donne l'inégalité.

L'inégalité de Bienaymé-Tchebichev s'utilise aussi sous la forme

$$P(|X - EX| > \rho\sigma_X) \leq \frac{1}{\rho^2}$$

($\rho > 0$).

6.3.2. Covariance et coefficient de corrélation

Proposition 6.16. *Si la variable aléatoire $X = (X_1, X_2)$ à valeurs dans \mathbb{R}^2 admet une densité f_X régulière et si X_1 et X_2 possèdent un moment d'ordre deux, la variable aléatoire produit $X_1 X_2$ admet une moyenne.*

Démonstration. D'après la proposition 6.6, les marginales admettent une densité. Une démarche analogue à celle utilisée pour démontrer l'assertion (i) de la proposition 6.10 assure que l'on peut appliquer le théorème de Fubini et écrire :

$$\begin{aligned} \int_{\mathbb{R}^2} x_1^2 f_X(x_1, x_2) dx_1 dx_2 &= \int_{-\infty}^{+\infty} x_1^2 \left[\int_{-\infty}^{+\infty} f_X(x_1, x_2) dx_2 \right] dx_1 \\ &= EX_1^2 < +\infty, \end{aligned}$$

et

$$\begin{aligned} \int_{\mathbb{R}^2} x_2^2 f_X(x_1, x_2) dx_1 dx_2 &= \int_{-\infty}^{+\infty} x_2^2 \left[\int_{-\infty}^{+\infty} f_X(x_1, x_2) dx_1 \right] dx_2 \\ &= EX_2^2 < +\infty. \end{aligned}$$

Il en résulte, puisque l'on a, pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$|x_1 x_2| f_X(x_1, x_2) \leq \frac{1}{2} (x_1^2 + x_2^2) f_X(x_1, x_2),$$

que l'application $(x_1, x_2) \mapsto |x_1 x_2| f_X(x_1, x_2)$ est Riemann-intégrable sur \mathbb{R}^2 et donc que la variable aléatoire $X_1 X_2$ admet une moyenne.

Comme dans le cas discret, la **covariance** et le **coefficient de corrélation** de deux variables aléatoires réelles servent à mesurer une certaine liaison entre ces variables aléatoires.

Définition 6.17. *Si la variable aléatoire $X = (X_1, X_2)$ à valeurs dans \mathbb{R}^2 admet une densité f_X et si X_1 et X_2 possèdent un moment d'ordre deux, le réel $E[(X_1 - EX_1)(X_2 - EX_2)]$ est appelé **covariance** de X_1 et X_2 et noté $\text{cov}(X_1, X_2)$. Le **coefficient de corrélation** de X_1 et X_2 est alors défini par :*

$$\rho_{X_1, X_2} = \frac{\text{cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}}.$$

Remarque. Sous ces hypothèses on a $\sigma_{X_1} \neq 0$ et $\sigma_{X_2} \neq 0$.

Proposition 6.18. *Si la variable aléatoire $X = (X_1, X_2)$ à valeurs dans \mathbb{R}^2 admet une densité f_X et si X_1 et X_2 possèdent un moment d'ordre deux, le coefficient de corrélation ρ_{X_1, X_2} vérifie :*

$$|\rho_{X_1, X_2}| \leq 1.$$

Démonstration. Après s'être assuré que l'inégalité de Schwarz est encore valable, on procède comme dans le cas des variables aléatoires discrètes.

6.4. Loi de Laplace-Gauss à deux dimensions

Nous appliquons ces notions à l'étude de la loi de Laplace-Gauss à deux dimensions.

Rappelons qu'une forme quadratique quelconque q sur \mathbb{R}^2 s'écrit :

$$q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2,$$

et que, si $a \neq 0$, sa décomposition en carrés s'écrit

$$q(x_1, x_2) = a \left[x_1 + \frac{b}{a}x_2 \right]^2 + x_2^2 \left(c - \frac{b^2}{a} \right).$$

On en déduit que, pour que q soit définie positive, il est nécessaire et suffisant qu'on ait :

$$a > 0, \quad c > 0, \quad c - \frac{b^2}{a} > 0.$$

On suppose cette condition réalisée dans la suite. Il est alors facile de démontrer, en utilisant la décomposition en carrés, que

$$I \equiv \int_{\mathbb{R}^2} \exp\left(-\frac{1}{2}q(x_1, x_2)\right) dx_1 dx_2 < +\infty.$$

Définition 6.19. *Une variable aléatoire $X = (X_1, X_2)$ à valeurs dans \mathbb{R}^2 de densité f_X définie par*

$$f_X(x_1, x_2) = K \exp\left(-\frac{1}{2}q(x_1, x_2)\right),$$

où $K = \frac{1}{I}$, est dite de Laplace-Gauss (ou normale) centrée à deux dimensions.

Nous allons calculer, à titre d'exemple, les **moments d'ordre un et deux** d'une telle variable aléatoire de Laplace-Gauss à deux dimensions et exprimer sa densité à l'aide de ces moments.

Calcul de K. En utilisant le théorème de Fubini et l'égalité

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = 1, \quad (6.10)$$

(m réel, $\sigma > 0$), on obtient

$$\begin{aligned} I &= \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}x_2^2\left(c - \frac{b^2}{a}\right)\right) \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{a}{2}\left(x_1 + \frac{b}{a}x_2\right)^2 dx_1\right)\right) dx_2 \\ &= \frac{\sqrt{2\pi}}{\sqrt{a}} \int_{-\infty}^{+\infty} \exp\left(-\frac{\left(c - \frac{b^2}{a}\right)}{2}x_2^2\right) dx_2 = \frac{\sqrt{2\pi}}{\sqrt{a\left(c - \frac{b^2}{a}\right)}}, \end{aligned}$$

et donc :

$$\boxed{K = \frac{\sqrt{ac - b^2}}{\sqrt{2\pi}}.}$$

Calcul de la moyenne EX_1 . On a

$$EX_1 = K \int_{\mathbb{R}^2} x_1 \exp\left(-\frac{1}{2}q(x_1, x_2)\right) dx_1 dx_2.$$

Si F est l'application de \mathbb{R}^2 dans \mathbb{R} définie par

$$F(x_1, x_2) = x_1 \exp\left(-\frac{1}{2}q(x_1, x_2)\right),$$

on constate que l'on a, pour tout $(x_1, x_2) \in (\mathbb{R}^+)^2$,

$$F(-x_1, -x_2) = -F(x_1, x_2),$$

et, pour tout $(x_1, x_2) \in \mathbb{R}^- \times \mathbb{R}^+$,

$$F(-x_1, x_2) = -F(x_1, -x_2).$$

On a donc :

$$\boxed{EX_1 = 0.}$$

On a de même

$$\boxed{EX_2 = 0.}$$

Calcul de EX_1X_2 . On a :

$$EX_1X_2 = K \int_{\mathbb{R}^2} x_1x_2 \exp\left(-\frac{1}{2}q(x_1, x_2)\right) dx_1dx_2.$$

La réduction de la forme quadratique q suggère le changement de variables sur \mathbb{R}^2 défini par :

$$\begin{cases} u = \sqrt{a} \left[x_1 + \frac{b}{a}x_2 \right] \\ v = x_2 \sqrt{c - \frac{b^2}{a}} \end{cases} \iff \begin{cases} x_1 = \frac{u}{\sqrt{a}} - \frac{b}{a\sqrt{c - \frac{b^2}{a}}}v \\ x_2 = \frac{v}{\sqrt{c - \frac{b^2}{a}}} \end{cases}$$

Le jacobien de ce difféomorphisme est

$$J = \begin{vmatrix} \frac{\partial x_1(u, v)}{\partial u} & \frac{\partial x_1(u, v)}{\partial v} \\ \frac{\partial x_2(u, v)}{\partial u} & \frac{\partial x_2(u, v)}{\partial v} \end{vmatrix} = \frac{1}{\sqrt{ac - b^2}},$$

et donc

$$EX_1X_2 = \frac{1}{2\pi} \int_{\mathbb{R}^2} \left(\frac{u}{\sqrt{a}} - \frac{bv}{a\sqrt{c - \frac{b^2}{a}}} \right) \left(\frac{v}{\sqrt{c - \frac{b^2}{a}}} \right) \exp\left(-\frac{1}{2}(u^2 + v^2)\right) du dv,$$

soit, en utilisant un argument de symétrie comme ci-dessus :

$$EX_1X_2 = \frac{1}{2\pi} \frac{-b}{ac - b^2} \int_{\mathbb{R}^2} v^2 \exp\left(-\frac{1}{2}(u^2 + v^2)\right) du dv.$$

Le théorème de Fubini et les égalités

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx = \int_{-\infty}^{+\infty} x^2 \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}$$

conduisent à l'égalité

$$\boxed{EX_1X_2 = \frac{-b}{ac - b^2} .}$$

Calcul de EX_1^2 . On a

$$EX_1^2 = K \int_{\mathbb{R}^2} x_1^2 \exp\left(-\frac{1}{2}q(x_1, x_2)\right) dx_1 dx_2.$$

Le même changement de variable et les mêmes arguments permettent d'écrire

$$EX_1^2 = \frac{1}{2\pi} \int_{\mathbb{R}^2} \left(\frac{u}{\sqrt{a}} - \frac{b}{a\sqrt{c - \frac{b^2}{a}}} v \right)^2 \exp\left(-\frac{1}{2}(u^2 + v^2)\right) du dv,$$

soit, après développement et réduction :

$$\begin{aligned} EX_1^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{u^2}{a} \exp\left(-\frac{1}{2}u^2\right) du \\ &\quad + \frac{1}{\sqrt{2\pi}} \frac{b^2}{a^2 \left(c - \frac{b^2}{a}\right)} \int_{-\infty}^{+\infty} v^2 \exp\left(-\frac{1}{2}v^2\right) dv \\ &= \frac{1}{a} + \frac{b^2}{a(ac - b^2)}. \end{aligned}$$

On en déduit la variance de X_1 :

$$\sigma_{X_1}^2 \equiv EX_1^2 = \frac{c}{ac - b^2}.$$

De même, la variance de X_2 est :

$$\sigma_{X_2}^2 \equiv EX_2^2 = \frac{a}{ac - b^2}.$$

Le coefficient de corrélation de X_1 et X_2 est alors donné par :

$$\rho_{X_1, X_2} = -\frac{b}{\sqrt{ac}}.$$

Remarquant que

$$1 - \rho_{X_1, X_2}^2 = \frac{ac - b^2}{ac},$$

un calcul algébrique simple conduit à l'expression suivante de la forme quadratique q en fonction des variances et coefficient de corrélation de X_1 et X_2 : pour tout $(x_1, x_2) \in \mathbb{R}^2$, on a

$$q(x_1, x_2) = \frac{1}{1 - \rho_{X_1, X_2}^2} \left(\frac{x_1^2}{\sigma_{X_1}^2} - 2\rho_{X_1, X_2} \frac{x_1}{\sigma_{X_1}} \frac{x_2}{\sigma_{X_2}} + \frac{x_2^2}{\sigma_{X_2}^2} \right).$$

On a de plus

$$K = \frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}\sqrt{1 - \rho_{X_1, X_2}^2}},$$

ce qui donne l'expression de la **densité d'une variable aléatoire de Laplace-Gauss centrée à deux dimensions**. Pour tout $(x_1, x_2) \in$

\mathbb{R}^2 , on a :

$$f_X(x_1, x_2) = \frac{\exp \left[-\frac{1}{2} \frac{1}{1 - \rho_{X_1, X_2}^2} \left(\frac{x_1^2}{\sigma_{X_1}^2} - 2 \rho_{X_1, X_2} \frac{x_1}{\sigma_{X_1}} \frac{x_2}{\sigma_{X_2}} + \frac{x_2^2}{\sigma_{X_2}^2} \right) \right]}{2\pi \sigma_{X_1} \sigma_{X_2} \sqrt{1 - \rho_{X_1, X_2}^2}}.$$

6.5. Indépendance de deux variables aléatoires réelles

Dans cette section, X_1 et X_2 sont deux variables aléatoires **réelles**. On note $X = (X_1, X_2)$ la variable aléatoire à valeurs dans \mathbb{R}^2 qui leur est associée. On étudie des **critères d'indépendance** des variables aléatoires X_1 et X_2 .

Proposition 6.20. (i) *Pour que les variables aléatoires X_1 et X_2 soient indépendantes, il faut et il suffit que, pour tout $(x_1, x_2) \in \mathbb{R}^2$, on ait :*

$$P(X_1 \leq x_1, X_2 \leq x_2) = P(X_1 \leq x_1) P(X_2 \leq x_2). \quad (6.11)$$

(ii) *Si les variables aléatoires X_1 et X_2 sont indépendantes et admettent des densités f_{X_1} et f_{X_2} , la variable aléatoire $X = (X_1, X_2)$ admet une densité f_X , fonction définie sur \mathbb{R}^2 par*

$$f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2).$$

(iii) *Inversement, si la variable aléatoire $X = (X_1, X_2)$ admet une densité f_X vérifiant⁸ : pour tout $(x_1, x_2) \in \mathbb{R}^2$,*

$$f_X(x_1, x_2) = f(x_1) g(x_2),$$

*où f et g sont des fonctions positives Riemann-intégrables sur \mathbb{R} , alors les variables aléatoires X_1 et X_2 sont **indépendantes**, et f et g sont, à un coefficient positif multiplicatif près, les **densités respectives** de X_1 et X_2 .*

Démonstration.

(i) Le sens direct est une conséquence de la définition de l'indépendance de variables aléatoires, la réciproque, qui ressort de la théorie de la mesure, est admise.

(ii) Sous ces hypothèses, la relation (6.11) s'écrit, pour tous réels a, b, c, d tels que $a < b$ et $c < d$,

8. On dit alors que la fonction f_X est produit direct des fonctions f et g .

$$P[a < X_1 \leq b, c < X_2 \leq d] = \left(\int_a^b f_{X_1}(u_1) du_1 \right) \left(\int_c^d f_{X_2}(u_2) du_2 \right),$$

soit, par le théorème de Fubini (noter que l'application $(x_1, x_2) \mapsto f_{X_1}(x_1)f_{X_2}(x_2)$ est régulière) :

$$P[a < X_1 \leq b, c < X_2 \leq d] = \int_a^b \int_c^d f_{X_1}(u_1) f_{X_2}(u_2) du_1 du_2,$$

ce qui démontre le résultat.

(iii) La variable aléatoire $X = (X_1, X_2)$ admettant une densité f_X (qui, par hypothèse, est régulière), les marginales X_1 et X_2 en admettent une, donnée respectivement sur \mathbb{R} par

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_X(x_1, x_2) dx_2,$$

et

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} f_X(x_1, x_2) dx_1.$$

Il en résulte, d'après l'hypothèse, que pour tout $x_1 \in \mathbb{R}$,

$$f_{X_1}(x_1) = f(x_1) \int_{-\infty}^{+\infty} g(x_2) dx_2$$

et, pour tout $x_2 \in \mathbb{R}$,

$$f_{X_2}(x_2) = g(x_2) \int_{-\infty}^{+\infty} f(x_1) dx_1.$$

Puisque l'on a

$$\int_{-\infty}^{+\infty} f_{X_1}(x_1) dx_1 = 1,$$

il vient, en intégrant :

$$\left(\int_{-\infty}^{+\infty} f(x_1) dx_1 \right) \left(\int_{-\infty}^{+\infty} g(x_2) dx_2 \right) = 1;$$

on a donc, pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$f_X(x_1, x_2) = f(x_1) g(x_2) = f_{X_1}(x_1) f_{X_2}(x_2).$$

On a alors, pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$\begin{aligned}
 P(X_1 \leq x_1, X_2 \leq x_2) &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_X(x_1, x_2) du_1 du_2 \\
 &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1}(u_1) f_{X_2}(u_2) du_1 du_2 \\
 &= \left(\int_{-\infty}^{x_1} f_{X_1}(u_1) du_1 \right) \left(\int_{-\infty}^{x_2} f_{X_2}(u_2) du_2 \right) \\
 &= P(X_1 \leq x_1) P(X_2 \leq x_2),
 \end{aligned}$$

ce qui démontre l'indépendance des variables aléatoires X_1 et X_2 .

Corollaire 6.21. Soient X_1 et X_2 deux variables aléatoires réelles indépendantes admettant des densités.

(i) Si Φ_1 et Φ_2 sont deux fonctions réelles telles que $\Phi_1(X_1)$ et $\Phi_2(X_2)$ admettent une moyenne, la variable aléatoire produit $\Phi_1(X_1)\Phi_2(X_2)$ admet aussi une moyenne et on a :

$$E[\Phi_1(X_1)\Phi_2(X_2)] = E[\Phi_1(X_1)] E[\Phi_2(X_2)]. \quad (6.12)$$

(ii) Si de plus les variables aléatoires X_1 et X_2 admettent une variance, on a $\text{cov}(X_1, X_2) = 0$ et donc :

$$\sigma_{X_1+X_2}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2. \quad (6.13)$$

La réciproque est fautive : on peut avoir $\text{cov}(X_1, X_2) = 0$, sans que les variables aléatoires X_1 et X_2 soient indépendantes.

Démonstration.

(i) La variable aléatoire $X = (X_1, X_2)$ admet une densité f_X produit direct de f_{X_1} et f_{X_2} . Il résulte du théorème de Fubini que l'on a :

$$\begin{aligned}
 &\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\Phi_1(u_1)| |\Phi_2(u_2)| f_{X_1}(u_1) f_{X_2}(u_2) du_1 du_2 \\
 &= \left(\int_{-\infty}^{+\infty} |\Phi_1(u_1)| f_{X_1}(u_1) du_1 \right) \left(\int_{-\infty}^{+\infty} |\Phi_2(u_2)| f_{X_2}(u_2) du_2 \right) < +\infty,
 \end{aligned}$$

ce qui démontre que la variable aléatoire $|\Phi_1(X_1)\Phi_2(X_2)|$ admet une moyenne. Le théorème de Fubini, appliqué sans valeurs absolues, donne la relation (6.12).

(ii) En particulier, on a

$$E[X_1 X_2] = E[X_1] E[X_2],$$

c'est-à-dire, $\text{cov}(X_1, X_2) = 0$. On obtient alors l'égalité (6.13) en développant le carré $(\overset{\circ}{X}_1 + \overset{\circ}{X}_2)^2$ et en utilisant la linéarité de l'espérance.

6.6. Somme de variables aléatoires réelles indépendantes

On étudie la loi de la somme de deux variables aléatoires indépendantes et admettant une densité.

Proposition 6.22. *Si X_1 et X_2 sont deux variables aléatoires réelles indépendantes admettant des densités f_{X_1} et f_{X_2} , la variable aléatoire $X = X_1 + X_2$ admet une densité f_X donnée par la convolution des fonctions f_{X_1} et f_{X_2} , c'est-à-dire la fonction définie pour tout $x \in \mathbb{R}$ par*

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X_1}(x_1) f_{X_2}(x - x_1) dx_1 = \int_{-\infty}^{+\infty} f_{X_1}(x - x_2) f_{X_2}(x_2) dx_2. \quad (6.14)$$

Démonstration. La variable aléatoire (X_1, X_2) admet une densité $f_{(X_1, X_2)}$ produit direct de f_{X_1} et f_{X_2} ; on a donc, pour tout $x \in \mathbb{R}$,

$$P(X_1 + X_2 \leq x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbf{1}_{D_x}(x_1, x_2) f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2,$$

où D_x est le domaine de \mathbb{R}^2 défini par :

$$D_x = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 \leq x\}.$$

Faisant le changement de variables de \mathbb{R}^2 sur lui-même, de jacobien 1, défini par

$$\begin{cases} u = x_1 \\ v = x_1 + x_2 \end{cases} \iff \begin{cases} x_1 = u \\ x_2 = v - u, \end{cases}$$

il vient, pour tout $x \in \mathbb{R}$:

$$P(X_1 + X_2 \leq x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbf{1}_{]-\infty, x]}(v) f_{X_1}(u) f_{X_2}(v - u) du dv.$$

Il en résulte, par le théorème de Fubini, que, pour tout $x \in \mathbb{R}$, on a

$$P(X_1 + X_2 \leq x) = \int_{-\infty}^x \left(\int_{-\infty}^{+\infty} f_{X_1}(u) f_{X_2}(v - u) du \right) dv,$$

ce qui démontre le résultat.

Exemple 6.4. Soient X_1 et X_2 deux variables aléatoires réelles **indépendantes** de même **loi exponentielle** de paramètre $p > 0$; déterminer la loi de la somme $X = X_1 + X_2$ par sa densité f_X . (La variable aléatoire X peut représenter la somme de deux temps d'attente indépendants, de même loi exponentielle).

On rappelle qu'alors, pour tout $x \in \mathbb{R}$, on a :

$$f_{X_1}(x) = f_{X_2}(x) = \mathbf{1}_{\mathbb{R}^+}(x) p \exp(-px).$$

La variable aléatoire X admet une densité donnée par, pour tout $x \in \mathbb{R}$,

$$f_X(x) = \int_{-\infty}^{+\infty} \mathbf{1}_{\mathbb{R}^+}(x_1) p \exp(-px_1) \mathbf{1}_{\mathbb{R}^+}(x-x_1) p \exp(-p(x-x_1)) dx_1,$$

soit :

$$f_X(x) = p^2 \mathbf{1}_{\mathbb{R}^+}(x) \exp(-px) \int_0^x dx_1.$$

La variable aléatoire X admet donc la densité f_X définie par, pour tout $x \in \mathbb{R}$,

$$f_X(x) = p^2 \mathbf{1}_{\mathbb{R}^+}(x) x \exp(-px).$$

6.7. Densités conditionnelles

Avertissement. Ce paragraphe est a priori destiné aux candidats préparant l'oral de l'Agrégation.

Nous ne considérons ici que des variables aléatoires (X_1, X_2) à valeurs dans \mathbb{R}^2 et admettant une **densité** $f_{(X_1, X_2)}$. Dans ce cas, on sait que, pour tout $x_1 \in \mathbb{R}$, on a $P(X_1 = x_1) = 0$, puisque X_1 admet une densité. Il est donc impossible de définir, comme nous l'avons fait pour une variable aléatoire discrète, une probabilité conditionnelle $P(X_2 \leq x_2 \mid X_1 = x_1)$. On définit toutefois par analogie une notion de **densité conditionnelle**⁹.

Définition 6.23. Soit (X_1, X_2) une variable aléatoire à valeurs dans \mathbb{R}^2 admettant une **densité** $f_{(X_1, X_2)}$ régulière¹⁰. On définit, pour tout $x_1 \in \mathbb{R}$ tel que $f_{X_1}(x_1) \neq 0$, la **densité conditionnelle de X_2 sachant $X_1 = x_1$** comme la fonction $f_{X_2}^{X_1=x_1}$ définie sur \mathbb{R} par, pour tout $x_2 \in \mathbb{R}$

$$f_{X_2}^{X_1=x_1}(x_2) = \frac{f_{(X_1, X_2)}(x_1, x_2)}{f_{X_1}(x_1)}. \quad (6.15)$$

9. La théorie générale des probabilités conditionnelles sera présentée dans le deuxième tome.

10. Voir note de la proposition 6.6.

Proposition 6.24. *Sous les hypothèses de la définition précédente, et si le support de f_{X_1} , à savoir l'adhérence de l'ensemble $\{x_1 \in \mathbb{R} \mid f_{X_1}(x_1) \neq 0\}$, est un intervalle I , on a, pour tous réels a_1, a_2, b_1, b_2 tels que $a_1 < b_1$, $[a_1, b_1] \subset I$ et $a_2 < b_2$:*

$$\boxed{P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = \int_{a_1}^{b_1} \left[\int_{a_2}^{b_2} f_{X_2}^{X_1=x_1}(x_2) dx_2 \right] f_{X_1}(x_1) dx_1.}$$

(6.16)

Démonstration. Puisque, pour tout $x_1 \in [a_1, b_1]$, on a $f_{X_1}(x_1) \neq 0$, il résulte du théorème de Fubini et de la définition (6.15) que :

$$\begin{aligned} P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) &= \int_{a_1}^{b_1} \left[\int_{a_2}^{b_2} f_{(X_1, X_2)}(x_1, x_2) dx_2 \right] dx_1 \\ &= \int_{a_1}^{b_1} \left[\int_{a_2}^{b_2} f_{X_2}^{X_1=x_1}(x_2) dx_2 \right] f_{X_1}(x_1) dx_1. \end{aligned}$$

Inversement, si X_1 admet une densité f_{X_1} de support un intervalle I , et s'il existe une **famille de densités** $(f^x(\cdot))_{x \in I}$ telles que, pour tous réels a_1, a_2, b_1, b_2 vérifiant $a_1 < b_1$, $[a_1, b_1] \subset I$ et $a_2 < b_2$, on ait l'égalité :

$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = \int_{a_1}^{b_1} \left[\int_{a_2}^{b_2} f^{x_1}(x_2) dx_2 \right] f_{X_1}(x_1) dx_1,$$

la variable aléatoire (X_1, X_2) admet une **densité** $f_{(X_1, X_2)}$ de support contenu dans $I \times \mathbb{R}$ et on a, pour tout $(x_1, x_2) \in I \times \mathbb{R}$,

$$\boxed{f_{(X_1, X_2)}(x_1, x_2) = f_{X_1}(x_1) f^{x_1}(x_2).}$$

La variable aléatoire X_2 admet alors une **densité conditionnelle sachant** $X_1 = x_1$ donnée par f^{x_1} .

Exemple 6.5.

- Soit $X = (X_1, X_2)$ une variable aléatoire de loi de **Laplace-Gauss** à deux dimensions (centrée). On rappelle que sa densité f_X est définie par : pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$f_X(x_1, x_2) = K \exp\left(-\frac{1}{2}q(x_1, x_2)\right),$$

où la forme quadratique q est définie par : pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$q(x_1, x_2) = \frac{1}{1 - \rho^2} \left(\frac{x_1^2}{\sigma_1^2} - 2\rho \frac{x_1 x_2}{\sigma_1 \sigma_2} + \frac{x_2^2}{\sigma_2^2} \right).$$

Le coefficient K vaut

$$K = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}};$$

les coefficients sont tels que : $\sigma_1 > 0$, $\sigma_2 > 0$ et $|\rho| < 1$. La forme quadratique q peut s'écrire encore : pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$q(x_1, x_2) = \frac{1}{1-\rho^2} \left[\left(\frac{x_2}{\sigma_2} - \rho \frac{x_1}{\sigma_1} \right)^2 + (1-\rho^2) \frac{x_1^2}{\sigma_1^2} \right].$$

Il en résulte que la marginale X_1 admet une densité f_{X_1} qui vérifie, pour tout $x_1 \in \mathbb{R}$,

$$\begin{aligned} f_{X_1}(x_1) &= K \exp\left(-\frac{x_1^2}{2\sigma_1^2}\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2(1-\rho^2)\sigma_2^2} \left(x_2 - \rho \frac{\sigma_2}{\sigma_1} x_1\right)^2\right) dx_2 \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{x_1^2}{2\sigma_1^2}\right). \end{aligned}$$

La densité conditionnelle de X_2 sachant $X_1 = x_1$ est donc donnée par : pour tout $x_2 \in \mathbb{R}$,

$$f_{X_2}^{X_1=x_1}(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)\sigma_2^2} \left(x_2 - \rho \frac{\sigma_2}{\sigma_1} x_1\right)^2\right],$$

c'est-à-dire que, pour tout $x_1 \in \mathbb{R}$, la fonction $f_{X_2}^{X_1=x_1}$ est la densité d'une loi gaussienne sur \mathbb{R} , à savoir la loi $\mathcal{N}\left(\rho \frac{\sigma_2}{\sigma_1} x_1, (1-\rho^2)\sigma_2^2\right)$.

- **Inversement**, si la variable aléatoire $X = (X_1, X_2)$ est telle que la loi de X_1 est la loi de Gauss $\mathcal{N}(0, \sigma_1^2)$ et si la variable aléatoire X_2 admet une densité conditionnelle sachant $X_1 = x_1$ donnée par : pour tout $x_2 \in \mathbb{R}$,

$$f^{x_1}(x_2) = \frac{1}{\sqrt{2\pi}a} \exp\left(-\frac{1}{2a^2} (x_2 - bx_1)^2\right),$$

où $a > 0$, la variable aléatoire $X = (X_1, X_2)$ suit une loi de **Laplace-Gauss centrée** à deux dimensions.

6.8. Annexe. L'intégrale de Riemann dans \mathbb{R}^n

Voici, sans démonstrations, les principales définitions et les propriétés les plus importantes de l'intégrale de Riemann dans \mathbb{R}^n .

6.8.1. Définition de l'intégrale de Riemann dans \mathbb{R}^n

a. Intégrale d'une fonction sur un pavé. Soit $P = [a_1, b_1] \times \cdots \times [a_n, b_n]$ un pavé compact dans \mathbb{R}^n .

La notion de fonction intégrable au sens de Riemann sur P , et l'intégrale d'une fonction intégrable au sens de Riemann sur P se définissent exactement comme en dimension 1, en remplaçant les intervalles par des pavés, et la longueur d'un intervalle par le volume d'un pavé.

On commence par définir **l'intégrale d'une fonction en escalier**. Une fonction en escalier (sur un pavé P), est par définition une fonction de la forme

$$f = \sum_{i=1}^k \lambda_i \mathbf{1}_{A_i}, \quad (6.17)$$

où les A_i sont des pavés fermés bornés contenus dans P . Noter que les fonctions indicatrices des pavés ouverts, semi-ouverts, etc. sont des fonctions en escalier. L'intégrale de la fonction en escalier (6.17) est par définition

$$\int_P f(x) dx = \sum_{i=1}^k \lambda_i \text{vol } A_i, \quad (6.18)$$

(La représentation d'une fonction en escalier sous la forme (6.17) n'est pas unique, mais la somme figurant au second membre de (6.18) ne dépend pas de la représentation choisie.)

On dit qu'une fonction f bornée définie sur P est **intégrable au sens de Riemann**, ou **Riemann-intégrable**, sur P si, pour tout $\varepsilon > 0$, il existe des fonctions en escalier φ et ψ définies sur P telles que

$$\varphi \leq f \leq \psi$$

et

$$\int_P (\psi(x) - \varphi(x)) dx < \varepsilon.$$

Si f est Riemann-intégrable sur P , les nombres

$$\sup_{\varphi \leq f} \int_P \varphi(x) dx \quad \text{et} \quad \inf_{\psi \geq f} \int_P \psi(x) dx,$$

où φ et ψ sont des fonctions en escalier, sont finis, et ont la même valeur. Cette valeur est par définition **l'intégrale de f sur P** . Elle est notée

$$\int_P f(x) dx \quad \text{ou} \quad \int_P f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

b. Propriétés. Les propriétés élémentaires de l'intégrale sont les mêmes qu'en dimension 1 : la somme, le produit de deux fonctions Riemann-intégrables sont Riemann-intégrables, etc. Une fonction Riemann-intégrable est bornée (par définition, en fait). Une fonction continue

sur P est Riemann-intégrable. Une fonction continue par morceaux est Riemann-intégrable si les « morceaux » sont suffisamment simples : par exemple, en dimension 2, polygones, disques, et plus généralement ensembles de la forme

$$a \leq x \leq b, \quad g(x) \leq y \leq k(x)$$

(g, h fonctions continues sur $[a, b]$), ou

$$c \leq y \leq d, \quad k(y) \leq x \leq l(y)$$

(k, l fonctions continues sur $[c, d]$), et réunions d'ensembles de cette forme.

c. Intégrale sur une partie bornée quarrable de \mathbb{R}^n . Soit A est une partie bornée de \mathbb{R}^n . Fixons un pavé compact P contenant A . On dit que A est **quarrable** si la fonction caractéristique $\mathbf{1}_A$ est intégrable sur P . On définit alors le volume de A en posant

$$\text{vol}(A) = \int_P \mathbf{1}_A(x) dx.$$

Supposons A quarrable, et soit f une fonction bornée sur A . On dit que f est Riemann-intégrable sur A si la fonction f^* obtenue en prolongeant f par 0 en-dehors de A est intégrable sur P , et on pose $\int_A f(x) dx = \int_P f^*(x) dx$. Toutes ces notions ne dépendent pas du pavé P choisi.

Par exemple, en dimension 2, les ensembles « de forme simple » définis ci-dessus sont quarrables.

L'intérieur $\overset{\circ}{A}$ et la fermeture \bar{A} d'un ensemble borné quarrable sont aussi quarrables, et on a $\text{vol}(\overset{\circ}{A}) = \text{vol}(A) = \text{vol}(\bar{A})$. On a de même

$$\int_{\overset{\circ}{A}} f(x) dx = \int_A f(x) dx.$$

Autrement dit, dans le calcul de l'intégrale, on peut négliger la partie de A contenue dans la frontière. Cette observation est importante pour l'application de la formule de changement de variables.

d. Intégrale sur \mathbb{R}^n (intégrale généralisée). Soit f une fonction positive définie sur \mathbb{R}^n , et Riemann-intégrable sur tout pavé. On dit que f est Riemann-intégrable sur \mathbb{R}^n si la borne supérieure des intégrales $\int_P f(x) dx$, où P décrit l'ensemble des pavés compacts de \mathbb{R}^n est finie. On pose alors

$$\int_{\mathbb{R}^n} f(x) dx = \sup_{P \text{ pavé compact}} \int_P f(x) dx.$$

Soit (A_k) est une suite croissante quelconque d'ensembles bornés quarrables telle que $\bigcup_k A_k = \mathbb{R}^n$ (on peut prendre par exemple pour A_k le pavé $[-k, k] \times \cdots \times [-k, k]$, ou la boule euclidienne de rayon k centrée à l'origine). Alors, pour que f soit Riemann-intégrable sur \mathbb{R}^n , il faut et il suffit¹¹ que la suite des intégrales $\int_{A_k} f(x) dx$ soit bornée, et on a dans ce cas

$$\int_{\mathbb{R}^n} f(x) dx = \lim_k \uparrow \int_{A_k} f(x) dx.$$

Si f est une fonction de signe quelconque définie sur \mathbb{R}^n , telle que $|f|$ soit Riemann-intégrable, alors, quelle que soit la suite (A_k) d'ensembles bornés quarrables telle que $\bigcup_k A_k = \mathbb{R}^n$, la suite des intégrales $\int_{A_k} f(x) dx$ est convergente¹². De plus, sa limite ne dépend pas de la suite (A_k) choisie. On pose alors

$$\int_{\mathbb{R}^n} f(x) dx = \lim_k \int_{A_k} f(x) dx.$$

d. Intégrale sur une partie non bornée quarrable de \mathbb{R}^n . Enfin, on peut définir l'intégrale généralisée de f sur une partie A non nécessairement bornée de \mathbb{R}^n , pourvu que l'intersection de A avec chaque pavé soit quarrable. On dit simplement dans ce cas que A est **quarrable**. Si A est donc une partie non bornée quarrable de \mathbb{R}^n , on dit que f est Riemann-intégrable sur A si la fonction \bar{f} obtenue en prolongeant f par 0 en-dehors de A est Riemann-intégrable sur \mathbb{R}^n . On pose alors

$$\int_A f(x) dx = \int_{\mathbb{R}^n} \bar{f}(x) dx.$$

Remarque importante. Pour une fonction f dont la valeur absolue $|f|$ n'est pas Riemann-intégrable, **on ne définit pas d'intégrale « semi-convergente » lorsque $d > 1$** . En effet, en dimension > 1 le comportement des intégrales $\int_{A_k} f(x) dx$ dépend très fortement du choix des A_k , dont la forme est arbitraire, et peut être choisie très biscornue. On peut en fait toujours choisir les A_k de manière à obtenir comme limite $+\infty$, ou $-\infty$, ou un nombre réel arbitraire (tel nombre réel que l'on voudra, écrivait-on au début du siècle). Dans certains cas, la limite peut être différente suivant qu'on prend pour A_k le pavé de côté $2k$ centré en 0 ou la boule de rayon k et de centre 0.

11. La démonstration de ce fait n'est élémentaire que si on suppose de plus que A_k est contenu dans l'intérieur de A_{k+1} pour $k = 1, 2, \dots$

12. Ici encore, la démonstration n'est élémentaire que si on suppose de plus que chacun des A_k est contenu dans l'intérieur de A_{k+1} .

6.8.2. Le théorème de Fubini

Le théorème de Fubini permet de comparer les intégrales définies sur un produit et les intégrales itérées. Par souci de simplicité, on ne s'occupe que du cas $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Les énoncés pour $\mathbb{R}^n = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ ($n = n_1 + n_2$) sont les mêmes.

a. Cas d'une fonction continue sur un pavé $[a, b] \times [c, d]$ La version la plus utile du théorème de Fubini est la suivante.

Soit f une fonction continue définie sur le pavé $[a, b] \times [c, d]$ dans \mathbb{R}^2 . Il est bien connu que les fonctions

$$x \mapsto \int_c^d f(x, y) dy \quad \text{et} \quad y \mapsto \int_a^b f(x, y) dx$$

sont continues respectivement sur $[a, b]$ et $[c, d]$. On a

$$\begin{aligned} \int_{[a,b] \times [c,d]} f(x, y) dx dy &= \int_a^b \left(\int_c^d f(x, y) dy \right) dx \\ &= \int_c^d \left(\int_a^b f(x, y) dx \right) dy \end{aligned} \quad (6.19)$$

b. Cas général. Il n'y a pas d'énoncé aussi commode valide dans le cas où f est seulement supposée intégrable au sens de Riemann. En effet, si f est supposée intégrable sur $[a, b] \times [c, d]$ on ne peut pas garantir que la fonction $y \mapsto f(x, y)$ soit, pour tout x , intégrable sur $[c, d]$, et même si c'est le cas, il n'est pas assuré que la fonction

$$x \mapsto \int_c^d f(x, y) dy,$$

qui est alors définie pour tout $x \in [a, b]$, soit Riemann-intégrable sur $[a, b]$. Les mêmes difficultés se présentent dans le cas d'une intégrale étendue au plan tout entier, et, dans ce cas, même si f est supposée continue.

Le « théorème de Fubini » pour l'intégrale de Riemann s'énonce ainsi.

Soit f une fonction définie et Riemann-intégrable sur \mathbb{R}^2 . Si pour tout x la fonction $y \mapsto f(x, y)$ est Riemann-intégrable sur \mathbb{R} (ce qui, avec les conventions ci-dessus, signifie que $y \mapsto f(x, y)$ est Riemann-intégrable sur tout intervalle fermé borné, et que l'intégrale $\int_{-\infty}^{+\infty} |f(x, y)| dy$ est convergente) et si la fonction

$$x \mapsto \int_{-\infty}^{+\infty} |f(x, y)| dy$$

est Riemann-intégrable sur \mathbb{R} , alors on a

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x, y) dy \right) dx. \quad (6.20)$$

L'énoncé correspondant dans le cas d'un pavé s'en déduit aussitôt. On dispose bien sûr des énoncés analogues où les rôles de x et de y sont échangés, ce qui permet, moyennant la vérification de certaines conditions, l'interversion des intégrations.

Noter que, si f est continue sur le plan, ou seulement continue « par morceaux simples », il suffit pour montrer que f est intégrable-Riemann de vérifier que, pour une suite croissante (P_k) de pavés telle que $\bigcup_k P_k = \mathbb{R}^2$, les intégrales

$$\int_k |f(x, y)| dx dy,$$

forment une suite bornée.

6.8.3. La formule de changement de variable

Soit A et B deux parties quarrables de \mathbb{R}^2 , et soit $\phi : B \rightarrow A$ une application continue qui réalise un **difféomorphisme** de l'intérieur de B sur l'intérieur de A . Alors, pour que f soit Riemann-intégrable sur A il faut et il suffit que la fonction

$$u \mapsto f(\phi(u)) |\det \phi'(u)|$$

le soit sur B et on a

$$\int_A f(x) dx = \int_B f(\phi(u)) |\det \phi'(u)| du. \quad (6.21)$$

On dit que le second membre de (6.21) s'obtient à partir du premier membre au moyen du **changement de variable** $x = \phi(u)$ (ou

$$\begin{cases} x_1 = \phi_1(u_1, u_2) \\ x_2 = \phi_2(u_1, u_2) \end{cases}$$

si l'on veut faire intervenir les coordonnées).

Dans la formule (6.21), la notation $\phi'(u)$ désigne la matrice jacobienne de ϕ au point $u = (u_1, u_2)$:

$$\phi'(u) = \begin{pmatrix} \frac{\partial \phi_1}{\partial u_1}(u_1, u_2) & \frac{\partial \phi_1}{\partial u_2}(u_1, u_2) \\ \frac{\partial \phi_2}{\partial u_1}(u_1, u_2) & \frac{\partial \phi_2}{\partial u_2}(u_1, u_2) \end{pmatrix}$$

qu'on écrit aussi plus simplement

$$\begin{pmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{pmatrix}.$$

Son déterminant

$$\det \phi'(u) = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{vmatrix}$$

est appelé **Jacobien** de ϕ au point u . Bien noter que c'est la **valeur absolue du Jacobien** qui intervient dans (6.21).

À titre d'exemple, voici le calcul de l'intégrale

$$I = \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx.$$

Une méthode consiste à calculer de deux façons différentes l'intégrale double

$$\iint_{\mathbb{R}^2} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) dx dy. \quad (6.22)$$

Appliquons d'abord le théorème de Fubini énoncé ci-dessus, en prenant

$$f(x, y) = \exp\left(-\frac{1}{2}(x^2 + y^2)\right) = \exp\left(-\frac{1}{2}x^2\right) \exp\left(-\frac{1}{2}y^2\right).$$

Pour tout x , la fonction $y \mapsto f(x, y)$ est intégrable sur \mathbb{R} (par exemple parce qu'on a $\exp\left(-\frac{1}{2}y^2\right) \leq (1 + \frac{1}{2}y^2)^{-1}$), et on a

$$\int_{\mathbb{R}} f(x, y) dy = \exp\left(-\frac{1}{2}x^2\right) \left(\int_{\mathbb{R}} \exp\left(-\frac{1}{2}y^2\right) dy \right).$$

La fonction $x \mapsto \int_{\mathbb{R}} f(x, y) dy$ est intégrable sur \mathbb{R} (pour la même raison). Après avoir noté que f est positive, on peut utiliser le théorème de Fubini pour obtenir

$$\begin{aligned} \iint_{\mathbb{R}^2} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) dx dy &= \int_{\mathbb{R}} \left[\exp\left(-\frac{1}{2}x^2\right) \left(\int_{\mathbb{R}} \exp\left(-\frac{1}{2}y^2\right) dy \right) \right] dx \\ &= \left(\int_{\mathbb{R}} \exp\left(-\frac{1}{2}x^2\right) dx \right) \left(\int_{\mathbb{R}} \exp\left(-\frac{1}{2}y^2\right) dy \right) = I^2. \end{aligned} \quad (6.23)$$

Calculons maintenant l'intégrale double (6.22) en passant en coordonnées polaires. De façon précise, effectuons dans (6.22) le changement de variable

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases} \quad 0 \leq r < +\infty, \quad 0 \leq \theta < 2\pi. \quad (6.24)$$

L'application $\phi : [0, +\infty[\times [0, 2\pi[\rightarrow \mathbb{R}^2$ qui définit ce changement de variable réalise un difféomorphisme de $]0, +\infty[\times]0, 2\pi[$ sur le plan privé du demi-axe $[0, +\infty[\times \{0\}$. Le jacobien de ϕ peut se calculer en différenciant (6.24) :

$$\begin{cases} dx = \cos \theta dr - r \sin \theta d\theta \\ dy = \sin \theta dr + r \cos \theta d\theta. \end{cases}$$

On a

$$J(\phi) = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r.$$

L'application de (6.21) donne

$$\iint_{\mathbb{R}^2} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) dx dy = \iint_{[0, +\infty[\times [0, 2\pi[} \exp\left(-\frac{1}{2}r^2\right) r dr d\theta.$$

En appliquant à l'intégrale du second membre le théorème de Fubini, on obtient

$$\begin{aligned} \iint_{\mathbb{R}^2} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) dx dy &= \left(\int_{[0, +\infty[} \exp\left(-\frac{1}{2}r^2\right) r dr \right) \left(\int_{[0, 2\pi[} d\theta \right) \\ &= 2\pi \left[\exp\left(-\frac{1}{2}r^2\right) \right]_0^{+\infty} = 2\pi. \end{aligned} \quad (6.25)$$

La comparaison de (6.23) et de (6.25) nous donne $I^2 = 2\pi$, soit

$$\boxed{\int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}x^2\right) dx = \sqrt{2\pi}.} \quad (6.26)$$

Remarque. Une démonstration souvent présentée évite le recours à l'utilisation du théorème de Fubini sur un domaine non borné : on calcule d'une part l'intégrale, notée I_R , de la fonction $\exp\left(-\frac{1}{2}(x^2 + y^2)\right)$ sur le carré $[-R, R]$, d'autre part, en passant en coordonnées polaires, l'intégrale, notée J_R , de cette fonction sur le disque de centre 0 et de rayon R , et on fait tendre R vers l'infini après avoir observé que $J_R \leq I_R \leq J_{R\sqrt{2}}$.

Exercices

Exercice 6.1. Loi triangulaire et indépendance. Soient deux réels $a > 0$ et $\alpha > 0$. Soit f la fonction définie sur \mathbb{R}^+ par : pour tout $x \in \mathbb{R}^+$,

$$f(x) = \alpha \left(x \mathbf{1}_{\left]0, \frac{a}{2}\right[}(x) + (a-x) \mathbf{1}_{\left] \frac{a}{2}, a\right[}(x) \right).$$

1. Calculer la constante α pour que f soit une densité de probabilité. On choisit dorénavant cette valeur pour α .

2. Soient X une variable aléatoire de densité f et un réel $b \in \left]0, \frac{a}{2}\right[$; calculer les probabilités $P\left(X > \frac{a}{2}\right)$ et $P\left(\frac{a}{2} - b < X \leq \frac{a}{2} + b\right)$.

3. Démontrer que pour tout $b \in \left]0, \frac{a}{2}\right[$, les événements $A = \left(X > \frac{a}{2}\right)$ et $B = \left(\frac{a}{2} - b < X \leq \frac{a}{2} + b\right)$ sont **indépendants**.

Solution.

1. On a

$$\frac{1}{\alpha} \int_{-\infty}^{+\infty} f(x) dx = \int_0^{\frac{a}{2}} x dx + \int_{\frac{a}{2}}^a (a-x) dx = 2 \left. \frac{x^2}{2} \right|_0^{\frac{a}{2}},$$

soit

$$\alpha = \frac{4}{a^2}.$$

2. On a

$$P\left(X > \frac{a}{2}\right) = \alpha \int_{\frac{a}{2}}^a (a-x) dx$$

soit encore :

$$P\left(X > \frac{a}{2}\right) = \frac{1}{2}.$$

De même, on a

$$P\left(\frac{a}{2} - b < X \leq \frac{a}{2} + b\right) = \alpha \left(\int_{\frac{a}{2}-b}^{\frac{a}{2}} x dx + \int_{\frac{a}{2}}^{\frac{a}{2}+b} (a-x) dx \right),$$

ce qui donne, en effectuant le changement de variable $y = a - x$ dans la deuxième intégrale :

$$P\left(\frac{a}{2} - b < X \leq \frac{a}{2} + b\right) = 2\alpha \int_{\frac{a}{2}-b}^{\frac{a}{2}} x dx = 2\alpha \left. \frac{x^2}{2} \right|_{\frac{a}{2}-b}^{\frac{a}{2}} = \alpha b(a-b),$$

soit :

$$P\left(\frac{a}{2} - b < X \leq \frac{a}{2} + b\right) = \frac{4b(a-b)}{a^2}.$$

3. On a :

$$P(A \cap B) = P\left(\frac{a}{2} < X \leq \frac{a}{2} + b\right) = \alpha \int_{\frac{a}{2}}^{\frac{a}{2}+b} (a-x) dx,$$

soit, par changement de variable $y = a - x$:

$$P(A \cap B) = \alpha \int_{\frac{a}{2}-b}^{\frac{a}{2}} y dy = \frac{\alpha}{2} [b(a-b)],$$

ce qui démontre que l'on a :

$$P(A \cap B) = P(A) P(B),$$

c'est-à-dire que **les événements A et B sont indépendants.**

Exercice 6.2. Loi exponentielle et temps d'arrivée. La date d'arrivée T du premier client après l'ouverture d'un guichet est une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{A}, P) , de loi exponentielle de paramètre $p > 0$.

1. Calculer la probabilité $P\left(T > \frac{1}{p}\right)$.

2. On fixe $\varepsilon > 0$. On considère pour $k \in \mathbb{N}$, les « tranches horaires » $[k\varepsilon, (k+1)\varepsilon[$ et les événements $A_k = \{T \in [k\varepsilon, (k+1)\varepsilon[$, correspondant à la propriété « le client arrive dans l'intervalle de temps $[k\varepsilon, (k+1)\varepsilon[$ ». Calculer la probabilité $P(T \in [k\varepsilon, (k+1)\varepsilon[)$.

3. Soit X la variable aléatoire à valeurs dans \mathbb{N} définie par : pour tout $\omega \in \Omega$,

$$X(\omega) = k \iff \omega \in A_k.$$

Quelle est la loi de la variable aléatoire X ?

4. Calculer, pour tout $t > 0$ et tout $h > 0$, la probabilité $P(t < T)$ et la probabilité conditionnelle $P(T > t+h \mid T > t)$.

Solution.

1. La loi de T étant la loi exponentielle $\exp(p)$, on a

$$P\left(T > \frac{1}{p}\right) = \int_{\frac{1}{p}}^{+\infty} p \exp(-px) dx = -\exp(-px) \Big|_{\frac{1}{p}}^{+\infty}$$

soit

$$P\left(T > \frac{1}{p}\right) = e^{-1} \simeq 0,367.$$

2. On a

$$\begin{aligned}
 P(T \in [k\varepsilon, (k+1)\varepsilon]) &= \int_{k\varepsilon}^{(k+1)\varepsilon} p \exp(-px) dx \\
 &= -\exp(-px) \Big|_{k\varepsilon}^{(k+1)\varepsilon} \\
 &= e^{-k\varepsilon} - e^{-(k+1)\varepsilon},
 \end{aligned}$$

soit encore :

$$P(T \in [k\varepsilon, (k+1)\varepsilon]) = e^{-k\varepsilon} (1 - e^{-\varepsilon}).$$

3. Puisque, pour tout $k \in \mathbb{N}$, on a

$$(X = k) = (T \in [k\varepsilon, (k+1)\varepsilon]),$$

il résulte de la question précédente que la loi de X est la loi géométrique $\mathcal{G}_N(1 - e^{-\varepsilon})$. Un phénomène modélisé en temps continu par une loi exponentielle doit être modélisé par une loi géométrique si on remplace le temps continu par un temps discret.

4. On a

$$P(T > t) = \int_t^{+\infty} p \exp(-px) dx = -\exp(-px) \Big|_t^{+\infty},$$

soit

$$P(T > t) = \exp(-pt).$$

Alors, puisque pour tout $t > 0$ et tout $h > 0$, on a

$$(T > t+h) \cap (T > t) = (T > t+h)$$

il vient :

$$P(T > t+h) = \exp(-p(t+h)).$$

Par définition de la probabilité conditionnelle, on a

$$P(T > t+h \mid T > t) = \frac{P(T > t+h, T > t)}{P(T > t)}$$

soit :

$$P(T > t+h \mid T > t) = \exp(-ph) \equiv P(T > h).$$

On dit que la loi exponentielle est sans mémoire. La probabilité d'attendre encore pendant un temps h , sachant qu'on a déjà attendu pendant le temps t est indépendante de t . Elle ne décroît pas (ni ne croît) avec t ! D'autres phénomènes sont régis par une loi exponentielle, comme la durée de vie d'un atome radioactif. Savoir qu'un atome radioactif s'est formé il y a une minute ou il y a 5 milliards d'années ne change en rien la probabilité de le voir se désintégrer dans la prochaine seconde.

On peut montrer que la propriété d'être sans mémoire caractérise les variables aléatoires à densité de loi exponentielle (c'est un exercice intéressant). On peut montrer de façon analogue que la condition $P(T > n+h \mid T \geq n) = P(T \geq h)$ caractérise les variables aléatoires de loi géométrique parmi les variables aléatoires à valeurs entières.

Exercice 6.3. Loi uniforme et décimales. Soit X une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{A}, P) , de loi uniforme sur l'intervalle $[0, 1]$. On définit les deux variables aléatoires D_1 et D_2 (première et deuxième décimale de X) par :

$$D_1 = [10X] \quad D_2 = [10^2X - 10D_1].$$

(On rappelle que $[x]$ désigne la partie entière du nombre réel x .)

1. Quelles sont les lois des variables aléatoires D_1 et D_2 ?
2. Démontrer que les variables aléatoires D_1 et D_2 sont indépendantes.

Remarque. On pourra généraliser cet exercice¹³ et démontrer que si D_n est la n -ième décimale de X la suite de variables aléatoires $(D_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires indépendantes de même loi uniforme sur l'ensemble $\{0, 1, 2, \dots, 9\}$.

Solution.

1. Pour tout entier $k \in \{0, 1, 2, \dots, 9\}$, on a

$$(D_1 = k) = (k \leq 10X < k + 1).$$

Puisque la variable aléatoire X est de loi uniforme sur l'intervalle $[0, 1]$, on a :

$$P(D_1 = k) = \frac{1}{10}.$$

De même, pour tout entier $k_2 \in \{0, 1, 2, \dots, 9\}$, on a

$$\begin{aligned} (D_2 = k_2) &= (k_2 \leq 10^2X - 10D_1 < k_2 + 1) \\ &= \bigcup_{k_1=0}^9 [(D_1 = k_1) \cap (k_2 \leq 10^2X - 10D_1 < k_2 + 1)] \\ &= \bigcup_{k_1=0}^9 [(D_1 = k_1) \cap \left(\frac{k_1}{10} + \frac{k_2}{10^2} \leq X < \frac{k_1}{10} + \frac{k_2}{10^2} + \frac{1}{10^2} \right)] \\ &= \bigcup_{k_1=0}^9 \left(\frac{k_1}{10} + \frac{k_2}{10^2} \leq X < \frac{k_1}{10} + \frac{k_2}{10^2} + \frac{1}{10^2} \right). \end{aligned}$$

Ces événements étant disjoints et de même probabilité 10^{-2} , il vient :

$$P(D_2 = k_2) = \frac{1}{10}.$$

2. De même, et plus brièvement, pour tous entiers $k_1, k_2 \in \{0, 1, 2, \dots, 9\}$, on a

$$\begin{aligned} (D_1 = k_1) \cap (D_2 = k_2) &= \left(\frac{k_1}{10} \leq X < \frac{k_1}{10} + \frac{1}{10} \right) \cap \left(\frac{k_1}{10} + \frac{k_2}{10^2} \leq X < \frac{k_1}{10} + \frac{k_2}{10^2} + \frac{1}{10^2} \right), \end{aligned}$$

13. Voir à ce sujet le deuxième tome.

soit

$$(D_1 = k_1) \cap (D_2 = k_2) = \left(\frac{k_1}{10} + \frac{k_2}{10^2} \leq X < \frac{k_1}{10} + \frac{k_2}{10^2} + \frac{1}{10^2} \right).$$

On a donc

$$P[(D_1 = k_1) \cap (D_2 = k_2)] = \frac{1}{10^2}.$$

On a montré que pour k_1 et k_2 dans $\{0, 1, 2, \dots, 9\}$, on a :

$$\boxed{P(D_1 = k_1, D_2 = k_2) = P(D_1 = k_1) P(D_2 = k_2)}.$$

Les variables aléatoires D_1 et D_2 sont donc indépendantes, de même loi uniforme sur l'ensemble $\{0, 1, 2, \dots, 9\}$.

Exercice 6.4. Lois uniforme et triangulaire ; convolution. Soient X et Y deux variables aléatoires définies sur un espace probabilisé (Ω, \mathcal{A}, P) , de loi uniforme sur l'intervalle $[0, 1]$ et indépendantes. Quelle est la loi de la variable aléatoire $Z = X + Y$?

Solution. La variable aléatoire Z admet une densité f_Z , convolution des densités de X et Y . Elle est donnée par : pour tout $z \in \mathbb{R}$,

$$f_Z(z) = \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1]}(x) \mathbf{1}_{[0,1]}(z-x) dx.$$

Puisque l'on a

$$\mathbf{1}_{[0,1]}(x) \mathbf{1}_{[0,1]}(z-x) = \mathbf{1}_{[0,1]}(z) \mathbf{1}_{[0,z]}(x) + \mathbf{1}_{[1,2]}(z) \mathbf{1}_{[z-1,1]}(x),$$

on peut écrire

$$f_Z(z) = \mathbf{1}_{[0,1]}(z) \int_0^z dx + \mathbf{1}_{[1,2]}(z) \int_{z-1}^1 dx.$$

La variable aléatoire Z admet donc la densité f_Z donnée par : pour tout $z \in \mathbb{R}$,

$$\boxed{f_Z(z) = z \mathbf{1}_{[0,1]}(z) + (2-z) \mathbf{1}_{[1,2]}(z)},$$

c'est-à-dire que Z suit une **loi triangulaire**.

Exercice 6.5. Loi d'une fonction de variable aléatoire et convolution.

Soit X une variable aléatoire réelle définie sur un espace probabilisé (Ω, \mathcal{A}, P) , de densité f continue par morceaux, et de fonction de répartition F . Soit Y la variable aléatoire $Y = X^2$.

1. Démontrer que Y admet une densité f_Y et l'exprimer à l'aide de f et F . Expliciter ce résultat dans le cas particulier où X suit la loi de Gauss $\mathcal{N}(0, 1)$.

2. Si X_1 et X_2 sont deux variables aléatoires réelles définies sur un espace probabilisé (Ω, \mathcal{A}, P) indépendantes et de loi de Gauss $\mathcal{N}(0, 1)$, justifier l'existence d'une densité pour la variable aléatoire $Z = X_1^2 + X_2^2$ et la calculer.

3. Même question pour la variable aléatoire $T = \sqrt{Z}$.

Solution.

1. Notons F_Y la fonction de répartition de la variable aléatoire Y . Remarquons d'abord que, pour tout $y \leq 0$, on a :

$$F_Y(y) \equiv P(Y \leq y) = 0.$$

Soit $y > 0$ quelconque ; on a

$$F_Y(y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f(u) du.$$

La fonction f étant continue par morceaux, la fonction de répartition F_Y est dérivable sur tout ouvert où f est continue et sa dérivée est la densité f_Y de Y ; elle est donnée par :

$$f_Y(y) = \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y})).$$

En résumé, Y admet une densité (continue par morceaux) donnée par :

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y})) & \text{si } y > 0 \\ 0 & \text{sinon.} \end{cases}$$

Si X suit la loi de Gauss $\mathcal{N}(0, 1)$, la fonction f est paire et on a :

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right) & \text{si } y > 0 \\ 0 & \text{sinon.} \end{cases}$$

Autrement dit, Y suit la loi du chi-deux à 1 degré de liberté.

2. Les variables aléatoires X_1^2 et X_2^2 sont indépendantes et leur somme Z admet donc une densité f_Z produit de convolution des densités de X_1^2 et X_2^2 . Il résulte de la question précédente que, pour tout $z \leq 0$, on a $f_Z(z) = 0$, et pour tout $z > 0$, on a :

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} \mathbf{1}_{\mathbb{R}^+}(y) \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right) \mathbf{1}_{\mathbb{R}^+}(z-y) \frac{1}{\sqrt{2\pi(z-y)}} \exp\left(-\frac{(z-y)}{2}\right) dy, \end{aligned}$$

soit, en réduisant et en faisant le changement de variables $y = zu$:

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi} \exp\left(-\frac{z}{2}\right) \int_0^z \frac{1}{\sqrt{y(z-y)}} dy \\ &= \frac{1}{2\pi} \exp\left(-\frac{z}{2}\right) \int_0^1 \frac{1}{\sqrt{u(1-u)}} du. \end{aligned}$$

On reconnaît la forme analytique de la densité d'une **loi exponentielle** $\exp\left(-\frac{z}{2}\right)$, ce qui donne d'ailleurs, sans calcul, le résultat élémentaire :

$$\int_0^1 \frac{1}{\sqrt{u(1-u)}} du = \pi.$$

On a alors :

$$f_Z(z) = \begin{cases} \frac{1}{2} \exp\left(-\frac{z}{2}\right) & \text{si } z > 0 \\ 0 & \text{sinon.} \end{cases}$$

On remarquera que cette loi est aussi la loi du chi-deux à 2 degrés de liberté.

3. Soit F_T la fonction de répartition de T ; pour tout $t \leq 0$, on a $F_T(t) = 0$, et pour tout $t > 0$, on a :

$$F_T(t) \equiv P(T \leq t) = P(Z \leq t^2) = \int_0^{t^2} \frac{1}{2} \exp\left(-\frac{z}{2}\right) dz.$$

La fonction F_T est donc dérivable, sauf peut-être en 0; la variable aléatoire T admet une densité f_T , dérivée de F_T , donnée par :

$$f_T(t) = \begin{cases} t \exp\left(-\frac{t^2}{2}\right) & \text{si } t > 0 \\ 0 & \text{sinon.} \end{cases}$$

Exercice 6.6. Calcul de moyenne. Soit X une variable aléatoire réelle définie sur un espace probablisé (Ω, \mathcal{A}, P) , de loi de Gauss $\mathcal{N}(0, 1)$. Vérifier que, pour tout $t \in \mathbb{R}$, la variable aléatoire $\exp(tX)$ admet une moyenne et la calculer.

Solution. Soit t réel. Pour tout $x \in \mathbb{R}$, on a :

$$\exp(tx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-t)^2}{2}\right) \exp\left(\frac{t^2}{2}\right),$$

ce qui démontre que la fonction $x \mapsto \exp(tx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ est Riemann-intégrable, et donc que la variable aléatoire $\exp(tX)$ admet une moyenne donnée par :

$$E[\exp(tX)] = \exp\left(\frac{t^2}{2}\right) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-t)^2}{2}\right) dx,$$

soit :

$$E[\exp(tX)] = \exp\left(\frac{t^2}{2}\right).$$

Exercice 6.7. Moments des lois continues classiques. Soit X une variable aléatoire réelle. Calculer sa moyenne et sa variance dans le cas où sa loi est successivement, la loi uniforme sur $[a, b]$, la loi exponentielle de

paramètre $p > 0$, la loi du chi-deux à n degrés de liberté. Si X est de loi de Cauchy, elle n'admet pas de moyenne et donc pas de variance.

Solution. Il suffit d'appliquer les formules et les calculs sont élémentaires ; nous ne donnons que les résultats.

– **Loi uniforme** :

$$\boxed{EX = \frac{a+b}{2} \quad \sigma_X^2 = \frac{(b-a)^2}{12}}$$

– **Loi exponentielle** $\exp(p)$:

$$\boxed{EX = \frac{1}{p} \quad \sigma_X^2 = \frac{1}{p^2}}$$

– **Loi du chi-deux** à n degrés de liberté :

$$\boxed{EX = n \quad \sigma_X^2 = 2n}$$

– **Loi de Cauchy** : la fonction $x \mapsto \frac{|x|}{\pi(1+x^2)}$ n'est pas intégrable sur \mathbb{R} ; la variable aléatoire X n'admet donc pas de moyenne.

Exercice 6.8. Loi de marginales et indépendance. Soit $X = (X_1, X_2)$ une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{A}, P) , de loi uniforme sur le carré $[0, 1] \times [0, 1]$. Démontrer que les variables aléatoires X_1 et X_2 sont indépendantes et de même loi uniforme sur $[0, 1]$.

Solution. Les marginales X_1 et X_2 admettent des densités données par :

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1] \times [0,1]}(x_1, x_2) dx_2 \quad \text{pour } x_1 \in \mathbb{R}$$

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1] \times [0,1]}(x_1, x_2) dx_1 \quad \text{pour } x_2 \in \mathbb{R}$$

Puisque l'on a

$$\mathbf{1}_{[0,1] \times [0,1]}(x_1, x_2) = \mathbf{1}_{[0,1]}(x_1) \mathbf{1}_{[0,1]}(x_2), \quad (6.27)$$

il vient pour x_1 et x_2 dans \mathbb{R} :

$$f_{X_1}(x_1) = \mathbf{1}_{[0,1]}(x_1) \quad \text{et} \quad f_{X_2}(x_2) = \mathbf{1}_{[0,1]}(x_2).$$

Il résulte alors de l'égalité (6.27) que la densité de X est produit direct des densités des variables aléatoires X_1 et X_2 et donc que ces dernières sont indépendantes.

Exercice 6.9. La covariance nulle n'implique pas l'indépendance. Soit $X = (X_1, X_2)$ une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{A}, P)

de loi uniforme sur le disque $\mathcal{D}(0, 1)$. Vérifier que les variables aléatoires X_1 et X_2 ne sont pas indépendantes. Calculer la covariance de X_1 et X_2 .

Solution. Les marginales X_1 et X_2 admettent des densités données pour tout x_1 et x_2 dans \mathbb{R} par :

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} \frac{1}{\pi} \mathbf{1}_{\mathcal{D}(0,1)}(x_1, x_2) dx_2$$

et

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} \frac{1}{\pi} \mathbf{1}_{\mathcal{D}(0,1)}(x_1, x_2) dx_1,$$

soit :

$$f_{X_1}(x_1) = \begin{cases} \int_0^{+\sqrt{1-x_1^2}} \frac{1}{\pi} dx_2 = \frac{2}{\pi} \sqrt{1-x_1^2} & \text{si } -1 < x_1 < 1 \\ 0 & \text{sinon,} \end{cases}$$

et

$$f_{X_2}(x_2) = \begin{cases} \int_0^{+\sqrt{1-x_2^2}} \frac{1}{\pi} dx_1 = \frac{2}{\pi} \sqrt{1-x_2^2} & \text{si } -1 < x_2 < 1 \\ 0 & \text{sinon.} \end{cases}$$

Le produit $f_{X_1}(x_1) f_{X_2}(x_2)$ est strictement positif sur le carré $]-1, 1[^2$ (et nul en dehors de ce carré). Il ne peut donc être égal à la densité de X . **Les variables aléatoires X_1 et X_2 ne sont donc pas indépendantes.** De plus, ces variables aléatoires ayant des densités paires ont une moyenne nulle. Enfin, par symétrie, on a :

$$EX_1X_2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\pi} x_1x_2 \mathbf{1}_{\mathcal{D}(0,1)}(x_1, x_2) dx_1 dx_2 = 0,$$

ce qui démontre que $\text{cov}(X_1, X_2) = 0$.

Chapitre 7

Approximation de lois. Loi faible des grands nombres

Avertissement. A priori, les candidats au concours interne du CAPES ne sont pas concernés par ce chapitre.

Les candidats au concours externe de l'Agrégation, principalement ceux qui ont choisi l'option Probabilités et Statistique, pourront considérer la lecture de ce chapitre comme une introduction aux notions d'approximation de lois développées ultérieurement (dans le tome II) en terme de convergence de suites de mesures bornées ; différentes notions de convergence de suites de variables aléatoires seront étudiées dans le cadre général.

7.1. Approximation de lois

L'objet de ce paragraphe est d'introduire aux problèmes d'**approximation de lois**.

Pour répondre à la demande des programmes du concours externe du CAPES et du concours interne de l'Agrégation, on ne parle pas ici de convergence en loi.

La loi binomiale, la loi hypergéométrique et les jeux de hasard font intervenir des factorielles. Quelques chiffres montreront l'intérêt de remplacer les lois de probabilité par des approximations¹.

$$10! = 3\,628\,800 \qquad 15! = 1\,307\,674\,368\,000,$$

et

$$52! = 80\,658\,175\,170\,943\,878\,571\,660\,636\,856\,403\,766\,975\,289\,505 \\ 440\,883\,277\,824\,000\,000\,000\,000.$$

7.1.1. Approximation de Poisson

Le **théorème de Poisson** donne une approximation de la loi binomiale $\mathcal{B}(n, p)$ lorsque n est « grand » et p « petit ».

1. Cette complexité numérique correspond bien sûr à une complexité dans les formules. Les logiciels récents de calcul symbolique, qui permettent en même temps de calculer avec des nombres de taille arbitraire, n'ont pas encore rendu cet argument tout à fait caduc. Il y a d'autres raisons pour étudier les approximations, comme on le verra dans la suite de ce chapitre, et notamment dans les exercices.

Théorème 7.1. (Théorème de Poisson.) Soit $(p_n)_{n \in \mathbb{N}}$ une suite de réels de l'intervalle $]0, 1[$ telle que $\lim_{n \rightarrow \infty} np_n = \lambda$ (où $\lambda > 0$). Considérons, pour chaque entier n , une variable aléatoire S_n de loi $\mathcal{B}(n, p_n)$. On a donc

$$P(S_n = k) = P_n(k) = \begin{cases} \binom{n}{k} p_n^k (1 - p_n)^{n-k} & \text{si } 0 \leq k \leq n \\ 0 & \text{sinon.} \end{cases}$$

Alors, pour tout entier $k \in \mathbb{N}$, la suite de terme général $P(S_n = k)$ est convergente et on a :

$$\lim_n P(S_n = k) = \exp(-\lambda) \frac{\lambda^k}{k!}.$$

Démonstration. Remarquons que l'on a, quand n tend vers l'infini,

$$p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right).$$

On a donc, pour tout $k \in \mathbb{N}$ fixé, et tout $n \geq k$:

$$P(S_n = k) = \frac{n(n-1)\dots(n-k+1)}{k!} \left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k}.$$

Mais on a

$$\begin{aligned} n(n-1)\dots(n-k+1) \left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k \\ = \frac{n(n-1)\dots(n-k+1)}{n^k} [\lambda + o(1)]^k \xrightarrow{n \rightarrow +\infty} \lambda^k. \end{aligned}$$

D'autre part, on sait que

$$\left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k} \xrightarrow{n \rightarrow +\infty} \exp(-\lambda),$$

ce qui démontre le résultat.

Remarque. Il existe un résultat plus précis qui fournit une **vitesse de convergence**, uniforme en k ; il est de démonstration technique et délicate (les lecteurs intéressés peuvent consulter par exemple le livre de Shiryaev, *Probability*). Nous ne ferons que le citer.

Si, pour tout $n \in \mathbb{N}$, on a $np_n = \lambda$, où λ est un réel strictement positif, on a l'estimation :

$$\sum_{k=0}^{+\infty} \left| P_n(k) - \exp(-\lambda) \frac{\lambda^k}{k!} \right| \leq \frac{2\lambda}{n} \min(2, \lambda).$$

Le théorème de **Poisson** donne une **approximation de la loi binomiale** lorsque le paramètre p est « **petit** ». Dans la pratique, il est d'usage de remplacer la loi binomiale par la loi de Poisson dès que n est assez grand, de l'ordre de 30, et que p est petit, de l'ordre de 0,1. Autrement dit, si X est une variable aléatoire de loi binomiale de **premier paramètre** n et de **moyenne** λ pas « **trop grande** », sa loi est approximativement la loi de Poisson $\mathcal{P}(\lambda)$.

Exemple 7.1. On veut déterminer numériquement la probabilité $P_n(k)$ pour que, parmi n personnes, k soient nées le premier janvier (on suppose qu'aucune n'est née le 29 février et que tous les autres jours sont équiprobables) ; on fait l'étude, dans les deux cas $n = 500$ et 600 , et pour les valeurs de k comprises entre 0 et 7.

Solution. Le nombre de personnes nées le premier janvier suit une loi binomiale $\mathcal{B}\left(n, \frac{1}{365}\right)$. Si $\mathcal{P}_\lambda(k) = \exp(-\lambda) \frac{\lambda^k}{k!}$, on rappelle que, pour tout $k \in \mathbb{N}^*$, on a la relation de récurrence :

$$\mathcal{P}_\lambda(k) = \mathcal{P}_\lambda(k-1) \frac{\lambda}{k}.$$

De même, on a

$$P_n(k+1) = P_n(k) \frac{n-k}{k+1} \frac{p}{1-p},$$

ce qui permet ici aussi un calcul par récurrence. Si $n = 500$, on a $np \equiv \lambda_1 = 1,36986$, et si $n = 600$, on a $np \equiv \lambda_2 = 1,64384$. Le tableau suivant permet de comparer la loi binomiale et son approximation :

k	$P_{500}(k)$	$\mathcal{P}_{\lambda_1}(k)$	$P_{600}(k)$	$\mathcal{P}_{\lambda_2}(k)$
0	0,2536	0,2541	0,1928	0,1932
1	0,3465	0,3481	0,3178	0,3177
2	0,2375	0,2384	0,2615	0,2611
3	0,1083	0,1088	0,1432	0,1431
4	0,0369	0,0372	0,0587	0,0588
5	0,0100	0,0102	0,0192	0,0193
6	0,00195	0,0023	0,0052	0,0053
7	0,0004	0,0004	0,0012	0,0012

On peut remarquer en passant que $\mathcal{P}_\lambda(k)$ est petit dès que $k \geq 5$. L'inégalité de Tchebichev, qui fournit une estimation supérieure de la dispersion d'une variable aléatoire autour de sa moyenne, nous donne, pour une variable aléatoire S suivant la loi $\mathcal{P}(\lambda)$ avec $\lambda = 1,36986$, compte tenu de ce que $ES = \sigma_S^2 = \lambda$, la majoration

$$P(S \geq 5) \leq P(S \leq ES + 2,6\sigma_S) \leq (2,6)^{-2} \leq 0,15,$$

environ 10 fois supérieure à la valeur de $P(S \geq 5)$. On voit pourquoi, malgré son utilité, l'inégalité de Tchebichev est qualifiée de grossière.

7.1.2. Approximation d'une loi binomiale par la loi de Gauss

Et quand p n'est pas trop petit, mais que n est grand, comment calculer la probabilité $P_n(k)$? La réponse est donnée par les théorèmes de **Moivre-Laplace**. Leur démonstration est hors programme. Pour le théorème local, elle s'appuie sur la formule de Stirling ($n! \sim n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}$ quand n tend vers l'infini), pour le théorème global, sur l'approximation d'une intégrale par des sommes de Riemann. Les lecteurs intéressés peuvent consulter par exemple le livre de Shiryaev, *Probability*.

Théorème 7.2. (Théorème de Moivre-Laplace local.) Avec les notations ci-dessus, on a, quand n tend vers l'infini,

$$P(S_n = k) \sim \frac{1}{\sqrt{2npq}} \exp\left(-\frac{(k - np)^2}{2npq}\right),$$

uniformément en k tel que $|k - np| = o\left[(npq)^{\frac{2}{3}}\right]$.

Théorème 7.3. (Théorème de Moivre-Laplace global.) ² Soit p tel que $0 < p < 1$. Soit S_n une variable aléatoire de loi $\mathcal{B}(n, p)$, et soit $\tilde{S}_n = \frac{S_n - E(S_n)}{\sigma(S_n)}$ la variable centrée réduite associée à S_n . Alors on a

$$\sup_{-\infty \leq a < b \leq +\infty} \left| P(a < \tilde{S}_n \leq b) - \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x^2}{2}\right) dx \right| \xrightarrow{n \rightarrow +\infty} 0.$$

La loi de \tilde{S}_n peut donc être approchée, quand n est grand, par la loi de Gauss centrée réduite $\mathcal{N}(0, 1)$. Un calcul immédiat permet d'en déduire une approximation de la loi de S_n elle-même.

2. Abraham De Moivre (1667-1754), mathématicien anglais d'origine française, précisa les principes du calcul des probabilités, la règle des probabilités composées. Il a aussi travaillé dans le domaine des équations aux différences finies et introduit les imaginaires en trigonométrie.

Corollaire 7.4. Si S_n est une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$, pour tous réels a, b tels que $a < b$, on a, en notant $q = 1 - p$:

$$\lim_n \left[P(a < S_n \leq b) - \frac{1}{\sqrt{2\pi}} \int_{\frac{a-np}{\sqrt{npq}}}^{\frac{b-np}{\sqrt{npq}}} \exp\left(-\frac{x^2}{2}\right) dx \right] = 0.$$

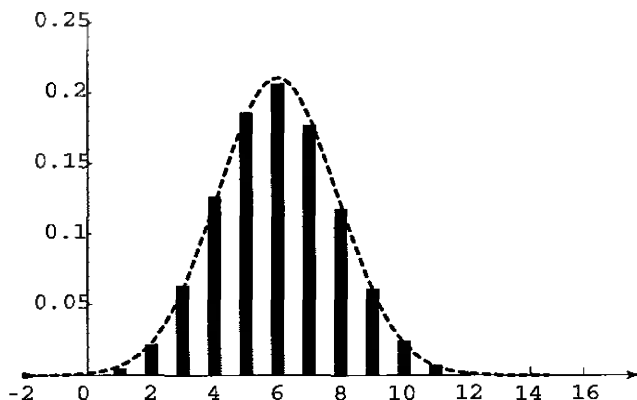


Figure 8. Comparaison de la loi binomiale $\mathcal{B}(15, 0,4)$ et de la loi de Gauss de même moyenne et même écart-type

Remarque. On peut démontrer (cas particulier d'un théorème de Berry-Esseen) que l'on a :

$$\sup_{x \in \mathbb{R}} |P(S_n \leq x) - \Phi(x)| \leq \frac{p^2 + q^2}{\sqrt{npq}},$$

où Φ est la fonction de répartition de la loi de Gauss $\mathcal{N}(0, 1)$. Cette approximation peut être mauvaise pour des valeurs voisines de 0 ou 1. Le théorème de Poisson est alors utilisable.

Rappelons que

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$$

et que si X suit la loi $\mathcal{N}(0, 1)$, on a

$$P(a < X \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x^2}{2}\right) dx = \Phi(b) - \Phi(a).$$

La fonction Φ est tabulée pour les $x > 0$ (si $x < 0$, on utilise le fait que : $\Phi(-x) = 1 - \Phi(x)$).

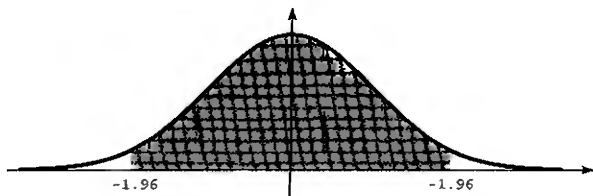


Figure 9. $P(-1,96 \leq X \leq 1,96) = 0,95$ si X suit la loi $\mathcal{N}(0, 1)$

Voici trois valeurs d'accroissements de la fonction Φ d'usage courant :

$$\begin{aligned} \Phi(1,64) - \Phi(-1,64) &= 0,9 & \Phi(1,96) - \Phi(-1,96) &= 0,95 \\ \Phi(3,09) - \Phi(-3,09) &= 0,99. \end{aligned}$$

Exemple 7.2. On jette 12000 fois un dé équilibré. On cherche la probabilité que le nombre de 6 obtenus soit compris entre 1800 et 2100.

Solution. Le nombre de 6 obtenus est une variable aléatoire de loi binomiale $\mathcal{B}\left(12000, \frac{1}{6}\right)$. Ici on a $a = 1800$, $b = 2100$, ce qui donne la valeur approximative de la probabilité cherchée :

$$\Phi\left(\frac{2100 - 2000}{\sqrt{12000 \frac{1}{6} \frac{5}{6}}}\right) - \Phi\left(\frac{1800 - 2000}{\sqrt{12000 \frac{1}{6} \frac{5}{6}}}\right) = \Phi(\sqrt{6}) - \Phi(-2\sqrt{6}) \simeq 0,992.$$

7.1.3. Approximation d'une loi hypergéométrique par une loi binomiale.

Rappelons sur un exemple la définition de la loi hypergéométrique. Un lac contient r poissons³, dont r_1 sont d'une espèce intéressante a . On pêche n poissons (on suppose que tous les poissons se laissent aussi facilement attraper). Soit X le nombre (aléatoire) de poissons de l'espèce a parmi les poissons attrapés. Alors on a

$$P(X = k) = \frac{\binom{r_1}{k} \binom{r - r_1}{n - k}}{\binom{r}{n}}$$

quand le second membre est défini, et $P(X = k) = 0$ sinon. On dit que X suit une loi hypergéométrique de paramètres n , r , r_1 .

3. Le lecteur pourra transformer cet exemple en un exemple traitant de sondages d'opinion.

Si on change de procédé, et qu'on rejette chaque poisson (vivant) après l'avoir pêché, le nombre Y de poissons pêchés de l'espèce a suit une loi binomiale $\mathcal{B}(n, \frac{r_1}{r})$: on a $P(Y = k) = \binom{n}{k} \left(\frac{r_1}{r}\right)^k \left(1 - \frac{r_1}{r}\right)^{n-k}$ si $0 \leq k \leq n$ et $P(Y = k) = 0$ sinon.

Il est naturel de penser que si r et r_1 sont grands devant n , il y a peu de différence entre les deux procédés. C'est ce que confirme l'énoncé suivant.

Théorème 7.5. *Considérons d'une part pour chaque $j \in \mathbb{N}$ une variable aléatoire X_j de loi hypergéométrique de paramètres n, r_j, r_j^1 , et d'autre part une variable aléatoire Y de loi binomiale $\mathcal{B}(n, p)$. Si*

$$\lim_{j \rightarrow +\infty} \frac{r_j^1}{r_j} = p,$$

où $p \in]0, 1[$, alors on a :

$$\lim_{j \rightarrow +\infty} P(X_j = k) = P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Démonstration. Dès que $k \leq r_j^1$ et $n - k \leq r_j - r_j^1$, on a, après simplification des coefficients binomiaux :

$$P(X_j = k) = \binom{n}{k} \prod_{l=0}^{k-1} \left(\frac{r_j^1 - l}{r_j - l} \right) \prod_{l=0}^{(n-k)-1} \left(\frac{r_j - r_j^1 - l}{r_j - l} \right),$$

ce qui assure facilement la convergence annoncée.

7.1.4. Théorème limite central

Les résultats concernant l'approximation de lois, ou plus précisément le comportement de certaines lois de probabilité quand un paramètre, réel ou entier, tend vers l'infini, sont souvent dénommés **théorèmes limite**. Le plus important de ces résultats a reçu le nom de **théorème limite central**. Nous en donnons ci-dessous une version élémentaire, sans démonstration, comme l'indique le programme, et nous montrons comment il s'applique dans la pratique.

Théorème 7.6. (Théorème limite central.) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires non constantes, indépendantes, de même loi, admettant un moment d'ordre deux, de moyenne EX_1 et d'écart-type σ_{X_1} . On définit, pour tout $n \in \mathbb{N}$, la variable aléatoire S_n par :

$$S_n = \frac{\sum_{j=1}^n X_j - n EX_1}{\sqrt{n} \sigma_{X_1}}$$

(S_n est la variable centrée réduite associée à $\sum_{j=1}^n X_j$). Alors, pour tout couple $(a, b) \in \overline{\mathbb{R}}^2$ tels que $a < b$, on a :

$$P(a < S_n \leq b) \xrightarrow[n \rightarrow +\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x^2}{2}\right) dx.$$

Remarque. Le théorème de Moivre-Laplace (global) est un cas particulier de ce théorème : en effet une variable aléatoire S_n de loi binomiale $\mathcal{B}(n, p)$ a même loi qu'une somme $X_1 + \dots + X_n$, où X_1, \dots, X_n sont des variables aléatoires de Bernoulli indépendantes de paramètre p .

Application. Si une variable aléatoire X s'écrit $\sum_{j=1}^n X_j$ où les variables aléatoires X_j vérifient les hypothèses du théorème, on pourra approximer, dans le cas où n est grand, la loi de X par la loi de Gauss $\mathcal{N}(EX, \sigma_X^2)$.

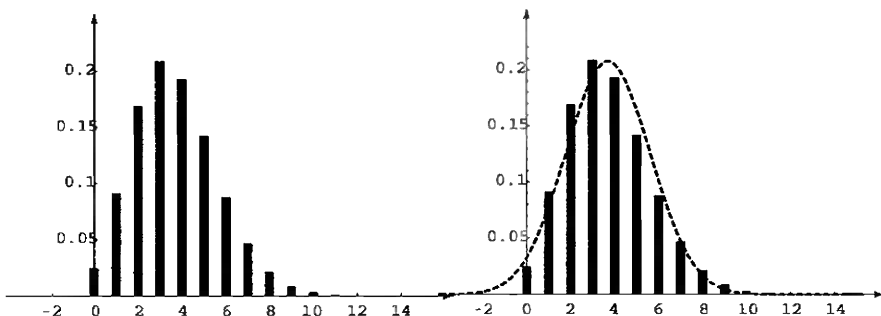


Figure 10. Comparaison de loi de Poisson $\mathcal{P}(\lambda)$ ($\lambda = 3, 7$) et de la loi de Gauss de même moyenne et de même variance

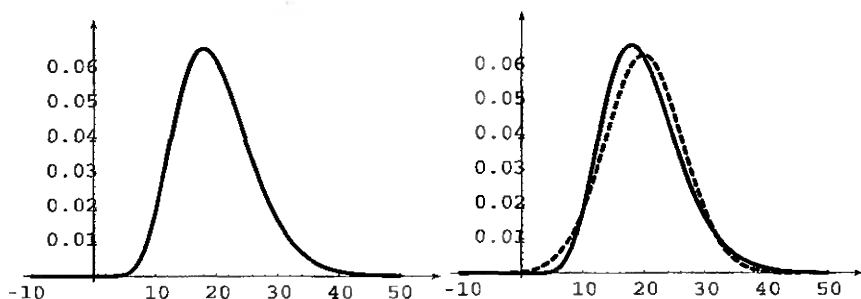


Figure 11. Comparaison de loi du chi-deux à 20 degrés de liberté et de la loi de Gauss de même moyenne et de même variance

Outre le cas des variables aléatoires de loi $\mathcal{B}(n, p)$, c'est aussi le cas des variables aléatoires de loi de Poisson $\mathcal{P}(\lambda)$ (même loi que la somme de n variables indépendantes de loi $\mathcal{P}(\lambda/n)$ – voir exemple 3.5) et des variables aléatoires de loi χ_n^2 (même loi que la somme de n variables aléatoires indépendantes de loi χ_1^2 – voir la définition).

Les approximations figurant dans le tableau suivant sont ainsi valables quand n et λ sont grands :

P_X	P_{X_j}	EX	σ_X^2
$\mathcal{B}(n, p)$	$\mathcal{B}(1, p)$	np	npq
$\mathcal{P}(\lambda)$	$\mathcal{P}\left(\frac{\lambda}{n}\right)$	λ	λ
χ_n^2	χ_1^2	n	$2n$

7.1.5. Convergence en loi

Nous terminons ce paragraphe, à la seule attention des candidats au concours interne de l'Agrégation, en donnant une définition de la convergence en loi.

Définition 7.7. Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles est dite **converger en loi** vers une variable aléatoire X si, en tout point x où la fonction de répartition F_X de la variable aléatoire X est continue, on a :

$$F_{X_n}(x) \longrightarrow F_X(x).$$

Remarques :

- On ne peut pas exiger la convergence en **tout** point : en effet, si $X_n = \frac{1}{n}$, il est naturel de dire que la suite $(X_n)_{n \in \mathbb{N}^*}$ **converge**

en loi vers la variable aléatoire $X = 0$. Or on a : $F_{X_n}(0) = P(X_n \leq 0) = 0$ et $F_X(0) = P(X \leq 0) = 1$.

- Il se peut que la suite des fonctions de répartition F_{X_n} converge ponctuellement vers une fonction qui ne soit pas une fonction de répartition. Voici un exemple d'une telle situation ; si $X_n = n$, on a :

$$F_{X_n}(x) = \begin{cases} 0 & \text{si } x < n \\ 1 & \text{si } x \geq n, \end{cases}$$

et donc, pour tout x fixé, $\lim_n F_{X_n}(x) = 0$.

- Les variables aléatoires X_n et la variable aléatoire X n'ont pas besoin d'être définies sur le même espace probabilisé ; il s'agit en fait d'une notion de convergence de la suite des lois de ces variables aléatoires.
- La conclusion du **théorème limite central** énoncé ci-dessus peut se reformuler ainsi : la suite de variables aléatoires S_n converge en loi vers la loi de Gauss $\mathcal{N}(0, 1)$.

7.2. Loi faible des grands nombres

L'étude de phénomènes aléatoires conduit souvent à étudier la suite des **moyennes arithmétiques** d'une suite de variables aléatoires **indépendantes de même loi** ; c'est en particulier le cas en statistique, lorsque, par exemple, on veut estimer un paramètre d'une loi de variable aléatoire liée à un phénomène, au vu d'une suite de réalisations de ce phénomène obtenues au cours d'expériences indépendantes. L'étude de la convergence de telles suites fait l'objet de résultats connus sous le nom de **lois des grands nombres**, lois **faibles** pour la notion de **convergence en probabilité** que nous allons définir ci-après, lois **fortes** qui sont relatives à la notion de **convergence presque sûre**, notion⁴ qui n'est pas au programme.

Définition 7.8. La suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ définies sur un espace probabilisé (Ω, \mathcal{A}, P) est dite **converger en probabilité** vers la variable aléatoire X si, pour tout $\varepsilon > 0$,

$$\lim_n P(|X_n - X| > \varepsilon) = 0.$$

Théorème 7.9. (Loi faible des grands nombres.) Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) ,

4. La notion de **convergence presque sûre** sera étudiée au deuxième tome.

indépendantes et admettant un moment d'ordre deux. On suppose la convergence des suites :

$$\boxed{\frac{1}{n} \sum_{j=1}^n EX_j \xrightarrow{n \rightarrow +\infty} m \quad \text{et} \quad \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2 \xrightarrow{n \rightarrow +\infty} 0.}$$

Alors, la suite des variables aléatoires $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ converge en probabilité vers m .

Démonstration. On a

$$E\bar{X}_n = \frac{1}{n} \sum_{j=1}^n EX_j.$$

Les variables aléatoires X_n étant indépendantes, on a aussi :

$$\sigma_{\bar{X}_n}^2 = \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2.$$

L'inégalité triangulaire conduit à l'inégalité :

$$|\bar{X}_n - m| \leq \left| \bar{X}_n - \frac{1}{n} \sum_{j=1}^n EX_j \right| + \left| \frac{1}{n} \sum_{j=1}^n EX_j - m \right|.$$

Mais, pour tout $\varepsilon > 0$, il existe $N(\varepsilon) \in \mathbb{N}^*$ tel que, pour tout $n \geq N(\varepsilon)$, on ait $\left| \frac{1}{n} \sum_{j=1}^n EX_j - m \right| \leq \frac{\varepsilon}{2}$. Pour tout $n \geq N(\varepsilon)$, on a donc l'inclusion des ensembles

$$\left(\left| \bar{X}_n - \frac{1}{n} \sum_{j=1}^n EX_j \right| \leq \frac{\varepsilon}{2} \right) \subset (|\bar{X}_n - m| \leq \varepsilon),$$

ou encore, pour les complémentaires, l'inclusion :

$$(|\bar{X}_n - m| > \varepsilon) \subset \left(\left| \bar{X}_n - \frac{1}{n} \sum_{j=1}^n EX_j \right| > \frac{\varepsilon}{2} \right).$$

L'inégalité de Bienaymé-Tchebichev permet d'écrire :

$$P \left(\left| \bar{X}_n - \frac{1}{n} \sum_{j=1}^n EX_j \right| > \frac{\varepsilon}{2} \right) \leq \frac{4}{\varepsilon^2} \sigma_{\bar{X}_n}^2 = \frac{4}{\varepsilon^2} \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2.$$

Il en résulte que, pour tout $n \geq N(\varepsilon)$, on a :

$$P(|\bar{X}_n - m| > \varepsilon) \leq \frac{4}{\varepsilon^2} \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2,$$

ce qui, en utilisant la seconde hypothèse, démontre le résultat.

Remarque. En particulier, les hypothèses du théorème précédent sont toutes satisfaites si les variables aléatoires X_n sont **indépendantes et de même loi** et si X_1 admet un moment d'ordre deux.

On étudie un cas particulier du théorème précédent (il lui est toutefois historiquement antérieur).

Théorème 7.10. (Théorème de Bernoulli.) Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements indépendants de même probabilité p . La suite des variables aléatoires $\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{A_j}$ converge en probabilité vers p .

Démonstration. Les variables aléatoires $\mathbf{1}_{A_n}$, $n \in \mathbb{N}^*$, sont indépendantes de même loi de **Bernoulli**; elles admettent un moment d'ordre deux, et on se trouve dans la situation de la remarque.

Ce théorème assure que, en étudiant un certain phénomène aléatoire, si on fait une suite d'expériences indépendantes, la suite des **fréquences relatives** d'apparition d'une certaine propriété liée à ce phénomène converge en probabilité (au sens de la probabilité P du modèle adopté) vers la probabilité de l'événement lié à cette propriété. C'est donc un théorème de cohérence du modèle probabiliste pour les adeptes de l'approche « fréquentiste » d'introduction de la notion de probabilité d'un événement, point de vue que nous n'avons pas spécialement adopté...!

Exercices

Exercice 7.1. Théorème limite central et loi de Poisson. Démontrer, en utilisant le théorème limite central, que :

$$\lim_n \exp(-n) \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}.$$

Solution. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi de Poisson de paramètre 1 et soit S_n la variable aléatoire $S_n = \sum_{j=1}^n X_j$.

La loi de S_n est la loi de Poisson de paramètre n ; on a donc :

$$P(S_n \leq n) = \exp(-n) \sum_{k=0}^n \frac{n^k}{k!}.$$

Mais, d'après le théorème de limite central, on a, avec les notations du théorème :

$$P(S_n \leq n) = P\left(\frac{S_n - n}{\sqrt{n}} \leq 0\right) \xrightarrow{n \rightarrow +\infty} \Phi(0) = \frac{1}{2},$$

ce qui démontre le résultat.

Exercice 7.2. Comparaison numérique : théorème limite central et inégalité de Tchebichev. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi admettant un moment d'ordre deux. On définit, pour tout $n \in \mathbb{N}^*$, la variable aléatoire S_n :

$$S_n = \frac{\sum_{j=1}^n X_j - n EX_1}{\sqrt{n} \sigma_{X_1}}.$$

Comparer, pour $k = 1, 2, 3$, les renseignements numériques fournis sur la probabilité $P(|S_n| \geq k)$ par le théorème limite central, puis par l'inégalité de Tchebichev.

Solution. Les variables aléatoires $\overset{\circ}{X}_j = \frac{X_j - EX_1}{\sigma_{X_1}}$ sont centrées réduites et on a :

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n \overset{\circ}{X}_j.$$

Les variables aléatoires $\overset{\circ}{X}_j$ étant indépendantes de même loi, on a :

$$\sigma_{S_n}^2 = \frac{1}{n} (n \sigma_{\overset{\circ}{X}_1}^2) = 1.$$

L'inégalité de Tchebichev donne donc, pour tout $\varepsilon > 0$:

$$P(|S_n| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}.$$

– pour $\varepsilon = 1$ l'inégalité de Tchebichev ne donne rien, alors que le théorème limite central dit que :

$$P(|S_n| \geq 1) \xrightarrow{n \rightarrow +\infty} 2(1 - \Phi(1)) = 0,31;$$

– pour $\varepsilon = 2$, l'inégalité de Tchebichev donne

$$P(|S_n| \geq 2) \leq 0,25,$$

alors que le théorème limite central dit que :

$$P(|S_n| \geq 2) \xrightarrow{n \rightarrow +\infty} 2(1 - \Phi(2)) = 0,0455.$$

– pour $\varepsilon = 3$, l'inégalité de Tchebichev donne

$$P(|S_n| \geq 3) \leq 0,1111,$$

alors que le théorème limite central dit que :

$$P (|S_n| \geq 3) \xrightarrow{n \rightarrow +\infty} 2(1 - \Phi(3)) = 0,0025.$$

Exercice 7.3. Taille d'échantillon et théorème limite central. Soit X une variable aléatoire de moyenne inconnue m et d'écart-type σ . Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi que X (on dit que (X_1, X_2, \dots, X_n) est un échantillon de taille n de X) ; on estime la moyenne m par la variable aléatoire $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$.

1. Justifier l'approximation, pour $\alpha > 0$ et n grand :

$$P (|\bar{X}_n - m| \geq \alpha) \simeq 1 - \left[\Phi \left(\frac{\sqrt{n}}{\sigma} \alpha \right) - \Phi \left(-\frac{\sqrt{n}}{\sigma} \alpha \right) \right]. \quad (7.1)$$

2. Quelle est la taille minimum de l'échantillon pour que, $\alpha > 0$ et $\beta \in]0, 1[$ étant fixés, on ait :

$$P (|\bar{X}_n - m| \geq \alpha) \leq \beta.$$

On traitera l'application numérique : $\sigma = 3$, $\alpha = 0,05\sigma$ et $\beta = 0,05$. On rappelle que : $\Phi^{-1}(0,975) = 1,96$.

Solution.

1. On pose

$$S_n = \frac{\sum_{j=1}^n X_j - nm}{\sqrt{n}\sigma} \equiv \sqrt{n} \frac{\bar{X}_n - m}{\sigma}.$$

On a alors l'égalité :

$$(|\bar{X}_n - m| \geq \alpha) = \left(|S_n| \geq \frac{\sqrt{n}}{\sigma} \alpha \right).$$

Mais, d'après le théorème limite central, si $a > 0$, on a :

$$P (|S_n| \geq a) \xrightarrow{n \rightarrow +\infty} 1 - [\Phi(a) - \Phi(-a)],$$

ce qui donne l'approximation (7.1) pour n grand.

2. Puisque l'on a :

$$1 - [\Phi(a) - \Phi(-a)] = 2 [1 - \Phi(a)],$$

on prend le plus petit entier n tel que

$$2 \left[1 - \Phi \left(\frac{\sqrt{n}}{\sigma} \alpha \right) \right] \leq \beta,$$

soit le plus petit entier tel que

$$n \geq \frac{\sigma^2}{\alpha^2} \left[\Phi^{-1} \left(1 - \frac{\beta}{2} \right) \right]^2.$$

Application numérique : la taille minimum n est le plus petit entier tel que

$$n \geq \frac{10^4}{25} [\Phi^{-1}(0,975)]^2,$$

soit :

$$n = 1537.$$

Exercice 7.4. Sondages et théorème de Moivre-Laplace. On veut estimer le pourcentage p de réponses positives à un référendum. Pour cela, on effectue un sondage sur n personnes et on estime p par la fréquence relative F_n de oui sur les personnes sondées. On cherche n_0 , plus petit entier n tel que la probabilité que F_n ne diffère de p de plus de $\alpha > 0$ soit inférieure à $\beta \in]0, 1[$.

Application numérique : on choisit $\beta = 0,05$ et $\alpha = 0,01$. On étudiera les deux cas suivants : i) on sait que $0 < p < 0,3$; ii) p est totalement inconnu. Par combien est divisée la taille de la population sondée si on choisit $\alpha = 0,05$?

Solution. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi de **Bernoulli** $\mathcal{B}(1, p)$; on estime p par la variable aléatoire

$$F_n = \frac{1}{n} \sum_{j=1}^n X_j. \text{ On est dans un cas particulier de l'exercice 7.3 auquel on}$$

réfère ; on a ici :

$$EX_1 = p \quad \text{et} \quad \sigma_{X_1} = \sqrt{pq},$$

où on pose $q = 1 - p$. Si on pose

$$Y_n = \frac{\sum_{j=1}^n X_j - np}{\sqrt{npq}} \equiv \sqrt{n} \frac{F_n - p}{\sqrt{pq}},$$

on a :

$$(|F_n - p| \geq \alpha) = \left(|Y_n| \geq \frac{\sqrt{n}}{\sigma_{X_1}} \alpha \right).$$

En appliquant le théorème limite central (cf la première question de l'exercice 7.3), n_0 est le plus petit n tel que :

$$n \geq \frac{pq}{\alpha^2} \left[\Phi^{-1} \left(1 - \frac{\beta}{2} \right) \right]^2. \quad (7.2)$$

Application numérique : si on choisit $\beta = 0,05$ et $\alpha = 0,01$, il vient :

$$n_0 = \left[\frac{pq}{10^{-2}} [\Phi^{-1}(0,975)]^2 \right] + 1,$$

soit :

$$n_0 = (196)^2 pq + 1.$$

Il en résulte que :

– si $p < 0,3$, puisque $\max_{p \in]0, 0,3[} p(1-p) = 0,3 \times 0,7$, on a :

$$n_0 = 8068.$$

– si p est inconnu, puisque $\max_{p \in]0,1[} p(1-p) = \frac{1}{4}$, on a :

$$n_0 = 9604.$$

Enfin, si on choisit $\beta = 0,05$ et $\alpha = 0,05$, il résulte de l'inégalité (7.2) que n_0 est divisé par 25, ce qui donne respectivement pour les valeurs de taille minimum : 323 et 384.

Autrement dit, dans le cas où p est inconnu (ou proche de 50%), si on veut que la probabilité d'une erreur $\geq 1\%$ soit inférieure à $1/20$, il faut interroger un échantillon de 10 000 personnes. Si on est moins exigeant, et qu'on accepte que la probabilité d'une erreur $\geq 5\%$ soit de l'ordre de $1/20$, on peut se contenter d'un échantillon 25 fois plus petit.

Exercice 7.5. Jeu de la roulette et théorème limite central. La probabilité de gagner une partie au jeu de la roulette est de $\frac{19}{37}$ (on se place du point de vue du casino) et la mise d'un joueur à une partie est d'un écu. Quel est le nombre minimum n_0 de parties qui doivent être jouées journalièrement pour que le casino gagne avec probabilité 0,5 au moins 1000 écus par jour ? Quelle est la probabilité d'une perte globale pour le casino durant ces n_0 parties ?

Solution. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi déterminée par

$$P(X_1 = 1) = \frac{18}{37} \quad \text{et} \quad P(X_1 = -1) = \frac{19}{37}.$$

La variable aléatoire X_n représentant le gain algébrique du casino à la n -ième partie, le gain algébrique du casino au cours des n premières parties est

$$G_n = \sum_{j=1}^n X_j. \quad \text{On a :}$$

$$EX_1 = -\frac{1}{37},$$

et puisque $EX_1^2 = 1$:

$$\sigma_{X_1}^2 = 1 - \left(\frac{1}{37}\right)^2.$$

Pour n grand, la variable aléatoire $\frac{G_n - \frac{n}{37}}{\sqrt{n}\sigma_{X_1}}$ suit approximativement une loi de Gauss centrée réduite, et puisque l'on a

$$(G_n \geq 1000) = \left(\frac{G_n - \frac{n}{37}}{\sqrt{n} \sigma_{X_1}} \geq \frac{1000 - \frac{n}{37}}{\sqrt{n} \sigma_{X_1}} \right)$$

Le nombre n_0 cherché est le plus petit nombre n de parties qui doivent être jouées journalièrement pour que l'on ait :

$$P(G_n \geq 1000) \equiv P \left(\frac{G_n - \frac{n}{37}}{\sqrt{n} \sigma_{X_1}} \geq \frac{1000 - \frac{n}{37}}{\sqrt{n} \sigma_{X_1}} \right) \geq \frac{1}{2},$$

c'est-à-dire le plus petit n tel que :

$$1000 - \frac{n}{37} \leq 0,$$

soit :

$$\boxed{n_0 = 37000.}$$

La probabilité d'une perte globale pour le casino durant ces n_0 parties est alors

$$P(G_{n_0} < 0) \equiv P \left(\frac{G_{n_0} - \frac{n_0}{37}}{\sqrt{n_0} \sigma_{X_1}} < \frac{-\frac{n_0}{37}}{\sqrt{n_0} \sigma_{X_1}} \right),$$

et, en approximant σ_{X_1} par 1 dans le quotient de droite :

$$P(G_{n_0} < 0) \simeq P \left(\frac{G_{n_0} - \frac{n_0}{37}}{\sqrt{n_0} \sigma_{X_1}} < -5,19 \right).$$

La variable aléatoire $\frac{G_{n_0} - \frac{n_0}{37}}{\sqrt{n_0} \sigma_{X_1}}$ suivant approximativement une loi de Gauss centrée réduite, on a :

$$\boxed{P(G_{n_0} < 0) \simeq 0.}$$

Tableau des lois de probabilité usuelles

Lois discrètes

Nom paramètres	$P(X = k)$	m	σ^2
Loi de Bernoulli $\mathcal{B}(1, p)$ ($0 < p < 1$).	$P(X = 1) = p,$ $P(X = 0) = q$	p	pq
<p>Une épreuve de Bernoulli est une expérience aléatoire avec deux résultats possibles, souvent appelés conventionnellement <i>succès</i> (probabilité p) et <i>échec</i> (probabilité $q = 1 - p$). La loi de Bernoulli $\mathcal{B}(1, p)$ est la loi de la variable aléatoire associée à une épreuve de Bernoulli qui prend la valeur 1 en cas de succès, 0 en cas d'échec. Cf. pp. 68, 117, 131.</p>			
Loi binomiale $\mathcal{B}(n, p)$ ($0 < p < 1, n$ entier > 0).	$P(X = k) = \binom{n}{k} p^k q^{n-k}$	np	npq
<p>Loi du nombre de succès dans une suite de n épreuves de Bernoulli indépendantes de paramètre p. Cf. pp. 17, 70, 120, 131, 139, 219, 222.</p>			
Loi hypergéométrique de paramètres n, r, r_1 .	$P(X = k) = \frac{\binom{r_1}{k} \binom{r-r_1}{n-k}}{\binom{r}{n}}$	$\frac{nr_1}{r}$	(1)
<p>Si d'une urne contenant r boules, dont r_1 boules rouges et $r - r_1$ boules blanches, on tire n boules, la loi du nombre de boules rouges obtenu est la loi hypergéométrique de paramètres n, r, r_1. (Si chaque boule est remplacée dans l'urne après avoir été tirée, on obtient à la place de la loi hypergéométrique une loi binomiale $\mathcal{B}(n, r_1/r)$.) La loi hypergéométrique intervient dans la modélisation des sondages et du contrôle de qualité. Cf. pp. 17, 224.</p>			
Loi de Poisson $\mathcal{P}(\lambda)$ ($\lambda > 0$)	$P(X = n) = \frac{\lambda^n}{n!} \exp(-\lambda)$	λ	λ
<p>Intervient dans la modélisation de phénomènes de comptage. Le nombre de désintégrations atomiques dans un corps radioactif pendant un intervalle de temps, le nombre d'arrivées à un guichet, le nombre de gouttes de pluie sur une surface donnée, suivent une loi de Poisson. La loi de Poisson s'obtient comme limite de lois binomiales. Cf. pp. 17, 64, 121, 132, 139, 219, 84.</p>			
Loi géométrique sur \mathbb{N} , $\mathcal{G}_{\mathbb{N}}(p)$ ($0 < p < 1$)	$P(X = n) = pq^n$	$\frac{q}{p}$	$\frac{q}{p^2}$
<p>Loi du nombre d'échecs rencontrés avant d'obtenir un succès, dans la répétition d'épreuves de Bernoulli indépendantes de paramètre p. Cf. pp. 16, 68, 80, 121, 131, 139.</p>			

Loi géométrique sur \mathbb{N}^* , $\mathcal{G}_N(p)$ ($0 < p < 1$)	$P(X = n) = pq^{n-1}$	$\frac{1}{p}$	$\frac{q}{p^2}$
Loi du nombre d'épreuves nécessaires pour obtenir un succès (en comptant celui-ci), dans la répétition d'épreuves de Bernoulli indépendantes de paramètre p . Cf. pp. 16, 68, 121, 131, 139.			
Loi binomiale négative $\mathcal{B}^-(r, p)$ ($0 < p < 1$, r entier > 0).	$P(X = k) = \binom{r+k-1}{k} p^r q^k$	$r \frac{q}{p}$	$r \frac{q}{p^2}$
Loi du nombre d'échecs rencontrés avant d'obtenir r succès, dans la répétition d'épreuves de Bernoulli indépendantes de paramètre p . Cf. pp. 78, 104, 140.			

⁽¹⁾ La variance de la loi hypergéométrique de paramètres n, r, r_1 est

$$\sigma^2 = n \frac{r-n}{r-1} \frac{r_1}{r} \left(1 - \frac{r_1}{r}\right).$$

Lois à densité

Nom paramètres	densité $f(x)$ (x réel)	m	σ^2
Loi uniforme sur $[a, b]$ ($a < b$)	$\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Cf. pp. 176, 214, 216.			
Loi de Gauss $\mathcal{N}(m, \sigma^2)$ (m réel, $\sigma^2 > 0$).	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{\sigma^2}\right)$	m	σ^2
Beaucoup de mesures suivent approximativement une loi normale (loi des erreurs). Le théorème limite central est la principale raison de l'intervention universelle de la loi normale dans les phénomènes naturels et autres. La loi normale est souvent employée comme modèle a priori, ou pour approcher des lois connues de même moyenne et de même variance. Cf. pp. 177, 188, 189, 222, 225.			
Loi de Cauchy.	$\frac{1}{\pi} \frac{1}{1+x^2}$	non défini	non défini
Cf. pp. 177, 217.			
Loi exponentielle $\exp(p)$ ($\lambda > 0$)	$\mathbf{1}_{\mathbb{R}^+}(x) p \exp(-px)$	$\frac{1}{p}$	$\frac{1}{p^2}$
Intervient dans la modélisation des temps d'attente et des durées de vie (atomes radioactifs, par exemple). Similaire aux lois géométriques. Cf. pp. 176, 200, 211, 217.			
Loi du chi-deux χ_n^2 (n entier > 0)	$\mathbf{1}_{\mathbb{R}^+}(x) \frac{1}{K_n} \exp\left(-\frac{x}{2}\right) x^{\frac{n}{2}-1}$	n	$2n$
Loi de $X_1^2 + \dots + X_n^2$ quand X_1, \dots, X_n sont des variables de loi $\mathcal{N}(0, 1)$ indépendantes. Utilisée en statistique dans le test d'adéquation d'une distribution empirique à une loi de probabilité donnée (test du chi-deux). Cf. pp. 178, 217.			

Bibliographie

- [1] Bouleau N. (1986), *Variables aléatoires et simulation*. Hermann, Paris.
- [2] Bouleau N. (1988), *Processus stochastiques et applications*. Hermann, Paris.
- [3] Brémaud P. (1984), *Introduction aux probabilités*, Springer-Verlag, Berlin, Heidelberg, New-York.
- [4] Cacoullos T. (1989), *Exercises in Probability*, Springer-Verlag, Berlin, Heidelberg, New-York.
- [5] Dacunha-Castelle D., Duflo M., *Probabilités et Statistiques*, tomes 1 (1982) et 2 (1983). Masson, Paris.
- [6] Dacunha-Castelle D., Duflo M., *Probabilités et Statistiques*, (exercices), tomes 1 et 2, Masson, Paris.
- [7] Durrett R. (1991), *Probability theory and examples*, Wadsworth & Brooks/Cole, Pacific Grove, California.
- [8] Gramain A. (1988), *Intégration*, Hermann, Paris.
- [9] Hammad P. (1984), *Cours de Probabilité*, Cujas, Paris.
- [10] Hammad P., Taranco A. (1984), *Exercices de probabilité*, Cujas, Paris.
- [11] Krickeberg K., Ziezold H. (1980). *Méthodes stochastiques. Introduction aux probabilités et à la statistique*, DIA/Belin, Paris.
- [12] Métivier M. (1972), *Notions fondamentales de la théorie des probabilités*, Dunod, Paris.
- [13] Ross S. (1983), *Stochastic processes*, Wiley, New-York, Chichester.
- [14] Shiryaev A. (1984, Second Edition 1996), *Probability*, Springer, Berlin, Heidelberg, New-York.
- [15] Stoyanov J. (1989), *Counterexamples in probability*, Wiley, New-York, Chichester.
- [16] Stromberg K. (1994), *Probability for Analysts*, Chapman & Hall, New-York, London.

Index

A

- algèbre des événements, 8
- anniversaires, 22
- approximation (de loi), 219
 - binomiale par loi de Gauss, 222
 - binomiale par loi de Poisson, 221
 - hypergéométrique par loi binomiale, 224
- attente (temps d'), 200

B

- Bayes
 - théorème de, 93, 99
- Bernoulli (loi de)
 - variance, 131
- Bernoulli (théorème de), 230
- Bernoulli (variable aléatoire de), 68
- Bernstein (polynômes de), 160
- Berry-Esseen (théorème de), 223
- binomial (modèle), 17
- binomiale (loi)
 - moyenne, 120
 - variance, 131
- binomiale (variable aléatoire)
 - marginale de v.a. trinomiale, 168
- binomiale négative (loi), 78, 80, 104, 140
- borélien, 175

C

- Cauchy (loi de), 177
- central (théorème limite), 226
- centrée (variable aléatoire), 127, 190
- centrée réduite
 - variable aléatoire discrète, 127
- centrée réduite (variable aléatoire), 190
- chi-deux (loi du), 178
- coefficient de corrélation, 134, 191
- complet (système de constituants), 90
- conditionnée (probabilité), 87
- conditionnelle (densité), 200, 201
- conditionnelle (loi), 94
- conditionnelle (probabilité), 88
- conditionnelles (probabilités)
 - en cascade, 89, 103, 104

- conjugués, 22
- convergence
 - en loi
 - en probabilité, 228
- convolution, 104, 214
 - de densités, 199
 - de deux lois binomiales, 70
 - de deux lois de Poisson, 65
 - de deux probabilités
- corrélation (coefficient de)
 - variables aléatoires à densité, 191
 - variables aléatoires discrètes, 134
- covariance
 - variables aléatoires à densité, 191, 218
 - variables aléatoires discrètes, 128

D

- décimales (loi uniforme), 213
- dénombrable, 13
- densité
 - d'une probabilité, 175
 - d'une variable aléatoire, 180
- densité conditionnelle, 200-202
- déterministe (phénomène), 4
- discret (espace probabilisable), 14
- discrète (variable aléatoire), 19
- droite achevée, 29
- droite de régression, 137

E

- écart-type
 - variable aléatoire discrète, 127
- échantillon, 232
- espace fondamental, 6
- espace probabilisable, 9
 - discret, 14
- espace probabilisé, 10
- espérance, voir moyenne
- Euler (indicateur d'), 73
- événement, 6
- éventualités, 6
- expérience aléatoire, 3, 6
- exponentielle (loi), 176
 - temps d'arrivée, 211

F

- famille sommable
 - de réels, 38, 40
 - de réels positifs, 29, 30, 34
- fonction
 - d'une v.a. à densité, 184, 186, 214
 - d'une v.a. discrète, 60, 118
 - de plusieurs v.a. discrètes, 60
 - de répartition, 151, 180, 214, 223
 - génératrice, 138-140, 142, 162, 165, 168
 - indicatrice, 67, 117
- fonction additive d'ensembles, 36
- formule
 - de Bayes, 93, 99
 - de Poincaré, 12, 24, 56
 - des probabilités totales, 90, 91, 99, 107
- Fubini, voir théorème de Fubini

G

- génératrice (fonction), 138-140, 142, 162, 165, 168
 - loi binomiale négative, 140
 - lois classiques, 139
- géométrique (loi)
 - moyenne, 121
 - variance, 131
- géométrique (modèle), 16
- germe (de probabilité), 15

H

- Hölder (inégalité de)
 - familles sommables, 123
- hypergéométrique (modèle), 17

I

- identités de Wald, 143, 144, 168
- indépendance
 - conditionnelle, 109
 - d'une famille d'événements, 54
 - d'une famille de variables aléatoires, 58
 - d'une famille de variables aléatoires discrètes, 59
 - de deux événements, 54
 - de deux variables aléatoires, 52, 57, 196

inégalité

- de Hölder, 123
- de Markov, 133
- de Minkowski, 123, 124
- de Schwarz, 124, 126
- de Tchebichev, 133, 134, 160, 162, 190, 221, 229, 231
- inverse (image), 19
- issue, 6

L

- Laplace-Gauss (loi de)
 - à deux dimensions, 192
 - à une dimension, 177
- Leibniz (formule de), 128
- loi
 - binomiale, 17, 70, 84, 94, 120, 131, 219, 222, 224
 - binomiale négative, 78, 80, 104, 140
 - conditionnelle, 94
 - d'une fonction de v.a., 184, 214
 - d'une somme, 61, 63, 64, 199, 200
 - de Bernoulli, 68, 85, 117, 145, 160, 230, 233
 - de Cauchy, 177
 - de Laplace-Gauss à deux dimensions, 192
 - de maximum, 80, 150, 151
 - de Poisson, 17, 64, 74, 84, 121, 168, 220, 230
 - du chi-deux, 178
 - exponentielle, 176
 - géométrique, 16, 48, 68, 80, 109, 212
 - hypergéométrique, 18, 224
 - normale, 177
 - normale centrée réduite, 179
 - triangulaire, 63, 210, 214
 - uniforme
 - sur un disque, 179
 - sur un ensemble fini, 15
 - sur un intervalle, 176
 - sur un rectangle, 179
- loi de probabilité, 20
- loi faible des grands nombres, 132, 228
- loterie, 23

M

- Méré (chevalier de), 91
- marginale, 183
 - loi d'une, 183, 217
- Markov (inégalité de)
 - variable aléatoire discrète, 133
- markovienne (suite), 98, 106
- Minkowski (inégalité de)
 - familles sommables, 123, 124
- moindres carrés, 136, 137
- Moivre-Laplace (théorème de), 222
- moments, 115, 122, 124, 125, 140, 146, 150, 165, 193, 216
- moyenne
 - v.a. à densité, 185, 186, 188, 193, 216
 - v.a. discrète, 115-118, 120, 121, 145, 150

N

- normale (loi)
 - à deux dimensions, 179
 - réelle, 177
- normale (variable aléatoire de loi)
 - moyenne, 188
 - variance, 189

P

- Pascal, 91
- pile ou face, 27, 99, 110, 145
- Poincaré (formule de), 12, 24, 56
- Poisson (loi de)
 - caractérisation, 168
 - moyenne, 121
 - variance, 132
- Poisson (modèle de), 17
- polynômes de Bernstein, 160
- probabilisable (espace), 9
- probabilisé (espace), 10
- probabilité, 10
- probabilité conditionnelle, 88, 89, 103, 104, 211
- probabilité produit, 66, 98

R

- réalisation, 6
- régression linéaire
 - variables aléatoires discrètes, 136
- roulette, 234

- répartition
 - continue, 177
 - discrète, 115
 - marginale, 183
 - multidimensionnelle, 179
 - normale, 177
 - uniforme, 16
- réels positifs, 22
- somme d'une famille sommable, 123
- sommation
 - réels positifs, 22
- somme d'une famille sommable, 123
- sondage, 233
- suite markovienne, 98, 106
- symétrique (loi), 63
- système complet de constituants, 90, 91, 93, 169

T

- Tchebichev (inégalité de)
 - variable aléatoire à densité, 190, 221, 229, 231
 - variable aléatoire discrète, 133, 134, 160, 162
- temps d'arrivée, 211
- théorème
 - d'approximation de Weierstrass, 161
 - de Bayes, 93, 99
 - de Bernoulli, 230
 - de Fubini (familles sommables), 44
 - de Moivre-Laplace, 222, 226, 233
 - de Poisson, 219, 221, 223
 - de transfert, 118, 146, 185, 186
 - limite central, 226, 228, 230-232, 234
- totales (formule des probabilités), 90, 107
- transfert (théorème de)
 - cas à densité, 185, 186
 - cas discret, 118, 146
- tribu
 - borélienne, 175
 - des événements, 175
- tribu, 8
- uniforme (probabilité sur un ensemble fini), 16
- uniforme (loi)
 - sur un disque, 179
 - sur un rectangle, 179

uniforme (probabilité)
sur un intervalle, 176
univers, 6

V

variable aléatoire, 18, 19
à densité, 175
centrée, 127, 190
centrée réduite, 127, 190
discrète, 19
variance
v.a. à densité, 188, 189
v.a. discrète, 127, 143

W

Wald (identités de), 143, 144, 168
Weierstrass (théorème
d'approximation de), 161

IMPRIMÉ EN GRANDE-BRETAGNE
PAR CAMBRIDGE UNIVERSITY PRESS
DÉPÔT LÉGAL NOVEMBRE 2007

Ce livre expose les bases de la théorie des probabilités : algèbre des événements, variables aléatoires, indépendance, probabilités conditionnelles, moments des variables aléatoires discrètes et continues, fonctions génératrices, théorèmes limites. Il comporte un très grand nombre d'exercices accompagnés de solutions détaillées.

Conçu à l'origine à l'intention des candidats au CAPES de mathématiques ou à l'agrégation interne, cet ouvrage s'est révélé très utile aux étudiants des premières années d'université. Il est le fruit de nombreuses années d'expérience des concours et de leur préparation.

Il est suivi d'un second tome, destiné aux étudiants en master de mathématiques et aux candidats à l'agrégation externe.

L'auteur insiste d'emblée, à juste titre, sur l'importance de la démarche de modélisation probabiliste. L'approche intuitive et concrète inhérente aux probabilités va ici de pair avec une exigence de rigueur et une grande précision dans la rédaction. La théorie est constamment illustrée par de nombreux exemples et contre-exemples.

Aucune connaissance préalable en probabilités n'est nécessaire, et certains préliminaires mathématiques (familles sommables, par exemple) sont traités en détail.

Spécialiste de probabilités, Jean-Yves Ouvrard a participé à maintes reprises au jury de l'agrégation de mathématiques, et s'est occupé activement de la préparation à ce concours à l'Université Joseph Fourier de Grenoble.

Collection enseignement des mathématiques

probabilités

2

master
agrégation

Jean-Yves Oувrard

C A S S I N I

1. J.-Y. Ouyvard, *Probabilités I*
3. M. Cottrell, V. Genon-Catalot, Ch. Duhamel, Th. Meyre, *Exercices de probabilités*
4. F. Rouvière, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*
5. J.-Y. Ouyvard, *Probabilités II*
6. G. Zémor, *Cours de cryptographie*
7. A. Szpirglas, *Exercices d'algèbre*
8. B. Perrin-Riou, *Algèbre, arithmétique et Maple*
10. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 1*
11. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 1*
12. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 2*
13. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 2*
14. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 3*
15. H. Krivine, *Exercices de mathématiques pour physiciens*
16. J. Jacod, Ph. Protter, *L'essentiel en théorie des probabilités*
17. M. Willem, *Analyse fonctionnelle élémentaire*
18. É. Amar, É. Matheron, *Analyse complexe*
19. B. Randé, *Problèmes corrigés. Concours 2002 et 2003 (MP)*
20. D. Perrin, *Mathématiques d'école*
21. B. Randé, *Problèmes corrigés. Concours 2004 (MP)*
22. P. Bourgade, *Olympiades internationales de mathématiques 1976-2005*
23. V. Prasolov, *Problèmes et théorèmes d'algèbre linéaire*
24. R. Sá Earp, E. Toubiana, *Introduction à la géométrie hyperbolique et aux surfaces de Riemann*
25. L. Di Menza, *Résolution numérique des équations aux dérivées partielles*
26. B. Candelpergher, *Calcul intégral*
27. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques, vol. 1*
28. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques, vol. 2*

JEAN-YVES OUVRARD

Probabilités

TOME II

Master – Agrégation

CASSINI

JEAN-YVES OUVRARD est maître de conférences à l'Université Joseph Fourier de Grenoble. Il est docteur d'État en mathématiques.

Troisième édition, 2009
ISBN 978-2-84225-144-4
© Cassini, Paris, 2000.

Première édition (2000) ISBN 2-84225-010-9,
deuxième édition corrigée (2004) ISBN 2-84225-086-7

Table des matières

Introduction	I
Chapitre 8. Lois et moments de variables aléatoires	3
8.1. Compléments de théorie de la mesure	3
8.2. Loi d'une variable aléatoire	9
8.3. Moments de variables aléatoires	15
Exercices	29
Chapitre 9. Indépendance de tribus, de variables aléatoires	39
9.1. Indépendance de familles d'événements et de variables aléatoires	39
9.2. Indépendance et événements asymptotiques	47
9.3. Quelques résultats liés à l'indépendance et au modèle de pile ou face	52
9.4. Convolution et loi de la somme de variables aléatoires indépen- dantes	61
Exercices	63
Chapitre 10. Convergences et lois des grands nombres	87
10.1. Convergence en probabilité et presque sûre	87
10.2. Convergence L^p et équi-intégrabilité	93
10.3. Séries de variables aléatoires indépendantes	98
10.4. Lois des grands nombres	101
Exercices	116
Chapitre 11. Probabilités et espérances conditionnelles	135
11.1. Noyaux et lois conditionnelles	135
11.2. Moments conditionnels	147
11.3. Espérance conditionnelle	150
11.3.1. L'espérance conditionnelle comme projecteur orthogo- nal dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$	151
11.3.2. Extension de la définition de l'espérance conditionnelle à $L^1(\Omega, \mathcal{A}, \mathbb{P})$	154
11.3.3. Extension de la définition de l'espérance conditionnelle à $\mathcal{M}^+(\mathcal{A})$	157

11.3.4. Théorèmes de convergence	159
11.3.5. Inégalité de Jensen	162
11.3.6. Calcul d'espérance conditionnelle	163
Exercices	164
Chapitre 12. Transformées de Fourier et fonctions caractéristiques	191
12.1. Définition et propriétés immédiates	191
12.2. Le théorème d'injectivité	193
12.3. Propriétés relatives à l'indépendance	200
12.4. Fonction caractéristique et moments	203
Exercices	212
Chapitre 13. Variables aléatoires gaussiennes	235
13.1. Définition et propriétés	236
13.2. Existence des mesures gaussiennes. Condition d'absolue continuité	238
13.3. Marginales	244
13.4. Régression ; le modèle linéaire	250
13.4.1. Estimation des paramètres de régression	252
13.4.2. Le modèle linéaire gaussien	259
Exercices	267
Chapitre 14. Convergence de mesures et convergence en loi	289
14.1. Convergence de mesures bornées sur \mathbb{R}^d	289
14.2. Convergence en loi	303
14.3. Théorème limite central	313
14.4. Estimation	320
Exercices	327
Chapitre 15. Processus et martingales discrets	349
15.1. Quelques exemples de processus	349
15.2. Processus et martingales : définitions	351
15.3. Temps d'arrêt	354
15.4. Premier théorème d'arrêt	358
15.5. Lemme maximal et martingales dans L^2	360
15.6. Décomposition de Doob	365
15.7. Convergence de martingales intégrables	369
15.8. Deuxième théorème d'arrêt	376
15.9. Convergence de sous- et surmartingales	378
Exercices	379

Chapitre 16. Chaînes de Markov	397
16.1. Introduction	397
16.2. Indépendance conditionnelle	401
16.3. Chaînes de Markov : propriétés générales	405
16.3.1. Propriété de Markov ; matrices de transition	405
16.3.2. Propriété de Markov simple ; lois fini-dimensionnelles	417
16.3.3. Loi initiale ; propriété de Markov forte	422
16.4. Visites à un état fixe	426
16.4.1. Étude de la suite des temps de passage en un point	428
16.4.2. Lois du nombre de visites d'un point et du premier temps de passage en ce point	430
16.5. Classification des états	435
16.5.1. Communication ; périodicité	435
16.5.2. Récurrence	440
16.5.3. Comportement asymptotique et classification	442
16.5.4. Critère analytique de récurrence	450
16.6. Calcul de la matrice potentiel et de $P_x(T_y^1 < +\infty)$	453
16.6.1. Calcul de la matrice potentiel	453
16.6.2. Calcul de $F(x, y) \equiv P_x(T_y^1 < +\infty)$	454
16.7. Mesures invariantes	457
16.8. Loi forte des grands nombres	470
16.8.1. Théorème de loi forte	470
16.8.2. Estimation de la matrice de transition	475
Exercices	477

Chapitre A. Résumé de théorie de la mesure **517**

A.1. Mesure et probabilité	517
A.2. Intégrale	521
A.3. Trois théorèmes de convergence	523
A.4. Mesure produit et théorème de Fubini	526

Index **531**

Liste des chapitres du premier tome

1. Phénomènes aléatoires et modèles probabilistes
2. Familles sommables de nombres réels
3. Indépendance
4. Probabilités et lois conditionnelles
5. Moments d'une variable aléatoire discrète
6. Variables aléatoires à densité
7. Approximation de lois. Loi faible des grands nombres

Introduction

*À mon ami François Brodeau,
avec qui j'ai longuement participé
au jury de l'agrégation.*

Le premier tome de cet ouvrage présentait la théorie élémentaire des probabilités. Nous abordons maintenant l'exposé moderne de cette théorie, qui repose entièrement sur la théorie de la mesure.

C'est grâce à elle que des fondements rigoureux ont pu être établis pour les probabilités au xx^e siècle, en faisant ainsi une véritable théorie mathématique. La première note de Lebesgue sur le problème de la mesure date de 1901, l'introduction du formalisme (Ω, \mathcal{A}, P) par Kolmogorov date de 1927.

Comme nous l'avons vu dans le premier tome, le simple emploi de ce formalisme permet de donner un sens précis à la notion d'événement ou de variable aléatoire, ou à celle d'indépendance. Ce n'était pas du tout évident avant Kolmogorov. Bel exemple de modélisation réussie... Mais nous avons éprouvé certaines difficultés, pour le cas des variables aléatoires à densité, à formuler l'analogue du théorème de transfert, ou à justifier les critères usuels d'indépendance (cf. ch. 6, p. 190, p. 201) : c'est qu'il faut pour cela toute la force de la théorie de la mesure ; de même pour formuler et démontrer un résultat de convergence presque sûre comme la loi forte des grands nombres. On n'imagine plus à notre époque d'enseignement des probabilités, y compris en vue des applications, en-dehors du cadre fourni par la théorie de la mesure.

Cet ouvrage a été conçu à l'intention des candidats au CAPES et à l'agrégation. Mais la forme que nous lui avons donnée lui permet d'être utilisé tout au long des études universitaires. Le premier tome sera utile aux étudiants de licence ou des classes préparatoires. Le présent volume pourra servir de manuel aux étudiants en master désireux d'approfondir leurs bases en probabilités.

Voici, dans la perspective des concours, un bref mode d'emploi de ces deux volumes. Le premier correspond aux programmes du CAPES et de l'agrégation interne. Le second répond aux besoins des candidats à l'agrégation externe. Dans la configuration actuelle du concours (programme de 1999) on demande à tous les candidats d'avoir une connaissance solide de la théorie, correspondant à nos chapitres 8 à 14. Certaines questions de modélisation, mettant en jeu des variables discrètes ou des lois de probabilité classiques, amèneront à se référer au premier tome. Les chapitres 15 (martingales) et 16 (chaînes de Markov) sont destinés aux candidats ayant choisi à l'oral l'option *Probabilités et statistique* (étant entendu que certaines notions sur les chaînes de Markov finies sont exigibles de tous les candidats).

Le candidat trouvera dans ce livre les matériaux pour construire lui-même sa leçon, aussi bien dans le corps des chapitres que dans les très nombreux exercices corrigés : on notera que dans notre exposé chaque notion introduite est suivie d'un

exemple. Le candidat devra aussi au cours de sa préparation rechercher dans d'autres sources des informations complémentaires et des exemples d'application, relatifs notamment aux thèmes proposés par le programme de l'année. Il devra en outre se former à l'utilisation d'un logiciel de calcul formel, et apprendre à illustrer avec ce logiciel les sujets exposés ; de nombreux exercices de ce livre pourront servir de support à cette illustration.

Pour rendre l'ouvrage plus vivant, nous avons donné, sous forme de notes, quelques renseignements biographiques sur les principaux mathématiciens qui ont contribué à l'élaboration de la théorie des probabilités ; notre source d'information principale a été le livre de B. Hauchecorne et D. Suratteau, *Des Mathématiciens de A à Z*, (Ellipses, 1996, Paris). Enfin, nous avons donné en annexe un *Résumé de théorie de la mesure*, mais nous avons maintenu dans le corps de l'exposé, au début du chapitre 8, certains résultats essentiels au probabiliste qui ne figurent pas toujours dans les cours d'intégration.

Je tiens à remercier les éditions Cassini : en rendant accessible cet ouvrage à un public motivé par la perspective d'un concours, mais aussi curieux d'apprendre et de réfléchir, elles me permettent d'apporter une aide, je l'espère fructueuse, à toute personne qui aura eu la patience de me suivre.

Je remercie tout particulièrement André Bellaïche, avec qui j'ai eu de longues et fructueuses confrontations sur cet ouvrage.

Enfin, je remercie les relecteurs de cet ouvrage : leurs remarques ont contribué au polissage du manuscrit et conduit à la forme définitive de ce livre. Je souhaite que le lecteur trouve ici matière à un travail agréable et enrichissant.

Chapitre 8

Lois et moments de variables aléatoires

Dans ce second tome, nous supposons connue la théorie de la mesure abstraite et de l'intégration. Le lecteur pourra trouver un résumé détaillé de cette théorie dans l'annexe figurant en fin de ce livre. À ces rappels, nous ajoutons ci-dessous quelques compléments, en général omis dans les cours d'intégration, mais indispensables en probabilités.

Dans la suite du chapitre nous donnons la présentation définitive, dans le cadre de la théorie de la mesure, des notions de loi et de moments d'une variable aléatoire.

8.1. Compléments de théorie de la mesure

Nous commençons par une étude du **principe de prolongement par mesurabilité**, très fréquemment utilisé en probabilités.

Définition 8.1. Une famille \mathcal{C} de parties d'un ensemble Ω est appelée π -système si elle est stable par intersection finie.

Une famille \mathcal{S} de parties d'un ensemble Ω est appelée λ -système si elle satisfait aux deux axiomes suivants :

(λ_1) pour toute suite croissante $(S_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{S} on a

$$\bigcup_{n \in \mathbb{N}} S_n \in \mathcal{S};$$

(λ_2) pour tous éléments A et B de \mathcal{S} tels que $A \subset B$, on a

$$B \setminus A \in \mathcal{S}.$$

Nous aurons besoin de la notion de π -système (resp. λ -système) engendré par une famille \mathcal{L} de parties de Ω . On observe d'abord que l'intersection d'une famille quelconque de π -systèmes (resp. de λ -systèmes) est un π -système (resp. un λ -système). De plus, $\mathcal{P}(\Omega)$ est à la fois un π -système et un λ -système; il existe donc un π -système (resp. un λ -système) contenant \mathcal{L} . On définit alors le π -système (resp. le λ -système) **engendré par \mathcal{L}** comme l'intersection de tous les π -systèmes (resp. les λ -systèmes) contenant \mathcal{L} .

On peut aussi caractériser le π -système (resp. λ -système) engendré par \mathcal{L} comme étant, au sens de l'inclusion, le *plus petit* π -système (resp. λ -système) contenant \mathcal{L} .

Remarque. Il faut noter que, tout comme dans le contexte des tribus, ce procédé de définition par « fermeture » n'est pas constructif : en général, on n'a pas d'expression explicite d'un élément générique du λ -système engendré par \mathcal{L} . Toutefois, le π -système engendré par \mathcal{L} est la famille de toutes les intersections **finies** d'éléments de \mathcal{L} (cette famille est le plus petit π -système contenant \mathcal{L}).

Exemples 8.1. Sur \mathbb{R} , les familles constituées

- des intervalles $]a, b[$, ($a \leq b$);
- des intervalles $[a, b]$, ($a \leq b$);
- des intervalles $[a, b]$, ($a \leq b$), et de l'ensemble vide;
- des demi-droites $[a, +\infty[$, ($a \in \mathbb{R}$);
- des demi-droites $]-\infty, a]$, ($a \in \mathbb{R}$)

sont des π -systèmes.

De même, sur \mathbb{R}^d , les familles constituées des ouverts, des ouverts bornés, des fermés, des pavés de la forme $\prod_{i=1}^d [a_i, b_i]$, ($a_i \leq b_i$) et de l'ensemble vide, sont des π -systèmes.

Dans la suite, à maintes reprises, on rencontrera des λ -systèmes. Pour fixer les idées, voici un exemple de λ -système qui n'est pas une tribu : si Ω est un ensemble non dénombrable, la famille de ses parties dénombrables est à la fois un π -système et un λ -système; toutefois, cette famille ne contient pas Ω , et plus généralement, elle n'est pas stable par passage au complémentaire : ce n'est donc pas une tribu.

Les relations entre ces différentes structures sont précisées par le lemme suivant.

Lemme 8.2. *Pour qu'un λ -système \mathcal{S} sur Ω soit une tribu sur Ω , il faut et il suffit que \mathcal{S} soit un π -système et que $\Omega \in \mathcal{S}$.*

Démonstration. La condition nécessaire est triviale. Démontrons la condition suffisante; si \mathcal{S} est à la fois un λ -système et π -système sur Ω qui contient Ω , \mathcal{S} est stable par rapport au complémentaire (puisque $\Omega \in \mathcal{S}$) et par union finie; pour ce dernier point, il suffit de remarquer que si A et B sont des éléments de \mathcal{S} , A^c et B^c le sont aussi; puisque l'on a

$$(A \cup B)^c = A^c \cap B^c,$$

et que \mathcal{S} est un π -système, on a $(A \cup B)^c \in \mathcal{S}$, et donc aussi $A \cup B \in \mathcal{S}$. Reste à démontrer que \mathcal{S} est stable par union dénombrable. Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'éléments de \mathcal{S} , on construit une suite $(B_n)_{n \in \mathbb{N}^*}$ croissante d'éléments

de \mathcal{S} ayant même réunion que la suite $(A_n)_{n \in \mathbb{N}^*}$: il suffit de définir B_n par récurrence en posant $B_1 = A_1$ et pour $n \geq 2$, $B_n = \bigcup_{j=1}^n A_j$. \square

Le lemme suivant est d'usage fréquent en probabilités. Il permet d'étendre une propriété possédée par une famille d'événements ayant la structure de λ -système à la tribu engendrée par cette famille.

Lemme 8.3 (Principe de prolongement par mesurabilité; version ensembliste). *Soit \mathcal{S} un λ -système sur Ω qui contient un π -système \mathcal{C} et tel que $\Omega \in \mathcal{S}$; alors \mathcal{S} contient la tribu $\sigma(\mathcal{C})$ engendrée par \mathcal{C} .*

Démonstration. Il suffit de démontrer que le λ -système Λ engendré par \mathcal{C} et Ω est égal à $\sigma(\mathcal{C})$.

Λ est un π -système; en effet, définissons, pour tout $A \in \mathcal{P}(\Omega)$, la famille d'ensembles

$$\Lambda_A = \{B \in \Lambda \mid B \cap A \in \Lambda\}.$$

Λ étant un λ -système, on vérifie qu'il en est de même de Λ_A . De plus : $\Lambda_A \subset \Lambda$. En particulier, pour tout $A \in \mathcal{C}$, puisque, par définition de Λ , $\Omega \in \Lambda_A$ et $\mathcal{C} \subset \Lambda_A$, on a $\Lambda_A = \Lambda$, la famille Λ étant le plus petit λ -système contenant \mathcal{C} et Ω ; on a donc :

$$\forall A \in \mathcal{C}, \forall B \in \Lambda \quad B \cap A \in \Lambda,$$

ce qui s'écrit : $\forall B \in \Lambda, \mathcal{C} \subset \Lambda_B$. Mais alors, $\forall B \in \Lambda$, la famille Λ_B est un λ -système contenant \mathcal{C} et Ω , donc : $\Lambda_B = \Lambda$. On vient de démontrer que, $\forall B \in \Lambda, \forall C \in \Lambda, B \cap C \in \Lambda$, c'est à dire que Λ est un π -système.

Il résulte du lemme précédent que Λ est une tribu et donc que : $\Lambda \supset \sigma(\mathcal{C})$. Mais $\sigma(\mathcal{C})$ étant un λ -système contenant \mathcal{C} et Ω , on a : $\Lambda \subset \sigma(\mathcal{C})$; l'égalité $\Lambda = \sigma(\mathcal{C})$ en résulte. \square

Voici une application importante de ce lemme :

Théorème 8.4 (Théorème d'unicité des mesures). *Soient μ_1 et μ_2 deux mesures positives sur l'espace probabilisable (Ω, \mathcal{A}) telles que*

$$\forall A \in \mathcal{C} \quad \mu_1(A) = \mu_2(A),$$

où \mathcal{C} est un π -système qui engendre la tribu \mathcal{A} .

1. *Si μ_1 et μ_2 sont bornées et de même masse, alors $\mu_1 = \mu_2$.*

2. *Si l'une des mesures μ_1 ou μ_2 est non bornée, et s'il existe une suite $(E_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{C} telle que $\Omega = \bigcup_{n \in \mathbb{N}} E_n$ et telle que*

$$\forall n \in \mathbb{N} \quad \mu_1(E_n) = \mu_2(E_n) < +\infty,$$

alors $\mu_1 = \mu_2$.

Démonstration. Si μ_1 et μ_2 sont bornées et de même masse, la famille $\mathcal{S} = \{A \in \mathcal{A} \mid \mu_1(A) = \mu_2(A)\}$ est alors un λ -système contenant \mathcal{C} et Ω ; elle contient donc la tribu \mathcal{A} engendrée par \mathcal{C} . Ceci démontre que $\mu_1 = \mu_2$.

Pour le cas où l'une des mesures est non bornée, on considère les restrictions de μ_1 et μ_2 aux E_n ; d'après la première partie, elles sont égales pour tout n . Par application de la formule de Poincaré (proposition 1.6., tome 1), qui reste valable pour des mesures finies (et c'est le cas pour chaque restriction de ces mesures aux E_n), les restrictions de μ_1 et μ_2 aux ensembles $F_n = \bigcup_{0 \leq j \leq n} E_j$ sont encore égales; les mesures μ_1 et μ_2 coïncident donc sur Ω , puisque la suite d'ensembles F_n est croissante de réunion Ω . \square

Remarque. Les hypothèses du théorème impliquent que μ_1 et μ_2 sont σ -finies. De plus, il résulte de ce théorème que si deux **probabilités** coïncident sur un π -système engendrant \mathcal{A} , elles sont égales.

Exemple d'application. Si deux mesures sur \mathbb{R} coïncident sur toutes les demi-droites $]-\infty, x]$, $x \in \mathbb{R}$, elles sont égales.

Le corollaire ci-dessous est très souvent utilisé en calcul des probabilités.

Notation. $\mathcal{C}_{\mathcal{K}}(\mathbb{R}^d)$ (resp. $\mathcal{C}_{\mathcal{K}}^+(\mathbb{R}^d)$) désigne l'ensemble des fonctions continues à support compact de \mathbb{R}^d dans \mathbb{R} (resp. et positives).

Corollaire 8.5. Soient μ_1 et μ_2 deux mesures positives sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ finies sur tout compact (on dit que ce sont des mesures de **Radon**). Si

$$\forall f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R}^d) \quad \int_{\mathbb{R}^d} f \, d\mu_1 = \int_{\mathbb{R}^d} f \, d\mu_2,$$

les mesures μ_1 et μ_2 sont égales.

Démonstration. La classe \mathcal{C} des ouverts bornés de \mathbb{R}^d est un π -système. Les mesures μ_1 et μ_2 coïncident sur \mathcal{C} ; en effet si $O \in \mathcal{C}$, il existe une suite croissante $(f_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{C}_{\mathcal{K}}^+(\mathbb{R}^d)$ convergeant simplement vers $\mathbf{1}_O$. D'après la propriété de Beppo Levi¹ et l'hypothèse, on a :

$$\mu_1(O) = \lim_n \int_{\mathbb{R}^d} f_n \, d\mu_1 = \lim_n \int_{\mathbb{R}^d} f_n \, d\mu_2 = \mu_2(O) < +\infty.$$

Il suffit d'appliquer le théorème précédent. \square

Voici maintenant une **version fonctionnelle** du lemme 8.3.

Théorème 8.6 (Principe de prolongement par mesurabilité; version fonctionnelle). Soient \mathcal{C} un π -système sur Ω et \mathcal{H} un espace vectoriel de fonctions réelles sur Ω tels que :

1. La propriété de **Beppo Levi** est aussi appelée théorème de convergence monotone.

(i) pour toute suite croissante $(h_n)_{n \in \mathbb{N}}$ d'éléments positifs de \mathcal{H} telle que $h \equiv \sup_n h_n$ soit fini (resp. borné), on a $h \in \mathcal{H}$;

(ii) $\mathbf{1}_\Omega \in \mathcal{H}$ et, pour tout $C \in \mathcal{C}$, $\mathbf{1}_C \in \mathcal{H}$;

alors \mathcal{H} contient toutes les fonctions réelles $\sigma(\mathcal{C})$ -mesurables (resp. $\sigma(\mathcal{C})$ -mesurables et bornées).

Démonstration. Il suffit de montrer que les fonctions $\mathbf{1}_A$, $A \in \sigma(\mathcal{C})$, sont dans \mathcal{H} . En effet, si c'est le cas, l'espace vectoriel \mathcal{H} contiendra toute fonction étagée $\sigma(\mathcal{C})$ -mesurable, puisqu'une telle fonction s'écrit $\sum_{i \in I} a_i \mathbf{1}_{A_i}$ (I fini, $a_i \in \mathbb{R}$ et $A_i \in \sigma(\mathcal{C})$). Toute fonction $\sigma(\mathcal{C})$ -mesurable positive finie (resp. bornée), étant limite croissante de fonctions étagées $\sigma(\mathcal{C})$ -mesurables, sera alors dans \mathcal{H} , en vertu de l'hypothèse (i). Enfin \mathcal{H} contiendra toute fonction $\sigma(\mathcal{C})$ -mesurable finie (resp. bornée) h , puisqu'une telle fonction se décompose sous la forme $h = h^+ - h^-$, où h^+ et h^- sont positives finies (resp. bornées) et $\sigma(\mathcal{C})$ -mesurables.

Il reste à démontrer que les fonctions $\mathbf{1}_A$, $A \in \sigma(\mathcal{C})$, sont dans \mathcal{H} , ce qui s'écrit $\mathcal{H} \supset \sigma(\mathcal{C})$, où on pose $\mathcal{H} = \{A \in \mathcal{P}(\Omega) \mid \mathbf{1}_A \in \mathcal{H}\}$. Par hypothèse, on a $\mathcal{H} \supset \mathcal{C}$ et $\Omega \in \mathcal{H}$. De plus \mathcal{H} est un λ -système car, d'une part, \mathcal{H} étant un espace vectoriel, pour tous S_1 et S_2 tels que $S_1 \supset S_2$, on a $\mathbf{1}_{S_1 \setminus S_2} = \mathbf{1}_{S_1} - \mathbf{1}_{S_2} \in \mathcal{H}$ et, d'autre part, en vertu de la première hypothèse, pour toute suite croissante $(S_n)_{n \in \mathbb{N}} \subset \mathcal{H}$, on a $\mathbf{1}_{\bigcup_{n \in \mathbb{N}} S_n} = \sup_n \mathbf{1}_{S_n} \in \mathcal{H}$. Il résulte du lemme 8.3 que $\sigma(\mathcal{C}) \subset \mathcal{H}$. Le théorème est démontré. \square

Nous terminons cette section en rappelant (sans en donner de démonstration) les énoncés de quelques théorèmes d'usage constant dans la suite.

Définition 8.7. Soit μ une mesure positive sur l'espace probabilisable (Ω, \mathcal{A}) . Soit f une fonction numérique mesurable positive définie sur cet espace. La mesure $\nu : A \mapsto \int_A f \, d\mu$ est dite **mesure de densité f par rapport à μ** et notée $f \cdot \mu$.

Définition 8.8. Une mesure ν sur (Ω, \mathcal{A}) est dite **absolument continue par rapport à μ** si, pour tout $A \in \mathcal{A}$ tel que $\mu(A) = 0$, on a $\nu(A) = 0$. On note : $\nu \ll \mu$.

Les mesures μ et ν sur (Ω, \mathcal{A}) sont dites **étrangères** s'il existe $N \in \mathcal{A}$ tel que $\mu(N) = 0$ et $\nu(N^c) = 0$. On note : $\nu \perp \mu$.

Exemple 8.2. La mesure de Lebesgue λ sur \mathbb{R} et la mesure de Dirac δ_0 en 0 sont étrangères, puisque $\lambda(\{0\}) = \delta_0(\{0\}^c) = 0$.

Si $\nu = f \cdot \mu$, on a bien sûr $\nu \ll \mu$. La réciproque fait l'objet du théorème de **Radon-Nikodym** (cf. par exemple Neveu ou Métivier pour une démonstration).

2. Cette notation est justifiée par la formule (8.1).

Théorème 8.9 (Théorème de Radon-Nikodym). Soit sur un espace probabilisable (Ω, \mathcal{A}) une mesure σ -finie μ et une mesure ν telles que $\nu \ll \mu$; alors il existe une fonction mesurable positive f (unique à une équivalence μ -p.p. près) telle que $\nu = f \cdot \mu$.

Proposition 8.10 (Intégration par rapport à une mesure à densité). Soient μ une mesure positive sur l'espace probabilisable (Ω, \mathcal{A}) et f une fonction numérique mesurable positive définie sur cet espace. Soit $\nu = f \cdot \mu$ la mesure de densité f par rapport à μ . Soit h une fonction mesurable sur (Ω, \mathcal{A}) .

– Si h est positive, on a :

$$\int_{\Omega} h \, d\nu = \int_{\Omega} h \cdot f \, d\mu. \quad (8.1)$$

– Si h est de signe quelconque, pour que h soit ν -intégrable il faut et il suffit que $h \cdot f$ soit μ -intégrable et, dans ce cas, l'égalité (8.1) est encore valable.

Définition 8.11. Soit T une application mesurable de l'espace probabilisable (E_1, \mathcal{E}_1) dans l'espace probabilisable (E_2, \mathcal{E}_2) . Soit μ_1 une mesure sur (E_1, \mathcal{E}_1) . La mesure sur (E_2, \mathcal{E}_2) définie par

$$\forall B \in \mathcal{E}_2 \quad \mu_2(B) = \mu_1[T^{-1}(B)]$$

est appelée **mesure image** de μ_1 par T et notée $T(\mu_1)$.

Théorème 8.12 (Théorème de la mesure image, ou théorème de transfert). Soit T une application mesurable de l'espace probabilisable (E_1, \mathcal{E}_1) dans l'espace probabilisable (E_2, \mathcal{E}_2) ; soit $T(\mu_1)$ la mesure image de μ_1 par T ; soit h une fonction mesurable sur (E_2, \mathcal{E}_2) .

– Si h est positive, on a :

$$\int_{E_2} h \, dT(\mu_1) = \int_{E_1} h \circ T \, d\mu_1. \quad (8.2)$$

– Si h est de signe quelconque, pour que h soit $T(\mu_1)$ -intégrable il faut et il suffit que $h \circ T$ soit μ_1 -intégrable et, dans ce cas, l'égalité (8.2) est encore valable.

Théorème 8.13 (Théorème de changement de variables). Soit T un diffeomorphisme d'un ouvert U de \mathbb{R}^d sur un ouvert V de \mathbb{R}^d , de classe C^1 . Soit f une fonction réelle mesurable définie sur U . Alors f est Lebesgue-intégrable sur U si et seulement si la fonction $v \mapsto |\det(T^{-1})'(v)| f[T^{-1}(v)]$ est Lebesgue-intégrable sur V . Dans ce cas, on a :

$$\int_U f(x) \, d\lambda_d(x) = \int_V |\det(T^{-1})'(v)| f[T^{-1}(v)] \, d\lambda_d(v). \quad (8.3)$$

Remarque. On dit souvent que le second membre de (8.3) s'obtient à partir du premier membre au moyen du changement de variable $v = T(x)$, ou $x = T^{-1}(v)$: v est la « nouvelle » variable, x l'« ancienne ». De plus, $\det(T^{-1})'(v)$ est souvent noté $\frac{Dv}{Dx}$ et est appelé **jacobien** du changement de variable.

8.2. Loi d'une variable aléatoire

Toutes les variables aléatoires seront définies sur un même espace probabilisé (Ω, \mathcal{A}, P) . Une **variable aléatoire** X à valeurs dans l'espace probabilisable (E, \mathcal{E}) est alors par définition une **application mesurable** de (Ω, \mathcal{A}) dans (E, \mathcal{E}) c'est à dire une application telle que :

$$\forall B \in \mathcal{E} \quad X^{-1}(B) \in \mathcal{A}.$$

Définition 8.14. On appelle **loi** (ou **loi de probabilité**) de la variable aléatoire X à valeurs dans l'espace probabilisable (E, \mathcal{E}) la mesure image P_X de P par X .

Afin d'étendre aux variables aléatoires à valeurs dans \mathbb{R}^d ($d > 1$) la notion de fonction de répartition, on introduit un ordre partiel sur \mathbb{R}^d en posant

$$x \leq y \quad \text{si et seulement si} \quad \forall i = 1, 2, \dots, d \quad x_i \leq y_i.$$

Pour $d = 1$, on retrouve l'ordre habituel (total) sur \mathbb{R} .

Définition 8.15. Soit X une variable aléatoire à valeurs dans $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$.

(a) On appelle **fonction de répartition** de X la fonction F_X de \mathbb{R}^d dans \mathbb{R}^+ définie par

$$\forall x \in \mathbb{R}^d \quad F_X(x) = P(X \leq x).$$

où \leq est l'ordre partiel usuel de \mathbb{R}^d .

(b) On dit que X admet la fonction f pour **densité** si sa loi P_X admet f pour densité par rapport à la mesure de Lebesgue λ_d sur \mathbb{R}^d .

Si une variable aléatoire X possède une densité f , toute fonction λ_d -presque partout égale à f est encore une densité de X et inversement toute densité de X est λ_d -presque partout égale à f . La densité de X est donc définie à l'égalité λ_d -p.p. près et on la confond souvent avec sa classe d'équivalence pour cette relation, qu'on note f_X . La densité de X vérifie donc

$$\forall A \in \mathcal{B}_{\mathbb{R}^d} \quad P_X(A) = \int_A f_X(x) d\lambda_d(x).$$

Le théorème d'unicité des mesures permet d'affirmer que pour que X admette une *densité*, il faut et il suffit qu'il existe une fonction *positive* f_X de $\mathcal{L}^1(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, \lambda_d)$ qui vérifie :

$$\forall x \in \mathbb{R}^d \quad F_X(x) = \int_{\{u \leq x\}} f_X(u) d\lambda_d(u). \quad (8.4)$$

En particulier si $d = 1$ et s'il existe une fonction **positive** f_X **intégrable au sens de Riemann** qui vérifie

$$\forall x \in \mathbb{R} \quad F_X(x) = \int_{]-\infty, x]} f_X(u) d\lambda(u),$$

en remarquant que, dans ce cas, cette intégrale est aussi une intégrale de Riemann, on retrouve la définition *élémentaire* d'une densité donnée antérieurement (chapitre 6).

Si la loi P_X est une mesure à **densité par rapport** à la **mesure dénombrement** de \mathbb{R}^d , la variable aléatoire est **discrète** (cette définition est un peu plus générale que celle donnée dans le premier tome). L'ensemble $\text{val}(X) = \{x \mid P(X = x) \neq 0\}$ est alors **dénombrable** et l'on a, δ_x désignant la mesure de Dirac en x ,

$$\forall A \in \mathcal{B}_{\mathbb{R}^d} \quad P_X(A) = \sum_{x \in \text{val}(X)} P(X = x) \delta_x(A),$$

ce qui, en termes de probabilités, s'écrit

$$P_X = \sum_{x \in \text{val}(X)} P(X = x) \delta_x.$$

On rappelle que, pour une telle variable aléatoire discrète X , on a $f \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, P_X)$ si et seulement si

$$\sum_{x \in \text{val}(X)} |f(x)| P(X = x) < +\infty$$

et que, s'il en est ainsi, on a

$$\int_{\mathbb{R}^d} f dP_X = \sum_{x \in \text{val}(X)} f(x) P(X = x).$$

Remarque importante. Si X est à valeurs dans $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, il résulte du corollaire 8.5 que sa loi est entièrement déterminée par la famille des intégrales $\int_{\Omega} f(X) dP \equiv \int_{\mathbb{R}^d} f dP_X$ où f parcourt $\mathcal{C}_{\mathcal{X}}^+(\mathbb{R}^d)$. Ceci fournit un *procédé très efficace pour étudier la loi* d'une variable aléatoire à valeurs dans $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, les théorèmes d'intégration s'appliquant tous sans problème.

Exemple 8.3. Soit X une variable aléatoire réelle de **loi gaussienne** $\mathcal{N}_{\mathbb{R}}(0, 1)$, c'est à dire de densité f_X définie par

$$\forall x \in \mathbb{R} \quad f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

On se propose d'étudier la loi de la variable aléatoire X^2 . Alors, pour tout $f \in \mathcal{C}_{\mathbb{X}}^+(\mathbb{R}^d)$, on a, par le théorème de transfert³ et celui d'intégration par rapport à une mesure à densité :

$$\begin{aligned} \int_{\Omega} f(X^2) dP &= \int_{\mathbb{R}} f(x^2) dP_X(x) \\ &= \int_{\mathbb{R}} f(x^2) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) d\lambda(x) \\ &= \frac{2}{\sqrt{2\pi}} \int_{\mathbb{R}^{*+}} f(x^2) \exp\left(-\frac{x^2}{2}\right) d\lambda(x). \end{aligned}$$

Le changement de variables associé au difféomorphisme T de $]0, +\infty[$ sur lui-même défini par $T(x) = x^2$, donne

$$\int_{\Omega} f(X^2) dP = \int_{\mathbb{R}^{*+}} f(y) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) y^{-\frac{1}{2}} d\lambda(y),$$

puisque, pour tout $y \in]0, +\infty[$, on a $T^{-1}(y) = \sqrt{y}$ et $(T^{-1})'(y) = \frac{1}{2}y^{-\frac{1}{2}}$. Il en résulte que

$$\int_{\Omega} f(X^2) dP = \int_{\mathbb{R}} fg d\lambda,$$

où g est la fonction définie sur \mathbb{R} *tout entier* par

$$\forall y \in \mathbb{R} \quad g(y) = \mathbf{1}_{\mathbb{R}^{*+}}(y) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) y^{-\frac{1}{2}}.$$

Ceci démontre que *la variable aléatoire admet la densité g* . La loi de probabilité de densité g est appelée loi $\gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. Nous reparlerons des **lois gamma** (cf. aussi le tableau de lois classiques).

Exemple 8.4. Soit $U = (U_1, U_2)$ une variable aléatoire à valeurs dans \mathbb{R}^2 de **loi normale** $\mathcal{N}_{\mathbb{R}^2}(0, \mathbf{1}_{\mathbb{R}^2})$, c'est à dire de densité f_U , fonction définie sur \mathbb{R}^2 par

$$\forall u \in \mathbb{R}^2 \quad f_U(u) = \frac{1}{2\pi} \exp\left(-\frac{\|u\|^2}{2}\right)$$

(il s'agit de la norme *euclidienne* usuelle). Soit g l'application de \mathbb{R}^2 dans \mathbb{R} définie par

3. Remarquer que *formellement* il suffit pour l'écrire de transformer la *grande* lettre désignant la variable aléatoire en la *petite* lettre correspondant aux *valeurs* prises par cette variable aléatoire.

$$\forall u \in \mathbb{R}^2 \quad g(u_1, u_2) = \begin{cases} \frac{u_1}{u_2} & \text{si } u_2 \neq 0 \\ 0 & \text{si } u_2 = 0, \end{cases}$$

et soit $X = g(U)$. On se propose d'étudier la loi de la variable aléatoire X . Pour tout $f \in \mathcal{C}_X^+(\mathbb{R}^2)$, il résulte du théorème de transfert et de celui d'intégration par rapport à une mesure à densité que (se rappeler que la mesure de Lebesgue d'une droite de \mathbb{R}^2 est nulle) :

$$\begin{aligned} \int_{\Omega} f(X) dP &= \int_{\mathbb{R}^2} (f \circ g)(x) dP_X(x) \\ &= \int_{\mathbb{R}^2 \setminus \{u_2=0\}} f\left(\frac{u_1}{u_2}\right) \frac{1}{2\pi} \exp\left(-\frac{u_1^2 + u_2^2}{2}\right) d\lambda_2(u_1, u_2). \end{aligned}$$

Soit T le difféomorphisme de $\mathbb{R}^2 \setminus \{u_2 = 0\}$ sur lui-même défini par

$$T(u_1, u_2) = \left(\frac{u_1}{u_2}, u_2\right);$$

son inverse T^{-1} est déterminé par les systèmes équivalents suivants

$$\begin{cases} x = \frac{u_1}{u_2} \\ y = u_2 \end{cases} \iff \begin{cases} u_1 = xy \\ u_2 = y \end{cases},$$

ce qui donne

$$T^{-1}(x, y) = (xy, y).$$

En faisant le changement de variables associé à T , de jacobien $\det(T^{-1})'(x, y)$, encore noté $\frac{D(u_1, u_2)}{D(x, y)}$, et qui vaut

$$\frac{D(u_1, u_2)}{D(x, y)} = \begin{vmatrix} y & x \\ 0 & 1 \end{vmatrix} = y,$$

il vient

$$\int_{\Omega} f(X) dP = \int_{\mathbb{R}^2 \setminus \{u_2=0\}} f(x) \frac{1}{2\pi} \exp\left(-\frac{(1+x^2)y^2}{2}\right) |y| d\lambda_2(x, y),$$

(ne surtout pas oublier la *valeur absolue* du jacobien), soit encore, puisque une droite est de mesure de Lebesgue nulle,

$$\int_{\Omega} f(X) dP = \int_{\mathbb{R}^2} f(x) \frac{1}{2\pi} \exp\left(-\frac{(1+x^2)y^2}{2}\right) |y| d\lambda_2(x, y).$$

D'après le théorème de Fubini, applicable automatiquement pour les fonctions mesurables *positives*, on a alors

$$\begin{aligned} \int_{\Omega} f(X) dP &= \int_{\mathbb{R}} f(x) \left(\int_{\mathbb{R}} \frac{1}{2\pi} \exp\left(-\frac{(1+x^2)y^2}{2}\right) |y| d\lambda(y) \right) d\lambda(x) \\ &= \int_{\mathbb{R}} f(x) \frac{1}{\pi} \frac{1}{1+x^2} d\lambda(x). \end{aligned}$$

La variable aléatoire X admet donc une densité f_X définie par

$$\forall x \in \mathbb{R}, \quad f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

c'est à dire que X suit la **loi de Cauchy**.

Remarque. La famille des parties de \mathbb{R}^d de la forme $\{y \in \mathbb{R}^d \mid y \leq x\}$ où $x \in \mathbb{R}^d$ forme un π -système. Si donc on connaît la fonction de répartition de la variable aléatoire X à valeurs dans \mathbb{R}^d , on connaît sa loi P_X sur ce π -système et on la connaît donc entièrement d'après le théorème d'égalité des mesures. La *fonction de répartition* est donc un outil pour identifier la loi d'une variable aléatoire. Sa définition est liée à une *structure d'ordre*; elle sera donc en particulier bien manipulable lorsque la variable aléatoire étudiée sera définie à l'aide d'opérations relatives à cet ordre (exemples : sup, inf, max, min).

La proposition suivante permet d'obtenir la **loi d'une variable aléatoire transformée d'une autre par un difféomorphisme**.

Proposition 8.16. *Soit X une variable aléatoire à valeurs dans \mathbb{R}^d et T un difféomorphisme de \mathbb{R}^d sur lui-même. Si X admet une densité f_X , la variable aléatoire $Y = T \circ X$, aussi notée $T(X)$, admet une densité f_Y définie par*

$$\forall y \in \mathbb{R}^d \quad f_Y(y) = |\det(T^{-1})'(y)| f_X[T^{-1}(y)].$$

Démonstration. Pour tout $f \in \mathcal{C}_X^+(\mathbb{R}^d)$ fixé, il résulte du théorème de transfert et de celui d'intégration par rapport à une mesure à densité que :

$$\int_{\Omega} f(Y) dP = \int_{\mathbb{R}^d} f \circ T(x) dP_X(x) = \int_{\mathbb{R}^d} f \circ T(x) f_X(x) d\lambda_d(x).$$

En faisant le changement de variables $y = T(x)$ de \mathbb{R}^d sur lui-même défini par le difféomorphisme T , on obtient :

$$\int_{\Omega} f(Y) dP = \int_{\mathbb{R}^d} f(y) |\det(T^{-1})'(y)| f_X[T^{-1}(y)] d\lambda_d(y),$$

d'où le résultat. □

Définition 8.17. *Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d = \prod_{i=1}^k \mathbb{R}^{d_i}$ et soit Π_i la projection canonique de \mathbb{R}^d sur \mathbb{R}^{d_i} . La variable aléatoire $X_i = \Pi_i \circ X$, à valeurs dans \mathbb{R}^{d_i} , est appelée **i -ième marginale** de X .*

Les propositions suivantes permettent le **calcul de la loi des marginales**. Pour alléger l'écriture, nous ne les énonçons que pour $k = 2$, le cas général se calquant sur ce cas particulier.

Proposition 8.18. Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

(a) Si X admet une **densité** f_X alors X_1 et X_2 admettent des densités f_{X_1} et f_{X_2} donnés par

$$\begin{aligned} \forall x_1 \in \mathbb{R}^{d_1} \quad f_{X_1}(x_1) &= \int_{\mathbb{R}^{d_2}} f_X(x_1, x_2) d\lambda_{d_2}(x_2) \\ \forall x_2 \in \mathbb{R}^{d_2} \quad f_{X_2}(x_2) &= \int_{\mathbb{R}^{d_1}} f_X(x_1, x_2) d\lambda_{d_1}(x_1). \end{aligned}$$

(b) Si X est une variable aléatoire **discrète**, X_1 et X_2 le sont aussi et on a :

$$\begin{aligned} \forall x_1 \in \text{val}(X_1) \quad P(X_1 = x_1) &= \sum_{x_2 \in \text{val}(X_2)} P(X_1 = x_1, X_2 = x_2) \\ \forall x_2 \in \text{val}(X_2) \quad P(X_2 = x_2) &= \sum_{x_1 \in \text{val}(X_1)} P(X_1 = x_1, X_2 = x_2). \end{aligned}$$

Démonstration. (a) Pour tout $f \in \mathcal{C}_{\mathcal{X}}^+(\mathbb{R}^{d_1})$ fixé, il résulte du théorème de transfert et de celui d'intégration par rapport à une mesure à densité que l'on a

$$\begin{aligned} \int_{\Omega} f(X_1) dP &= \int_{\mathbb{R}^d} (f \circ \Pi_1)(x) dP_X(x) \\ &= \int_{\mathbb{R}^d} (f \circ \Pi_1)(x_1, x_2) f_X(x_1, x_2) d\lambda_d(x_1, x_2), \end{aligned}$$

soit, d'après le théorème de Fubini (applicable car f est mesurable positive)

$$\int_{\Omega} f(X_1) dP = \int_{\mathbb{R}^{d_1}} f(x_1) \left(\int_{\mathbb{R}^{d_2}} f_X(x_1, x_2) d\lambda_{d_2}(x_2) \right) d\lambda_{d_1}(x_1),$$

d'où le résultat annoncé.

(b) On rappelle que si X est discrète, $\text{val}(X)$ est dénombrable et que l'on a

$$\forall x_1 \in \text{val}(X_1) \quad (X_1 = x_1) \stackrel{\text{p.p.s.}}{=} \bigcup_{x_2 \in \text{val}(X_2)} [(X_1 = x_1) \cap (X_2 = x_2)],$$

d'où le résultat. □

Remarque. Les propositions 8.16 et 8.18 sont souvent utilisées consécutivement. **L'exemple suivant fera bien ressentir cette association.**

Soit $X = (X_1, X_2)$ une variable aléatoire à valeurs dans \mathbb{R}^2 de densité f_X définie par

$$\forall (x_1, x_2) \in \mathbb{R}^2 \quad f_X(x_1, x_2) = p^2 \mathbf{1}_{\mathbb{R}^+}(x_1) \mathbf{1}_{\mathbb{R}^+}(x_2) \exp[-p(x_1 + x_2)],$$

où $p > 0$. On cherche la loi de la variable aléatoire $Y = 2X_1 - X_2$. Pour cela, on introduit la variable aléatoire (Y, X_2) , transformée de X par le difféomorphisme T de \mathbb{R}^2 sur lui-même défini par $T(x_1, x_2) = (2x_1 - x_2, x_2)$; Y en est la première marginale. La variable aléatoire (Y, X_2) admet la densité $f_{(Y, X_2)}$ définie par

$$\forall (y, x_2) \in \mathbb{R}^2 \quad f_{(Y, X_2)}(y, x_2) = \frac{1}{2} f_X\left(\frac{y + x_2}{2}, x_2\right).$$

La marginale Y admet donc la densité f_Y définie, pour tout $y \in \mathbb{R}$ par

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f_{(Y, X_2)}(y, x_2) d\lambda(x_2) \\ &= \frac{p}{2} \exp\left(-\frac{py}{2}\right) \int_{\mathbb{R}} p \mathbf{1}_{\mathbb{R}^+}(y + x_2) \mathbf{1}_{\mathbb{R}^+}(x_2) \exp\left[-\frac{3px_2}{2}\right] d\lambda(x_2) \\ &= \frac{p}{2} \exp\left(-\frac{py}{2}\right) \int_{\max(-y, 0)}^{+\infty} p \exp\left[-\frac{3px_2}{2}\right] dx_2 \\ &= \frac{p}{3} \exp\left(-\frac{py}{2}\right) \exp\left[-\frac{3p \max(-y, 0)}{2}\right], \end{aligned}$$

soit :

$$\forall y \in \mathbb{R} \quad f_Y(y) = \frac{p}{3} \left[\mathbf{1}_{\mathbb{R}^+}(y) \exp\left(-\frac{py}{2}\right) + \exp(y) \mathbf{1}_{\mathbb{R}^{*-}}(y) \right].$$

8.3. Moments de variables aléatoires

Les moments d'une variable aléatoire, quand ils existent, sont des paramètres qui donnent des renseignements sur la loi de cette variable aléatoire, et quelquefois, même, la déterminent complètement. Avant de donner une définition des moments et d'en étudier les propriétés, nous établissons les inégalités de **Hölder** et de **Minkowski** et en déduisons les premières propriétés des espaces \mathcal{L}^p .

Définition 8.19. Soit un réel $p \geq 1$; on note $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ l'ensemble des variables aléatoires X définies P-p.s., à valeurs dans \mathbb{R} ou \mathbb{R} et telles que $\int_{\Omega} |X|^p dP < +\infty$. Pour un tel X on note

$$\|X\|_p = \left(\int_{\Omega} |X|^p dP \right)^{1/p}.$$

On note $\mathcal{L}^{\infty}(\Omega, \mathcal{A}, P)$ l'ensemble des variables aléatoires X définies P-p.s., à valeurs dans \mathbb{R} ou \mathbb{R} et telles que $\sup \{x \mid P(|X| > x) > 0\} < +\infty$. Pour un tel X on note

$$\|X\|_{\infty} = \sup \{x \mid P(|X| > x) > 0\} = \inf \{x \mid P(|X| > x) = 0\}.$$

on dit alors que X est **essentiellement (ou P-p.s.) bornée**.

Remarque. Si $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$, X est P-p.s. finie.

Si $X \in \mathcal{L}^\infty(\Omega, \mathcal{A}, P)$, on a P-p.s. $|X| \leq \|X\|_\infty$.

Définition 8.20. Deux réels p et q sont **conjugués** s'ils sont strictement positifs et satisfont à l'égalité :

$$\frac{1}{p} + \frac{1}{q} = 1;$$

on a bien sûr $p > 1$ et $q > 1$. On définit de plus le conjugué de 1 comme étant $+\infty$.

Lemme 8.21. Soient p et q deux réels conjugués différents de 1. Pour tous $a, b \in \mathbb{R}^+ :$

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \quad (8.5)$$

Démonstration. La fonction $x \mapsto -\ln x$ étant **convexe** sur \mathbb{R}^{+*} , on a, pour tous $x, y > 0$

$$-\ln\left(\frac{x}{p} + \frac{y}{q}\right) \leq -\frac{1}{p} \ln x - \frac{1}{q} \ln y,$$

soit

$$\ln x^{1/p} + \ln y^{1/q} \leq \ln\left(\frac{x}{p} + \frac{y}{q}\right),$$

et donc :

$$x^{1/p} y^{1/q} \leq \frac{x}{p} + \frac{y}{q}.$$

Il suffit alors de choisir x et y tels que : $a = x^{1/p}$ et $b = y^{1/q}$ pour obtenir l'inégalité (8.5). \square

On déduit de ce lemme l'inégalité suivante :

Proposition 8.22 (Inégalité de Hölder). Soient p et q deux réels conjugués, finis ou non.

(a) Pour toutes variables aléatoires X, Y à valeurs dans $\overline{\mathbb{R}}^+$ et définies P-p.s., on a l'inégalité dans $\overline{\mathbb{R}}^+ :$

$$\int_{\Omega} XY \, dP \leq \left(\int_{\Omega} X^p \, dP \right)^{1/p} \left(\int_{\Omega} Y^q \, dP \right)^{1/q}. \quad (8.6)$$

(b) Si $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ et $Y \in \mathcal{L}^q(\Omega, \mathcal{A}, P)$, le produit XY est intégrable et on a l'inégalité de **Hölder**

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q. \quad (8.7)$$

Si $p = q = 2$, cette inégalité implique l'inégalité de **Schwarz** :

$$\left| \int_{\Omega} XY \, dP \right| \leq \left(\int_{\Omega} X^2 \, dP \right)^{1/2} \left(\int_{\Omega} Y^2 \, dP \right)^{1/2}.$$

Démonstration. (a) Cas de variables aléatoires positives. Si p et q sont finis; si l'un des deux termes du membre de droite de l'inégalité (8.6) est nul, par exemple le premier, alors $X = 0$ P-p.s. et on a alors $\int_{\Omega} XY dP = 0$. Si ces deux termes sont non nuls, il suffit de démontrer l'inégalité (8.6) lorsque les deux termes du membre de droite sont finis. Il résulte du lemme 8.21 que l'on a

$$\frac{X}{\|X\|_p} \frac{Y}{\|Y\|_q} \leq \frac{1}{p} \frac{X^p}{\|X\|_p^p} + \frac{1}{q} \frac{Y^q}{\|Y\|_q^q},$$

ce qui, en intégrant, démontre l'inégalité (8.6).

Si $p = 1$ et $q = +\infty$; on a $0 \leq Y \leq \|Y\|_{\infty}$ P-p.s., et donc $0 \leq XY \leq X \|Y\|_{\infty}$ P-p.s., ce qui, après intégration donne (8.6).

(b) Cas de variables aléatoires de signe quelconque. On applique l'inégalité (8.6) aux valeurs absolues. \square

On en déduit l'inégalité de **Minkowski**.

Proposition 8.23 (Inégalité de Minkowski). Soit un réel $p > 1$, fini ou non.

(a) Pour toutes variables aléatoires X, Y à valeurs dans $\overline{\mathbb{R}}^+$ et définies P-p.s., on a l'inégalité dans $\overline{\mathbb{R}}^+$:

$$\left(\int_{\Omega} (X + Y)^p dP \right)^{1/p} \leq \left(\int_{\Omega} X^p dP \right)^{1/p} + \left(\int_{\Omega} Y^p dP \right)^{1/p}. \quad (8.8)$$

(b) Si $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ et $Y \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$, la somme $X + Y$ est dans $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ et on a l'inégalité de **Minkowski** :

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p. \quad (8.9)$$

Démonstration. (a) Cas de variables aléatoires positives. Si $p > 1$ est fini; par linéarité, on a,

$$\int_{\Omega} (X + Y)^p dP = \int_{\Omega} [(X + Y)^{p-1} X] dP + \int_{\Omega} [(X + Y)^{p-1} Y] dP.$$

En appliquant l'inégalité (8.7) à chacun des facteurs du membre de droite, on obtient :

$$\begin{aligned} \int_{\Omega} (X + Y)^p dP &\leq \left(\int_{\Omega} (X + Y)^{q(p-1)} dP \right)^{1/q} \left(\int_{\Omega} X^p dP \right)^{1/p} \\ &\quad + \left(\int_{\Omega} (X + Y)^{q(p-1)} dP \right)^{1/q} \left(\int_{\Omega} Y^p dP \right)^{1/p}. \end{aligned}$$

Il suffit alors de remarquer que $q(p-1) = p$.

Si $p = +\infty$, on a

$$|X + Y| \leq |X| + |Y| \leq \|X\|_{\infty} + \|Y\|_{\infty} \quad \text{P-p.s.},$$

et donc

$$\|X + Y\|_{\infty} \leq \|X\|_{\infty} + \|Y\|_{\infty}.$$

(b) *Cas de variables aléatoires de signe quelconque.* Les variables aléatoires X et Y étant définies et finies P-p.s., il en est de même de $X + Y$; il suffit alors d'appliquer l'inégalité (8.9) aux valeurs absolues pour obtenir que $(\int_{\Omega} |X + Y|^p dP)^{1/p} < +\infty$. \square

De ces deux inégalités, on déduit des **propriétés des ensembles** $\mathcal{L}^p(\Omega, \mathcal{A}, P)$.

Proposition 8.24. (a) Si $p \geq 1$, $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ est un espace vectoriel semi-normé.

(b) Soient p et q deux entiers tels que : $1 \leq p \leq q \leq +\infty$; on a l'inclusion des ensembles $\mathcal{L}^q(\Omega, \mathcal{A}, P) \subset \mathcal{L}^p(\Omega, \mathcal{A}, P)$ et on a l'inégalité de semi-normes

$$\boxed{\|X\|_p \leq \|X\|_q} \quad (8.10)$$

Démonstration. (a) Pour le premier point, cela résulte de l'inégalité de Minkowski et de ce que l'on a, pour tout réel⁴ c , $\|cX\|_p = |c| \|X\|_p$.

(b) Il suffit d'étudier le cas où p et q sont distincts. Soient alors $X \in \mathcal{L}^q(\Omega, \mathcal{A}, P)$ et r le conjugué de $\frac{q}{p}$, c'est à dire $r = \frac{q}{q-p}$. L'inégalité de Hölder (8.6) permet d'écrire

$$\int_{\Omega} |X|^p \cdot 1 dP \leq \left(\int_{\Omega} (|X|^p)^{q/p} dP \right)^{p/q} \left(\int_{\Omega} 1^r dP \right)^{1/r} = \|X\|_q^p < +\infty,$$

ce qui démontre que $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ et l'inégalité (8.10). \square

Remarque. On a ainsi montré que si $1 \leq p \leq q \leq +\infty$, on a⁵ :

$$\boxed{\mathcal{L}^{\infty}(\Omega, \mathcal{A}, P) \subset \mathcal{L}^q(\Omega, \mathcal{A}, P) \subset \mathcal{L}^p(\Omega, \mathcal{A}, P) \subset \mathcal{L}^1(\Omega, \mathcal{A}, P)} \quad (8.11)$$

Si $p \geq 1$, l'application $X \mapsto \|X\|_p$ est une semi-norme sur $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ et on a $\|X\|_p = 0$ si et seulement si $X = 0$ P-p.s. L'espace vectoriel quotient de $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ par la relation d'équivalence d'égalité P-p.s. est noté $L^p(\Omega, \mathcal{A}, P)$; c'est alors un espace vectoriel normé dont la norme est obtenue par passage au quotient de la semi-norme $X \mapsto \|X\|_p$ (on parle de la « norme p » de X). Il est d'usage de noter de la même façon une variable aléatoire et sa **classe**; on fera de même pour la semi-norme et la norme quotient.

Définition 8.25. Soit X une variable aléatoire réelle définie sur l'espace probabilisé (Ω, \mathcal{A}, P) .

(a) Si $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ la quantité $\int_{\Omega} X dP$ est appelée **moyenne ou espérance mathématique** de X . Elle est notée $E(X)$ ou EX .

4. On rappelle la convention $0 \times (\pm\infty) = 0$.

5. Bien se souvenir que pour une mesure **non bornée** ces relations d'inclusion sont fausses !

(b) Si $X \in \mathcal{L}^\alpha(\Omega, \mathcal{A}, \mathbb{P})$, où $\alpha > 0$, la quantité $\int_{\Omega} X^\alpha d\mathbb{P}$ est appelée **moment d'ordre α** de X . C'est donc aussi la moyenne $E(X^\alpha)$ de la variable aléatoire X^α . En particulier, si $\alpha > 1$ et si $X \in \mathcal{L}^\alpha(\Omega, \mathcal{A}, \mathbb{P})$, la quantité $E[(X - EX)^\alpha]$ est appelée **moment centré d'ordre α** de X .

(c) Si $\alpha = 2$, le moment centré d'ordre 2 est appelé **variance** de X et noté σ_X^2 . Sa racine carrée positive σ_X est appelée **écart-type** de X .

Proposition 8.26. E est une forme linéaire (continue) sur l'espace vectoriel $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$.

Démonstration. C'est un résultat de la théorie de l'intégration (prop. A.21). \square

Définition 8.27. Si $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, la variable aléatoire $\overset{\circ}{X} = X - EX$ est appelée variable aléatoire **centrée** associée à X . Si $X \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, la variable aléatoire $\frac{X}{\sigma_X}$ (resp. $\frac{\overset{\circ}{X}}{\sigma_X}$) est appelée variable aléatoire **réduite** (resp. **centrée réduite**) associée à X .

Remarque. 1. Toutes ces définitions ne présument en rien de la forme de la loi de la variable aléatoire X ; elles recouvrent en particulier les définitions données dans les chapitres précédents pour les variables aléatoires discrètes et à densité (tome 1). **Nous renvoyons donc le lecteur à ces chapitres, tant pour les résultats classiques que pour les exercices concernant de telles variables aléatoires.**

2. Il résulte des relations d'inclusion (8.11) que **si une variable aléatoire admet un moment d'ordre $p \geq 1$, elle admet un moment de tout ordre ≥ 1 et $\leq p$.**

Calcul des moments

Si $X \in \mathcal{L}^\alpha(\Omega, \mathcal{A}, \mathbb{P})$, le théorème de transfert donne :

$$E(X^\alpha) = \int_{\mathbb{R}} x^\alpha dP_X(x).$$

En particulier,

– si X est une **variable aléatoire discrète**, puisqu'alors la loi de X vérifie

$$P_X = \sum_{x \in \text{val}(X)} P(X = x) \delta_x,$$

il vient :

$$E(X^\alpha) = \sum_{x \in \text{val}(X)} x^\alpha P(X = x);$$

- si X est une **variable aléatoire à densité** f_X , puisqu'alors la loi de X vérifie $P_X = f_X \cdot \lambda$, il vient, par le théorème d'intégration par rapport à une mesure à densité :

$$E(X^\alpha) = \int_{\mathbb{R}} x^\alpha f_X(x) d\lambda(x).$$

Proposition 8.28. Si $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$, sa variance vérifie :

$$\sigma_X^2 = E(X^2) - [EX]^2 \quad \text{et} \quad \forall(a, b) \in \mathbb{R} \quad \sigma_{aX+b}^2 = a^2\sigma_X^2.$$

Démonstration. Développer le carré et appliquer la linéarité de l'espérance. \square

L'écart-type est donc invariant par translation et *positivement* homogène.

Définition 8.29. Si X et Y appartiennent à $\mathcal{L}^2(\Omega, \mathcal{A}, P)$, il résulte de l'inégalité de Schwarz que la variable aléatoire $(X - EX)(Y - EY)$ appartient à $\mathcal{L}^1(\Omega, \mathcal{A}, P)$; la quantité $E[(X - EX)(Y - EY)]$ est appelée **covariance** de X et Y . Elle est notée $\text{cov}(X, Y)$.

Proposition 8.30. Si X et Y appartiennent à $\mathcal{L}^2(\Omega, \mathcal{A}, P)$, on a :

$$\text{cov}(X, Y) = E(XY) - (EX)(EY) \quad \text{et} \quad \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\text{cov}(X, Y).$$

Démonstration. Pour la première égalité, développer le produit et appliquer la linéarité de l'espérance, pour la seconde remarquer que :

$$\sigma_{X+Y}^2 = E\left[\overset{\circ}{(X + Y)}^2\right].$$

Développer alors le carré et appliquer la linéarité de l'espérance. \square

Nous généralisons ces notions au cas où la variable aléatoire est à **valeurs dans un espace vectoriel de dimension finie**; pour un tel espace F , on note de manière générique, F^* son dual algébrique et $\langle \cdot, \cdot \rangle$ la forme bilinéaire de dualité. Nous rappelons que toutes les normes sur F sont équivalentes; on note $\|\cdot\|$ l'une d'elle. L'espace F est muni de sa tribu borélienne \mathcal{F} (engendrée par les ouverts de F). Généralement, dans les applications, F sera un espace **euclidien**, la forme bilinéaire étant alors le produit scalaire; F sera alors identifié à son dual.

La présentation adoptée a pour but de définir des moments de manière intrinsèque. Le lecteur pourra se contenter de penser que F est l'espace \mathbb{R}^d muni du produit scalaire canonique.

Proposition 8.31. Soit X une variable aléatoire à valeurs dans F et $p \in [1, +\infty]$. Les assertions suivantes sont équivalentes :

- (i) $\|X\| \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$
 (ii) $\forall x^* \in F^* \quad \langle X, x^* \rangle \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$.

Démonstration. Si F^* est muni de la norme définie par

$$\|x^*\| = \sup_{\|x\| \leq 1} \langle x, x^* \rangle,$$

l'implication (i) donne (ii) résulte de l'inégalité :

$$\forall x^* \in F^* \quad |\langle X, x^* \rangle| \leq \|X\| \|x^*\|$$

Pour l'implication inverse, soit $(e_i)_{i=1, \dots, d}$ une base de F et $(e_i^*)_{i=1, \dots, d}$ la base duale dans F^* ; on a :

$$\|X\| \leq \sum_{i=1}^d |\langle X, e_i^* \rangle| \|e_i\|.$$

La fonction $x \mapsto x^p$ étant croissante sur \mathbb{R}^+ , il suffit d'appliquer l'inégalité de Minkowski. \square

Notation. $\mathcal{L}_F^p(\Omega, \mathcal{A}, P)$ est l'espace vectoriel des variables aléatoires à valeurs dans F telles que $\|X\| \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$.

Définition 8.32. Soit $X \in \mathcal{L}_F^1(\Omega, \mathcal{A}, P)$. La forme linéaire sur $F^* : x^* \mapsto \int_{\Omega} \langle X, x^* \rangle dP$ est appelée **moyenne** de X et notée $E(X)$. Identifiant F et son bidual, c'est à dire le dual algébrique⁶ de F^* , la moyenne $E(X)$ est l'unique élément de F qui satisfait à :

$$\boxed{\forall x^* \in F^* \quad \langle E(X), x^* \rangle = E \langle X, x^* \rangle.} \quad (8.12)$$

Remarque. 1. On note de la même façon l'opérateur moyenne sur $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ et $\mathcal{L}_F^1(\Omega, \mathcal{A}, P)$: si $F = \mathbb{R}$, ils coïncident. L'opérateur E est encore linéaire sur $\mathcal{L}_F^1(\Omega, \mathcal{A}, P)$.

2. Si $F = \mathbb{R}^d$ muni de sa base canonique, il résulte de (8.12) que $E(X)$ est le vecteur de \mathbb{R}^d de i^e composante $E(X_i)$.

Proposition 8.33. Soit $X \in \mathcal{L}_F^1(\Omega, \mathcal{A}, P)$; soient $A \in \mathcal{L}(F, G)$ et $b \in G$, où G est un autre espace vectoriel de dimension finie. Alors la variable aléatoire $AX + b \in \mathcal{L}_G^1(\Omega, \mathcal{A}, P)$ et :

$$\boxed{E(AX + b) = A E(X) + b.}$$

6. Le fait que F soit de dimension finie garantit que F et son bidual sont isomorphes ; ce n'est plus le cas en dimension infinie. Dans ce cas, la définition de la moyenne peut poser problème.

Démonstration. On a

$$\|AX + b\| \leq \|A\| \|X\| + \|b\| ,$$

ce qui démontre que $AX + b \in \mathcal{L}_G^1(\Omega, \mathcal{A}, \mathbf{P})$. De plus, en utilisant la définition du transposé de A , on a

$$\begin{aligned} \forall y^* \in G^* \quad \langle E(AX + b), y^* \rangle &= E \langle AX + b, y^* \rangle \\ &= E [\langle X, A^* y^* \rangle + \langle b, y^* \rangle] \\ &= \langle EX, A^* y^* \rangle + \langle b, y^* \rangle \\ &= \langle AE(X) + b, y^* \rangle , \end{aligned}$$

ce qui démontre le résultat. \square

Définition 8.34. Soit $X \in \mathcal{L}_F^2(\Omega, \mathcal{A}, \mathbf{P})$. La forme quadratique positive sur $F^* : x^* \rightarrow \int_{\Omega} \langle X - EX, x^* \rangle^2 d\mathbf{P}$ est appelée **variance** de X et notée $\sigma_X^2(\cdot)$. Elle est associée de manière bijective à l'opérateur linéaire auto-transposé positif $\Lambda_X \in \mathcal{L}(F^*, F)$ par la relation :

$$\boxed{\forall x^* \in F^* \quad \langle \Lambda_X x^*, x^* \rangle = \sigma_X^2(x^*) .}$$

Cet opérateur est appelé **opérateur d'auto-covariance** de X . Si $(e_i)_{i=1, \dots, d}$ est une base de F et $(e_i^*)_{i=1, \dots, d}$ la base duale dans F^* , la représentation matricielle C_X de l'opérateur Λ_X dans ces bases est appelée **matrice des covariances** de X . Elle vérifie :

$$\boxed{\forall i, j = 1, \dots, d \quad (C_X)_{ij} = \text{cov}(\langle X, e_i^* \rangle, \langle X, e_j^* \rangle) .}$$

Remarque. Si $F = \mathbb{R}^d$ est muni de sa base canonique, C_X est la matrice symétrique positive $d \times d$:

$$C_X = \begin{pmatrix} \sigma_{X_1}^2 & & & & & & & & \\ & \ddots & & & & & & & \\ & & \sigma_{X_i}^2 & \cdots & \text{cov}(X_j, X_i) & & & & \\ & & \vdots & \ddots & \vdots & & & & \\ & & \text{cov}(X_i, X_j) & \cdots & \sigma_{X_j}^2 & & & & \\ & & & & & \ddots & & & \\ & & & & & & \sigma_{X_d}^2 & & \end{pmatrix}$$

Proposition 8.35. Soit $X \in \mathcal{L}_F^2(\Omega, \mathcal{A}, \mathbf{P})$; soient $A \in \mathcal{L}(F, G)$ et $b \in G$, où G est un autre espace vectoriel normé de dimension finie. Alors la variable aléatoire $AX \in \mathcal{L}_G^2(\Omega, \mathcal{A}, \mathbf{P})$ et on a

$$\boxed{\Lambda_{AX+b} = A \Lambda_X A^* ,}$$

ce qui s'écrit, dans le cas de matrice de covariance :

$$\boxed{C_{AX+b} = AC_X A^* .}$$

Démonstration. On a

$$\|AX + b\|^2 \leq \frac{1}{2} (\|A\|^2 \|X\|^2 + \|b\|^2),$$

ce qui démontre que $AX + b \in \mathcal{L}_G^2(\Omega, \mathcal{A}, P)$. De plus, en utilisant la définition du transposé de A et la proposition 8.33, on a

$$\begin{aligned} \forall y^* \in G^* \quad \langle \Lambda_{AX+b} y^*, y^* \rangle &= E[\langle AX, y^* \rangle^2] \\ &= E[\langle \overset{\circ}{X}, A^* y^* \rangle]^2 \\ &= \langle \Lambda_X A^* y^*, A^* y^* \rangle \\ &= \langle A \Lambda_X A^* y^*, y^* \rangle ; \end{aligned}$$

par bilinéarisation, il en résulte que :

$$\forall x^*, y^* \in G^* \quad \langle \Lambda_{AX+b} x^*, y^* \rangle = \langle A \Lambda_X A^* x^*, y^* \rangle ,$$

ce qui démontre le résultat. \square

Nous donnons deux inégalités classiques, bien que grossières, qui permettent de donner quelques renseignements sur la concentration des valeurs prises par une variable aléatoire, en particulier autour de sa moyenne ; elles sont numériquement très mauvaises, ce qui ne surprend pas quand on étudie leur démonstration. **Ces inégalités servent surtout à démontrer des convergences en probabilité** (voir chapitre 10).

Proposition 8.36 (Inégalité de Markov). *Si $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ est positive, on a, pour tout $\varepsilon > 0$:*

$$P(X \geq \varepsilon) \leq \frac{EX}{\varepsilon} ,$$

et a fortiori

$$\boxed{P(X > \varepsilon) \leq \frac{EX}{\varepsilon} .}$$

En conséquence, si $X \in \mathcal{L}_F^1(\Omega, \mathcal{A}, P)$, on a, pour tout $\varepsilon > 0$:

$$\boxed{P(\|X\| > \varepsilon) \leq \frac{E\|X\|}{\varepsilon} .}$$

Démonstration. Si $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ est positive, soit $D = \{X \geq \varepsilon\}$. On a les minoration successives :

$$EX = \int_{\Omega} X dP \geq \int_D X dP \geq \varepsilon P(D) ,$$

d'où la première inégalité ; la seconde résulte alors de l'inclusion $(X > \varepsilon) \subset (X \geq \varepsilon)$.

Si $X \in \mathcal{L}_F^1(\Omega, \mathcal{A}, P)$, il suffit d'appliquer l'inégalité précédente à la variable aléatoire positive $\|X\|$. \square

Proposition 8.37 (Inégalité de Bienaymé-Tchebitchev). *Si $X \in \mathcal{L}_F^2(\Omega, \mathcal{A}, P)$, où F est un espace euclidien, on a, pour tout $\varepsilon > 0$:*

$$P(\|X - EX\| > \varepsilon) \leq \frac{\text{tr}(\Lambda_X)}{\varepsilon^2}.$$

En particulier, si $F = \mathbb{R}$, on a :

$$P(|X - EX| > \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2}.$$

Démonstration. Il suffit d'appliquer l'inégalité de Markov à la variable aléatoire positive $\|X - EX\|^2$ et au réel ε^2 et de remarquer que

$$(\|X - EX\| > \varepsilon) = (\|X - EX\|^2 > \varepsilon^2)$$

et que

$$E\|\overset{\circ}{X}\|^2 = \text{tr}(\Lambda_X) :$$

en particulier, si $F = \mathbb{R}$, on a :

$$E(\overset{\circ}{X})^2 = \sigma_X^2. \quad \square$$

L'inégalité de Markov a pour conséquence les inégalités suivantes qui mettent sur la voie des inégalités de **Bernstein**, amélioration de celle de Tchebitchev, et qui sont le départ de la théorie des grandes déviations.

Proposition 8.38. *Soient f une fonction de \mathbb{R} dans \mathbb{R} , croissante et strictement positive, et X une variable aléatoire réelle telles que $f \circ X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Alors, pour tout réel ε :*

$$P(X > \varepsilon) \leq \frac{E f(X)}{f(\varepsilon)}.$$

En particulier, si X est une variable aléatoire réelle telle que pour un $a > 0$ on ait $\exp(aX) \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, alors on a, pour tout réel ε :

$$P(X > \varepsilon) \leq \exp(-a\varepsilon) E[\exp(aX)].$$

Démonstration. Il suffit de remarquer que, puisque f est croissante, on a

$$(X > \varepsilon) \subset (f(X) \geq f(\varepsilon)),$$

et donc :

$$P(X > \varepsilon) \leq P[f(X) \geq f(\varepsilon)].$$

L'inégalité de Markov (avec inégalité large) appliquée à la variable aléatoire $f(X)$ et au réel positif $f(\varepsilon)$ permet de conclure. La seconde inégalité s'obtient à partir de la première en prenant $f(x) = \exp(ax)$. \square

On introduit la notion de **coefficient de corrélation** qui, comme on le verra ensuite permet de « mesurer » une certaine liaison entre des variables aléatoires.

Définition 8.39. Soient X et $Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ de variance non nulle. On appelle **coefficient de corrélation** de X et Y le réel

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Proposition 8.40. Soient X et $Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ de variance non nulle ; leur coefficient de corrélation $\rho_{X,Y}$ possède les propriétés suivantes :

(a) $|\rho_{X,Y}| \leq 1$;

(b) Pour que $|\rho_{X,Y}| = 1$, il faut et il suffit qu'il existe trois réels a, b, c non tous nuls tels que :

$$P(aX + bY + c = 0) = 1.$$

Démonstration. (a) L'inégalité de Schwarz permet d'écrire :

$$|E(\overset{\circ}{X}\overset{\circ}{Y})| \leq E|\overset{\circ}{X}\overset{\circ}{Y}| \leq (E\overset{\circ}{X}^2)^{\frac{1}{2}}(E\overset{\circ}{Y}^2)^{\frac{1}{2}},$$

ce qui démontre l'inégalité annoncée.

(b) Si $|\rho_{X,Y}| = 1$, le polynôme du second degré en λ , $E[\overset{\circ}{X} + \lambda\overset{\circ}{Y}]^2$, a son discriminant réduit nul et admet donc une racine double λ_0 ; on a alors $E[\overset{\circ}{X} + \lambda_0\overset{\circ}{Y}]^2 = 0$. Il en résulte que $P(\overset{\circ}{X} + \lambda_0\overset{\circ}{Y} = 0) = 1$.

Réciproquement, supposons qu'il existe trois réels a, b, c non tous nuls tels que

$$P(aX + bY + c = 0) = 1. \quad (8.13)$$

Si $c \neq 0$, a et b sont différents de 0 ; en effet, si par exemple $a = 0$, on a alors $P(bY + c = 0) = 1$ et donc $\sigma_{bY+c}^2 = \sigma_0^2 = 0$, soit encore $b^2\sigma_Y^2 = 0$ et donc $b = 0$, ce qui est impossible d'après (8.13). Dans ce cas on a

$$P(X = \alpha Y + \beta) = 1, \quad (8.14)$$

où $\alpha \neq 0$.

Si $c = 0$, on a $a \neq 0$ ou $b \neq 0$. Si par exemple $a \neq 0$, l'égalité (8.14) est encore satisfaite avec $\beta = 0$ (si c'était b , on ferait un calcul analogue).

Dans ces deux cas on a donc

$$\text{cov}(X, Y) = E[(\alpha\overset{\circ}{Y})\overset{\circ}{Y}] = \alpha \sigma_Y^2 \quad \text{et} \quad \sigma_X^2 = \sigma_{\alpha Y + \beta}^2 = \alpha^2 \sigma_Y^2,$$

ce qui donne $\rho_{X,Y} = \frac{\alpha \sigma_Y^2}{|\alpha| \sigma_Y^2}$, et donc : $|\rho_{X,Y}| = 1$. □

Problème de régression linéaire

Les variables aléatoires X et $Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ étant données, on cherche une « meilleure » approximation de Y comme fonction affine de X au sens des moindres carrés, à savoir une solution en le couple $(a, b) \in \mathbb{R}^2$ du problème de minimisation

$$\inf\{\Phi(a, b) \mid (a, b) \in \mathbb{R}^2\},$$

où

$$\Phi(a, b) = E[Y - (aX + b)]^2.$$

Ce problème est appelé, improprement d'ailleurs, **problème de régression linéaire**.

Il n'y a formellement pas un mot à changer à l'analyse que nous avons faite au tome 1 pour les variables aléatoires discrètes. Nous redonnons toutefois la résolution de ce problème dans ce cadre général. On a :

$$\begin{aligned} \Phi(a, b) &= E[\overset{\circ}{Y} - a\overset{\circ}{X} + (EY - aEX - b)]^2 \\ &= E[\overset{\circ}{Y} - a\overset{\circ}{X}]^2 + [EY - aEX - b]^2 \end{aligned}$$

Pour tout a fixé, cette quantité est minimum pour $b = \hat{b}_a = EY - aEX$. Reste à minimiser en a la quantité

$$\Phi(a, \hat{b}_a) \equiv \sigma_Y^2 - 2a \operatorname{cov}(X, Y) + a^2 \sigma_X^2;$$

une valeur \hat{a} minimisant ce polynôme du second degré est l'unique solution de l'équation

$$-\operatorname{cov}(X, Y) + a \sigma_X^2 = 0.$$

La solution du problème de régression linéaire est le couple (\hat{a}, \hat{b}_a) défini par

$$\begin{cases} \hat{a} &= \frac{\operatorname{cov}(X, Y)}{\sigma_X^2} &= \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \\ \hat{b}_a &= EY - \hat{a} EX &= EY - EX \cdot \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}. \end{cases}$$

La droite D d'équation

$$(y - EY) - \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (x - EX) = 0$$

est appelée **droite de régression linéaire de Y en X** . La « meilleure » approximation de Y comme fonction affine de X au sens des moindres carrés est

$$EY + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (X - EX)$$

et on a $P[(X, Y) \in D] = 1$ si et seulement si $\Phi(\hat{a}, \hat{b}_a) = 0$.

Cas particulier. Si la variable aléatoire est de loi uniforme sur l'ensemble des n points du plan $\{(x_i, y_i)\}_{1 \leq i \leq n}$ alors $\Phi(a, b) = \frac{1}{n} \sum_{i=1}^n [y_i - (ax_i + b)]^2$. On retrouve la droite d'approximation des moindres carrés des physiciens. (Exercice : déterminer alors l'équation de la droite D).

Les lois les plus courantes

Ci-dessous un tableau des lois μ les plus courantes, avec leurs moyenne m et variance σ^2 quand elles en admettent ; on donne aussi leur transformée de Fourier $\hat{\mu}$ (voir plus loin pour une définition).

Lois discrètes

Nom (paramètres)	Mesure de probabilité μ	$\hat{\mu}(t)$	m	σ^2
Loi de Bernoulli $\mathcal{B}(1, p)$ ($0 < p < 1, q = 1 - p$).	$p \delta_1 + q \delta_0$	$pe^{it} + q$	p	pq
Loi binomiale $\mathcal{B}(n, p)$ ($0 < p < 1, n$ entier > 0).	$\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \delta_k$	$(pe^{it} + q)^n$	np	npq
Loi de Poisson $\mathcal{P}(\lambda)$ ($\lambda > 0$)	$\sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \delta_n$	$\exp[\lambda(e^{it} - 1)]$	λ	λ
Loi géométrique sur \mathbb{N} , $\mathcal{G}_{\mathbb{N}}(p)$ ($0 < p < 1$)	$\sum_{n=0}^{\infty} pq^n \delta_n$	$\frac{p}{1 - qe^{it}}$	$\frac{q}{p}$	$\frac{q}{p^2}$
Loi géométrique sur \mathbb{N}^* , $\mathcal{G}_{\mathbb{N}^*}(p)$ ($0 < p < 1$)	$\sum_{n=1}^{\infty} pq^{n-1} \delta_n$	$\frac{pe^{it}}{1 - qe^{it}}$	$\frac{1}{p}$	$\frac{q}{p^2}$

Lois à densité $\mu = f \cdot \lambda$

Nom (paramètres)	Densité $f(x)$ ($x \in \mathbb{R}$)	$\hat{\mu}(t)$ ($t \in \mathbb{R}$)	m	σ^2
Loi uniforme sur $[a, b]$ ($a < b$)	$\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$	$\frac{e^{itb} - e^{ita}}{it(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Loi de Cauchy	$\frac{1}{\pi} \frac{1}{1+x^2}$	$e^{- t }$	n'existent pas	

Lois à densité $\mu = f \cdot \lambda$ (suite)

Nom (paramètres)	Densité $f(x)$ ($x \in \mathbb{R}$)	$\hat{\mu}$	m	σ^2
Loi de Gauss $\mathcal{N}(m, \sigma^2)$ (m réel, $\sigma^2 > 0$)	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	$e^{imt} - \frac{t^2\sigma^2}{2}$	m	σ^2
Première loi de Laplace	$\frac{1}{2} e^{- x }$	$\frac{1}{1+t^2}$	0	2
Loi exponentielle $\exp(p)$ ($p > 0$)	$\mathbf{1}_{\mathbb{R}^+}(x) p e^{-px}$	$\frac{1}{1 - \frac{it}{p}}$	$\frac{1}{p}$	$\frac{1}{p^2}$
Loi Gamma $\gamma(a, p)$ ($a > 0, p > 0$)	$\mathbf{1}_{\mathbb{R}^+}(x) \frac{p^a}{\Gamma(a)} e^{-px} x^{a-1}$	$(1 - \frac{it}{p})^{-a}$	$\frac{a}{p}$	$\frac{a}{p^2}$
Loi du chi-deux à n degrés de liberté χ_n^2	$\mathbf{1}_{\mathbb{R}^+}(x) \frac{e^{-\frac{x}{2}} x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}$	$(1-2it)^{-\frac{n}{2}}$	n	$2n$
Loi Bêta de 1 ^{re} espèce ($a > 0, b > 0$)	$\frac{1}{\mathbf{B}(a, b)} x^{a-1} (1-x)^{b-1}$ ($x \in [0, 1]$)		$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
Loi Bêta de 2 ^e espèce ($a > 0, b > 0$)	$\frac{1}{\mathbf{B}(a, b)} \frac{x^{a-1}}{(1+x)^{a+b}}$ ($x \in \mathbb{R}^+$)		$\frac{a}{b-1}$ si $b > 1$; n'existe pas si $b \leq 1$	$\frac{a(a+b-1)}{(b-1)^2(b-2)}$ si $b > 2$; n'existe pas si $b \leq 2$
Loi de Student à n degrés de liberté	$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} (1 + \frac{t^2}{n})^{-\frac{n+1}{2}}$		0	$\frac{n}{n-2}$ si $n > 2$; n'existe pas si $n = 2$

Remarque. Une loi exponentielle $\exp(p)$ est une loi $\gamma(1, p)$. Une loi du chi-deux à n degrés de liberté est une loi $\gamma(\frac{n}{2}, \frac{1}{2})$. Une loi uniforme sur $[0, 1]$ est une loi Bêta de première espèce $\mathbf{B}(1, 1)$.

Rappels : fonctions eulériennes \mathbf{B} (Bêta) et Γ (Gamma). On a

$$\Gamma(a) = \int_0^{+\infty} \exp(-x) x^{a-1} dx \quad (a > 0)$$

$$\Gamma(a) = (a-1) \Gamma(a-1) \quad (a > 1)$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

$$\mathbf{B}(a, b) = \int_0^{+\infty} \frac{x^{a-1}}{(1+x)^{a+b}} dx = \int_0^1 x^{a-1} (1-x)^{b-1} dx. \quad (a, b > 0)$$

$$\mathbf{B}(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)}.$$

Exercices

Toutes les variables aléatoires introduites sont définies sur un même espace probabilisé (Ω, \mathcal{A}, P) .

Exercice 8.1. Résultat fondamental pour la simulation de lois de probabilité. Soit X une variable aléatoire de fonction de répartition F . On définit la fonction G de la variable réelle par

$$\forall t \in \mathbb{R} \quad G(t) = \inf\{x \mid F(x) \geq t\};$$

cette fonction G est appelée **pseudo-inverse** de F .

1. Démontrer successivement que

- (a) si F est continue, on a, pour tout $t \in]0, 1[$, $F[G(t)] = t$;
- (b) si F est strictement croissante, on a, pour tout $x \in \mathbb{R}$, $G[F(x)] = x$;
- (c) si F est continue et strictement croissante, F est bijective de \mathbb{R} sur $]0, 1[$ et on a $G = F^{-1}$.

2. Démontrer que si F est continue et strictement croissante, $F(X)$ suit la loi uniforme sur $[0, 1]$.

3. Démontrer que si Y suit la loi uniforme sur $[0, 1]$, la variable aléatoire $G(Y)$ admet F comme fonction de répartition.

Solution.

1. On sépare bien l'influence des hypothèses de monotonie et de continuité.

- (a) Soit, pour tout $t \in]0, 1[$, l'ensemble $A_t = \{x \mid F(x) \geq t\}$. C'est une *demi-droite* car, F étant croissante, pour tout $x_0 \in A_t$ et tout $y \geq x_0$, on a $F(y) \geq F(x_0) \geq t$, et donc, $y \in A_t$. De plus, on a

$$F[G(t)] \geq t;$$

en effet, puisque $G(t) = \inf A_t$ et que A_t est une demi-droite, pour tout $y > G(t)$, on a $y \in A_t$ et donc $F(y) \geq t$. La fonction F étant *continue à droite*, il suffit de prendre une suite de réels $y_n \in A_t$ convergeant en décroissant vers $G(t)$, pour obtenir :

$$F[G(t)] = \lim_n \searrow F(y_n) \geq t.$$

Si F est continue, on a de plus :

$$F[G(t)] \leq t;$$

en effet, par définition de G , pour tout $y < G(t)$, on a $F(y) < t$. La fonction F étant *continue à gauche*, il suffit de prendre une suite de réels $y_n < G(t)$ convergeant en croissant vers $G(t)$, pour obtenir :

$$F[G(t)] = \lim_n \nearrow F(y_n) \leq t.$$

Au total, si F est continue, on a, pour tout $t \in]0, 1[$, $F[G(t)] = t$.

(b) Pour tout $x \in \mathbb{R}$, par définition de G , on a

$$G[F(x)] = \inf\{y \mid F(y) \geq F(x)\} \leq x;$$

de plus, si F est *strictement* croissante, pour tout y tel que $F(y) \geq F(x)$, on a $y \geq x$ (sinon, on aurait $y < x$, et donc $F(y) < F(x)$), ce qui démontre que $G[F(x)] \geq x$. Il en résulte que $G[F(x)] = x$.

(c) Si F est *continue et strictement croissante*, on a à la fois :

$$\forall t \in]0, 1[\quad F[G(t)] = t \quad \text{et} \quad \forall x \in \mathbb{R} \quad G[F(x)] = x.$$

Il en résulte que F est bijective de \mathbb{R} sur $]0, 1[$ et que : $G = F^{-1}$.

2. Si F est continue, pour tout $y \in]0, 1[$, on a $F[G(y)] = y$ et donc, en utilisant la croissance stricte de F ,

$$P[F(X) \leq y] = P[F(X) \leq F[G(y)]] = P[X \leq G(y)] = F[G(y)] = y.$$

Puisque de plus on a

$$P[F(X) \leq y] = \begin{cases} 0 & \text{si } y < 0 \\ 1 & \text{si } y > 1, \end{cases}$$

$F(X)$ suit la loi uniforme sur $[0, 1]$.

3. On a l'équivalence : $F(x) \geq t \iff x \geq G(t)$; donc, si Y suit la loi uniforme sur $[0, 1]$, on a :

$$\forall x \in \mathbb{R} \quad P[G(Y) \leq x] = P[Y \leq F(x)] = F(x),$$

ce qui démontre que $G(Y)$ *admet* F *comme fonction de répartition*.

Remarque. Les résultats de cet exercice permettent théoriquement de **simuler toute loi sur \mathbb{R}** à partir d'une variable aléatoire de loi uniforme. En effet, un appel à la fonction « random » (ou « rand » ou autre appellation, suivant les langages) d'un ordinateur est censé donner une réalisation y d'une variable aléatoire Y de loi uniforme sur $[0, 1]$, ce nombre « aléatoire », ou au hasard, étant fabriqué par un générateur uniforme. Si on veut simuler une variable aléatoire réelle de fonction de répartition F , on calcule (lorsque c'est faisable) sa pseudo-inverse G , et $G(y)$ est alors une réalisation de la variable aléatoire $G(Y)$ de fonction de répartition F . Cette méthode peut être numériquement très lourde, ou même impraticable; des **méthodes spécifiques** existent pour simuler un certain nombre de lois classiques, comme on le met en évidence dans les exercices suivants.

Exercice 8.2. Simulation de lois de variables aléatoires discrètes. Soit X une variable aléatoire réelle discrète prenant les valeurs d'une suite strictement croissante $(x_n)_{n \in \mathbb{N}}$ telle que, pour tout $n \in \mathbb{N}$, $P(X = x_n) = p_n \geq 0$, avec $\sum_{n=0}^{+\infty} p_n = 1$. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$, et Y la variable aléatoire définie par

$$Y = x_0 \mathbf{1}_{(U < p_0)} + \sum_{n=1}^{+\infty} x_n \mathbf{1}_{(p_0 + p_1 + \dots + p_{n-1} < U < p_0 + p_1 + \dots + p_n)}.$$

Vérifier que X et Y ont même loi.

Solution. On a

$$P(Y = x_0) = P(U < p_0) = p_0$$

et, pour $n \geq 1$,

$$P(Y = x_n) = P(p_0 + p_1 + \dots + p_{n-1} < U < p_0 + p_1 + \dots + p_n) = p_n.$$

Remarque. Le résultat de cet exercice permet donc de simuler toute **loi discrète** sur \mathbb{R} à partir d'une variable aléatoire de loi uniforme.

Exercice 8.3. Simulation de la loi exponentielle. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$, et X la variable aléatoire définie par $X = -\frac{1}{p} \ln(U)$, où $p > 0$. Déterminer la loi de X .

Solution. En appliquant le théorème de transfert, puis en effectuant le changement de variable associé au difféomorphisme de $]0, 1[$ sur \mathbb{R}^{+*} défini par $v = -\frac{1}{p} \ln(u)$, il vient :

$$\begin{aligned} \forall f \in \mathcal{C}_X^+(\mathbb{R}) \quad E f(X) &= E f\left[-\frac{1}{p} \ln(U)\right] \\ &= \int_{\mathbb{R}} f\left[-\frac{1}{p} \ln(u)\right] \mathbf{1}_{]0, 1[}(u) d\lambda(u) \\ &= \int_{\mathbb{R}^{+*}} f[v] \exp(-pv) p d\lambda(v), \end{aligned}$$

ce qui démontre que P_X est la loi $\exp(p)$. *Remarque.* Le résultat de cet exercice est couramment utilisé pour simuler une **loi exponentielle** à partir d'une variable aléatoire de loi uniforme, sans avoir recours à la méthode générale (mais plus lourde) d'inversion de la fonction de répartition.

Exercice 8.4. Loïs normales dans \mathbb{R}^2 , lois exponentielle et de Hotelling. Soit $X = (X_1, X_2)$ une variable aléatoire à valeurs dans \mathbb{R}^2 de loi normale $\mathcal{N}_{\mathbb{R}^2}(0, \mathbf{I})$, c'est à dire admettant une densité f_X donnée par

$$\forall x \in \mathbb{R}^2 \quad f_X(x) = \frac{1}{2\pi} \exp\left(-\frac{\|x\|^2}{2}\right),$$

où $\|\cdot\|$ désigne la norme euclidienne usuelle.

1. Déterminer la loi de la variable aléatoire $\|X\|^2$.

2. Soit $D = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = x_2\}$. Démontrer que la variable aléatoire T définie par

$$T = \begin{cases} \left(\frac{X_1 + X_2}{X_1 - X_2}\right)^2 & \text{si } X \notin D \\ () & \text{sinon} \end{cases}$$

admet une densité ; la calculer (la loi de T est appelée **loi de Hotelling**⁷).

7. Cette loi apparaît dans l'étude du test de comparaison d'une moyenne expérimentale à une moyenne théorique (cf., par exemple, C. Fourgeaud et A. Fuchs, *Statistique*, Dunod, p. 129)

Solution.

1. Par application des théorèmes de transfert et d'intégration par rapport à une mesure à densité, on a, pour tout $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R})$,

$$\begin{aligned} E f(\|X\|^2) &= \int_{\mathbb{R}^2} f(\|x\|^2) \frac{1}{2\pi} \exp\left(-\frac{\|x\|^2}{2}\right) d\lambda_2(x) \\ &= \int_{\mathbb{R}^2 \setminus \mathbb{R}^+ \times \{0\}} f(\|x\|^2) \frac{1}{2\pi} \exp\left(-\frac{\|x\|^2}{2}\right) d\lambda_2(x). \end{aligned}$$

Effectuons le changement de variables en coordonnées polaires associé au difféomorphisme de $\mathbb{R}^+ \times]0, 2\pi[$ sur $\mathbb{R}^2 \setminus \mathbb{R}^+ \times \{0\}$ défini par

$$x = \rho \cos \theta \quad y = \rho \sin \theta,$$

de jacobien ρ , et appliquons le théorème de Fubini :

$$\begin{aligned} E f(\|X\|^2) &= \int_{\mathbb{R}^+ \times]0, 2\pi[} f(\rho^2) \frac{1}{2\pi} \exp\left(-\frac{\rho^2}{2}\right) \rho d(\lambda \otimes \lambda)(\rho, \theta) \\ &= \int_{\mathbb{R}^+} f(\rho^2) \exp\left(-\frac{\rho^2}{2}\right) \rho d\lambda(\rho). \end{aligned}$$

Par un dernier changement de variables associé au difféomorphisme de \mathbb{R}^+ sur lui-même défini par $u = \rho^2$, on obtient :

$$\forall f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R}) \quad E f(\|X\|^2) = \int_{\mathbb{R}} f(u) \mathbf{1}_{\mathbb{R}^+}(u) \frac{1}{2} \exp\left(-\frac{u}{2}\right) d\lambda(u),$$

ce qui démontre que $\|X\|^2$ a la loi exponentielle $\exp(-\frac{1}{2})$.

2. Par application des théorèmes de transfert et d'intégration par rapport à une mesure à densité, on a pour tout $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R})$ (puisque $\lambda_2(D) = 0$) :

$$E f(T) = \int_{\mathbb{R}^2 \setminus D} f\left[\left(\frac{x_1 + x_2}{x_1 - x_2}\right)^2\right] \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) d\lambda_2(x_1, x_2).$$

Effectuons le changement de variables associé au difféomorphisme de $\mathbb{R}^2 \setminus D$ sur $\mathbb{R}^2 \setminus (\{0\} \times \mathbb{R})$ défini par

$$\begin{cases} u = \frac{x_1 + x_2}{x_1 - x_2} \\ v = x_1 + x_2 \end{cases} \iff \begin{cases} x_1 = \frac{1}{2}(v + \frac{v}{u}) \\ x_2 = \frac{1}{2}(v - \frac{v}{u}), \end{cases}$$

de jacobien

$$\frac{D(x_1, x_2)}{D(u, v)} = \det \begin{pmatrix} -\frac{v}{2u^2} & \frac{1}{2}(1 + \frac{1}{u}) \\ \frac{v}{2u^2} & \frac{1}{2}(1 - \frac{1}{u}) \end{pmatrix} = -\frac{v}{2u^2},$$

et appliquons le théorème de Fubini; on obtient, pour tout $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R})$,

$$\begin{aligned} E f(T) &= \frac{1}{2\pi} \int_{\mathbb{R}^2 \setminus \{0\} \times \mathbb{R}} f(u^2) \exp\left[-\frac{1}{4}v^2\left(1 + \frac{1}{u^2}\right)\right] \frac{v}{2u^2} d\lambda \otimes \lambda(u, v) \\ &= \frac{1}{4\pi} \int_{\mathbb{R}^*} \frac{f(u^2)}{u^2} \left[\int_{\mathbb{R}} |v| \exp\left[-\frac{1}{4}v^2\left(1 + \frac{1}{u^2}\right)\right] d\lambda(v) \right] d\lambda(u). \end{aligned}$$

Mais, par comparaison des intégrales de Lebesgue et Riemann généralisées, on a

$$\begin{aligned} \int_{\mathbb{R}} |v| \exp\left[-\frac{1}{4}v^2\left(1 + \frac{1}{u^2}\right)\right] d\lambda(v) &= 2 \int_0^{+\infty} v \exp\left[-\frac{1}{4}v^2\left(1 + \frac{1}{u^2}\right)\right] dv \\ &= \int_0^{+\infty} \frac{8}{1 + \frac{1}{u^2}} \exp(-w) dw \\ &= \frac{8}{1 + \frac{1}{u^2}}, \end{aligned}$$

d'où, pour tout $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R})$,

$$E f(T) = \frac{1}{\pi} \int_{\mathbb{R}^*} f[u^2] \frac{2}{u^2 + 1} d\lambda(u) = \frac{1}{\pi} \int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}^{+*}}(t) f(t) \frac{1}{(t+1)\sqrt{t}} d\lambda(t).$$

Ceci démontre que T admet la densité f_T donnée par

$$\forall t \in \mathbb{R} \quad f_T(t) = \frac{1}{\pi} \mathbf{1}_{\mathbb{R}^{+*}}(t) \frac{1}{(t+1)\sqrt{t}};$$

c'est une loi Bêta de deuxième espèce : $B\left(\frac{1}{2}, \frac{1}{2}\right)$.

Exercice 8.5. Moments d'ordre α et théorème de Fubini. Soient X une variable aléatoire positive et G la fonction définie par

$$\forall x \in \mathbb{R} \quad G(x) = P(X > x).$$

Démontrer que pour que X admette un moment d'ordre $\alpha \geq 1$, il faut et il suffit que la fonction $x \mapsto x^{\alpha-1}G(x)$ soit Lebesgue-intégrable sur \mathbb{R}^+ et que, dans ce cas, on a :

$$E(X^\alpha) = \alpha \int_{\mathbb{R}^+} x^{\alpha-1} G(x) d\lambda(x).$$

Solution. D'après le théorème de Fubini pour les fonctions mesurables positives, on a

$$\begin{aligned} \int_{\mathbb{R}^+} x^{\alpha-1} G(x) d\lambda(x) &= \int_{\mathbb{R}^+} x^{\alpha-1} \left[\int_{\Omega} \mathbf{1}_{(X>x)} dP \right] d\lambda(x) \\ &= \int_{\Omega} \left[\int_{\mathbb{R}^+} x^{\alpha-1} \mathbf{1}_{(X>x)} d\lambda(x) \right] dP \\ &= \int_{\Omega} \frac{X^\alpha}{\alpha} dP, \end{aligned}$$

d'où le résultat.

Exercice 8.6. Formulation équivalente de l'inégalité de Hölder. Soient p, q et r des réels positifs tels que $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$; démontrer que

1. Pour toutes variables aléatoires X, Y à valeurs dans $\overline{\mathbb{R}}^+$ et définies P-p.s., on a l'inégalité dans $\overline{\mathbb{R}}^+$:

$$\left(\int_{\Omega} (XY)^r dP \right)^{1/r} \leq \left(\int_{\Omega} X^p dP \right)^{1/p} \left(\int_{\Omega} Y^q dP \right)^{1/q}. \quad (8.15)$$

2. Si $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ et $Y \in \mathcal{L}^q(\Omega, \mathcal{A}, P)$, on a $XY \in \mathcal{L}^r(\Omega, \mathcal{A}, P)$ et on a l'inégalité

$$\|XY\|_r \leq \|X\|_p \|Y\|_q. \quad (8.16)$$

En déduire que si p, q et r sont des réels positifs tels que $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$, si $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$, $Y \in \mathcal{L}^q(\Omega, \mathcal{A}, P)$ et $Z \in \mathcal{L}^r(\Omega, \mathcal{A}, P)$, alors $XYZ \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ et on a l'inégalité :

$$\|XYZ\|_1 \leq \|X\|_p \|Y\|_q \|Z\|_r. \quad (8.17)$$

Solution.

1. Il résulte de l'inégalité (8.6) appliquée aux variables aléatoires X^r et Y^r avec les réels conjugués $\frac{p}{r}$ et $\frac{q}{r}$ que l'on a

$$\int_{\Omega} X^r Y^r dP \leq \left(\int_{\Omega} (X^r)^{p/r} dP \right)^{r/p} \left(\int_{\Omega} (Y^r)^{q/r} dP \right)^{r/q},$$

ce qui, en élevant à la puissance $\frac{1}{r}$, donne l'inégalité (8.15).

2. Si $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ et $Y \in \mathcal{L}^q(\Omega, \mathcal{A}, P)$, il en résulte que

$$\int_{\Omega} |XY|^r dP \leq \left(\int_{\Omega} |X|^{p/r} dP \right)^{r/p} \left(\int_{\Omega} |Y|^q dP \right)^{r/q} < +\infty,$$

et donc que $XY \in \mathcal{L}^r(\Omega, \mathcal{A}, P)$; l'inégalité (8.16) en résulte.

Remarque. Cette formulation est donc équivalente à celle donnant l'inégalité de Hölder (pour la réciproque, prendre $r = 1$).

Enfin, si p, q et r sont des réels positifs tels que $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$, définissons α par la relation $\frac{1}{\alpha} = \frac{1}{q} + \frac{1}{r}$. Soient $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$, $Y \in \mathcal{L}^q(\Omega, \mathcal{A}, P)$ et $Z \in \mathcal{L}^r(\Omega, \mathcal{A}, P)$; d'après ce qui précède, on a $YZ \in \mathcal{L}^{\alpha}(\Omega, \mathcal{A}, P)$ et

$$\|YZ\|_{\alpha} \leq \|Y\|_q \|Z\|_r.$$

Les réels α et p étant conjugués, le produit $X(YZ)$ est intégrable, l'inégalité de Hölder donne

$$\|X(YZ)\|_1 \leq \|X\|_p \|YZ\|_{\alpha},$$

ce qui, en vertu de l'inégalité précédente démontre l'inégalité (8.17).

Exercice 8.7. Variance, opérateur de covariance et support de loi. Soit $X \in \mathcal{L}_F^2(\Omega, \mathcal{A}, P)$ une variable aléatoire à valeur dans un espace euclidien F , d'opérateur de covariance Λ_X . Démontrer d'abord que si $F = \mathbb{R}$, on a :

$$\boxed{X = EX \text{ P-p.s.} \iff \sigma_X = 0.}$$

En déduire que dans le cas général, on a :

$$\boxed{\text{P-p.s.} \quad (X - EX) \in (\ker \Lambda_X)^\perp.}$$

Solution. Si $F = \mathbb{R}$, on a :

$$\sigma_X^2 = 0 \iff E(X - EX)^2 = 0 \iff X = EX \text{ P-p.s.}$$

Dans le cas général, on a alors :

$$x \in \ker \Lambda_X \iff E(\langle \overset{\circ}{X}, x \rangle^2) = 0 \iff \langle \overset{\circ}{X}, x \rangle = 0 \text{ P-p.s.}$$

Soit alors $(e_i)_{i=1, \dots, l}$ une base de $\ker \Lambda_X$. Pour tout $i = 1, \dots, l$, il existe un ensemble de probabilité nulle N_i tel que, pour tout $\omega \notin N_i$, on ait : $\langle \overset{\circ}{X}(\omega), e_i \rangle = 0$. Soit $N = \bigcup_{i=1}^l N_i$; on a $P(N) = 0$ et

$$\forall \omega \notin N \quad \forall i = 1, \dots, l \quad \langle \overset{\circ}{X}(\omega), e_i \rangle = 0,$$

et donc :

$$\boxed{\forall \omega \notin N \quad \overset{\circ}{X}(\omega) \in (\ker \Lambda_X)^\perp.}$$

Exercice 8.8. Généralisation du problème de régression linéaire au cas de variables aléatoires à valeurs dans un espace euclidien. Soient deux variables aléatoires $X \in \mathcal{L}_F^2(\Omega, \mathcal{A}, P)$ et $Y \in \mathcal{L}_G^2(\Omega, \mathcal{A}, P)$ où F et G sont des espaces euclidiens ; on suppose que l'opérateur d'auto-covariance Λ_X est inversible. On cherche une « meilleure » approximation de Y comme fonction affine de X au sens des moindres carrés, à savoir une solution en le couple $(A, b) \in \mathcal{L}(F, G) \times G$ du problème de minimisation :

$$\inf(\Phi(A, b) \mid (A, b) \in \mathcal{L}(F, G) \times G),$$

où :

$$\Phi(A, b) = E \|Y - (AX + b)\|^2.$$

On introduira l'opérateur d'intercovariance de X et Y , unique opérateur $\Lambda_{X,Y} \in \mathcal{L}(F, G)$ vérifiant :

$$\forall (x, y) \in F \times G \quad \langle \Lambda_{X,Y} x, y \rangle = E[\langle \overset{\circ}{X}, x \rangle \langle \overset{\circ}{Y}, y \rangle],$$

et on remarquera que : $\Lambda_{X,Y} = (\Lambda_{Y,X})^*$.

Solution. On a $\overset{\circ}{Y} - A\overset{\circ}{X}$ étant centrée :

$$\begin{aligned}\Phi(A, b) &= E\|\overset{\circ}{Y} - A\overset{\circ}{X} + (EY - AEX - b)\|^2 \\ &= E\|\overset{\circ}{Y} - A\overset{\circ}{X}\|^2 + \|EY - AEX - b\|^2 + 2E\langle \overset{\circ}{Y} - A\overset{\circ}{X}, EY - AEX - b \rangle \\ &= E\|\overset{\circ}{Y} - A\overset{\circ}{X}\|^2 + \|EY - AEX - b\|^2\end{aligned}$$

Pour tout A fixé, cette quantité est minimum pour $\widehat{b}_A = EY - AEX$. Reste à minimiser en A :

$$\Phi(A, \widehat{b}_A) \equiv E\|\overset{\circ}{Y} - A\overset{\circ}{X}\|^2.$$

Mais on a :

$$\begin{aligned}E\|\overset{\circ}{Y} - A\overset{\circ}{X}\|^2 &= \text{tr} E[(\overset{\circ}{Y} - A\overset{\circ}{X})(\overset{\circ}{Y} - A\overset{\circ}{X})^*] \\ &= \text{tr} E[\overset{\circ}{Y}\overset{\circ}{Y}^* + A(\overset{\circ}{X}\overset{\circ}{X}^*)A^* - (\overset{\circ}{Y}\overset{\circ}{X}^*)A^* - A\overset{\circ}{X}\overset{\circ}{Y}^*] \\ &= \text{tr} [\Lambda_Y + A\Lambda_X A^* - \Lambda_{X,Y} A^* - A\Lambda_{Y,X}].\end{aligned}$$

Si Ψ est l'application (différentiable) de $\mathcal{L}(F, G)$ dans \mathbb{R} définie par

$$\Psi(A) = \text{tr} [\Lambda_Y + A\Lambda_X A^* - \Lambda_{X,Y} A^* - A\Lambda_{Y,X}],$$

son application dérivée est donnée par

$$\begin{aligned}\forall H \in \mathcal{L}(F, G) \quad \Psi'(A)(H) &= \text{tr} [H\Lambda_X A^* + A\Lambda_X H^* - \Lambda_{X,Y} H^* - H\Lambda_{Y,X}] \\ &= 2 \text{tr} [(A\Lambda_X - \Lambda_{X,Y})H^*].\end{aligned}$$

Un point stationnaire \widehat{A} est donné par

$$\widehat{A} = \Lambda_{X,Y} \Lambda_X^{-1}.$$

Ce point stationnaire correspond bien à un **minimum**. En effet, on a :

$$\forall H \in \mathcal{L}(F, G) \quad \Psi''(A)(H)(H) = 2 \text{tr} [H\Lambda_X H^*] \geq 0.$$

De plus, si $(e_i)_{i=1, \dots, d}$ est une base orthonormée de G, on a

$$\text{tr} [H\Lambda_X H^*] = \sum_{j=1}^d \langle H\Lambda_X H^* e_j, e_j \rangle = \sum_{j=1}^d \langle \Lambda_X H^* e_j, H^* e_j \rangle = \sum_{j=1}^d \|\Lambda_X^{\frac{1}{2}} H^* e_j\|^2;$$

puisque Λ_X est inversible, il en résulte que $\Psi''(A)(H)(H) = 0$ si et seulement si $H^* e_i = 0$ pour tout $i = 1, \dots, d$, autrement dit, si et seulement si H^* , donc H, est nul. Cela assure que \widehat{A} correspond à un **minimum**.

La solution du problème de régression linéaire est le couple

$$(\widehat{A}, \widehat{b}_A) = (\Lambda_{X,Y} \Lambda_X^{-1}, EY - \Lambda_{X,Y} \Lambda_X^{-1}(EX)).$$

La surface D (sous-espace affine) d'équation

$$(y - EY) - \Lambda_{X,Y} \Lambda_X^{-1}(x - EX) = 0$$

est appelée **surface de régression linéaire de Y en X**. La « meilleure » approximation de Y comme fonction affine de X au sens des moindres carrés est

$$EY + \Lambda_{X,Y} \Lambda_X^{-1} (X - EX)$$

et on a $P\{(X, Y) \in D\} = 1$ si et seulement si $\Phi(\widehat{A}, \widehat{b}_A) = 0$.

Chapitre 9

Indépendance de tribus, de variables aléatoires

L'indépendance est une notion fondamentale en calcul des probabilités et en statistique. Elle sera toujours relative à un espace probabilisé fixé (Ω, \mathcal{A}, P) . Toutes les variables aléatoires seront définies sur cet espace.

Les notions élémentaires sur l'indépendance étudiées dans le chapitre 3 du tome I sont supposées connues.

9.1. Indépendance de familles d'événements et de variables aléatoires

Définition 9.1. Soit (Ω, \mathcal{A}, P) un espace probabilisé.

(a) Les événements $A \in \mathcal{A}$ et $B \in \mathcal{A}$ sont **indépendants** si :

$$P(A \cap B) = P(A)P(B).$$

(b) Deux **familles** d'événements \mathcal{A}_1 et \mathcal{A}_2 sont **indépendantes** si tout élément de \mathcal{A}_1 est indépendant de tout élément de \mathcal{A}_2 .

Remarque. On vérifiera (cf. chapitre 3, tome 1) que pour que $A \in \mathcal{A}$ et $B \in \mathcal{A}$ soient indépendants, il faut et il suffit que les tribus $\sigma(\{A\})$ et $\sigma(\{B\})$ soient indépendantes. Le théorème suivant généralise cette propriété et est d'un usage courant.

Théorème 9.2. Soient \mathcal{C}_1 et \mathcal{C}_2 deux π -**systèmes** contenus dans \mathcal{A} ; on note \mathcal{F}_1 (resp. \mathcal{F}_2) la **tribu engendrée** par \mathcal{C}_1 (resp. \mathcal{C}_2). Pour que les **familles d'événements** \mathcal{F}_1 et \mathcal{F}_2 soient **indépendantes** il faut et il suffit que les familles d'événements \mathcal{C}_1 et \mathcal{C}_2 le soient.

Démonstration. La condition nécessaire est évidente d'après la définition de l'indépendance de deux familles d'événements.

Pour la condition suffisante, supposons que \mathcal{C}_1 et \mathcal{C}_2 soient *indépendantes*. Nous allons utiliser le principe de *prolongement par mesurabilité* dans sa version ensembliste (lemme 8.3). Soit donc \mathcal{D} l'ensemble des événements indépendants de tout élément de \mathcal{C}_2 :

$$\mathcal{D} = \{D \in \mathcal{A} \mid \forall A_2 \in \mathcal{C}_2 \quad P(D \cap A_2) = P(D)P(A_2)\}.$$

Par hypothèse, $\mathcal{C}_1 \subset \mathcal{D}$ et $\Omega \in \mathcal{D}$. Démontrons que \mathcal{D} est un λ -système.

– Si D_1 et $D_2 \in \mathcal{D}$ sont tels que $D_1 \supset D_2$, on a, pour tout $A_2 \in \mathcal{C}_2$,

$$\begin{aligned} P[(D_1 \setminus D_2) \cap A_2] &= P[(D_1 \cap A_2) \setminus (D_2 \cap A_2)] \\ &= P[(D_1 \cap A_2)] - P[(D_2 \cap A_2)], \end{aligned}$$

et puisque D_1 et $D_2 \in \mathcal{D}$,

$$\begin{aligned} P[(D_1 \setminus D_2) \cap A_2] &= P(D_1)P(A_2) - P(D_2)P(A_2) \\ &= P(D_1 \setminus D_2)P(A_2). \end{aligned}$$

Comme de plus $D_1 \setminus D_2 \in \mathcal{A}$, il vient : $D_1 \setminus D_2 \in \mathcal{D}$.

– Par ailleurs, pour toute suite croissante $(D_n)_{n \in \mathbb{N}} \subset \mathcal{D}$, on a

$$\forall n \in \mathbb{N} \quad \forall A_2 \in \mathcal{C}_2 \quad P(D_n \cap A_2) = P(D_n)P(A_2).$$

et donc, par limite monotone,

$$\forall A_2 \in \mathcal{C}_2 \quad P\left[\left(\bigcup_{n \in \mathbb{N}} D_n\right) \cap A_2\right] = P\left(\bigcup_{n \in \mathbb{N}} D_n\right)P(A_2).$$

Comme de plus $\bigcup_{n \in \mathbb{N}} D_n \in \mathcal{A}$, il vient : $\bigcup_{n \in \mathbb{N}} D_n \in \mathcal{D}$.

Il résulte alors du principe de *prolongement par mesurabilité* que $\mathcal{D} \supset \mathcal{F}_1$, ce qui s'écrit :

$$\forall F_1 \in \mathcal{F}_1 \quad \forall A_2 \in \mathcal{C}_2 \quad P(F_1 \cap A_2) = P(F_1)P(A_2). \quad (9.1)$$

Soit alors \mathcal{E} l'ensemble des événements indépendants de tout élément de \mathcal{F}_1 :

$$\mathcal{E} = \{E \in \mathcal{A} \mid \forall F_1 \in \mathcal{F}_1 \quad P(F_1 \cap E) = P(F_1)P(E)\}.$$

La relation (9.1) s'écrit $\mathcal{E} \supset \mathcal{C}_2$. On a $\Omega \in \mathcal{E}$ et \mathcal{E} est bien sûr encore un λ -système; il résulte alors encore du principe de *prolongement par mesurabilité* que $\mathcal{E} \supset \mathcal{F}_2$, ce qui démontre le théorème. \square

Notation et rappel. Soit X une application de Ω dans E ; si \mathcal{G} est une famille de parties de E , on note $X^{-1}(\mathcal{G})$ la famille de parties de Ω :

$$\{X^{-1}(G) \in \mathcal{P}(\Omega) \mid G \in \mathcal{G}\}.$$

En particulier, si \mathcal{G} est une tribu sur E , la famille $X^{-1}(\mathcal{G})$ est une tribu sur Ω dite **tribu engendrée par l'application X** .

Définition 9.3. Soient X_i , $i = 1, 2$, deux variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) à valeurs dans des espaces probabilisables respectifs (E_i, \mathcal{E}_i) . Les **variables aléatoires** X_1 et X_2 sont **indépendantes** si les tribus $X_1^{-1}(\mathcal{E}_1)$ et $X_2^{-1}(\mathcal{E}_2)$ engendrées respectivement par les applications X_1 et X_2 sont indépendantes.

Remarque d'usage courant. Soient, pour $i = 1, 2$, deux variables aléatoires X_i définies sur l'espace probabilisé (Ω, \mathcal{A}, P) à valeurs dans des espaces probabilisables respectifs (E_i, \mathcal{E}_i) et f_i deux applications mesurables de (E_i, \mathcal{E}_i) dans l'espace probabilisable (F_i, \mathcal{F}_i) . Si les variables aléatoires X_i sont indépendantes, il en est de même des variables aléatoires $f_i \circ X_i$ (souvent notées $f_i(X_i)$).

Exemple. Si les variables aléatoires X_i sont à valeurs dans \mathbb{R}^{d_i} et sont indépendantes, toute marginale de X_1 est indépendante de toute marginale de X_2 .

Proposition 9.4. Avec les notations précédentes, soit, pour $i = 1, 2$, un π -système \mathcal{C}_i engendrant \mathcal{E}_i . Pour que les variables aléatoires X_1 et X_2 soient **indépendantes** il faut et il suffit que les π -systèmes $X_i^{-1}(\mathcal{C}_i)$, $i = 1, 2$, soient indépendants.

Démonstration. On rappelle que les tribus engendrées par les applications X_i , $i = 1, 2$, satisfont la relation

$$\sigma[X_i^{-1}(\mathcal{C}_i)] = X_i^{-1}[\sigma(\mathcal{C}_i)] \equiv X_i^{-1}(\mathcal{E}_i).$$

La famille d'événements $X_i^{-1}(\mathcal{C}_i)$ étant un π -système, la proposition résulte alors immédiatement du théorème 9.2. \square

On obtient alors le **critère général d'indépendance** suivant pour deux variables aléatoires en termes de **lois** de variables aléatoires.

Corollaire 9.5. Soient X_i , $i = 1, 2$ deux variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) à valeurs dans des espaces probabilisables respectifs (E_i, \mathcal{E}_i) . Pour que les variables aléatoires X_1 et X_2 soient indépendantes il faut et il suffit que :

$$\boxed{P_{X_1, X_2} = P_{X_1} \otimes P_{X_2}}, \quad (9.2)$$

où $P_{X_1} \otimes P_{X_2}$ désigne la probabilité produit des probabilités P_{X_1} et P_{X_2} , lois de X_1 et X_2 .

Démonstration. Par définition, pour que X_1 et X_2 soient indépendantes, il faut et il suffit que

$$\forall A_1 \in \mathcal{E}_1 \quad \forall A_2 \in \mathcal{E}_2 \quad P(X_1^{-1}(A_1) \cap X_2^{-1}(A_2)) = \prod_{i=1}^2 P(X_i^{-1}(A_i)).$$

Mais, pour tout $A_1 \in \mathcal{E}_1$ et tout $A_2 \in \mathcal{E}_2$, on a

$$P(X_1^{-1}(A_1) \cap X_2^{-1}(A_2)) = P[(X_1, X_2)^{-1}(A_1 \times A_2)] = P_{X_1, X_2}[A_1 \times A_2]$$

et

$$\prod_{i=1}^2 P(X_i^{-1}(A_i)) = P_{X_1} \otimes P_{X_2}(A_1 \times A_2).$$

Il en résulte que pour que X_1 et X_2 soient indépendantes il faut et il suffit que

$$\forall A_1 \in \mathcal{E}_1 \quad \forall A_2 \in \mathcal{E}_2 \quad P_{X_1, X_2} [A_1 \times A_2] = P_{X_1} \otimes P_{X_2} (A_1 \times A_2),$$

ce qui, en vertu du théorème d'unicité des mesures (théorème 8.4), est équivalent à

$$P_{X_1, X_2} = P_{X_1} \otimes P_{X_2},$$

puisque l'ensemble des pavés $A_1 \times A_2$, où $A_1 \in \mathcal{E}_1$ et $A_2 \in \mathcal{E}_2$ est un π -système qui engendre la tribu produit $\mathcal{E}_1 \otimes \mathcal{E}_2$. \square

Ce critère prend la forme fonctionnelle suivante :

Proposition 9.6. Soient X_i , $i = 1, 2$ deux variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) à valeurs dans des espaces probabilisables respectifs (E_i, \mathcal{E}_i) . Les trois assertions suivantes sont équivalentes :

(i) Les variables aléatoires X_1 et X_2 sont indépendantes ;

(ii) Pour toutes fonctions f_i réelles **positives** \mathcal{E}_i -mesurables ($i = 1, 2$),

on a

$$\int_{\Omega} \left[\prod_{i=1}^2 f_i \circ X_i \right] dP = \prod_{i=1}^2 \int_{\Omega} f_i \circ X_i dP ;$$

(iii) Pour toutes fonctions f_i réelles **bornées** \mathcal{E}_i -mesurables ($i = 1, 2$),

on a

$$\int_{\Omega} \left[\prod_{i=1}^2 f_i \circ X_i \right] dP = \prod_{i=1}^2 \int_{\Omega} f_i \circ X_i dP,$$

ce qui s'écrit encore

$$\boxed{E [f_1(X_1) f_2(X_2)] = E [f_1(X_1)] E [f_2(X_2)] .}$$

Démonstration. Remarquons d'abord que pour les fonctions considérées dans (ii) et (iii), on a, par le théorème de transfert,

$$\int_{\Omega} \left[\prod_{i=1}^2 f_i \circ X_i \right] dP = \int_{E_1 \times E_2} \left[\prod_{i=1}^2 f_i(x_i) \right] dP_{X_1, X_2}(x_1, x_2)$$

et

$$\prod_{i=1}^2 \int_{\Omega} f_i \circ X_i dP = \prod_{i=1}^2 \int_{E_i} f_i(x_i) dP_{X_i}(x_i).$$

On démontre alors successivement les implications :

(i) \Rightarrow (ii) et (iii) : il suffit d'appliquer la relation (9.2) et d'appliquer le théorème de Fubini.

(iii) \Rightarrow (ii) : la relation est en particulier vraie pour des f_i positives bornées \mathcal{E}_i -mesurables. Pour des f_i positives \mathcal{E}_i -mesurables quelconques,

considérer les suites de fonctions *positives bornées* \mathcal{E}_i -mesurables $\min(f_i, n)$ et appliquer la propriété de Beppo Levi.

(iii) \Rightarrow (i) : en prenant $f_i = \mathbf{1}_{A_i}$, où $A_i \in \mathcal{E}_i$, on obtient de suite la relation (9.2), ce qui est équivalent à l'indépendance de X_1 et X_2 . \square

Du critère *général* d'indépendance, on déduit les critères *particuliers* suivants :

Corollaire 9.7. Soient X_i , $i = 1, 2$ deux variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs respectivement dans $(\mathbb{R}^{d_1}, \mathcal{B}_{\mathbb{R}^{d_1}})$.

1. Critère d'indépendance en termes de fonctions de répartition. Pour que les variables aléatoires X_1 et X_2 soient indépendantes il faut et il suffit que :

$$\forall x_1 \in \mathbb{R}^{d_1} \quad \forall x_2 \in \mathbb{R}^{d_2} \quad F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2). \quad (9.3)$$

2. Critère d'indépendance en termes de densités.

(a) Si les variables aléatoires X_1 et X_2 admettent des densités respectives f_{X_1} et f_{X_2} et sont indépendantes, alors la variable aléatoire (X_1, X_2) admet une densité f_{X_1, X_2} **produit direct** de f_{X_1} et f_{X_2} , c'est-à-dire qui vérifie :

$$\forall x_1 \in \mathbb{R}^{d_1} \quad \forall x_2 \in \mathbb{R}^{d_2} \quad f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2). \quad (9.4)$$

(b) Inversement si la variable aléatoire (X_1, X_2) admet une densité f_{X_1, X_2} produit direct de deux fonctions intégrables positives f_1 et f_2 , c'est-à-dire qui vérifie la relation :

$$\forall x_1 \in \mathbb{R}^{d_1} \quad \forall x_2 \in \mathbb{R}^{d_2} \quad f_{X_1, X_2}(x_1, x_2) = f_1(x_1)f_2(x_2) \quad (9.5)$$

alors f_1 et f_2 sont, à un facteur positif près, les densités respectives de X_1 et X_2 , et les variables aléatoires X_1 et X_2 sont indépendantes.

3. Critère d'indépendance pour des variables aléatoires discrètes.

Si les variables aléatoires X_1 et X_2 sont discrètes, il en est de même de la variable aléatoire (X_1, X_2) . Pour que les variables aléatoires X_1 et X_2 soient indépendantes il faut et il suffit que :

$$\forall x_1 \in X_1(\Omega) \quad \forall x_2 \in X_2(\Omega) \quad \mathbb{P}[X_1 = x_1, X_2 = x_2] = \mathbb{P}(X_1 = x_1) \mathbb{P}(X_2 = x_2). \quad (9.6)$$

Démonstration. **1.** Si les variables aléatoires X_1 et X_2 sont indépendantes, la relation (9.3) résulte du critère général. Inversement, si la relation (9.3) est satisfaite, les probabilités \mathbb{P}_{X_1, X_2} et $\mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$ coïncident sur le π -système

$$\{\{u_1 \leq x_1\} \times \{u_2 \leq x_2\} \mid x_1 \in \mathbb{R}^{d_1}, x_2 \in \mathbb{R}^{d_2}\}$$

qui engendre la tribu produit $\mathcal{B}_{\mathbb{R}^{d_1}} \otimes \mathcal{B}_{\mathbb{R}^{d_2}}$; elles sont égales, et les variables aléatoires X_1 et X_2 sont donc indépendantes.

2. Soit λ_i la mesure de Lebesgue sur \mathbb{R}^{d_i} . Dans les deux cas considérés, on a $P_{X_i} = f_{X_i} \cdot \lambda_i$, et donc, par le théorème de Fubini,

$$P_{X_1} \otimes P_{X_2} = f_{X_1} f_{X_2} \cdot \lambda_1 \otimes \lambda_2. \quad (9.7)$$

(a) Si donc X_1 et X_2 sont indépendantes, le corollaire 9.5 et la relation précédente assurent que (X_1, X_2) admet une densité f_{X_1, X_2} produit direct de f_{X_1} et f_{X_2} .

(b) Inversement, si la relation (9.5) est satisfaite, la variable aléatoire X_1 (resp. X_2) admet une densité f_{X_1} (resp. f_{X_2}) donnée par

$$\forall x_1 \in \mathbb{R}^{d_1} \quad f_{X_1}(x_1) = f_1(x_1) \int_{\mathbb{R}^{d_2}} f_2(x_2) d\lambda_2(x_2),$$

respectivement,

$$\forall x_2 \in \mathbb{R}^{d_2} \quad f_{X_2}(x_2) = f_2(x_2) \int_{\mathbb{R}^{d_1}} f_1(x_1) d\lambda_1(x_1).$$

Intégrant par exemple la première égalité, il vient

$$1 = \left(\int_{\mathbb{R}^{d_1}} f_1(x_1) d\lambda_1(x_1) \right) \left(\int_{\mathbb{R}^{d_2}} f_2(x_2) d\lambda_2(x_2) \right).$$

Il en résulte que

$$\forall x_1 \in \mathbb{R}^{d_1} \quad \forall x_2 \in \mathbb{R}^{d_2} \quad f_{X_1}(x_1) f_{X_2}(x_2) = f_1(x_1) f_2(x_2),$$

ce qui, en vertu de l'hypothèse, démontre que

$$P_{X_1, X_2} = f_{X_1} f_{X_2} \cdot \lambda_1 \otimes \lambda_2. \quad (9.8)$$

En rapprochant les relations (9.7) et (9.8), il vient

$$P_{X_1, X_2} = P_{X_1} \otimes P_{X_2},$$

ce qui démontre l'indépendance des variables aléatoires X_1 et X_2 . **À noter qu'alors la relation (9.4) est encore vraie.**

3. Si les variables aléatoires X_1 et X_2 sont discrètes, la condition nécessaire est évidente. Démontrons que la condition est suffisante. Si la relation (9.6) est vraie, pour tout $A_1 \in \mathcal{B}_{\mathbb{R}^{d_1}}$ et tout $A_2 \in \mathcal{B}_{\mathbb{R}^{d_2}}$, on a

$$P_{X_1, X_2}(A_1 \times A_2) = \sum_{\substack{x_1 \in X_1(\Omega) \\ x_2 \in X_2(\Omega)}} P[(X_1 = x_1) \cap (X_2 = x_2)] \delta_{x_1} \otimes \delta_{x_2}(A_1 \times A_2),$$

et donc, par le théorème de Fubini.

$$P_{X_1, X_2}(A_1 \times A_2) = \left(\sum_{x_1 \in X_1(\Omega)} P(X_1 = x_1) \delta_{x_1}(A_1) \right) \\ \times \left(\sum_{x_2 \in X_2(\Omega)} P(X_2 = x_2) \delta_{x_2}(A_2) \right),$$

soit encore

$$P_{X_1, X_2}(A_1 \times A_2) = P_{X_1}(A_1) P_{X_2}(A_2).$$

Le théorème d'unicité des mesures (théorème 8.4) assure l'égalité des probabilités

$$P_{X_1, X_2} = P_{X_1} \otimes P_{X_2}.$$

Ceci démontre l'indépendance des variables aléatoires X_1 et X_2 . \square

Les propriétés suivantes liant indépendance et moments de variables aléatoires sont d'un usage constant.

Proposition 9.8. Soient X_1 et X_2 deux variables aléatoires réelles définies sur l'espace probabilisé (Ω, \mathcal{A}, P) et indépendantes.

(a) Si X_1 et X_2 admettent une moyenne, il en est de même de la variable aléatoire $X_1 X_2$ et on a

$$E(X_1 X_2) = E(X_1) E(X_2).$$

(b) Si X_1 et X_2 admettent un moment d'ordre deux, on a :

$$\text{cov}(X_1, X_2) = 0 \quad \text{et} \quad \sigma_{X_1 + X_2}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2.$$

Démonstration. (a) D'après la proposition 9.6, on a

$$\int_{\Omega} |X_1 X_2| dP = \left(\int_{\Omega} |X_1| dP \right) \left(\int_{\Omega} |X_2| dP \right) < +\infty.$$

Il résulte alors du théorème de transfert que

$$\int_{\Omega} X_1 X_2 dP = \int_{\mathbb{R}} x_1 x_2 dP_{X_1, X_2}(x_1, x_2).$$

L'indépendance des variables aléatoires X_1 et X_2 étant équivalente à l'égalité $P_{X_1, X_2} = P_{X_1} \otimes P_{X_2}$, le théorème de Fubini permet d'écrire que

$$\int_{\Omega} X_1 X_2 dP = \int_{\mathbb{R}} x_1 x_2 d(P_{X_1} \otimes P_{X_2})(x_1, x_2) \\ = \prod_{i=1}^2 \int_{\mathbb{R}} x dP_{X_i}(x) \\ = E(X_1) E(X_2).$$

(b) Appliquer ce dernier résultat et le calcul général de la variance d'une somme de variables aléatoires (proposition 8.30). \square

Corollaire 9.9. Soient $X_i \in \mathcal{L}_E^2(\Omega, \mathcal{A}, P)$, $i = 1, 2$, où E est un espace euclidien. Si les variables aléatoires X_1 et X_2 sont indépendantes, on a la relation suivante sur les opérateurs de covariance :

$$\Lambda_{X_1+X_2} = \Lambda_{X_1} + \Lambda_{X_2}.$$

Si de plus $E = \mathbb{R}^d$, cette relation donne en termes de matrices de covariance :

$$C_{X_1+X_2} = C_{X_1} + C_{X_2}.$$

Démonstration. Pour tout $x \in E$, les variables aléatoires $\langle X_1, x \rangle$ et $\langle X_2, x \rangle$ sont indépendantes. On a alors

$$\langle \Lambda_{X_1+X_2}, x, x \rangle \equiv \sigma_{\langle X_1+X_2, x \rangle}^2 = \sigma_{\langle X_1, x \rangle}^2 + \sigma_{\langle X_2, x \rangle}^2 \equiv \langle \Lambda_{X_1}, x, x \rangle + \langle \Lambda_{X_2}, x, x \rangle,$$

d'où le résultat. \square

On généralise maintenant la notion d'indépendance d'événements, de familles d'événements, de variables aléatoires à des familles quelconques indexées sur un ensemble I .

Définition 9.10. Soit $(A_j)_{j \in I}$ une famille d'événements. Les événements A_j ($j \in I$) sont indépendants si

$$\forall J \in \mathcal{P}_f(I) \quad P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

où $\mathcal{P}_f(I)$ est l'ensemble des parties finies de I .

Remarque. On parle aussi, dans ce cas, d'unc « famille d'événements indépendants ». Ceci est une notion d'indépendance globale ; nous renvoyons au chapitre 3, tome 1, pour la notion d'indépendance n à n et ses relations avec cette dernière.

Définition 9.11. Soit $(\mathcal{A}_i)_{i \in I}$ une famille de familles d'événements. On dit que les familles d'événements \mathcal{A}_i ($i \in I$) sont indépendantes si, pour tout choix d'un A_i dans \mathcal{A}_i ($i \in I$), les événements A_i ($i \in I$) sont indépendants.

Remarque. Dans cette définition, les familles \mathcal{A}_i n'ont aucune structure particulière. Toutefois, cette notion est particulièrement utilisée dans le cas où ces familles d'événements sont des π -systèmes ou des tribus (on parlera par exemple d'une suite de sous-tribus de \mathcal{A} indépendantes). **Le théorème 9.2 se généralise alors à une famille quelconque de π -systèmes indépendants.**

Théorème 9.12. Soit une famille $(\mathcal{C}_i)_{i \in I}$ de π -systèmes contenus dans \mathcal{A} et indépendants; on note \mathcal{F}_i la tribu engendrée par \mathcal{C}_i . Soit $\{I_j\}_{j \in J}$ une partition quelconque de I . Pour tout $j \in J$, on note \mathcal{A}_j la tribu engendrée par la famille d'événements $\bigcup_{i \in I_j} \mathcal{C}_i$, c'est-à-dire la plus petite¹ tribu contenant $\bigcup_{i \in I_j} \mathcal{C}_i$. Alors les tribus \mathcal{A}_j , $j \in J$, sont indépendantes. En particulier les tribus \mathcal{F}_i engendrées par les \mathcal{C}_i ($i \in I$) sont indépendantes.

Démonstration. Nous n'en donnons que les grandes lignes. Il est clair qu'il suffit de démontrer ce résultat dans le cas où I est fini, et donc dans le cas où $I = \{1, 2, \dots, n\}$; une récurrence permet de le faire. Pour éviter les problèmes d'indexation, nous nous contenterons de démontrer que si \mathcal{C}_i , $i = 1, 2, 3$, sont des π -systèmes indépendants les tribus engendrées par \mathcal{C}_1 et $\mathcal{C}_2 \cup \mathcal{C}_3$ sont indépendantes. La famille d'événements $\mathcal{C}_2 \cup \mathcal{C}_3$ n'est plus un π -système; soit \mathcal{C}_4 le π -système engendré par $\mathcal{C}_2 \cup \mathcal{C}_3$ et Ω , c'est-à-dire $\{\mathcal{C}_2 \cap \mathcal{C}_3 \mid \mathcal{C}_2 \in \mathcal{C}_2 \cup \{\Omega\} \text{ et } \mathcal{C}_3 \in \mathcal{C}_3 \cup \{\Omega\}\}$; on a

$$\mathcal{C}_2 \cup \mathcal{C}_3 \subset \mathcal{C}_4 \subset \sigma(\mathcal{C}_2 \cup \mathcal{C}_3),$$

et donc, $\sigma(\mathcal{C}_4) = \sigma(\mathcal{C}_2 \cup \mathcal{C}_3)$. Il est clair que les π -systèmes \mathcal{C}_1 et \mathcal{C}_4 sont indépendants; par le théorème 9.2, il en est de même des tribus engendrées. \square

Définition 9.13. La famille $(X_i)_{i \in I}$ de variables aléatoires à valeurs respectivement dans des espaces probabilisables (E_i, \mathcal{E}_i) , $i \in I$, est une famille de variables aléatoires indépendantes si les tribus $X_i^{-1}(\mathcal{E}_i)$, $i \in I$, engendrées par ces variables aléatoires sont indépendantes.

On dit alors souvent plus brièvement que « les variables aléatoires X_i ($i \in I$) sont indépendantes ».

Tous les critères étudiés précédemment pour des familles où $|I| = 2$ se généralisent facilement au cas où I est fini. Pour le cas où I est quelconque, il faut définir la notion de mesure produit sur $\prod_{i \in I} E_i$. Cela est possible par exemple si $E_i = \mathbb{R}$ pour tout $i \in I$ par le théorème de prolongement de **Kolmogorov**, mais ce résultat ne figure pas au programme de l'agrégation (cf. Annexe, théorème A.39 et corollaire A.40, pour un énoncé²).

9.2. Indépendance et événements asymptotiques

On étudie dans cette section deux théorèmes célèbres qui sont d'usage fréquent dans l'étude de la convergence presque sûre de suites ou de séries de variables aléatoires, étude que nous ferons ultérieurement.

1. Au sens de l'inclusion entre familles de parties.

2. Pour une démonstration de ce théorème, on pourra consulter le livre de Jacques Neveu, *Bases mathématiques du Calcul des Probabilités*, Masson & Cie (1964).

Définition 9.14. Soit, sur l'ensemble Ω , une suite $(\mathcal{A}_n)_{n \in \mathbb{N}}$ de sous-tribus de la tribu \mathcal{A} . On note $\bigvee_{n \in \mathbb{N}} \mathcal{A}_n$ la tribu engendrée par $\bigcup_{n \in \mathbb{N}} \mathcal{A}_n$, c'est-à-dire la plus petite tribu (au sens de l'inclusion entre familles de parties) contenant $\bigcup_{n \in \mathbb{N}} \mathcal{A}_n$. On appelle **tribu asymptotique** la tribu

$$\mathcal{A}_\infty = \bigcap_{n \in \mathbb{N}} \left(\bigvee_{p \geq n} \mathcal{A}_p \right).$$

Les éléments de \mathcal{A}_∞ sont appelés **événements asymptotiques**. Une variable aléatoire \mathcal{A}_∞ -mesurable est dite **asymptotique**.

Exemple 9.1. Tribu asymptotique associée à une suite de variables aléatoires. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs respectivement dans l'espace probabilisable (E_n, \mathcal{E}_n) .

On prend pour \mathcal{A}_p la tribu $X_p^{-1}(\mathcal{E}_p)$ engendrée par la variable aléatoire X_p , encore notée $\sigma(X_p)$. C'est la tribu des événements de la forme $(X_p \in A_p)$, $(A_p \in \mathcal{E}_p)$: pour un ω donné, on sait donc dire si un tel événement est réalisé ou non, dès que l'on connaît la valeur $X_p(\omega)$.

La tribu $\bigvee_{p \geq n} X_p^{-1}(\mathcal{E}_p)$ (plus petite tribu sur Ω rendant mesurables simultanément toutes les applications X_p , $p \geq n$), encore notée $\sigma(X_p \mid p \geq n)$, est, par définition, la tribu engendrée par la famille de variables aléatoires $(X_p)_{p \geq n}$; elle est constituée d'événements dont la réalisation, pour un ω donné, dépend de la suite $(X_p(\omega), X_{p+1}(\omega), \dots)$ (il faut toutefois se garder de penser que cette dépendance peut être explicitée).

La tribu asymptotique est alors

$$\mathcal{A}_\infty = \bigcap_{n \in \mathbb{N}} \sigma(X_p \mid p \geq n).$$

Un événement est dans \mathcal{A}_∞ si sa réalisation, pour un ω donné, tout en dépendant de la suite $(X_0(\omega), X_1(\omega), \dots)$ ne dépend pas des n premières valeurs, et ce, quel que soit l'entier n . C'est le cas, par exemple, lorsque les X_n sont à valeurs réelles, de l'événement

$$\{ \text{la suite } (X_n)_{n \in \mathbb{N}} \text{ converge dans } \mathbb{R} \}.$$

Montrons précisément que cet événement est asymptotique. Rappelons d'abord qu'une suite numérique $(a_n)_{n \in \mathbb{N}}$ converge dans $\overline{\mathbb{R}}$ si et seulement si on a

$$\liminf_n a_n = \limsup_n a_n ;$$

elle converge donc dans \mathbb{R} si et seulement si

$$-\infty < \liminf_n a_n = \limsup_n a_n < +\infty.$$

Ainsi, on a l'égalité des événements

$$\left\{ \text{la suite } (X_n)_{n \in \mathbb{N}} \text{ converge dans } \mathbb{R} \right\} = \left\{ \limsup_n X_n = \liminf_n X_n \in \mathbb{R} \right\};$$

il suffit donc de démontrer que les variables aléatoires $\limsup_n X_n$, et $\liminf_n X_n$ sont asymptotiques. Il suffit de traiter le cas de $\limsup_n X_n$. Pour cela, notons \mathcal{B}_n la tribu $\sigma(X_p \mid p \geq n)$; on remarque que, pour tout p et pour tout $n \geq p$, $\sup_{k \geq n} X_k$ est \mathcal{B}_p -mesurable. Il en résulte que, pour tout p , $\lim_n (\sup_{k \geq n} X_k)$ est \mathcal{B}_p -mesurable, autrement dit, que $\limsup_n X_n$ est \mathcal{A}_∞ -mesurable.

On en déduit aussitôt que l'événement (la série $\sum X_n$ converge dans \mathbb{R}) est aussi asymptotique. Une autre justification est d'ailleurs aussi de dire que, d'après le critère de Cauchy, on a

$$\left(\sum_{n \in \mathbb{N}} X_n \text{ converge} \right) = \left(\lim_{n,m} \sum_{k=n}^m X_k = 0 \right) \in \bigcap_{n \in \mathbb{N}} \mathcal{B}_n = \mathcal{A}_\infty.$$

Théorème 9.15 (Loi du tout ou rien ou loi de 0, 1). *Soit, sur l'espace de probabilité (Ω, \mathcal{A}, P) , une suite $(\mathcal{A}_n)_{n \in \mathbb{N}}$ de sous-tribus de la tribu \mathcal{A} , indépendantes, et soit \mathcal{A}_∞ la tribu asymptotique associée. On a alors :*

$$\forall A \in \mathcal{A}_\infty \quad P(A) = 0 \text{ ou } 1.$$

Démonstration. Il résulte du théorème 9.12 que, pour tous entiers M et N tels que $M < N$, les tribus \mathcal{A}_n , $n \leq M$, et $\bigvee_{p \geq N} \mathcal{A}_p$ sont indépendantes. Mais, pour tout N , $\mathcal{A}_\infty \subset \bigvee_{p \geq N} \mathcal{A}_p$, donc, pour tout M , les tribus \mathcal{A}_n , $n \leq M$, et \mathcal{A}_∞ sont indépendantes; il en est alors de même des tribus \mathcal{A}_n , $n \in \mathbb{N}$, et \mathcal{A}_∞ . Mais alors les tribus $\bigvee_{p \geq 1} \mathcal{A}_p$ et \mathcal{A}_∞ sont indépendantes et, puisque $\mathcal{A}_\infty \subset \bigvee_{p \geq 1} \mathcal{A}_p$, la tribu \mathcal{A}_∞ est indépendante d'elle-même. En particulier, on a :

$$\forall A \in \mathcal{A}_\infty \quad P(A \cap A) = P(A)P(A),$$

d'où le résultat. □

Corollaire 9.16. *Avec les mêmes notations que dans le théorème précédent, toute variable aléatoire \mathcal{A}_∞ -mesurable est presque sûrement constante.*

Exemple 9.2. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles **indépendantes** définies sur (Ω, \mathcal{A}, P) . Alors la série $\sum_{n \in \mathbb{N}} X_n$ converge ou diverge presque sûrement, c'est-à-dire :

$$P\left(\sum_{n \in \mathbb{N}} X_n \text{ converge}\right) = 0 \text{ ou } 1.$$

En effet, on a vu (exemple 9.1) que l'événement $(\sum_{n \in \mathbb{N}} X_n \text{ converge})$ est asymptotique; la loi du tout ou rien donne le résultat.

Remarque. En conséquence, si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires réelles **indépendantes** définies sur (Ω, \mathcal{A}, P) , pour démontrer que la série $\sum_{n \in \mathbb{N}} X_n$ converge presque sûrement, il suffit de démontrer que

$$P\left(\sum_{n \in \mathbb{N}} X_n \text{ converge}\right) > 0.$$

Remarque. Un cas particulier de l'exemple précédent est celui où $X_n = \varepsilon_n/n$, les variables aléatoires ε_n ($n \in \mathbb{N}$) étant **indépendantes**, à valeurs 1 ou -1 , de même loi donnée par

$$P(\varepsilon_n = 1) = P(\varepsilon_n = -1) = \frac{1}{2}.$$

La loi du tout ou rien nous dit que l'une de ces deux assertions est vraie (sans préciser laquelle) :

- (i) La série de terme général $\frac{\varepsilon_n}{n}$ est P-p.s. convergente.
- (ii) La série de terme général $\frac{\varepsilon_n}{n}$ est P-p.s. divergente.

Rappel. Soit $(A_n)_{n \in \mathbb{N}}$ une suite de parties de Ω . On définit les deux parties de Ω , **limite supérieure et inférieure**³ de la suite d'ensembles $(A_n)_{n \in \mathbb{N}}$ par

$$\limsup_n A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{p \geq n} A_p \quad \liminf_n A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{p \geq n} A_p.$$

L'ensemble $\limsup_n A_n$ est l'ensemble des $\omega \in \Omega$ qui appartiennent à une infinité de A_n . L'ensemble $\liminf_n A_n$ est l'ensemble des $\omega \in \Omega$ qui, à partir d'un certain rang (dépendant de ω , bien sûr) appartiennent à *tous* les A_n . On a donc l'inclusion :

$$\liminf_n A_n \subset \limsup_n A_n ;$$

de plus, on a les égalités :

$$\left(\limsup_n A_n\right)^c = \liminf_n A_n^c \quad \text{et} \quad \left(\liminf_n A_n\right)^c = \limsup_n A_n^c.$$

Ces ensembles jouent un grand rôle dans l'étude des convergences de suites de variables aléatoires, l'outil fondamental étant le lemme de Borel-Cantelli.

Lemme 9.17 (Lemme de Borel-Cantelli). Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements.

(a) On a l'implication :

$$\sum_{n=0}^{+\infty} P(A_n) < +\infty \implies P(\limsup_n A_n) = 0.$$

3. Autres notations : $\limsup_n A_n = \overline{\lim} A_n$ et $\liminf_n A_n = \underline{\lim} A_n$.

(b) Si les événements A_n sont **indépendants**, on a :

$$\sum_{n=0}^{+\infty} P(A_n) = +\infty \implies P(\limsup_n A_n) = 1.$$

Démonstration. (a) Pour tout $n \in \mathbb{N}$, on a

$$P(\limsup_n A_n) \leq P\left(\bigcup_{p \geq n} A_p\right) \leq \sum_{p \geq n} P(A_p),$$

ce qui donne le résultat, le membre de droite étant le reste d'ordre n d'une série convergente.

(b) On a :

$$P(\limsup_n A_n) = 1 - P(\liminf_n A_n^c).$$

Puisque les événements A_n^c sont aussi indépendants, on a, d'après les propriétés d'une probabilité pour les suites monotones (propriété dite de monotonie séquentielle) :

$$\begin{aligned} P(\liminf_n A_n^c) &= \lim_n \lim_q P\left(\bigcap_{p=n}^q A_p^c\right) \\ &= \lim_n \lim_q \prod_{p=n}^q P(A_p^c) \\ &= \lim_n \lim_q \prod_{p=n}^q [1 - P(A_p)]. \end{aligned}$$

Mais, puisque $\exp(-x) \geq 1 - x$, on a :

$$0 \leq \prod_{p=n}^q [1 - P(A_p)] \leq \exp\left(-\sum_{p=n}^q P(A_p)\right);$$

le membre de droite convergeant vers 0 lorsque q tend vers l'infini, on a

$$\lim_q \prod_{p=n}^q [1 - P(A_p)] = 0,$$

ce qui démontre le résultat. □

Remarque. La première implication du lemme de Borel-Cantelli est toujours vraie (il n'est pas besoin de supposer les A_n indépendants); par contre sa réciproque est fautive : prendre l'espace probabilisé $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$ où λ est la restriction de la mesure de Lebesgue à $[0, 1]$ et, pour tout $n \in \mathbb{N}^*$, $A_n =]0, 1/n]$. On a $\limsup_n A_n = \emptyset$, et donc $P(\limsup_n A_n) = 0$, et pourtant $\sum_{n=1}^{+\infty} P(A_n) = +\infty$.

Ce même exemple montre de plus que la deuxième implication ne peut être vraie en général, ce qui explique l'hypothèse supplémentaire d'indépendance.

9.3. Quelques résultats liés à l'indépendance et au modèle de pile ou face

Nous avons vu comment construire un modèle probabiliste décrivant un jeu de pile ou face en n coups (avec une pièce équilibrée). Mathématiquement, cela revient à la construction d'un espace probabilisé (Ω, \mathcal{A}, P) et de n variables aléatoires indépendantes X_i ($i = 1, \dots, n$) de même loi uniforme sur $[0, 1[$ définies sur cet espace : on peut prendre $\Omega = \{0, 1\}^n$, muni de la probabilité uniforme, et pour X_i la projection sur le i -ième facteur : $(x_1, \dots, x_i, \dots, x_n) \mapsto x_i$.

Le problème analogue dans le cas d'une suite infinie de lancers devrait conduire naturellement à prendre pour espace des événements $\Omega = \{0, 1\}^{\mathbb{N}^*}$. Toutefois l'existence sur cet espace d'une mesure de probabilité dont pour chaque i la i -ième marginale soit la loi uniforme sur $\{0, 1\}$ n'est pas évidente. Si on veut l'obtenir par application d'un résultat général, il faut employer le théorème de Kolmogorov (cf. corollaire A.40).

Mais le problème posé possède une autre solution, a priori plus élémentaire, où on prend notamment pour Ω l'intervalle $[0, 1[$, muni de la mesure de Lebesgue. Cette solution, que nous allons étudier ci-dessous, n'est pas en fait si éloignée que cela de la précédente : à chaque $x \in [0, 1[$, on peut en effet faire correspondre une suite appartenant à $\{0, 1\}^{\mathbb{N}^*}$ en écrivant le développement de x en base 2 :

$$x = 0, x_1 x_2 \dots x_n \dots$$

(il y a ambiguïté pour les rationnels de la forme $p/2^q$, car ces nombres possèdent deux développements en base 2, par exemple $1/2 = 0,1000\dots = 0,0111\dots$, de même que les rationnels de la forme $p/10^q$ possèdent deux développements décimaux : $0,70000\dots = 0,69999\dots$; on convient ici de choisir celui des deux développements qui s'achève par une suite de zéros). On définit ensuite sur $\Omega = [0, 1[$ (muni de sa tribu borélienne et de la mesure de Lebesgue) des variables aléatoires D_n ($n \in \mathbb{N}^*$) en prenant pour $D_n(x)$ la n -ième décimale dyadique de x . On montre alors (voir prop. 9.18 ci-dessous) que ces variables aléatoires sont indépendantes et de loi uniforme sur $\{0, 1\}$, ce qui fournit une solution au problème posé.

Développement dyadique d'un réel $x \in [0, 1[$. Définissons pour tout $x \in [0, 1[$ les suites de terme général $D_n(x)$ et $R_n(x)$ par :

$$R_0(x) = x$$

et, pour $n \in \mathbb{N}^*$,

$$D_n(x) = [2R_{n-1}(x)] \quad R_n(x) = 2R_{n-1}(x) - D_n(x).$$

Par construction,

$$D_n(x) \in \{0, 1\} \quad \text{et} \quad R_n(x) \in [0, 1[$$

et un raisonnement par récurrence conduit immédiatement à :

$$\forall n \in \mathbb{N}^* \quad x = \sum_{j=1}^n \frac{D_j(x)}{2^j} + \frac{1}{2^n} R_n(x);$$

En faisant tendre n vers l'infini, on obtient :

$$x = \sum_{j=1}^{+\infty} \frac{D_j(x)}{2^j}. \quad (9.9)$$

De façon générale, lorsqu'on a

$$x = \sum_{j=1}^{+\infty} \frac{x_j}{2^j}. \quad (9.10)$$

on écrit symboliquement

$$x = 0, x_1 x_2 \dots x_j \dots \quad (9.10')$$

et on dit que le second membre de (9.10) ou de (9.10') est un *développement dyadique* de x . Nous avons donc montré que *tout* $x \in [0, 1[$ possède une *développement dyadique*.

Le développement dyadique d'un réel n'est pas unique ; on a en effet

$$\forall n \in \mathbb{N}^* \quad \frac{1}{2^{n-1}} = \sum_{j=n}^{+\infty} \frac{1}{2^j} \quad (9.11)$$

si bien que, pour tout $k \in \mathbb{N}^*$ et toute suite finie $(x_j)_{1 \leq j \leq k} \in \{0, 1\}^k$, on a

$$\sum_{j=1}^k \frac{x_j}{2^j} + \frac{1}{2^{k+1}} = \sum_{j=1}^k \frac{x_j}{2^j} + \sum_{j=k+2}^{+\infty} \frac{1}{2^j}. \quad (9.12)$$

soit

$$0, x_1 x_2 \dots x_k 10000 \dots = 0, x_1 x_2 \dots x_k 01111 \dots \quad (9.12')$$

Puisque tout entier impair peut s'écrire sous la forme $x_1 2^k + x_2 2^{k-1} + \dots + x_k 2 + 1$ (écriture d'un entier en base 2), on vérifie aisément que les nombres de la forme (9.12) sont exactement les rationnels de la forme $p/2^q$ (appelés rationnels dyadiques) de l'intervalle $]0, 1[$. On dispose donc

pour les rationnels dyadiques de deux développements : le premier, appelé *développement fini*, s'achève par une suite ininterrompue de 0, le second s'achève par une suite ininterrompue de 1.

Il n'y a pas d'autre cas de non-unicité : supposons en effet

$$0, x_1 x_2 \dots x_j \dots = 0, y_1 y_2 \dots y_j \dots$$

Soit k le premier entier tel que $x_k \neq y_k$. En échangeant au besoin les deux membres de l'égalité précédente, on peut supposer que $x_k = 1$, $y_k = 0$. On a alors

$$\frac{1}{2^k} + \sum_{j=k+1}^{\infty} \frac{x_j}{2^j} = \sum_{j=k+1}^{\infty} \frac{y_j}{2^j}$$

On déduit de la relation (9.11) que la seule possibilité est $x_j = 0$ pour tout $j \geq k+1$, $y_j = 1$ pour tout $j \geq k+1$, autrement dit qu'on se trouve dans la situation (9.12).

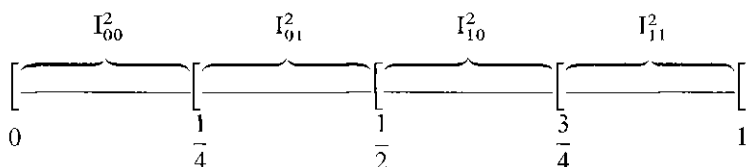
Retenons pour finir que pour $x \in [0, 1[$ la suite de terme général $D_n(x)$ fournit les chiffres du développement décimal de x lorsque celui-ci est unique. Lorsque $x \in]0, 1[$ est rationnel dyadique, elle fournit les chiffres du développement *fini*. En effet, un calcul simple où on pose $d_n = D_n(x)$ pour alléger l'écriture montre que $R_n(x) = 0, d_{n+1} d_{n+2} \dots$. Puisque $0,111\dots = 1$ et que $R_n(x) < 1$, le développement $x = 0, d_1 d_2 \dots d_n \dots$ ne peut pas être formé à partir d'un certain rang d'une suite ininterrompue de 1.

Proposition 9.18. *Soit l'espace probabilisé $([0, 1[, \mathcal{B}_{[0,1[}, \mathbb{P})$ où \mathbb{P} est la restriction de la mesure de Lebesgue à $[0, 1[$. Sur cet espace, la suite $(D_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires indépendantes de même loi de Bernoulli $\mathcal{B}(1, \frac{1}{2}) \equiv \frac{1}{2}(\delta_0 + \delta_1)$. De plus, pour tout $n \in \mathbb{N}^*$, la variable aléatoire R_n est de loi uniforme sur $[0, 1[$ et les variables aléatoires R_n et (D_1, D_2, \dots, D_n) sont indépendantes.*

Démonstration. • Pour tout $n \in \mathbb{N}^*$ et pour tout n -uple $\underline{\varepsilon}_n = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) \in \{0, 1\}^n$, notons $\underline{I}_{\underline{\varepsilon}_n}^n$ l'intervalle dyadique

$$\underline{I}_{\underline{\varepsilon}_n}^n = \left[\sum_{j=1}^n \frac{\varepsilon_j}{2^j}, \sum_{j=1}^n \frac{\varepsilon_j}{2^j} + \frac{1}{2^n} \right[.$$

Cet intervalle est constitué des réels de $[0, 1[$ dont le développement dyadique commence par $0, \varepsilon_1 \varepsilon_2 \dots \varepsilon_n$. Par exemple, pour $n = 2$, on a :



On a

$$\bigcap_{j=1}^n (D_j = \varepsilon_j) = \underline{\mathbf{1}}_{\varepsilon_n}^n,$$

si bien que :

$$P\left[\bigcap_{j=1}^n (D_j = \varepsilon_j)\right] = \frac{1}{2^n}.$$

Mais alors, pour toute partie non vide J de $\{1, 2, \dots, n\}$, on obtient, en sommant sur tous les $\varepsilon_j \in J^c$:

$$P\left[\bigcap_{j \in J} (D_j = \varepsilon_j)\right] = \frac{1}{2^{|J|}}.$$

En particulier, pour tout $j \in \{1, 2, \dots, n\}$, on a

$$P(D_j = \varepsilon_j) = \frac{1}{2}.$$

On obtient alors que

$$P\left[\bigcap_{j \in J} (D_j = \varepsilon_j)\right] = \prod_{j \in J} P(D_j = \varepsilon_j),$$

c'est-à-dire, puisque n et J sont arbitraires, que les variables aléatoires D_j forment une suite de variables aléatoires indépendantes de loi de Bernoulli $\mathcal{B}(1, \frac{1}{2})$.

• On a, en notant \mathbb{I} l'application $x \mapsto x$ de $[0, 1[$ dans \mathbb{R} :

$$R_n = 2^n \mathbb{I} - \sum_{j=1}^n 2^{n-j} D_j.$$

Alors, pour tout $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R})$ et tout $\underline{\varepsilon}_n = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) \in \{0, 1\}^n$, on a

$$\begin{aligned} E\left[f(R_n) \prod_{j=1}^n \mathbf{1}_{(D_j = \varepsilon_j)}\right] &= E\left[f\left(2^n \mathbb{I} - \sum_{j=1}^n 2^{n-j} \varepsilon_j\right) \prod_{j=1}^n \mathbf{1}_{(D_j = \varepsilon_j)}\right] \\ &= \int_{\mathbb{R}} \underline{\mathbf{1}}_{\frac{1}{2^n}}^n(x) f\left(2^n x - \sum_{j=1}^n 2^{n-j} \varepsilon_j\right) d\lambda(x), \end{aligned}$$

soit, en faisant le changement de variables dans l'intégrale de Lebesgue défini par $y = 2^n x - \sum_{j=1}^n 2^{n-j} \varepsilon_j$,

$$\begin{aligned} E\left[f(R_n) \prod_{j=1}^n \mathbf{1}_{(D_j = \varepsilon_j)}\right] &= \int_{\mathbb{R}} \underline{\mathbf{1}}_{\frac{1}{2^n}}^n\left(\frac{1}{2^n} y + \sum_{j=1}^n \frac{\varepsilon_j}{2^j}\right) f(y) \frac{1}{2^n} d\lambda(y) \\ &= \frac{1}{2^n} \int_{\mathbb{R}} \mathbf{1}_{[0, 1[}(y) f(y) d\lambda(y), \end{aligned}$$

soit encore

$$E\left[f(R_n) \prod_{j=1}^n \mathbf{1}_{(D_j = \varepsilon_j)}\right] = P\left[\bigcap_{j=1}^n (D_j = \varepsilon_j)\right] \left[\int_{\mathbb{R}} \mathbf{1}_{[0,1]}(y) f(y) d\lambda(y)\right], \quad (9.13)$$

et donc, en sommant sur $\underline{\varepsilon}_n \in \{0, 1\}^n$ dans chacun des membres de (9.13) :

$$E[f(R_n)] = \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(y) f(y) d\lambda(y).$$

Ceci démontre que R_n est de loi uniforme sur $[0, 1[$. De plus, pour tout $f \in \mathcal{C}_X^+(\mathbb{R})$, pour toute partie J de $\{1, 2, \dots, n\}$ et pour tout $(\varepsilon_j)_{j \in J} \in \{0, 1\}^J$, il vient en sommant dans chacun des membres de (9.13) sur tous les ε_j , $j \in J^c$:

$$E\left[f(R_n) \prod_{j \in J} \mathbf{1}_{(D_j = \varepsilon_j)}\right] = \left[\prod_{j \in J} P(D_j = \varepsilon_j)\right] E[f(R_n)],$$

ce qui démontre que R_n et (D_1, D_2, \dots, D_n) sont indépendantes. \square

Remarque. La suite de variables aléatoires (R_n) ne constitue pas une famille de variables aléatoires indépendantes : on peut par exemple observer que si c'était le cas, puisque

$$D_n = -R_n + 2R_{n-1}$$

et que R_{n-1} et R_n admettent une densité, il en serait de même de D_n (voir section suivante, prop. 9.23), ce qui est faux !

En corollaire, nous donnons une démonstration constructive de l'existence d'une suite de variables aléatoires réelles **indépendantes de lois arbitraires données**⁴.

Corollaire 9.19. Soit $(\mu_j)_{j \in \mathbb{N}^*}$ une suite de probabilités sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Il existe une suite de variables aléatoires réelles $(X_j)_{j \in \mathbb{N}^*}$ définies sur l'espace probabilisé $([0, 1[, \mathcal{B}_{[0,1]}, P)$, où P est la probabilité restriction de la mesure de Lebesgue à $[0, 1[$, indépendantes et telle que, pour tout $j \in \mathbb{N}^*$, X_j soit de loi μ_j .

Démonstration. On commence par prouver l'existence d'une suite de variables aléatoires indépendantes de même loi uniforme sur l'intervalle $[0, 1[$. Le cas général s'en déduit aisément.

4. La première présentation mathématiquement rigoureuse des suites de variables aléatoires indépendantes — et en particulier du jeu de pile ou face — est due à Hugo Steinhaus (1923, 1930), qui proposa de considérer ces variables aléatoires comme des fonctions mesurables définies sur $[0, 1[$. Le travail de Steinhaus précède de quelques années la publication par Kolmogorov de sa construction axiomatique de la théorie des probabilités, fondée sur la théorie de la mesure et l'emploi d'espaces probabilisés arbitraires (Ω, \mathcal{A}, P) (1929, 1933).

Reprenons les notations de la proposition 9.18. Les fonctions D_n ($n \in \mathbb{N}^*$) sont considérées comme des variables aléatoires définies sur $[0, 1[$ muni de sa tribu borélienne et de la mesure de Lebesgue.

Soit $(N_j)_{j \in \mathbb{N}^*}$ une suite de sous-ensembles infinis formant une partition de \mathbb{N}^* :

$$\mathbb{N}^* = \bigsqcup_{j \in \mathbb{N}^*} N_j$$

et soit φ_j la suite obtenue en prenant dans l'ordre croissant les éléments de N_j . Pour obtenir une telle partition, on peut partir d'une bijection $\Phi : \mathbb{N}^* \times \mathbb{N}^* \rightarrow \mathbb{N}^*$, par exemple la bijection donnée par $\Phi(j, k) = \frac{(j+k-2)(j+k-1)}{2} + k$. On note φ_j la suite croissante $k \mapsto \Phi(j, k)$ et on appelle N_j l'image de \mathbb{N}^* par cette suite.

Pour $j \in \mathbb{N}^*$, on pose

$$Y_j = \sum_{k=1}^{\infty} \frac{1}{2^k} D_{\varphi_j(k)}.$$

(En d'autres termes, on répartit les décimales dyadiques de x en une infinité de sous-suites et on forme un nouveau réel, noté $Y_j(x)$ avec chacune de ces sous-suites.)

Les variables aléatoires Y_j , $j \in \mathbb{N}$, sont *indépendantes*. En effet chaque Y_j est mesurable par rapport à la tribu $\sigma(D_n, n \in N_j)$ et ces tribus sont indépendantes puisque les N_j forment une partition de \mathbb{N}^* et que les D_n , $n \in \mathbb{N}^*$ sont indépendantes (cf. th. 9.12).

Pour tout $j \in \mathbb{N}^*$ la loi de Y_j est uniforme sur $[0, 1[$. En effet, posons pour $n \in \mathbb{N}^*$

$$Y_{j,n} = \sum_{k=1}^n \frac{1}{2^k} D_{\varphi_j(k)}.$$

La loi d'une somme de variables aléatoires indépendantes ne dépend que des lois de ces variables aléatoires (voir plus loin, prop. 9.22). Comme les variables aléatoires $D_1, D_2, \dots, D_k, \dots$ sont indépendantes et ont toutes la même loi, la loi de $Y_{j,n}$ est donc la même que celle de

$$Z_n = \sum_{k=1}^n \frac{1}{2^k} D_k.$$

Notons que $Z = \lim_n \nearrow Z_n$ est en fait la fonction identité sur $[0, 1[$. Puisque $Y_j = \lim_n \nearrow Y_{j,n}$, on a $(Y_j \leq y) = \lim_n \searrow (Y_{j,n} \leq y)$. De même $(Z \leq y) = \lim_n \searrow (Z_n \leq y)$. Par suite on obtient pour $y \in [0, 1[$

$$P(Y_j \leq y) = \lim_n P(Y_{j,n} \leq y) = \lim_n P(Z_n \leq y) = P(Z \leq y) = y.$$

Enfin, si F_j est la fonction de répartition de la probabilité μ_j définie par

$$\forall x \in \mathbb{R} \quad F_j(x) = \mu_j([-\infty, x]).$$

si G_j est sa « pseudo-inverse » définie par

$$\forall t \in \mathbb{R} \quad G_j(t) = \inf\{x \mid F_j(x) \geq t\},$$

et si $X_j = G_j(Y_j)$, il résulte de l'exercice 1 du chapitre 8 (résultat fondamental pour la simulation de lois de probabilité) que la loi de X_j est μ_j , ce qui restait à démontrer. \square

Complément. Probabilités produit sur $\{0, 1\}^{\mathbb{N}^*}$.

Modélisation du jeu de pile ou face au moyen d'un espace de suites.

Il est en fait possible, à partir du modèle fondé sur $[0, 1[$, de construire un modèle où l'espace fondamental est l'espace des suites $\{0, 1\}^{\mathbb{N}^*}$, qui est le modèle « naturel » auquel nous faisons allusion au début de cette section.

Notons D l'application de $[0, 1[$ dans $\{0, 1\}^{\mathbb{N}^*}$ définie par :

$$\forall x \in [0, 1[\quad D(x) = (D_n(x))_{n \in \mathbb{N}^*}$$

Il résulte de la relation (9.9) que D est *injective*. En vertu de l'analyse que nous avons faite du développement dyadique, D a pour image $\{0, 1\}^{\mathbb{N}^*}$ privé du sous-ensemble Ω_1 formé des suites qui valent 1 à partir d'un certain rang.

On munit $\{0, 1\}^{\mathbb{N}^*}$ de la « tribu produit » \mathcal{A} , produit des tribus des parties sur chacune des composantes ; plus précisément \mathcal{A} est la tribu engendrée par la famille des parties de la forme :

$$\prod_{n \in \mathbb{N}^*} A_n \equiv \left\{ \omega \in \{0, 1\}^{\mathbb{N}^*} \mid \forall n \in \mathbb{N}^* \quad \omega_n \in A_n \right\},$$

où A_n est une partie de $\{0, 1\}$ égale à $\{0, 1\}$ sauf pour un nombre fini d'indices n .

L'application D de $([0, 1[, \mathcal{B}_{[0,1[})$ dans $(\{0, 1\}^{\mathbb{N}^*}, \mathcal{A})$ est mesurable : en effet il suffit de constater que pour tout $I \in \mathcal{P}_f(\mathbb{N}^*)$, pour tout $n \in I$, et pour tout $A_n = \{\varepsilon_n\}$ où $\varepsilon_n \in \{0, 1\}$, on a :

$$D^{-1}\left(\prod_{n \in \mathbb{N}^*} A_n\right) = \bigcap_{n \in I} (D_n = \varepsilon_n) \in \mathcal{B}_{[0,1[}.$$

On note alors Q la probabilité sur $(\Omega, \mathcal{A} \cap \Omega)$, image de P par l'application mesurable D .

En prenant pour Z_j la projection de $\{0, 1\}^{\mathbb{N}^*}$ sur le j -ième facteur, on obtient une suite de variables aléatoires définies sur $(\{0, 1\}^{\mathbb{N}^*}, \mathcal{A}, Q)$, *indépendantes* et de même loi $\mathcal{B}(1, 1/2)$. On a en effet $D_j = Z_j \circ D$ pour $j \in \mathbb{N}^*$. Il en résulte, par définition même de la probabilité Q , que $Q(Z_j = \varepsilon) = P(D_j = \varepsilon)$, et plus généralement que $Q(Z_{j_1} = \varepsilon_{j_1}, \dots, Z_{j_n} = \varepsilon_{j_n}) = P(D_{j_1} = \varepsilon_{j_1}, \dots, D_{j_n} = \varepsilon_{j_n})$. Par suite, Z_j a même loi que D_j et les Z_j sont indépendants puisque les D_j le sont.

On peut noter que Q est exactement la mesure produit des probabilités uniformes sur les facteurs $\{0, 1\}$ (voir cor. A.40 pour une définition) : les propriétés à vérifier pour le montrer traduisent exactement le fait que les variables aléatoires Z_j , $j \in \mathbb{N}^*$ sont indépendantes et de même loi $\mathcal{B}(1, 1/2)$.

Exemple d'utilisation. Dans une partie infinie de pile ou face, la probabilité de voir se réaliser une infinité de fois une suite $(\varepsilon_1, \dots, \varepsilon_n)$ est égale à 1. En effet, l'événement considéré s'écrit $\limsup_n A_n$, où

$$A_j = \{(Z_{j+1}, Z_{j+2}, \dots, Z_{j+n}) = (\varepsilon_1, \dots, \varepsilon_n)\};$$

or si

$$B_j = \{(Z_{jn+1}, Z_{jn+2}, \dots, Z_{(j+1)n}) = (\varepsilon_1, \dots, \varepsilon_n)\},$$

les événements B_j sont Q -indépendants et

$$\limsup_n B_n \subset \limsup_n A_n.$$

De plus, puisque $Q(B_j) = 1/2^n$, on a

$$\sum_{j=1}^{+\infty} Q(B_j) = +\infty$$

et, par le lemme de Borel-Cantelli, $Q(\limsup_n B_n) = 1$. A fortiori, on a $Q(\limsup_n A_n) = 1$.

Remarque. Pour répondre à la question, on notera qu'il suffit d'avoir un modèle permettant de parler d'une suite infinie de variables aléatoires indépendantes de même loi $\mathcal{B}(1, 1/2)$.

Modèle canonique pour une suite de variables aléatoires de Bernoulli indépendantes.

Si maintenant on considère une suite de variables aléatoires de lois données μ_j , on peut transporter à $\{0, 1\}^{\mathbb{N}^*}$ la solution du problème « construire une suite de variables aléatoires indépendantes de loi μ_j » obtenue au corollaire 9.19 : les variables aléatoires X_j que nous avons construites étant définies sur l'espace $[0, 1[$, il suffit, après avoir observé que l'application D^{-1} définit sur $\{0, 1\}^{\mathbb{N}^*} \setminus \Omega_1$ est mesurable, de considérer les variables $X_j \circ D^{-1}$ (on les prolonge arbitrairement sur l'ensemble Ω_1 , qui est de Q -probabilité nulle).

On obtient bien un modèle où l'espace fondamental est un espace de suites, mais il n'est pas réellement « adapté » au problème : les variables aléatoires X_j ne sont pas comme dans le modèle de pile ou face ci-dessus en relation avec les projections sur les espaces facteurs (dites aussi *applications coordonnées*).

Voici, en supposant pour simplifier que les lois données μ_j soient des lois sur $\{0, 1\}$ (lois de Bernoulli de paramètre p_j), comment construire un modèle mieux adapté. Ce modèle conviendra par exemple à la description d'un jeu de pile ou face avec une pièce non équilibrée, les p_j étant alors tous égaux à p .

Soit $X: [0, 1] \rightarrow \{0, 1\}^{\mathbb{N}^*}$ l'application qui à x associe la suite $(X_j(x))_{j \in \mathbb{N}^*}$. Notons que cette application dépend du choix de la suite $\mu = (\mu_j)_{j \in \mathbb{N}^*}$. L'application X est mesurable : vu la définition de la tribu \mathcal{A} sur $\{0, 1\}^{\mathbb{N}^*}$, il suffit de vérifier que pour tout n , l'application $x \mapsto (X_1(x) \dots, X_n(x))$ est mesurable, ce qui est le cas.

Soit P_μ l'image par X de la probabilité P . Alors, si $\{0, 1\}^{\mathbb{N}^*}$ est muni de la tribu \mathcal{A} et de la probabilité P_μ , les projections $Z_j = \text{pr}_j$ sont indépendantes et de loi μ_j ($j \in \mathbb{N}^*$). Comme ci-dessus, par définition même de la probabilité image, on a $P_\mu(Z_{j_1} = \varepsilon_{j_1}, \dots, Z_{j_n} = \varepsilon_{j_n}) = P(X_{j_1} = \varepsilon_{j_1}, \dots, X_{j_n} = \varepsilon_{j_n})$, d'où il résulte que Z_j a même loi que X_j et que les Z_j sont indépendants.

On a montré du même coup que P_μ est la probabilité produit des probabilités μ_j , $j \in \mathbb{N}^*$.

En particulier, la loi P_μ ne dépend que des μ_j , et non des X_j dont la construction fait intervenir des choix arbitraires. C'est pourquoi on peut appeler le modèle que nous venons de construire (espace fondamental $(\{0, 1\}^{\mathbb{N}^*}, \mathcal{A}, P_\mu)$ et variables aléatoires Z_j , $j \in \mathbb{N}^*$) *modèle canonique* pour la réalisation d'une suite de variables aléatoires de Bernoulli indépendantes et de lois données. On peut observer que nous avons démontré le théorème de Kolmogorov dans un cas particulier.

Constatons aussi qu'une fois de plus on a obtenu pour la solution d'un problème de modélisation deux solutions équivalentes (et cette fois assez différentes) : la présente solution et celle du corollaire 9.19.

Remarque. Considérons le cas où les μ_j sont toutes égales à $\mathcal{B}(1, p)$ avec $0 < p < 1$. On écrit alors P_p au lieu de P_μ (on a par exemple $P_{1/2} = Q$).

La mesure P_p est diffuse, et étrangère à Q si $p \neq \frac{1}{2}$. On a en effet pour $\omega \in \{0, 1\}$ et pour tout $n \in \mathbb{N}^*$

$$\begin{aligned} P_p(\{\omega\}) &= P_p(Z_1 = \omega_1, \dots, Z_n = \omega_n, \dots) \\ &\leq P_p(Z_1 = \omega_1, \dots, Z_n = \omega_n) \leq \rho^n \end{aligned}$$

où $\rho = \sup(p, 1 - p)$. D'où $P_p(\{\omega\}) \leq \lim_n \rho^n = 0$. Pour montrer que P_p et Q sont étrangères, le plus simple est d'exhiber un événement qui soit à la fois de probabilité 1 pour P_p et de probabilité 0 pour Q . C'est le cas de l'événement $\lim_n \frac{1}{n}(Z_1 + \dots + Z_n) = p$ (loi forte des grands nombres, th. 10.20).

On tire de là la construction d'une mesure sur $[0, 1]$ diffuse et étrangère

à la mesure de Lebesgue. Les probabilités P_p et Q attribuant l'une et l'autre la mesure 0 au sous-ensemble dénombrable Ω_1 , on peut restreindre ces probabilités à $\{0, 1\}^{\mathbb{N}^*} \setminus \Omega_1$. D'autre part D définit une bijection de $[0, 1[$ sur ce dernier ensemble et l'inverse de cette bijection, que nous noterons D^{-1} , est mesurable. En effet, la tribu $\mathcal{B}_{[0,1]}$ étant engendrée par les intervalles dyadiques de la forme $[\frac{k}{2^n}, \frac{k+1}{2^n}[$, il suffit de remarquer que l'image inverse d'un tel intervalle par D^{-1} est l'ensemble des $\omega \in \Omega$ dont les n premières coordonnées sont égales aux n premiers chiffres du développement dyadique de $\frac{k}{2^n}$. En considérant maintenant les images par D^{-1} des probabilités Q et P_p on obtient d'une part la mesure de Lebesgue sur $[0, 1[$, d'autre part une probabilité diffuse sur $[0, 1[$, nécessairement étrangère à la mesure de Lebesgue.

9.4. Convolution et loi de la somme de variables aléatoires indépendantes

Définition 9.20. Soient μ_1 et μ_2 deux mesures bornées (resp. deux probabilités) sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Soit S l'application somme sur \mathbb{R}^d . La mesure image de $\mu_1 \otimes \mu_2$ par S est appelée **produit de convolution** de μ_1 et μ_2 et notée $\mu_1 * \mu_2$. C'est une mesure bornée (resp. une probabilité) sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$.

Proposition 9.21. Pour toute fonction mesurable positive f sur \mathbb{R}^d on a :

$$\int_{\mathbb{R}^d} f d(\mu_1 * \mu_2) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x_1 + x_2) d(\mu_1 \otimes \mu_2)(x_1, x_2). \quad (9.14)$$

De plus, $f \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, \mu_1 * \mu_2)$ si et seulement si l'application $(x_1, x_2) \mapsto f(x_1 + x_2)$ est $(\mu_1 \otimes \mu_2)$ -intégrable. Dans ce cas l'égalité précédente est encore satisfaite.

Démonstration. La démonstration est standard. Si $f = \mathbf{1}_A$, où $A \in \mathcal{B}_{\mathbb{R}^d}$, c'est la définition de $\mu_1 * \mu_2$. L'égalité (9.14) est alors vraie par linéarité pour toute fonction étagée positive, puis par la propriété de Beppo Levi, pour toute fonction mesurable positive (prendre une suite croissante de fonctions étagées positives convergeant vers f). Le reste de la démonstration est encore standard : pour l'intégrabilité, prendre les valeurs absolues et appliquer (9.14); enfin, décomposer f en ses parties positive et négative. \square

Proposition 9.22. Soient sur l'espace probabilisé (Ω, \mathcal{A}, P) deux variables aléatoires indépendantes X_1 et X_2 à valeurs dans \mathbb{R}^d . La loi de $X_1 + X_2$ est le **produit de convolution** des lois de X_1 et X_2 :

$$P_{X_1 + X_2} = P_{X_1} * P_{X_2}. \quad (9.15)$$

Démonstration. Puisque X_1 et X_2 sont indépendantes, on a $P_{X_1, X_2} = P_{X_1} \otimes P_{X_2}$. Il en résulte que :

$$P_{X_1 + X_2} = P_{S \circ (X_1, X_2)} = S [P_{X_1} \otimes P_{X_2}] = P_{X_1} * P_{X_2}. \quad \square$$

Proposition 9.23. *Si, outre les hypothèses de la proposition 9.22, on suppose que X_1 et X_2 admettent des densités respectives f_{X_1} et f_{X_2} , alors $X_1 + X_2$ admet une densité $f_{X_1 + X_2}$ définie par*

$$\forall y \in \mathbb{R}^d \quad f_{X_1 + X_2}(y) = \int_{\mathbb{R}^d} f_{X_1}(x_1) f_{X_2}(y - x_1) d\lambda_d(x_1) \\ = \int_{\mathbb{R}^d} f_{X_1}(y - x_2) f_{X_2}(x_2) d\lambda_d(x_2).$$

On dit que $f_{X_1 + X_2}$ est le **produit de convolution** des fonctions f_{X_1} et f_{X_2} .

Démonstration. Pour tout $A \in \mathcal{B}_{\mathbb{R}^d}$, on a, d'après la proposition 9.22 et le fait que X_1 et X_2 admettent des densités et sont indépendantes :

$$P_{X_1 + X_2}(A) = \int_{\mathbb{R}^d} \mathbf{1}_A(x_1 + x_2) f_{X_1}(x_1) f_{X_2}(x_2) d\lambda_d \otimes \lambda_d(x_1, x_2).$$

Faisons le changement de variables

$$\begin{cases} y_1 = x_1 + x_2 \\ y_2 = x_2. \end{cases}$$

Le difféomorphisme associé étant de jacobien 1, il vient

$$P_{X_1 + X_2}(A) = \int_{\mathbb{R}^d} \mathbf{1}_A(y_1) f_{X_1}(y_1 - y_2) f_{X_2}(y_2) d\lambda_d \otimes \lambda_d(y_1, y_2),$$

soit, d'après le théorème de Fubini,

$$P_{X_1 + X_2}(A) = \int_A \left[\int_{\mathbb{R}^d} f_{X_1}(y_1 - y_2) f_{X_2}(y_2) d\lambda_d(y_2) \right] d\lambda_d(y_1),$$

ce qui démontre le résultat. □

Exemple 9.3. Soient X_1 et X_2 deux variables aléatoires réelles définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , indépendantes, de lois respectives $\gamma(a_1, p)$ et $\gamma(a_2, p)$, où a_1, a_2 et p sont des réels strictement positifs. La loi de $X_1 + X_2$ est la loi $\gamma(a_1 + a_2, p)$.

En effet, pour $i = 1, 2$, on a

$$\forall x \in \mathbb{R} \quad f_{X_i}(x) = \mathbf{1}_{\mathbb{R}^+(x)} \frac{p^{a_i}}{\Gamma(a_i)} \exp(-px) x^{a_i - 1},$$

et $X_1 + X_2$ admet pour densité le produit de convolution des fonctions f_{X_1} et f_{X_2} ; or, on a

$$f_{X_1}(y - x_2) f_{X_2}(x_2) = \mathbf{1}_{\mathbb{R}^+}(y - x_2) \mathbf{1}_{\mathbb{R}^+}(x_2) \frac{p^{a_1} p^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \exp(-py) (y - x_2)^{a_1 - 1} x_2^{a_2 - 1},$$

et, comme on a⁵

$$\mathbf{1}_{\mathbb{R}^+}(y - x_2) \mathbf{1}_{\mathbb{R}^+}(x_2) = \mathbf{1}_{\mathbb{R}^+}(y) \mathbf{1}_{[0, y]}(x_2),$$

il vient :

$$f_{X_1 + X_2}(y) = \mathbf{1}_{\mathbb{R}^+}(y) \frac{p^{a_1 + a_2}}{\Gamma(a_1) \Gamma(a_2)} \exp(-py) \int_0^y (y - x_2)^{a_1 - 1} x_2^{a_2 - 1} dx_2,$$

soit, après changement de variables,

$$f_{X_1 + X_2}(y) = \mathbf{1}_{\mathbb{R}^+}(y) \frac{p^{a_1 + a_2}}{\Gamma(a_1) \Gamma(a_2)} \exp(-py) y^{a_1 + a_2 - 1} \int_0^1 (1 - u)^{a_1 - 1} u^{a_2 - 1} du;$$

Mais, puisque l'application $y \mapsto \mathbf{1}_{\mathbb{R}^+}(y) \frac{p^{a_1 + a_2}}{\Gamma(a_1 + a_2)} \exp(-py) y^{a_1 + a_2 - 1}$ est une densité de probabilité, la fonction $f_{X_1 + X_2}$ est égale à cette densité (le coefficient normalisateur est unique !) et $X_1 + X_2$ suit la loi $\gamma(a_1 + a_2, p)$. On obtient de plus ainsi la relation :

$$\int_0^1 (1 - u)^{a_1 - 1} u^{a_2 - 1} du = \frac{\Gamma(a_1) \Gamma(a_2)}{\Gamma(a_1 + a_2)}.$$

Exercices

Sauf mention spéciale, toutes les variables aléatoires seront définies sur un même espace probabilisé (Ω, \mathcal{A}, P) .

Exercice 9.1. Variables aléatoires gaussiennes indépendantes; indépendance du rayon et de l'angle polaire. X et Y sont deux variables aléatoires réelles indépendantes de même loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$. On note Φ l'application de $\mathbb{R}^{+*} \times]0, 2\pi[$ sur $\mathbb{R}^2 \setminus \mathbb{R}^+ \times \{0\}$ (transformation polaire) définie par

$$\forall (\rho, \theta) \in \mathbb{R}^{+*} \times]0, 2\pi[\quad \Phi(\rho, \theta) = (\rho \cos \theta, \rho \sin \theta),$$

et Ψ son inverse. On définit alors l'application $\widehat{\Psi}$ de \mathbb{R}^2 dans $\mathbb{R} \times]0, 2\pi[$ par

$$\widehat{\Psi}(x, y) = \begin{cases} \Psi(x, y) & \text{si } (x, y) \in \mathbb{R}^2 \setminus \mathbb{R}^+ \times \{0\} \\ (0, 0) & \text{sinon.} \end{cases}$$

5. Le maniement méticuleux des fonctions indicatrices évitera bien des erreurs dans la détermination du support des lois.

On définit la variable aléatoire (R, Θ) par

$$(R, \Theta) = \widehat{\Psi}(X, Y).$$

Déterminer la loi de la variable aléatoire (R, Θ) par sa densité ; justifier l'indépendance des variables aléatoires R et Θ d'une part, et des variables aléatoires $S \equiv R^2$ et Θ d'autre part. Spécifier la loi de S .

Solution. Pour tout $f \in \mathcal{C}_c^+(\mathbb{R}^2)$, on a, par le théorème de transfert :

$$E[f(R, \Theta)] = E[f \circ \widehat{\Psi}(X, Y)] = \int_{\mathbb{R}^2} f \circ \widehat{\Psi}(x, y) dP_{(X, Y)}(x, y).$$

Les variables aléatoires X et Y sont indépendantes et admettent des densités ; on a donc :

$$E[f(R, \Theta)] = \int_{\mathbb{R}^2} f \circ \widehat{\Psi}(x, y) f_X(x) f_Y(y) d\lambda_2(x, y).$$

Puisque, dans le plan, une demi-droite est de mesure de Lebesgue nulle, on a alors :

$$\begin{aligned} E[f(R, \Theta)] &= \int_{\mathbb{R}^2 \setminus \mathbb{R}^+ \times \{0\}} f \circ \Psi(x, y) f_X(x) f_Y(y) d\lambda_2(x, y) \\ &= \int_{\mathbb{R}^2 \setminus \mathbb{R}^+ \times \{0\}} f \circ \Psi(x, y) \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) d\lambda_2(x, y). \end{aligned}$$

Effectuons le changement de coordonnées polaires défini par le difféomorphisme Ψ , de jacobien r ; il vient

$$E[f(R, \Theta)] = \int_{\mathbb{R}^+ \times]0, 2\pi[} f(r, \theta) \frac{1}{2\pi} \exp\left(-\frac{r^2}{2}\right) r d\lambda_2(r, \theta),$$

soit encore :

$$E[f(R, \Theta)] = \int_{\mathbb{R}^2} f(r, \theta) \mathbf{1}_{\mathbb{R}^+ \times]0, 2\pi[}(r, \theta) \frac{1}{2\pi} \exp\left(-\frac{r^2}{2}\right) r d\lambda_2(r, \theta).$$

On en déduit que (R, Θ) admet la densité $f_{(R, \Theta)}$ donnée par

$$\forall (r, \theta) \in \mathbb{R}^2 \quad f_{(R, \Theta)}(r, \theta) = \mathbf{1}_{]0, 2\pi[}(\theta) \frac{1}{2\pi} \mathbf{1}_{\mathbb{R}^+}(r) \exp\left(-\frac{r^2}{2}\right) r.$$

La densité $f_{(R, \Theta)}$ étant produit direct de deux fonctions mesurables positives, les variables aléatoires R et Θ sont indépendantes et de densités respectives f_R et f_Θ données par

$$\forall r \in \mathbb{R} \quad f_R(r) = \mathbf{1}_{\mathbb{R}^+}(r) \exp\left(-\frac{r^2}{2}\right) r.$$

$$\forall \theta \in \mathbb{R} \quad f_\Theta(\theta) = \mathbf{1}_{]0, 2\pi[}(\theta) \frac{1}{2\pi}.$$

La loi de Θ est la loi uniforme sur $]0, 2\pi[$.

Les variables aléatoires S et Θ sont encore indépendantes comme fonctions mesurables de telles variables aléatoires.

Enfin, pour tout $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R})$, en suivant les mêmes étapes que précédemment, on a

$$E[f(S)] = E[f(R^2)] = \int_{\mathbb{P}} f(r^2) dP_{\mathbb{R}}(r),$$

soit, puisque la mesure de Lebesgue d'un singleton est nulle,

$$E[f(S)] = \int_{\mathbb{R}^{++}} f(r^2) \exp\left(-\frac{r^2}{2}\right) r d\lambda(r),$$

et, en effectuant le changement de coordonnées défini par $s = r^2$, difféomorphisme de \mathbb{R}^{++} sur lui-même de jacobien $\frac{1}{2\sqrt{s}}$:

$$E[f(S)] = \int_{\mathbb{R}} f(s) \mathbf{1}_{\mathbb{R}^{++}}(s) \frac{1}{2} \exp\left(-\frac{s}{2}\right) d\lambda(s),$$

c'est-à-dire que S suit la loi exponentielle $\exp(\frac{1}{2})$.

Exercice 9.2. Sur la voie de la simulation de deux variables aléatoires indépendantes de loi gaussienne $\mathcal{N}_{\mathbb{P}}(0, 1)$. S et Θ sont deux variables aléatoires réelles indépendantes de loi respective la loi exponentielle $\exp(1/2)$ et la loi uniforme sur $[0, 2\pi]$. On définit les deux variables aléatoires réelles X et Y par

$$X = \sqrt{S} \cos \Theta \quad Y = \sqrt{S} \sin \Theta.$$

Quelle est la loi de la variable aléatoire (X, Y) ?

Solution. Remarquons que $P(S \geq 0) = 1$; (X, Y) est donc définie presque sûrement. Pour tout $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R}^2)$, on a.

$$E[f(X, Y)] = E\left[f(\sqrt{S} \cos \Theta, \sqrt{S} \sin \Theta)\right],$$

soit, par le théorème de transfert,

$$E[f(X, Y)] = \int_{\mathbb{R}^2} f(\sqrt{s} \cos \theta, \sqrt{s} \sin \theta) dP_{(S, \Theta)}(s, \theta).$$

Les variables aléatoires S et Θ sont indépendantes et admettent des densités; on a donc

$$E[f(X, Y)] = \int_{\mathbb{R}^2} f(\sqrt{s} \cos \theta, \sqrt{s} \sin \theta) f_S(s) f_{\Theta}(\theta) d\lambda_2(s, \theta),$$

soit :

$$E[f(X, Y)] = \int_{\mathbb{R}^{++} \times]0, 2\pi[} f(\sqrt{s} \cos \theta, \sqrt{s} \sin \theta) \frac{1}{2\pi} \frac{1}{2} \exp\left(-\frac{s}{2}\right) d\lambda_2(s, \theta).$$

Effectuons le changement de variables associé au difféomorphisme H de $\mathbb{R}^{++} \times]0, 2\pi[$ sur $\mathbb{R}^2 \setminus \mathbb{R}^+ \times \{0\}$ défini par

$$\forall (s, \theta) \in \mathbb{R}^{++} \times]0, 2\pi[\quad x = \sqrt{s} \cos \theta \quad y = \sqrt{s} \sin \theta.$$

Puisque

$$H'(s, \theta) = \begin{vmatrix} \frac{1}{2\sqrt{s}} \cos \theta & -\sqrt{s} \sin \theta \\ \frac{1}{2\sqrt{s}} \sin \theta & \sqrt{s} \cos \theta \end{vmatrix}$$

et que

$$(H^{-1})'(x, y) = [H'(H^{-1}(x, y))]^{-1},$$

le jacobien du difféomorphisme vaut :

$$\det(H^{-1})'(x, y) = \frac{1}{\det H'(H^{-1}(x, y))} = 2.$$

Il vient alors

$$E[f(X, Y)] = \int_{\mathbb{R}^2 \setminus \{x=0\}} f(x, y) \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) d\lambda_2(x, y),$$

soit, puisque toute demi-droite est de mesure de Lebesgue nulle

$$E[f(X, Y)] = \int_{\mathbb{R}^2} f(x, y) \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) d\lambda_2(x, y).$$

Ceci démontre que (X, Y) admet la densité $f_{(X, Y)}$ donnée par

$$\forall (x, y) \in \mathbb{R}^2 \quad f_{(X, Y)}(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right).$$

Il en résulte que X et Y sont deux variables aléatoires réelles indépendantes de même loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$.

Remarque. On a vu en exercice que si U est une variable aléatoire de loi uniforme sur $]0, 1[$, la variable aléatoire $-2 \ln U$ suit la loi exponentielle $\exp(1/2)$; cette remarque et l'exercice donnent alors une méthode de simulation de deux variables aléatoires indépendantes de loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$.

Exercice 9.3. Loi et moments de maximum de variables aléatoires de loi exponentielle. On se donne une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes de même loi exponentielle $\exp(1)$. On définit pour tout $n \in \mathbb{N}^*$ les variables aléatoires Y_n et Z_n par

$$Y_n = \max_{1 \leq i \leq n} X_i \quad Z_n = \sum_{i=1}^n \frac{X_i}{i}.$$

Démontrer par récurrence que Y_n et Z_n ont même loi; en déduire les moyenne et variance de Y_n .

Solution.

- La fonction de répartition de Y_n est donnée par, pour tout $y \in \mathbb{R}$,

$$F_{Y_n}(y) = P\left[\bigcap_{i=1}^n (X_i \leq y)\right] = \begin{cases} \left(\int_0^y \exp(-u) du\right)^n & \text{si } y > 0 \\ 0 & \text{si } y \leq 0. \end{cases}$$

La variable aléatoire Y_n admet donc une densité donnée, sauf en 0, par la dérivée de F_{Y_n} , soit :

$$\forall y \in \mathbb{R} \quad f_{Y_n}(y) = \mathbf{1}_{\mathbb{R}^+}(y) n \exp(-y) (1 - \exp(-y))^{n-1}.$$

• On a $Y_1 = Z_1$: ces variables aléatoires ont donc même loi. Supposons que Y_n et Z_n aient même loi ; on a :

$$Z_{n+1} = Z_n + \frac{X_{n+1}}{n+1}.$$

Les variables aléatoires Z_n et $X_{n+1}/(n+1)$ sont indépendantes ; Z_{n+1} a alors une densité convolution de celles de Z_n et $X_{n+1}/(n+1)$: or (petit calcul classique), la loi de $X_{n+1}/(n+1)$ est la loi $\exp(n+1)$, on a donc pour tout $z \in \mathbb{R}$:

$$f_{Z_{n+1}}(z) = \int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}^+}(y) n \exp(-y) (1 - \exp(-y))^{n-1} \left[(n+1) \mathbf{1}_{\mathbb{R}^+}(z-y) \exp(-(n+1)(z-y)) \right] d\lambda(y),$$

soit, puisque $\mathbf{1}_{\mathbb{R}^+}(y) \mathbf{1}_{\mathbb{R}^+}(z-y) = \mathbf{1}_{\mathbb{R}^+}(z) \mathbf{1}_{[0,z]}(y)$,

$$f_{Z_{n+1}}(z) = \mathbf{1}_{\mathbb{R}^+}(z) n(n+1) \exp(-(n+1)z) \int_0^z \exp(ny) (1 - \exp(-y))^{n-1} dy.$$

Mais on a

$$\begin{aligned} \int_0^z \exp(ny) (1 - \exp(-y))^{n-1} dy &= \int_0^z \exp(y) (\exp(y) - 1)^{n-1} dy \\ &= \frac{(\exp(y) - 1)^n}{n} \Big|_0^z \\ &= \frac{(\exp(z) - 1)^n}{n}, \end{aligned}$$

ce qui démontre que

$$f_{Z_{n+1}}(z) = \mathbf{1}_{\mathbb{R}^+}(z) (n+1) \exp(-z) (1 - \exp(-z))^n = f_{Y_{n+1}}(z),$$

c'est-à-dire que : $f_{Z_{n+1}} = f_{Y_{n+1}}$.

• On a alors $EY_n = EZ_n = \sum_{i=1}^n \frac{1}{i} EX_i$, soit :

$$EY_n = \sum_{i=1}^n \frac{1}{i}.$$

Les variables aléatoires Y_n et Z_n ont aussi même variance et, les variables aléatoires X_i étant indépendantes, on a $\sigma_{Y_n}^2 = \sigma_{Z_n}^2 = \sum_{i=1}^n \frac{1}{i^2} \sigma_{X_i}^2$, soit :

$$\sigma_{Y_n}^2 = \sum_{i=1}^n \frac{1}{i^2}.$$

Exercice 9.4. Lois du chi-deux et de Student. $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires indépendantes de même loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$. On définit pour tout $n \in \mathbb{N}^*$ les variables aléatoires Y_n, T'_n et T_n par

$$Y_n = \sum_{j=1}^n X_j^2 \quad T'_n = \frac{X_{n+1}}{\sqrt{Y_n}} \quad T_n = \sqrt{n} T'_n.$$

Démontrer par récurrence que Y_n admet une densité f_{Y_n} donnée par

$$\forall t \in \mathbb{R} \quad f_{Y_n}(y) = \mathbf{1}_{\mathbb{R}^+}(y) \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \exp\left(-\frac{y}{2}\right) y^{\frac{n}{2}-1}. \quad (9.16)$$

Calculer les moyenne et variance de Y_n . Démontrer que T'_n admet une densité que l'on calculera. En déduire que T_n admet une densité f_{T_n} donnée par

$$\forall t \in \mathbb{R} \quad f_{T_n}(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{(1 + \frac{t^2}{n})^{\frac{n+1}{2}}}. \quad (9.17)$$

La loi de Y_n est la loi du **chi-deux à n degrés de liberté**, celle de T_n est la **loi de Student de paramètre n** .

Sans utiliser la densité de T_n , déterminer pour quelles valeurs de n les moyenne et variance de T_n existent et, dans ce cas, les calculer.

Solution. • Par la même méthode standard que dans les exercices précédents on montre que, pour tout $f \in \mathcal{C}_c^+(\mathbb{R})$:

$$E f(X_1^2) = \int_{\mathbb{R}} f(x) \mathbf{1}_{\mathbb{R}^+}(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x}{2}\right) x^{-\frac{1}{2}} d\lambda(x),$$

ce qui démontre que Y_1 admet une densité donnée par la relation (9.16) à l'ordre 1. Supposons que Y_n admet une densité f_{Y_n} donnée par la relation (9.16) à l'ordre n . Les variables aléatoires Y_n et X_{n+1}^2 sont indépendantes et on a $Y_{n+1} = Y_n + X_{n+1}^2$; Y_{n+1} a alors une densité convolution de celles de Y_n et X_{n+1}^2 , et, comme de plus X_{n+1}^2 et X_1^2 ont même loi, il vient

$$\forall y \in \mathbb{R} \quad f_{Y_{n+1}}(y) = \int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}^+}(x) \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \exp\left(-\frac{x}{2}\right) x^{\frac{n}{2}-1} \left[\mathbf{1}_{\mathbb{R}^+}(y-x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y-x}{2}\right) (y-x)^{-\frac{1}{2}} \right] d\lambda(x).$$

Il en résulte que, pour tout $y \in \mathbb{R}$:

$$f_{Y_{n+1}}(y) = \mathbf{1}_{\mathbb{R}^+}(y) \exp\left(-\frac{y}{2}\right) \frac{1}{\sqrt{2\pi}} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \int_{[0, y]} x^{\frac{n}{2}-1} (y-x)^{-\frac{1}{2}} d\lambda(x).$$

Mais, pour $y > 0$, par le changement de variables $x = uy$, on a

$$\int_{[0, y]} x^{\frac{n}{2}-1} (y-x)^{-\frac{1}{2}} d\lambda(x) = y^{\frac{n+1}{2}-1} \int_{[0, 1]} u^{\frac{n}{2}-1} (1-u)^{-\frac{1}{2}} d\lambda(u).$$

On a donc

$$\forall y \in \mathbb{R} \quad f_{Y_{n+1}}(y) = K \mathbf{1}_{\mathbb{R}^+}(y) \exp\left(-\frac{y}{2}\right) y^{\frac{n+1}{2}-1}, \quad (9.18)$$

où

$$K = \frac{1}{\sqrt{2\pi}} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_{(0,1)} u^{\frac{n}{2}-1} (1-u)^{-\frac{1}{2}} d\lambda(u).$$

Pour identifier K , il suffit d'écrire que $f_{Y_{n+1}}$ est une densité, d'intégrer sur \mathbb{R} et de faire le changement de variables $t = y/2$. On obtient le résultat demandé.

• On a

$$EY_1 = E(X_1^2) = \sigma_{X_1}^2 + [EX_1]^2 = 1,$$

et donc, par linéarité,

$$EY_n = \sum_{j=1}^n E(X_j^2) = n.$$

Les variables aléatoires X_j^2 étant indépendantes, on a

$$\sigma_{Y_n}^2 = \sum_{j=1}^n \sigma_{X_j^2}^2;$$

or on a

$$\sigma_{X_j^2}^2 = E(X_1^4) - [EX_1^2]^2.$$

et, un calcul simple conduisant à $E(X_1^4) = 3$, il vient :

$$\sigma_{Y_n}^2 = 2n.$$

• Remarquons que, puisque Y_n admet une densité, T'_n est définie \mathbb{P} -presque sûrement. De plus, Y_n et X_{n+1} sont indépendantes. On a alors, pour tout $f \in \mathcal{C}_K^+(\mathbb{R})$,

$$Ef(T'_n) = \int_{(Y_n \neq 0)} f\left(\frac{X_{n+1}}{\sqrt{Y_n}}\right) d\mathbb{P},$$

et, par le théorème de transfert,

$$Ef(T'_n) = \int_{\mathbb{R} \times \mathbb{R}^*} f\left(\frac{x}{\sqrt{y}}\right) f_{X_{n+1}}(x) f_{Y_n}(y) d\lambda_2(x, y),$$

soit :

$$Ef(T'_n) = \int_{\mathbb{R} \times \mathbb{R}^*} f\left(\frac{x}{\sqrt{y}}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \exp\left(-\frac{y}{2}\right) y^{\frac{n}{2}-1} d\lambda_2(x, y).$$

Faisons le changement de variables associé au difféomorphisme de l'ouvert $\mathbb{R} \times \mathbb{R}^{+*}$ sur lui-même défini par

$$\begin{cases} t = \frac{x}{\sqrt{y}} \\ z = y \end{cases} \iff \begin{cases} x = t\sqrt{z} \\ y = z. \end{cases}$$

et de jacobien :

$$\frac{D(x, y)}{D(t, z)} = \begin{vmatrix} \sqrt{z} & t \\ 0 & 1 \end{vmatrix} = \sqrt{z}.$$

Il vient

$$E f(T'_n) = \int_{\mathbb{R} \times \mathbb{R}^{+*}} f(t) \frac{1}{\sqrt{2\pi}} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \exp\left(-\frac{z}{2}(1+t^2)\right) z^{\frac{n+1}{2}-1} d\lambda_2(t, z),$$

et, par le théorème de Fubini,

$$E f(T'_n) = \int_{\mathbb{R}} f(t) \left[\int_{\mathbb{R}^{+*}} \frac{1}{\sqrt{2\pi}} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \exp\left(-\frac{z}{2}(1+t^2)\right) z^{\frac{n+1}{2}-1} d\lambda(z) \right] d\lambda(t),$$

soit, après le nouveau changement de variables défini par $u = \frac{z}{2}(1+t^2)$ et après réduction :

$$E f(T'_n) = \int_{\mathbb{R}} f(t) \frac{1}{\sqrt{\pi}} \frac{1}{\Gamma(\frac{n}{2})} \frac{1}{(1+t^2)^{\frac{n+1}{2}}} \left[\int_{\mathbb{R}^{+*}} \exp(-u) u^{\frac{n+1}{2}-1} d\lambda(u) \right] d\lambda(t).$$

Ceci démontre que T'_n admet une densité donnée par

$$\forall t \in \mathbb{R} \quad f_{T'_n}(t) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{(1+t^2)^{\frac{n+1}{2}}}.$$

Enfin, puisque $T_n = \sqrt{n} T'_n$, on a

$$\forall t \in \mathbb{R} \quad f_{T_n}(t) = f_{T'_n}\left(\frac{t}{\sqrt{n}}\right) \frac{1}{\sqrt{n}}.$$

ce qui démontre la relation (9.17).

• Les variables aléatoires $|X_{n+1}|$ et $1/\sqrt{Y_n}$ sont indépendantes et positives ; on alors écrire, dans \mathbb{R}^{+} ,

$$\int_{\Omega} |T_n| dP = \sqrt{n} \left(\int_{\Omega} |X_{n+1}| dP \right) \left(\int_{\Omega} \frac{1}{\sqrt{Y_n}} dP \right).$$

Le premier facteur du membre de droite est fini ; quant au second, par le théorème de transfert, il s'écrit

$$\int_{\Omega} \frac{1}{\sqrt{Y_n}} dP = \int_{\mathbb{R}^{+*}} \frac{1}{\sqrt{y}} f_{Y_n}(y) d\lambda(y) :$$

il en résulte que $\int_{\Omega} |T_n| dP < +\infty$ si et seulement si $n > 1$. Dans ce cas on a, toujours par indépendance,

$$ET_n = \sqrt{n} E(X_{n+1}) E\left(\frac{1}{\sqrt{Y_n}}\right),$$

soit :

$$\boxed{ET_n = 0.}$$

De même, les variables aléatoires X_{n+1}^2 et $\frac{1}{Y_n}$ sont indépendantes et positives; on peut donc écrire dans \mathbb{R}^+ :

$$\int_{\Omega} T_n^2 dP = n \left(\int_{\Omega} X_{n+1}^2 dP \right) \left(\int_{\Omega} \frac{1}{Y_n} dP \right).$$

Le premier facteur du membre de droite est fini; quant au second, par le théorème de transfert, il s'écrit

$$\int_{\Omega} \frac{1}{Y_n} dP = \int_{\mathbb{R}^{+*}} \frac{1}{y} f_{Y_n}(y) d\lambda(y).$$

Il en résulte que $\int_{\Omega} T_n^2 dP < +\infty$ si et seulement si $n > 2$. Si $n > 2$, on a, toujours par indépendance :

$$ET_n^2 = n E(X_{n+1}^2) E\left(\frac{1}{Y_n}\right).$$

Comme on a

$$E\left(\frac{1}{Y_n}\right) = \int_0^{+\infty} \frac{1}{y} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \exp\left(-\frac{y}{2}\right) y^{\frac{n}{2}-1} dy,$$

il vient, après changement de variables,

$$E\left(\frac{1}{Y_n}\right) = \frac{1}{n-2},$$

et donc :

$$\boxed{\sigma_{T_n}^2 = \frac{n}{n-2}.}$$

Exercice 9.5. Loi de la somme de deux variables aléatoires indépendantes, l'une à densité, l'autre discrète. X et Y sont deux variables aléatoires indépendantes de loi respective la loi géométrique $\mathcal{G}_{\mathbb{N}}(p)$, où $0 < p < 1$, et la loi exponentielle $\exp(1)$. Étudier la loi de $Z = X + Y$.

Solution. Pour tout $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R})$, tenant compte successivement de l'indépendance de X et Y, et donc de celle des variables aléatoires $\mathbf{1}_{(X=n)}$ et $f(n+Y)$, puis du théorème de transfert, on a

$$\begin{aligned} Ef(Z) &= \sum_{n \in \mathbb{N}} \int_{(X=n)} f(n+Y) dP \\ &= \sum_{n \in \mathbb{N}} \left[P(X=n) \int_{\Omega} f(n+Y) dP \right] \\ &= \sum_{n \in \mathbb{N}} \left[P(X=n) \int_{\mathbb{R}^+} f(n+y) \mathbf{1}_{\mathbb{P}^+}(y) \exp(-y) d\lambda(y) \right]. \end{aligned}$$

On note $q = 1 - p$; on obtient alors, en faisant un changement de variables dans chaque intégrale.

$$\begin{aligned} E f(Z) &= \sum_{n \in \mathbb{N}} \left[P(X = n) \int_{\mathbb{R}} f(z) \mathbf{1}_{\mathbb{R}^+}(z - n) \exp(-(z - n)) d\lambda(z) \right] \\ &= \int_{\mathbb{R}} f(z) \mathbf{1}_{\mathbb{R}^+}(z) \left[\sum_{n=0}^{\lfloor z \rfloor} p(qe)^n \right] \exp(-z) d\lambda(z). \end{aligned}$$

La variable aléatoire Z a donc une densité f_Z donnée par

$$\forall z \in \mathbb{R} \quad f_Z(z) = p \mathbf{1}_{\mathbb{R}^+}(z) \frac{1 - (qe)^{\lfloor z \rfloor + 1}}{1 - qe} \exp(-z).$$

Exercice 9.6. L'indépendance de variables aléatoires n'est pas toujours intuitive!

Lois gamma et bêta. X et Y sont deux variables aléatoires indépendantes de loi respective $\gamma(a, p)$ et $\gamma(b, p)$ où a, b, p sont des réels strictement positifs. On définit les variables aléatoires U, T, Z par

$$U = X + Y \quad T = \frac{X}{X + Y} \quad Z = \frac{X}{Y}.$$

Étudier la loi de la variable aléatoire (U, T) et en déduire l'indépendance des variables aléatoires U et T . Préciser les lois des variables aléatoires U, T et Z .

Solution. Les variables aléatoires X et Y sont indépendantes et admettent des densités; $X + Y$ admet donc aussi une densité, si bien que $P(X + Y = 0) = 1$ et qu'ainsi, l'application (U, T) est définie presque sûrement. Pour tout $f \in \mathcal{C}_c^+(\mathbb{R}^2)$, le théorème de transfert permet alors d'écrire

$$E[f(U, T)] = \int_{\mathbb{R}^2 \setminus \Delta} f\left(x + y, \frac{x}{x + y}\right) f_X(x) f_Y(y) d\lambda_2(x, y),$$

où Δ est la droite $\{(x, y) \in \mathbb{R}^2 \mid x + y = 0\}$, soit encore

$$E[f(U, T)] = K \int_{(\mathbb{R}^{++})^2} f\left(x + y, \frac{x}{x + y}\right) \exp(-p(x + y)) x^{a-1} y^{b-1} d\lambda_2(x, y),$$

où K est une constante strictement positive qui dépend de a, b, p . Effectuons le changement de variables associé au difféomorphisme de l'ouvert $(\mathbb{R}^{++})^2$ sur $\mathbb{R}^{++} \times]0, 1[$ défini par

$$\begin{cases} u = x + y \\ t = \frac{x}{x + y} \end{cases} \iff \begin{cases} x = ut \\ y = u(1 - t), \end{cases}$$

de jacobien

$$\frac{D(x, y)}{D(u, t)} = \begin{vmatrix} t & u \\ 1 - t & -u \end{vmatrix} = -u.$$

Il vient

$$E f(U, T) = K \int_{\mathbb{R}^{+*} \times]0, 1[} f(u, t) \exp(-u)(ut)^{a-1} [u(1-t)]^{b-1} | - u | d\lambda_2(u, t).$$

soit encore :

$$E f(U, T) = K \int_{\mathbb{R}^2} f(u, t) \mathbf{1}_{\mathbb{R}^{+*}}(u) e^{-u} u^{a+b-1} \mathbf{1}_{]0, 1[}(t) t^{a-1} (1-t)^{b-1} d\lambda_2(u, t).$$

Il en résulte que (U, T) admet une densité $f_{(U, T)}$ définie en tout $(u, t) \in \mathbb{R}^2$ par

$$f_{(U, T)}(u, t) = K (\mathbf{1}_{\mathbb{R}^{+*}}(u) \exp(-u) u^{a+b-1}) (\mathbf{1}_{]0, 1[}(t) t^{a-1} (1-t)^{b-1}).$$

La fonction $f_{(U, T)}$ est produit direct de deux fonctions mesurables positives; les variables aléatoires U et T sont donc indépendantes. De plus **la loi de U est la loi $\gamma(a+b, p)$** et celle de T la **loi bêta $B(a, b)$ de première espèce** (portée par $]0, 1[$). Quant à la loi de Z , il suffit de remarquer que

$$Z = \frac{T}{1-T},$$

et d'appliquer la méthode de calcul de loi maintenant standard. On trouve que Z admet une densité f_Z donnée par

$$\forall z \in \mathbb{R} \quad f_Z(z) = \mathbf{1}_{\mathbb{R}^+}(z) \frac{1}{B(a, b)} \frac{z^{a-1}}{(1+z)^{a+b}}.$$

La loi de Z est la **loi bêta $B(a, b)$ de deuxième espèce** (portée par \mathbb{R}^+). Évidemment, les variables aléatoires U et Z sont encore indépendantes!

Exercice 9.7. Une caractérisation des lois exponentielles. Soient X et Y deux variables aléatoires réelles, *indépendantes* de même loi μ ; on suppose qu'elles admettent une densité f (autrement dit $\mu = f \cdot \lambda$, où λ est la mesure de Lebesgue sur \mathbb{R}) telle que $f(x) > 0$ si et seulement si $x \in \mathbb{R}^+$. On définit les variables aléatoires U et W par

$$U = \min(X, Y) \quad W = |X - Y|.$$

1. Démontrer que la variable aléatoire (U, W) admet une densité que l'on exprimera en fonction de f (pour cela, on pourra calculer, pour toute fonction g définie sur \mathbb{R}^2 , mesurable positive et éventuellement bornée, l'intégrale $\int_{\Omega} g(U, W) dP$). En déduire que les variables aléatoires U et W admettent des densités f_U et f_W .

2. Pour simplifier, on suppose de plus que f est bornée et que sa restriction à \mathbb{R}^+ est continue. Démontrer que U et W admettent des densités f_U et f_W dont les restrictions à \mathbb{R}^+ sont continues. Déduire des résultats précédents que μ est **une loi exponentielle si et seulement si U et W sont indépendantes**.

3. Démontrer qu'en fait, même si on ne suppose plus que f est bornée et que sa restriction à \mathbb{R}^+ est continue, μ est **une loi exponentielle si et seulement si U et W sont indépendantes**.

Solution.

1. Pour tout $g \in \mathcal{C}_c^+(\mathbb{R}^2)$, l'indépendance de X et Y et le théorème de transfert permettent d'écrire⁶ :

$$\begin{aligned} E[g(U, W)] &= \int_{\mathbb{R}^2} g(x \wedge y, |x - y|) f_X(x) f_Y(y) d\lambda_2(x, y) \\ &= \int_{(x \leq y)} g(x, y - x) f_X(x) f_Y(y) d\lambda_2(x, y) \\ &\quad + \int_{(x > y)} g(y, x - y) f_X(x) f_Y(y) d\lambda_2(x, y). \end{aligned}$$

Mais, puisque $f_X = f_Y = f$ et que $\lambda_2(\{x = y\}) = 0$, on a

$$\int_{(x > y)} g(y, x - y) f_X(x) f_Y(y) d\lambda_2(x, y) = \int_{(x \geq y)} g(y, x - y) f_X(y) f_Y(x) d\lambda_2(x, y),$$

et donc :

$$E[g(U, W)] = 2 \int_{(x \leq y)} g(x, y - x) f(x) f(y) d\lambda_2(x, y).$$

Effectuons le changement de variables associé au difféomorphisme de l'ouvert \mathbb{R}^2 sur lui-même défini par

$$\begin{cases} u = x \\ w = y - x \end{cases} \iff \begin{cases} x = u \\ y = u + w. \end{cases}$$

de jacobien 1 ; il vient

$$E[g(U, W)] = 2 \int_{(0 \leq w)} g(u, w) f(u) f(u + w) d\lambda_2(u, w).$$

ce qui démontre que la variable aléatoire (U, W) admet une densité $f_{(U, W)}$ donnée par

$$\forall (u, w) \in \mathbb{R}^2 \quad f_{(U, W)}(u, w) = 2 \mathbf{1}_{\mathbb{R}^+}(w) f(u) f(u + w).$$

Les variables aléatoires U et W admettent alors des densités f_U et f_W données par

$$\forall u \in \mathbb{R} \quad f_U(u) = 2 f(u) \int_{\mathbb{R}^+} f(u + w) d\lambda(w),$$

et

$$\forall w \in \mathbb{R} \quad f_W(w) = 2 \mathbf{1}_{\mathbb{R}^+}(w) \int_{\mathbb{R}} f(u) f(u + w) d\lambda(u).$$

2. Il en résulte que pour que les variables aléatoires U et W soient indépendantes il faut et il suffit que pour λ_2 -presque tout $(u, w) \in \mathbb{R} \times \mathbb{R}^+$, on ait

$$f(u) f(u + w) = 2 \left(f(u) \int_{\mathbb{R}^+} f(u + \alpha) d\lambda(\alpha) \right) \left(\int_{\mathbb{R}} f(\alpha) f(\alpha + w) d\lambda(\alpha) \right). \quad (9.19)$$

6. On note classiquement $x \wedge y = \min(x, y)$ et $x \vee y = \max(x, y)$.

On vérifie facilement qu'il en est ainsi lorsque μ est une loi exponentielle.

Inversement, supposons que la relation (9.19) soit satisfaite pour λ_2 -presque tout $(u, w) \in \mathbb{R} \times \mathbb{R}^+$. Puisque $f(u) > 0$ dès que $u \geq 0$, on a, après changement de variables, pour λ_2 -presque tout $(u, w) \in \mathbb{R}^+ \times \mathbb{R}^+$,

$$f(u+w) = 2 \left(\int_{[u, +\infty[} f(\alpha) d\lambda(\alpha) \right) \left(\int_{\mathbb{R}} f(\alpha) f(\alpha+w) d\lambda(\alpha) \right). \quad (9.20)$$

On suppose de plus que f est bornée et que sa restriction à \mathbb{R}^+ est continue. On va montrer qu'en fait, cette égalité est vraie **pour tout** $(u, w) \in \mathbb{R}^+ \times \mathbb{R}^+$: puisque f est bornée et de restriction à \mathbb{R}^+ continue, l'application $w \mapsto \int_{\mathbb{R}} f(\alpha) f(\alpha+w) d\lambda(\alpha)$ est continue sur \mathbb{R}^+ (d'après le théorème de convergence dominée) ; les applications f et $u \mapsto \int_{[u, +\infty[} f(\alpha) d\lambda(\alpha)$ étant de plus continues sur \mathbb{R}^+ , l'application

$$(u, w) \mapsto f(u+w) - 2 \left(\int_{[u, +\infty[} f(\alpha) d\lambda(\alpha) \right) \left(\int_{\mathbb{R}} f(\alpha) f(\alpha+w) d\lambda(\alpha) \right)$$

est continue sur $\mathbb{R}^+ \times \mathbb{R}^+$. Il en résulte⁷ que l'égalité (9.20) est vraie pour tout $(u, w) \in \mathbb{R}^+ \times \mathbb{R}^+$. En particulier, on peut y prendre $w = 0$, ce qui donne, pour tout $u \in \mathbb{R}^+$,

$$f(u) = 2C \int_{[u, +\infty[} f(\alpha) d\lambda(\alpha), \quad (9.21)$$

où

$$C = \int_{\mathbb{R}} [f(\alpha)]^2 d\lambda(\alpha) > 0$$

(l'égalité (9.21) assure que l'on a $C > 0$, puisque $f(u) > 0$ si $u \geq 0$). Puisque f est continue sur \mathbb{R}^+ , l'application $u \mapsto \int_{[u, +\infty[} f(\alpha) d\lambda(\alpha)$ est dérivable sur \mathbb{R}^+ de dérivée f ; on a alors, d'après l'égalité (9.21),

$$\forall u \in \mathbb{R}^+ \quad f'(u) = -2Cf(u).$$

La solution générale non nulle de cette équation différentielle est donnée, pour tout $u \in \mathbb{R}^+$, par

$$f(u) = p \exp(-2Cu), \quad \text{où } p > 0 :$$

la solution qui fait de f une densité de probabilité est alors donnée par, pour tout $u \in \mathbb{R}^+$,

$$f(u) = 2C \exp(-2Cu) ;$$

c'est-à-dire que μ **est une loi exponentielle**.

Lemme. Une fonction g continue sur $\mathbb{R}^+ \times \mathbb{R}^+$ nulle λ_d -presque partout est nulle partout sur $\mathbb{R}^+ \times \mathbb{R}^+$.

En effet, puisque g est en particulier continue sur $\mathbb{R}^{+*} \times \mathbb{R}^{+*}$, l'ensemble $O = \{t \in \mathbb{R}^{+*} \times \mathbb{R}^{+*} \mid g(t) \neq 0\}$ est un ouvert de $\mathbb{R}^{+*} \times \mathbb{R}^{+*}$. Si O est vide, g est nulle sur $\mathbb{R}^{+*} \times \mathbb{R}^{+*}$; sinon, il existe une boule ouverte (donc de mesure strictement positive) contenue dans O , ce qui contredit le fait que g soit nulle λ_d -presque partout. Ainsi, g est nulle sur $\mathbb{R}^{+*} \times \mathbb{R}^{+*}$. Il résulte alors de la continuité de g sur $\mathbb{R}^+ \times \mathbb{R}^+$ que g est aussi nulle sur les axes.

7. Cf. le lemme ci-dessous.

3. On ne suppose plus que f est bornée et que sa restriction à \mathbb{R}^+ est continue. Si les variables aléatoires U et V sont indépendantes, pour λ_2 -presque tout $(u, w) \in \mathbb{R}^+ \times \mathbb{R}^+$, on a toujours l'égalité (9.20) qui s'écrit encore

$$f(u + w) = G(u) f_W(w). \quad (9.22)$$

où G est la fonction **continue** définie sur \mathbb{R}^+ par, pour tout $u \geq 0$,

$$G(u) = \int_{[u, +\infty[} f(\alpha) d\lambda(\alpha). \quad (9.23)$$

On a alors, pour λ -presque tout $w \in \mathbb{R}^+$,

$$G(w) = \int_{[w, +\infty[} f(\alpha) d\lambda(\alpha) = \int_{\mathbb{R}^+} f(u + w) d\lambda(u) = \int_{\mathbb{R}^+} G(u) f_W(w) d\lambda(u),$$

soit

$$G(w) = m f_W(w), \quad (9.24)$$

où $m = \int_{\mathbb{R}^+} G(u) d\lambda(u)$.

Le support de la densité f_W étant contenu dans \mathbb{R}^+ , il résulte de l'égalité (9.24) que la fonction G est intégrable sur \mathbb{R}^+ , d'intégrale non nulle. Ainsi, on a, d'après l'égalité (9.22), pour λ_2 -presque tout $(u, w) \in \mathbb{R}^+ \times \mathbb{R}^+$,

$$f(u + w) = \frac{G(u)G(w)}{m}. \quad (9.25)$$

Tenait compte de la définition de G , on a alors, pour tout $u \in \mathbb{R}^+$ et **tout** $v \in \mathbb{R}^+$,

$$G(u + v) = \int_{[u+v, +\infty[} f(w) d\lambda(w) = \int_{[v, +\infty[} f(u + w) d\lambda(w),$$

et donc, d'après l'égalité (9.25), pour λ -**presque tout** $u \in \mathbb{R}^+$ et **tout** $v \in \mathbb{R}^+$,

$$G(u + v) = \int_{[v, +\infty[} \frac{G(u)G(w)}{m} d\lambda(w) = \frac{G(u)}{m} \int_{[v, +\infty[} G(w) d\lambda(w).$$

Par continuité de G , on a alors pour **tout** $u \in \mathbb{R}^+$ et **tout** $v \in \mathbb{R}^+$,

$$G(u + v) = \frac{G(u)}{m} \int_{[v, +\infty[} G(w) d\lambda(w).$$

Puisque $G(0) = 1$, on a alors pour tout $v \in \mathbb{R}^+$,

$$G(v) = \frac{1}{m} \int_{[v, +\infty[} G(w) d\lambda(w),$$

ce qui implique que G est dérivable (G est continue) et que

$$G'(v) = -\frac{G(v)}{m}.$$

En tenant compte de ce que $G(0) = 1$, il vient pour tout $v \in \mathbb{R}^+$,

$$G(v) = \exp\left(-\frac{v}{m}\right);$$

d'après l'égalité (9.23), on a alors, pour λ -presque tout $u \in \mathbb{R}^+$,

$$f(u) = -G'(u) = \frac{1}{m} \exp\left(-\frac{u}{m}\right),$$

ce qui démontre que μ est encore la loi exponentielle.

Exercice 9.8. Loi de Dirichlet et statistiques d'ordre. Soit $(X_i)_{1 \leq i \leq n}$ une suite finie de variables aléatoires réelles indépendantes de même loi μ et admettant une densité f continue par morceaux; on note F leur fonction de répartition. On définit l'application r de \mathbb{R}^n dans lui-même par

$$\forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \quad r(x_1, x_2, \dots, x_n) = (x_{(1)}, x_{(2)}, \dots, x_{(n)}),$$

où les $x_{(i)}$ sont les réels x_i rangés par valeur croissante, c'est-à-dire les réels définis par

$$\{x_1, x_2, \dots, x_n\} = \{x_{(1)}, x_{(2)}, \dots, x_{(n)}\} \quad \text{et} \quad x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

On définit la variable aléatoire $(X_{(1)}, X_{(2)}, \dots, X_{(n)}) = r(X_1, X_2, \dots, X_n)$. Si $1 \leq k \leq n$, $X_{(k)}$ est appelée la k^{e} **statistique d'ordre**. Calculer la fonction de répartition de $X_{(k)}$ en fonction de F et f et justifier l'existence d'une densité $f_{X_{(k)}}$ pour la variable aléatoire $X_{(k)}$; l'expliciter. On identifiera la loi de $X_{(k)}$ dans le cas particulier où μ est la loi uniforme sur $[0, t]$, avec $t > 0$.

Déterminer la loi de la variable aléatoire $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ dans le cas particulier où μ est la loi uniforme sur $[0, t]$, avec $t > 0$. Cette loi est appelée loi de **Dirichlet**.

Solution. On remarque d'abord que, puisque la variable aléatoire (X_1, X_2, \dots, X_n) admet une densité,

$$P(X_{(1)} < X_{(2)} < \dots < X_{(n)}) = 1.$$

Alors, pour tout $y \in \mathbb{R}$, on a

$$P(X_{(k)} \leq y) = P\left[\bigcup_{\substack{J \in \mathcal{P}_J \{1, 2, \dots, n\} \\ |J| \geq k}} B_J\right] = \sum_{i=k}^n P\left[\bigoplus_{\substack{I \in \mathcal{P}_I \{1, 2, \dots, n\} \\ |I|=i}} B_I\right],$$

où

$$B_J = \left[\bigcap_{j \in I} (X_j \leq y)\right] \cap \left[\bigcap_{j \in J^c} (X_j > y)\right],$$

en ayant remarqué que si $|J| = |J'|$ on a $B_J \cap B_{J'} = \emptyset$. Mais, pour i fixé, les probabilités $P(B_J)$ sont les mêmes pour toutes les parties J telles que $|J| = i$; il en résulte que :

$$\forall y \in \mathbb{R} \quad F_{X_{(k)}}(y) = \sum_{i=k}^n \binom{n}{i} [F(y)]^i [1 - F(y)]^{n-i}.$$

La fonction F est dérivable de dérivée f ; il en est de même de $F_{X_{(k)}}$. La variable aléatoire $X_{(k)}$ admet donc une densité $f_{X_{(k)}}$, dérivée de $F_{X_{(k)}}$, ce qui donne, pour tout $y \in \mathbb{R}$:

$$f_{X_{(k)}}(y) = \sum_{i=k}^n i \binom{n}{i} f(y) [F(y)]^{i-1} [1 - F(y)]^{n-i} \\ - f(y) \sum_{i=k}^{n-1} (n-i) \binom{n}{i} [F(y)]^i [1 - F(y)]^{n-i-1} ,$$

soit, en tenant compte des relations

$$i \binom{n}{i} = n \binom{n-1}{i-1} \quad \text{et} \quad (n-i) \binom{n}{i} = n \binom{n-1}{i}$$

et en faisant le changement d'indice $j = i - 1$ dans la première somme :

$$f_{X_{(k)}}(y) = n f(y) \sum_{j=k-1}^{n-1} \binom{n-1}{j} [F(y)]^j [1 - F(y)]^{n-1-j} \\ - n f(y) \sum_{i=k}^{n-1} \binom{n-1}{i} [F(y)]^i [1 - F(y)]^{n-i-1} .$$

Après simplification, il vient

$$\forall y \in \mathbb{R} \quad f_{X_{(k)}}(y) = n f(y) \left[\binom{n-1}{k-1} [F(y)]^{k-1} [1 - F(y)]^{n-k} \right] .$$

En particulier, si μ est la loi uniforme sur $[0, t]$, ou a

$$\forall y \in \mathbb{R} \quad f_{X_{(k)}}(y) = \frac{n!}{(k-1)!(n-k)!} \mathbf{1}_{[0,t]}(y) \frac{1}{t} \left[\frac{y}{t} \right]^{k-1} \left[1 - \frac{y}{t} \right]^{n-k} .$$

autrement dit, la variable aléatoire $X_{(k)}$ suit la **loi bêta de première espèce** sur l'intervalle $[0, t]$ de paramètres k et $n - k + 1$.

Déterminons la loi de la variable aléatoire $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ dans le cas particulier où μ est la loi uniforme sur $[0, t]$ où $t > 0$.

Notons \mathcal{S}_n le simplexe de \mathbb{R}^n , $\{x_1 \leq x_2 \leq \dots \leq x_n\}$. Pour toute permutation $\sigma \in \Sigma_n$, définissons l'isométrie \mathcal{P}_σ de \mathbb{R}^n sur lui-même par, pour tout $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$,

$$\mathcal{P}_\sigma(x_1, x_2, \dots, x_n) = (x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) .$$

On a l'égalité

$$\mathbb{R}^n = \bigcup_{\sigma \in \Sigma_n} \mathcal{P}_\sigma^{-1}(\mathcal{S}_n) . \quad (9.26)$$

La variable aléatoire (X_1, X_2, \dots, X_n) admet la densité $f_{\mathcal{C}_n} = \frac{1}{t^n} \mathbf{1}_{[0,t]^n}$. Pour tout $g \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R}^n)$, on a

$$E[g(X_{(1)}, X_{(2)}, \dots, X_{(n)})] = E[g \circ r(X_1, X_2, \dots, X_n)] .$$

soit par le théorème de transfert,

$$\begin{aligned} & \mathbb{E} [g(X_{(1)}, X_{(2)}, \dots, X_{(n)})] \\ &= \int_{\mathbb{R}^n} (g \circ r)(x_1, x_2, \dots, x_n) f_{C_n}(x_1, x_2, \dots, x_n) d\lambda_n(x_1, x_2, \dots, x_n). \end{aligned}$$

En notant $\underline{x}_n = (x_1, x_2, \dots, x_n)$, et en tenant compte de l'égalité (9.26), on a alors

$$\begin{aligned} \mathbb{E} [g(X_{(1)}, X_{(2)}, \dots, X_{(n)})] &= \sum_{\sigma \in \Sigma_n} \int_{\mathcal{P}_\sigma^{-1}(\mathcal{S}_n)} (g \circ r)(\underline{x}_n) f_{C_n}(\underline{x}_n) d\lambda_n(\underline{x}_n) \\ &= \sum_{\sigma \in \Sigma_n} \int_{\mathbb{R}^n} \mathbf{1}_{\mathcal{S}_n}[\mathcal{P}_\sigma(\underline{x}_n)] (g \circ r)(\underline{x}_n) f_{C_n}(\underline{x}_n) d\lambda_n(\underline{x}_n). \end{aligned}$$

Le changement de variables $\underline{y}_n = \mathcal{P}_\sigma(\underline{x}_n)$ défini par l'isométrie \mathcal{P}_σ (de jacobien ± 1) donne alors :

$$\begin{aligned} & \mathbb{E} [g(X_{(1)}, X_{(2)}, \dots, X_{(n)})] = \\ & \sum_{\sigma \in \Sigma_n} \int_{\mathbb{R}^n} \mathbf{1}_{\mathcal{S}_n}(\underline{y}_n) (g \circ r) [\mathcal{P}_\sigma^{-1}(\underline{y}_n)] f_{C_n} [\mathcal{P}_\sigma^{-1}(\underline{y}_n)] d\lambda_n(\underline{y}_n). \end{aligned}$$

La fonction $(g \circ r) \cdot f_{C_n}$ étant invariante par \mathcal{P}_σ^{-1} , les intégrales figurant dans le deuxième membre de cette égalité sont indépendantes de σ ; on a donc :

$$\mathbb{E} [g(X_{(1)}, X_{(2)}, \dots, X_{(n)})] = n! \int_{\mathcal{S}_n} (g \circ r)(\underline{y}_n) f_{C_n}(\underline{y}_n) d\lambda_n(\underline{y}_n).$$

Tenant compte de la valeur de la densité f_{C_n} , on en déduit que la variable aléatoire $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ admet une densité $f_{(X_{(1)}, X_{(2)}, \dots, X_{(n)})}$ donnée par

$$f_{(X_{(1)}, X_{(2)}, \dots, X_{(n)})} = \frac{n!}{t^n} \mathbf{1}_{(0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq t)}.$$

Exercice 9.9. Loi multinomiale. Soit $k \in \mathbb{N}^*$ fixé. Pour tout $n \in \mathbb{N}^*$ on considère une partition $(A_i^n)_{1 \leq i \leq k}$ de Ω , où $A_i^n \in \mathcal{A}$. On suppose que les familles, indexées sur n , constituées par les éléments de ces partitions sont *indépendantes*. On suppose de plus que :

$$\forall n \in \mathbb{N}^*, \forall i = 1, 2, \dots, k \quad \mathbb{P}(A_i^n) = p_i,$$

où $p_i > 0$ et $\sum_{i=1}^k p_i = 1$. On définit les variables aléatoires X^n et Y^n à valeurs dans \mathbb{R}^k par

$$X^n = \begin{pmatrix} \mathbf{1}_{A_1^n} \\ \vdots \\ \mathbf{1}_{A_k^n} \end{pmatrix} \quad \text{et} \quad Y^n = \sum_{j=1}^n X^j.$$

On note $S = \{e_1, e_2, \dots, e_k\}$ la base canonique de \mathbb{R}^k et

$$D_n = \left\{ y \in \{0, 1, 2, \dots, n\}^k \mid \sum_{j=1}^k y_j = n \right\}.$$

Déterminer $P(X^n = e_j)$ pour tout j tel que $1 \leq j \leq k$, et en déduire, pour tout $y \in D_n$, la probabilité $P(Y^n = y)$. La loi de Y^n est appelée loi multinomiale de paramètres $n, p_1, p_2, \dots, p_{k-1}$ et notée $M(n; p_1, p_2, \dots, p_{k-1})$; elle généralise la loi binomiale. Déterminer les moyennes et matrice de covariance de Y^n .

Solution. On a

$$P(X^n = e_j) = P(A_j^n) = p_j.$$

Puisque

$$\sum_{i=1}^k Y_i^n = \sum_{i=1}^k \sum_{j=1}^n X_i^j = \sum_{j=1}^n \left(\sum_{i=1}^k \mathbf{1}_{A_i^j} \right) = \sum_{j=1}^n \mathbf{1}_\Omega = n,$$

on a $Y^n(\Omega) \subset D_n$. Soit $y \in D_n$; notant $\underline{J}^{y_1, y_2, \dots, y_k}$ l'ensemble des partitions de $\{0, 1, 2, \dots, n\}$ à k éléments J_1, J_2, \dots, J_k tels que $|J_1| = y_1, |J_2| = y_2, \dots, |J_k| = y_k$, on a

$$P(Y^n = y) = \bigcup_{\{J_1, J_2, \dots, J_k\} \in \underline{J}^{y_1, y_2, \dots, y_k}} \left[\bigcap_{J_1 \in J_1} (X^{j_1} = e_1) \cap \bigcap_{J_2 \in J_2} (X^{j_2} = e_2) \dots \cap \bigcap_{J_k \in J_k} (X^{j_k} = e_k) \right],$$

ce qui donne, en tenant compte de l'indépendance des variables aléatoires X^j :

$$P(Y^n = y) = \sum_{\{J_1, J_2, \dots, J_k\} \in \underline{J}^{y_1, y_2, \dots, y_k}} p_1^{y_1} p_2^{y_2} \dots p_k^{y_k}.$$

Mais on a

$$\text{card}(\underline{J}^{y_1, y_2, \dots, y_k}) = \binom{n}{y_1} \binom{n-y_1}{y_2} \binom{n-(y_1+y_2)}{y_3} \dots \binom{n-(y_1+y_2+\dots+y_{k-2})}{y_{k-1}}$$

c'est-à-dire, en tenant compte l'égalité $y_1 + y_2 + \dots + y_{k-1} = n - y_k$,

$$\text{card}(\underline{J}^{y_1, y_2, \dots, y_k}) = \frac{n!}{y_1! y_2! \dots y_k!}$$

(on reconnaît le coefficient multinomial). On a donc

$$\forall y \in D_n \quad P(Y^n = y) = \frac{n!}{y_1! y_2! \dots y_k!} p_1^{y_1} p_2^{y_2} \dots p_k^{y_k}.$$

La loi de Y^n s'écrit :

$$P_{Y^n} = \sum_{y \in D_n} P(Y^n = y) \delta_y.$$

On a

$$EX^n = \begin{pmatrix} E\mathbf{1}_{A_1^n} \\ \vdots \\ E\mathbf{1}_{A_k^n} \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix}.$$

et donc :

$$EY^n = \begin{pmatrix} np_1 \\ \vdots \\ np_k \end{pmatrix}.$$

La matrice de covariance C_{X^n} a pour termes

$$(C_{X^n})_{i,i} = \sigma_{X_i^n}^2 = E(\mathbf{1}_{A_i^n}^2) - (E\mathbf{1}_{A_i^n})^2 = p_i(1 - p_i),$$

et, si $i \neq j$:

$$(C_{X^n})_{i,j} = \text{cov}(X_i^n, X_j^n) = E(\mathbf{1}_{A_i^n} \mathbf{1}_{A_j^n}) - [E(\mathbf{1}_{A_i^n})][E(\mathbf{1}_{A_j^n})] = -p_i p_j.$$

Y^n étant somme des variables aléatoires *indépendantes* X^j , sa matrice de covariance est la somme des matrices de covariance des X^j soit :

$$C_{Y^n} = n C_{X^n}.$$

Exercice 9.10. Un résultat lié au théorème limite central. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles *indépendantes* de même loi, admettant un moment d'ordre 2, *centrées*. Pour $n \in \mathbb{N}^*$, on définit la variable aléatoire Z_n par

$$Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j.$$

On admettra que si la suite $(Z_n)_{n \in \mathbb{N}^*}$ converge presque sûrement vers une variable aléatoire Z , il résulte du théorème limite central (voir le chapitre sur les convergences en loi) que la loi de Z est la loi $\mathcal{N}_{\mathbb{R}}(0, 1)$. En déduire que la suite $(Z_n)_{n \in \mathbb{N}^*}$ ne peut pas converger presque sûrement.

Solution. Définissons, pour tout $n \in \mathbb{N}^*$, la tribu $\mathcal{A}_n = \sigma(X_p \mid p \geq n)$ et la tribu asymptotique $\mathcal{A}_\infty = \bigcap_{n \in \mathbb{N}^*} \mathcal{A}_n$; pour tout n_0 fixé, et tout $p \geq n_0$, notons $Y_{n_0,p} = \frac{1}{\sqrt{p}} \sum_{j=n_0}^p X_j$; en notant $(U_n \rightarrow)$ l'ensemble où la suite de variables aléatoires $(U_n)_{n \in \mathbb{N}^*}$ converge, on a $(Y_{n_0,p} \rightarrow) \in \mathcal{A}_{n_0}$; mais on a $(Z_p \rightarrow) = (Y_{n_0,p} \rightarrow)$ et donc, pour tout n_0 fixé, $(Z_p \rightarrow) \in \mathcal{A}_{n_0}$. Il en résulte que $(Z_p \rightarrow) \in \mathcal{A}_\infty$. Mais alors, si la suite $(Z_n)_{n \in \mathbb{N}^*}$ converge presque sûrement vers une variable aléatoire Z , Z est presque sûrement égale à une variable aléatoire \mathcal{A}_∞ -mesurable et donc, par la loi du tout ou rien, égale presque sûrement à une constante, ce qui est en contradiction avec le fait que la loi de Z est la loi $\mathcal{N}_{\mathbb{R}}(0, 1)$.

Exercice 9.11. Une application du lemme de Borel-Cantelli. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles *indépendantes*, démontrer l'équivalence :

$$P(\sup X_n < +\infty) = 1 \iff \exists A > 0 \text{ tel que } \sum_{j=1}^{+\infty} P(X_n > A) < +\infty.$$

Solution. S'il existe $A > 0$ tel que $\sum_{j=1}^{+\infty} P(X_n > A) < +\infty$, il résulte du lemme de Borel-Cantelli que $P(\limsup_n (X_n > A)) = 0$, soit encore $P(\liminf_n (X_n \leq A)) = 1$.

Il résulte alors de l'inclusion

$$\liminf_n (X_n \leq A) \subset (\sup X_n < +\infty)$$

que

$$P(\sup X_n < +\infty) = 1.$$

Supposons maintenant que, pour tout $A > 0$, on ait $\sum_{j=1}^{+\infty} P(X_n > A) = +\infty$. Les variables aléatoires X_n étant indépendantes, il résulte de la « réciproque » du lemme de Borel-Cantelli que

$$\forall A > 0 \quad P(\limsup_n (X_n > A)) = 1,$$

et donc, puisque \mathbb{N}^* est dénombrable, que

$$P\left[\bigcap_{A \in \mathbb{N}^*} (\limsup_n (X_n > A))\right] = 1.$$

On a la suite d'implications :

$$\begin{aligned} \omega \in \bigcap_{A \in \mathbb{N}^*} (\limsup_n (X_n > A)) \\ \iff \forall A \in \mathbb{N}^*, \forall n \in \mathbb{N}^*, \exists p \geq n \text{ tel que } X_p(\omega) > A \\ \implies \forall A \in \mathbb{N}^* \quad \sup X_n(\omega) > A \implies \sup X_n(\omega) = +\infty. \end{aligned}$$

Il en résulte que

$$P(\sup X_n = +\infty) = 1.$$

on encore

$$P(\sup X_n < +\infty) = 0.$$

Si donc $P(\sup X_n < +\infty) = 1$, il existe $A > 0$ tel que $\sum_{j=1}^{+\infty} P(X_n > A) < +\infty$.

Exercice 9.12. Lemme de Borel-Cantelli et convergence presque sûre. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes, de même loi. Démontrer l'équivalence :

$$P\left(\frac{X_n}{n} \rightarrow 0\right) = 1 \iff \int_{\Omega} |X_1| dP < +\infty.$$

Remarque. Cette propriété est utilisée lors de l'étude de lois fortes des grands nombres.

Solution. Rappelons d'abord (voir exercice du chapitre 8 tome 2) qu'à l'aide du théorème de Fubini, on démontre l'égalité, dans \mathbb{R}^+ ,

$$\int_{\Omega} |X_1| dP = \int_{\mathbb{R}^+} P(|X_1| > x) d\lambda(x).$$

On a donc, pour tout $\varepsilon > 0$,

$$\int_{\Omega} |X_1| dP = \sum_{n=0}^{+\infty} \int_{[n\varepsilon, (n+1)\varepsilon[} P(|X_1| > x) d\lambda(x).$$

Puisque l'application $x \rightarrow P(|X_1| > x)$ est *décroissante*, il en résulte la double inégalité :

$$\varepsilon \sum_{n=0}^{+\infty} P(|X_1| > (n+1)\varepsilon) \leq \int_{\Omega} |X_1| dP \leq \varepsilon \sum_{n=0}^{+\infty} P(|X_1| > n\varepsilon). \quad (9.27)$$

En particulier, en prenant $\varepsilon = 1$, il vient

$$\int_{\Omega} |X_1| dP < +\infty \iff \sum_{n=1}^{+\infty} P(|X_1| > n) < +\infty,$$

et, puisque les variables aléatoires X_n ont même loi,

$$\int_{\Omega} |X_1| dP < +\infty \iff \sum_{n=1}^{+\infty} P(|X_n| > n) < +\infty.$$

Mais, les variables aléatoires X_n étant indépendantes, il résulte du lemme de Borel-Cantelli que l'on a les équivalences

$$\int_{\Omega} |X_1| dP < +\infty \iff \sum_{n=1}^{+\infty} P(|X_n| > n) < +\infty \iff P(\limsup_n (|X_n| > n)) = 0,$$

et

$$\int_{\Omega} |X_1| dP = +\infty \iff \sum_{n=1}^{+\infty} P(|X_n| > n) = +\infty \iff P(\limsup_n (|X_n| > n)) = 1.$$

Ainsi, on a les implications :

$$P\left(\frac{X_n}{n} \rightarrow 0\right) = 1 \implies P(\limsup_n (|X_n| > n)) = 0 \iff \int_{\Omega} |X_1| dP < +\infty.$$

Inversement, si $\int_{\Omega} |X_1| dP < +\infty$, il résulte de (9.27) que, *pour tout* $\varepsilon > 0$, on a

$$\sum_{n=0}^{+\infty} P[|X_1| > (n+1)\varepsilon] < +\infty,$$

et donc aussi, puisque les variables aléatoires X_n ont même loi :

$$\sum_{n=0}^{+\infty} P[|X_n| > (n+1)\varepsilon] < +\infty.$$

Il résulte alors du lemme de Borel-Cantelli que

$$\forall \varepsilon > 0 \quad \mathbb{P}\left[\limsup_n (|X_n| > (n+1)\varepsilon)\right] = 0;$$

on a alors, \mathbb{Q}^{+*} étant dénombrable,

$$\mathbb{P}\left[\bigcup_{\varepsilon \in \mathbb{Q}^{+*}} \limsup_n (|X_n| > (n+1)\varepsilon)\right] = 0.$$

Il s'en suit que

$$\mathbb{P}\left(\frac{X_n}{n} \rightarrow 0\right) \geq \mathbb{P}\left[\bigcap_{\varepsilon \in \mathbb{Q}^{+*}} \liminf_n \left(\frac{|X_n|}{n+1} \leq \varepsilon\right)\right] = 1,$$

et par conséquent

$$\boxed{\mathbb{P}\left(\frac{X_n}{n} \rightarrow 0\right) = 1.}$$

L'équivalence est démontrée.

Exercice 9.13. Duplication et symétrisation. À une variable aléatoire réelle X définie sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, on associe les applications \widehat{X} et X^s , définies sur le produit cartésien $\Omega \times \Omega$ et à valeurs respectivement dans \mathbb{R}^2 et \mathbb{R} , par : pour tout $(\omega, \omega') \in \Omega \times \Omega$,

$$\widehat{X}(\omega, \omega') = (X(\omega), X(\omega')) \quad \text{et} \quad X^s(\omega, \omega') = X(\omega) - X(\omega').$$

1. Vérifier que \widehat{X} et X^s sont des variables aléatoires définies sur l'espace probabilisé produit $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A}, \mathbb{P} \otimes \mathbb{P})$ et que les marginales \widehat{X}_1 et \widehat{X}_2 de \widehat{X} sont indépendantes de même loi que X .

2. Soit $p \geq 1$. Démontrer que si $X \in \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, la **symétrisée** X^s appartient à $\mathcal{L}^p(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A}, \mathbb{P} \otimes \mathbb{P})$. En particulier, si $X \in \mathcal{L}^2$, calculer les moyenne et variance de X^s .

3. Soit $(X_i)_{i \in I}$ une famille de variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, **indépendantes**; démontrer que les variables aléatoires \widehat{X}_i , (resp. les symétrisées X_i^s) associées sont $\mathbb{P} \otimes \mathbb{P}$ -indépendantes.

Remarque. Cette propriété est utilisée en particulier lors de l'étude de **convergence P-p.s. de séries de variables aléatoires indépendantes**.

Solution.

1. Pour tous boréliens A et B de \mathbb{R} , on a

$$\widehat{X}^{-1}(A \times B) = X^{-1}(A) \times X^{-1}(B) \in \mathcal{A} \otimes \mathcal{A},$$

ce qui démontre que \widehat{X} est une variable aléatoire; il en est alors de même de X^s , fonction mesurable de \widehat{X} . Par ailleurs, pour tous boréliens A et B de \mathbb{R} , on a

$$\widehat{X}_1^{-1}(A) \cap \widehat{X}_2^{-1}(B) = X^{-1}(A) \times X^{-1}(B),$$

ce qui donne, en jouant successivement sur la définition de la mesure produit, et sur l'égalité $P(\Omega) = 1$,

$$\begin{aligned} P \otimes P \left[\widehat{X}_1^{-1}(A) \cap \widehat{X}_2^{-1}(B) \right] &= P \left[X^{-1}(A) \right] \cdot P \left[X^{-1}(B) \right] \\ &= P \otimes P \left[X^{-1}(A) \times \Omega \right] \cdot P \otimes P \left[\Omega \times X^{-1}(B) \right] \\ &= P \otimes P \left[\widehat{X}_1^{-1}(A) \right] \cdot P \otimes P \left[\widehat{X}_2^{-1}(B) \right]. \end{aligned}$$

ce qui démontre la $P \otimes P$ -indépendance de \widehat{X}_1 et \widehat{X}_2 . De plus, en prenant pour B l'ensemble \mathbb{R} , on a

$$P \otimes P \left[\widehat{X}_1^{-1}(A) \right] = P \otimes P \left[\widehat{X}_1^{-1}(A) \cap \widehat{X}_2^{-1}(\mathbb{R}) \right] = P \left[X^{-1}(A) \right] \cdot P \left[X^{-1}(\mathbb{R}) \right],$$

soit

$$P \otimes P \left[\widehat{X}_1^{-1}(A) \right] = P \left[X^{-1}(A) \right].$$

ce qui démontre que **les variables aléatoires \widehat{X}_1 et X ont même loi (et donc aussi \widehat{X}_2).**

2. Il résulte du théorème de Fubini que

$$\begin{aligned} \int_{\Omega \times \Omega} |\widehat{X}_1|^p dP \otimes P &= \int_{\Omega} \left[\int_{\Omega} |\widehat{X}_1(\omega, \omega')|^p dP(\omega) \right] dP(\omega') \\ &= \int_{\Omega} \left[\int_{\Omega} |X(\omega)|^p dP(\omega) \right] dP(\omega') = \int_{\Omega} |X(\omega)|^p dP(\omega). \end{aligned}$$

ce qui démontre que si $X \in \mathcal{L}^p$, les variables aléatoires \widehat{X}_1 et \widehat{X}_2 sont dans \mathcal{L}^p , et donc aussi X^s .

3. Si $X \in \mathcal{L}^2$, puisque $X^s = \widehat{X}_1 - \widehat{X}_2$, et que \widehat{X}_1 et \widehat{X}_2 sont de même loi, donc de mêmes moments, que X , on a

$$\boxed{EX^s = E\widehat{X}_1 - E\widehat{X}_2 = 0.}$$

Puisque de plus \widehat{X}_1 et \widehat{X}_2 sont indépendantes, on a

$$\boxed{\sigma_{X^s}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 = 2\sigma_X^2.}$$

4. Il suffit de le faire pour I fini. Pour tous boréliens A_i et B_i , $i \in I$, de \mathbb{R} , on a

$$\begin{aligned} P \otimes P \left[\bigcap_{i \in I} \widehat{X}_i^{-1}(A_i \times B_i) \right] &= P \otimes P \left[\left(\bigcap_{i \in I} X_i^{-1}(A_i) \right) \times \left(\bigcap_{i \in I} X_i^{-1}(B_i) \right) \right] \\ &= P \left(\bigcap_{i \in I} X_i^{-1}(A_i) \right) \cdot P \left(\bigcap_{i \in I} X_i^{-1}(B_i) \right), \end{aligned}$$

soit, par indépendance des X_j ,

$$\begin{aligned} P \otimes P \left[\bigcap_{i \in I} \widehat{X}_i^{-1}(A_i \times B_i) \right] &= \prod_{i \in I} P(X_i^{-1}(A_i)) \cdot \prod_{j \in I} P(X_j^{-1}(B_j)) \\ &= \prod_{i \in I} P \otimes P(X_i^{-1}(A_i) \times X_i^{-1}(B_i)) \\ &= \prod_{i \in I} P \otimes P(\widehat{X}_i^{-1}(A_i \times B_i)). \end{aligned}$$

ce qui démontre l'indépendance des \widehat{X}_j ; les symétrisées X_j^s étant fonction mesurables des \widehat{X}_j sont aussi $\mathbf{P} \otimes \mathbf{P}$ -indépendantes.

Chapitre 10

Convergences et lois des grands nombres

Dans la première partie de ce chapitre, on étudie les notions de convergence en probabilité, presque sûre et L^p ainsi que les relations entre ces divers modes de convergence; la notion d'équi-intégrabilité est introduite à cette fin. La deuxième partie traite des lois faible et forte des grands nombres.

10.1. Convergence en probabilité et presque sûre

Dans ce paragraphe, toutes les variables aléatoires sont définies sur le même espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{R}^d , $d \geq 1$, ou dans $\overline{\mathbb{R}}$.

On note indifféremment $|\cdot|$ la valeur absolue dans \mathbb{R} , (éventuellement prolongée à $\overline{\mathbb{R}}$; cf. tome 1, p. 33) ou une norme sur \mathbb{R}^d .

Définition 10.1. (a) Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires converge presque sûrement vers une variable aléatoire X s'il existe un ensemble $C \in \mathcal{A}$ de probabilité 1 sur lequel la suite converge ponctuellement (ou simplement) vers X . On note :

$$X_n \xrightarrow{p.p.s.} X.$$

(b) Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires converge en probabilité vers une variable aléatoire X si, pour tout $\varepsilon > 0$, la suite de terme général $P(|X_n - X| > \varepsilon)$ converge vers 0. On note :

$$X_n \xrightarrow{p} X.$$

(c) Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires converge presque sûrement (resp. en probabilité) s'il existe une variable aléatoire X telle que cette suite converge presque sûrement (resp. en probabilité) vers X .

Notation. On note $(X_n \rightarrow)$ (resp. $(X_n \rightarrow X)$) l'ensemble des ω pour lequel la suite $(X_n(\omega))_{n \in \mathbb{N}}$ converge (resp. converge vers $X(\omega)$).

Remarque. 1. Si une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires converge presque sûrement (resp. en probabilité), la limite X est P-p.s. unique. C'est clair pour

la première convergence. Pour la seconde, soient X et X' deux limites en probabilité; pour tout $\varepsilon > 0$ et tout $n \in \mathbb{N}$, on a, par l'inégalité triangulaire,

$$(|X - X'| > \varepsilon) \subset (|X - X_n| > \frac{\varepsilon}{2}) \cup (|X_n - X'| > \frac{\varepsilon}{2}),$$

et par conséquent :

$$P(|X - X'| > \varepsilon) \leq P(|X - X_n| > \frac{\varepsilon}{2}) + P(|X_n - X'| > \frac{\varepsilon}{2}).$$

En passant à la limite, il vient

$$\forall \varepsilon > 0 \quad P(|X - X'| > \varepsilon) = 0.$$

On a le résultat en remarquant que

$$(X \neq X') = \bigcup_{n \in \mathbb{N}^*} \left(|X - X'| > \frac{1}{n} \right).$$

2. La convergence en probabilité $X_n \xrightarrow{P} X$ s'écrit de manière quantifiée :

$$\forall \varepsilon > 0, \forall \delta > 0 \exists N(\varepsilon, \delta) \in \mathbb{N} \text{ tel que } n \geq N(\varepsilon, \delta) \Rightarrow P(|X_n - X| > \varepsilon) \leq \delta.$$

Ceci est équivalent à l'assertion :

$$\forall \varepsilon > 0 \quad \exists N(\varepsilon) \in \mathbb{N} \quad \text{tel que } n \geq N(\varepsilon) \Rightarrow P(|X_n - X| > \varepsilon) \leq \varepsilon.$$

Il est trivial que la première assertion implique la seconde. Réciproquement, supposons vraie la seconde assertion et soit $\varepsilon > 0$ et $\delta > 0$. Si $\delta \geq \varepsilon$, il suffit de prendre $N(\varepsilon, \delta) = N(\varepsilon)$; si $\delta < \varepsilon$, on prend $N(\varepsilon, \delta) = N(\delta)$. On a alors, pour tout $n \geq N(\delta)$,

$$P(|X_n - X| > \delta) \leq \delta,$$

et le résultat vient de l'inégalité

$$(|X_n - X| > \varepsilon) \subset (|X_n - X| > \delta).$$

3. Si l'espace des valeurs prises par les variables aléatoires est \mathbb{R}^d avec $d \geq 2$, **le choix de la norme est indifférent** (on le voit facilement en exprimant que toutes les normes sont équivalentes). De plus, pour que $X_n \xrightarrow{P} X$ il faut et il suffit que, pour tout $j = 1, 2, \dots, d$, on ait $X_n^j \xrightarrow{P} X^j$, où X_n^j désigne la $j^{\text{ème}}$ composante de X_n . La condition nécessaire est triviale : la condition suffisante résulte des inégalités suivantes (on choisit la norme max) :

$$P\left(\max_{1 \leq j \leq d} |X_n^j - X^j| > \varepsilon\right) \leq P\left(\bigcup_{j=1}^d (|X_n^j - X^j| > \varepsilon)\right) \leq \sum_{j=1}^d P(|X_n^j - X^j| > \varepsilon).$$

Les conditions suffisantes de convergence P-p.s. données ci-dessous sont d'un usage fréquent.

Théorème 10.2. *S'il existe une série à termes positifs de terme général ε_n convergente et telle que*

$$\sum_{n=0}^{+\infty} P(|X_{n+1} - X_n| > \varepsilon_n) < +\infty,$$

la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement.

Démonstration. Le lemme de Borel-Cantelli assure que

$$P\left[\limsup_n (|X_{n+1} - X_n| > \varepsilon_n)\right] = 0.$$

L'ensemble $C = \liminf_n (|X_{n+1} - X_n| \leq \varepsilon_n)$ est alors de probabilité 1 et, pour tout $\omega \in C$ la série de terme général $|X_{n+1}(\omega) - X_n(\omega)|$ est convergente et donc aussi la suite de terme général $X_n(\omega)$. \square

Remarque. Ce théorème sera en particulier utilisé dans la comparaison de la convergence en probabilité et de la convergence presque sûre (th. 10.4).

Théorème 10.3. *Si X est une variable aléatoire telle que, pour tout $\varepsilon > 0$,*

$$\sum_{n=0}^{+\infty} P(|X_n - X| > \varepsilon) < +\infty,$$

la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers X .

Démonstration. Le lemme de Borel-Cantelli assure que pour tout $\varepsilon \in \mathbb{Q}^{*+}$

$$P\left[\limsup_n (|X_n - X| > \varepsilon)\right] = 0,$$

et donc aussi, puisque \mathbb{Q}^{*+} est dénombrable,

$$P\left[\bigcup_{\varepsilon \in \mathbb{Q}^{*+}} \limsup_n (|X_n - X| > \varepsilon)\right] = 0.$$

L'ensemble $C = \bigcap_{\varepsilon \in \mathbb{Q}^{*+}} \liminf_n (|X_n - X| \leq \varepsilon)$ est alors de probabilité 1 ; or cet ensemble n'est autre que l'ensemble des ω pour lesquels la suite de terme général $X_n(\omega)$ converge vers $X(\omega)$. \square

Remarque. Le théorème 10.3 fournit une condition suffisante mais non nécessaire de convergence P-p.s. En effet, considérons sur l'espace probabilisé $([0, 1], \mathcal{B}_{[0,1]}, P)$, où P est la mesure de Lebesgue sur $[0, 1]$, les variables aléatoires $X_n = \mathbf{1}_{[0, 1/n]}$; pour tout $\varepsilon > 0$, on a $P(|X_n| > \varepsilon) = 1/n$, et donc

$\sum_{j=1}^{+\infty} \mathbb{P}(|X_n| > \varepsilon) = +\infty$, alors que la suite de terme général X_n converge P-p.s. vers 0. Une réciproque partielle est étudiée dans l'exercice 2.

On compare maintenant les différents modes de convergence.

Théorème 10.4. (a) Si la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement, elle converge en probabilité et les limites sont P-p.s. égales.

(b) Si la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X , il existe une sous-suite $(X_{n_j})_{j \in \mathbb{N}}$ qui converge presque sûrement vers X .

(c) Pour que la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X il faut et il suffit qu'elle soit **de Cauchy pour la convergence en probabilité**, c'est à dire que, pour tout $\varepsilon > 0$, la suite double de terme général $\mathbb{P}(|X_n - X_m| > \varepsilon)$ converge vers 0.

Démonstration. (a) En effet, soit X une limite presque sûre de la suite $(X_n)_{n \in \mathbb{N}}$; pour tout $\varepsilon > 0$, on a

$$(X_n \rightarrow X) \subset \liminf_n (|X_n - X| \leq \varepsilon),$$

soit, en prenant les probabilités des complémentaires.

$$0 \leq \limsup_n \mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}[\limsup_n (|X_n - X| > \varepsilon)] \\ \leq \mathbb{P}[(X_n \rightarrow X)^c] = 0.$$

Il en résulte que $X_n \xrightarrow{P} X$.

(b) Si la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X , elle est de Cauchy pour la convergence en probabilité; cela résulte de ce que pour tout $\varepsilon > 0$ et tous $n, m \in \mathbb{N}$, on a

$$(|X_n - X_m| > \varepsilon) \subset (|X - X_n| > \frac{\varepsilon}{2}) \cup (|X_m - X| > \frac{\varepsilon}{2}).$$

On a d'ailleurs ainsi montré la condition nécessaire de la troisième assertion. Montrons maintenant que si $(X_n)_{n \in \mathbb{N}}$ est de Cauchy pour la convergence en probabilité, il existe une sous-suite $(X_{n_j})_{j \in \mathbb{N}}$ qui converge presque sûrement. Pour cela, on construit la suite d'entiers n_j en posant $n_0 = 1$ et, pour tout $j \in \mathbb{N}^*$,

$$n_j = \inf \left[n > n_{j-1} \mid \forall p, q \geq n \quad \mathbb{P}(|X_p - X_q| > \frac{1}{2^j}) < \frac{1}{3^j} \right].$$

Puisque la suite $(X_n)_{n \in \mathbb{N}}$ est de Cauchy pour la convergence en probabilité, la suite ainsi construite tend en croissant vers $+\infty$: on a de plus

$$\sum_{n=0}^{+\infty} \mathbb{P}(|X_{n_{j+1}} - X_{n_j}| > \frac{1}{2^j}) < +\infty.$$

Il résulte du théorème 10.2 que la suite $(X_{n_j})_{j \in \mathbb{N}}$ converge presque sûrement.

En résumé, si la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité, elle est de Cauchy et on en extrait une sous-suite qui converge P-p.s.

(c) Reste à démontrer que **si $(X_n)_{n \in \mathbb{N}}$ est de Cauchy pour la convergence en probabilité, elle converge en probabilité.** Soit X la limite presque sûre de la suite extraite $(X_{n_j})_{j \in \mathbb{N}}$; d'après l'assertion (a) démontrée ci-dessus, cette sous-suite converge en probabilité vers X ; de plus pour tous entiers n et j ,

on a

$$P(|X_n - X| > \varepsilon) \leq P\left(|X_n - X_{n_j}| > \frac{\varepsilon}{2}\right) + P\left(|X_{n_j} - X| > \frac{\varepsilon}{2}\right).$$

La convergence en probabilité de la suite $(X_n)_{n \in \mathbb{N}}$ vers X résulte alors de ce qu'elle est de Cauchy et de la convergence en probabilité vers X de la suite $(X_{n_j})_{j \in \mathbb{N}}$. \square

Voici un exemple de suite convergeant en probabilité mais pas presque sûrement : $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires *indépendantes* à valeurs 0 ou 1 telle que

$$\forall n \in \mathbb{N}^* \quad P(X_n = 1) = \frac{1}{n} \quad \text{et} \quad P(X_n = 0) = 1 - \frac{1}{n}.$$

La suite $(X_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers 0 puisque, pour tout $\varepsilon > 0$, on a

$$P(|X_n| > \varepsilon) = P(X_n = 1) = \frac{1}{n}.$$

Elle ne converge pas P-p.s. En effet, si elle converge P-p.s., c'est vers 0, ce qui n'est pas, comme le montre l'argument suivant : les événements $(X_n = 1)$ sont indépendants et satisfont à l'égalité

$$\sum_{n=1}^{+\infty} P(X_n = 1) = +\infty.$$

Le lemme de Borel-Cantelli assure alors que

$$P\left(\limsup_n (X_n = 1)\right) = 1.$$

ce qui signifie que la suite prend P-p.s. une infinité de fois la valeur 1 et ne peut converger P-p.s. vers 0.

Remarque. Il est évident que si f est une fonction continue de \mathbb{R}^d dans \mathbb{R}^k et si la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers X , la suite $(f(X_n))_{n \in \mathbb{N}}$ converge presque sûrement vers $f(X)$. Nous allons établir un résultat analogue pour la convergence en probabilité.

Proposition 10.5. *Soit f est une fonction continue de \mathbb{R}^d dans \mathbb{R}^k ; si la suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires à valeurs dans \mathbb{R}^d converge en probabilité vers X , la suite $(f(X_n))_{n \in \mathbb{N}}$ converge en probabilité vers $f(X)$.*

Démonstration. Soit $\delta > 0$ quelconque. Choisissons $a > 0$ tel que $P_X[B(0, a)^c] \leq \delta/2$ (c'est possible puisque $\lim_n P_X[B(0, n)^c] = P_X(\emptyset) = 0$). Écrivons que f est uniformément continue sur la boule fermée $B_f(0, 2a)$: $\forall \varepsilon > 0, \exists \eta(\varepsilon) > 0$ tel que $\|x - y\| \leq \eta(\varepsilon)$ et $x, y \in B_f(0, 2a) \Rightarrow \|f(x) - f(y)\| \leq \varepsilon$.

En particulier, en vertu de l'inégalité triangulaire, pour $\varepsilon > 0$ fixé *quelconque*, on a

$$\|x\| \leq a \quad \text{et} \quad \|x - y\| \leq \eta(\varepsilon) \wedge a \Rightarrow \|f(x) - f(y)\| \leq \varepsilon,$$

soit, en prenant la contraposée de cette implication :

$$\|f(x) - f(y)\| > \varepsilon \Rightarrow \|x\| > a \quad \text{ou} \quad \|x - y\| > \eta(\varepsilon) \wedge a.$$

Il en résulte que, pour tout $n \in \mathbb{N}$, on a l'inclusion :

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset \{\|X\| > a\} \cup \{\|X_n - X\| > \eta(\varepsilon) \wedge a\}.$$

La suite $(X_n)_{n \in \mathbb{N}}$ convergeant en probabilité vers X , il existe un entier N tel que, pour tout $n \geq N$, on ait

$$P(\|X_n - X\| > \eta(\varepsilon) \wedge a) \leq \frac{\delta}{2}.$$

On a alors, pour tout $n \geq N$,

$$P(\|f(X_n) - f(X)\| > \varepsilon) \leq P(\|X\| > a) + P(\|X_n - X\| > \eta(\varepsilon) \wedge a) \leq \delta.$$

On a montré que, pour tout $\varepsilon > 0$, $\lim_n P(\|f(X_n) - f(X)\| > \varepsilon) = 0$. \square

Exemple 10.1. Soient deux suites $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ de variables aléatoires à valeurs dans \mathbb{R}^d convergeant en probabilité respectivement vers X et Y .

Alors $(X_n, Y_n) \xrightarrow{P} (X, Y)$: en effet, la suite $((X_n, Y_n))_{n \in \mathbb{N}}$ à valeurs dans \mathbb{R}^{2d} converge en probabilité vers (X, Y) et le produit scalaire est continu.

Par le même argument, si $d = 1$, on a $\max(X_n, Y_n) \xrightarrow{P} \max(X, Y)$.

Remarque. Il apparaît que dans tout ce qui vient d'être exposé, on peut changer les variables aléatoires sur un ensemble de probabilité nulle sans altérer ni les définitions, ni les résultats, ce qui suggère une « théorie » des convergences sur les **classes de variables aléatoires**. Plus précisément, soit X une application de $D_X \in \mathcal{A}$ à valeurs dans \mathbb{R}^d ou \mathbb{R} : on dit que X est une variable aléatoire **définie P-p.s.** si X est mesurable par rapport à la tribu trace $D_X \cap \mathcal{A}$ de D_X sur \mathcal{A} et si $P(D_X) = 1$. On dira que les variables aléatoires définies P-p.s. X et Y sont **égales P-p.s.** si $P(\{\omega \in D_X \cap D_Y \mid X(\omega) = Y(\omega)\}) = 1$. Suivant que l'espace des valeurs prises par les variables aléatoires est \mathbb{R}^d ou \mathbb{R} , on définit alors l'ensemble \mathcal{G} des variables aléatoires définies P-p.s. (resp. définies P-p.s. et P-p.s. finies) :

c'est un espace vectoriel; le sous-ensemble K des variables aléatoires P-p.s. égales à 0 en est un sous-espace vectoriel. L'égalité P-p.s. est une relation d'équivalence; l'ensemble quotient $L^0(\Omega, \mathcal{A}, P)$ de G par cette relation d'équivalence est l'espace vectoriel quotient de G par K et est appelé **ensemble des (classes de) variables aléatoires définies P-p.s. (et, dans le cas de \mathbb{R} , P-p.s. finies)**. Tout ce qui a été dit sur les convergences se transporte alors à $L^0(\Omega, \mathcal{A}, P)$. On peut définir sur $L^0(\Omega, \mathcal{A}, P)$ une **métrique** qui le rende complet et telle que la convergence de suite au sens de cette métrique soit équivalente à la convergence en probabilité (cf. exercice 1, chapitre 10).

10.2. Convergence L^p et équi-intégrabilité

Dans ce paragraphe, toutes les variables aléatoires sont définies sur le même espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{R}^d , $d \geq 1$, ou \mathbb{R} .

Si X est une variable aléatoire intégrable, le théorème de convergence dominée implique que

$$\lim_{a \rightarrow +\infty} \int_{(|X| > a)} |X| dP = 0.$$

La notion d'**équi-intégrabilité** généralise cette propriété à une famille quelconque de variables aléatoires en lui donnant un caractère uniforme.

Définition 10.6. La famille de variables aléatoires $(X_i)_{i \in I}$, où I est un ensemble quelconque, est **équi-intégrable** si

$$\lim_{a \rightarrow +\infty} \sup_{i \in I} \int_{(|X_i| > a)} |X_i| dP = 0.$$

On donne une condition suffisante d'équi-intégrabilité.

Proposition 10.7. Si la famille $(X_i)_{i \in I}$ est uniformément bornée par une variable aléatoire positive et intégrable X , c'est à dire si

$$\forall i \in I \quad |X_i| \leq X \quad \text{P-p.s.},$$

elle est équi-intégrable. En particulier, toute **famille finie** de variables aléatoires intégrables est équi-intégrable.

Démonstration. On a $(|X_i| > a) \subset (|X| > a)$ pour tout $i \in I$ et pour tout $a > 0$. Il en résulte que, pour tout $a > 0$, on a

$$\forall i \in I \quad \int_{(|X_i| > a)} |X_i| dP \leq \int_{(|X| > a)} |X| dP,$$

et donc aussi

$$\sup_{i \in I} \int_{(|X_i| > a)} |X_i| dP \leq \int_{(|X| > a)} |X| dP;$$

ceci démontre que la famille est équi-intégrable, le membre de droite tendant vers 0 quand a tend vers l'infini.

Si I est fini, la variable aléatoire positive $X = \max_{i \in I} |X_i|$ est intégrable ; il suffit d'appliquer la première partie. \square

On donne maintenant une condition nécessaire et suffisante d'équi-intégrabilité. Il nous faut auparavant définir la notion d'équi-continuité.

Définition 10.8. Soit I est un ensemble quelconque. La famille de variables aléatoires $(X_i)_{i \in I}$ est équi-continue si

$$\forall \varepsilon > 0 \quad \exists \eta > 0 \quad \text{tel que} \quad P(A) \leq \eta \implies \sup_{i \in I} \int_A |X_i| dP \leq \varepsilon.$$

Remarque. Cette notion n'est autre que la notion habituelle d'équi-continuité en un point d'une famille de fonctions. En effet, il est classique de définir l'**algèbre métrique** \mathcal{A} , c'est à dire l'ensemble \mathcal{A} muni de l'écart (ou pseudo-distance) défini par l'application $(A, B) \mapsto P(A \Delta B)$; $P(A)$ représente alors la « distance » de A à \emptyset et, dans cette optique, il s'agit en fait de l'équi-continuité en \emptyset de la famille de fonctions $A \mapsto \int_A |X_i| dP$.

Proposition 10.9. La famille de variables aléatoires $(X_i)_{i \in I}$ est équi-intégrable si et seulement si elle est équi-continue et bornée dans L^1 , c'est à dire telle que $\sup_{i \in I} \int_{\Omega} |X_i| dP < +\infty$.

Démonstration. • *Condition nécessaire.* Supposons la famille équi-intégrable. Pour tout $A \in \mathcal{A}$ et tout $a > 0$, on a, pour tout $i \in I$,

$$\begin{aligned} \int_A |X_i| dP &\leq \int_{A \cap \{|X_i| \leq a\}} |X_i| dP + \int_{A \cap \{|X_i| > a\}} |X_i| dP \\ &\leq aP(A) + \sup_{i \in I} \int_{\{|X_i| > a\}} |X_i| dP ; \end{aligned}$$

on a donc, pour tout $A \in \mathcal{A}$ et tout $a > 0$,

$$\sup_{i \in I} \int_A |X_i| dP \leq aP(A) + \sup_{i \in I} \int_{\{|X_i| > a\}} |X_i| dP.$$

En prenant pour A l'ensemble Ω , on obtient que la famille est bornée dans L^1 . Par ailleurs, $\varepsilon > 0$ étant donné, on choisit $a > 0$ tel que

$$\sup_{i \in I} \int_{\{|X_i| > a\}} |X_i| dP \leq \frac{\varepsilon}{2} \quad \text{puis} \quad \eta = \frac{\varepsilon}{2a} ; \quad \text{alors, dès que } P(A) \leq \eta, \quad \text{on a}$$

$$\sup_{i \in I} \int_A |X_i| dP \leq \varepsilon \quad \text{et la famille est équi-continue.}$$

• *Condition suffisante.* Supposons la famille équi-continue et bornée dans L^1 . Il résulte de l'inégalité de Markov (cf. chapitre 8) que l'on a, pour tout

$a > 0$ et tout $i \in I$,

$$P(|X_i| > a) \leq \frac{1}{a} \int |X_i| dP \leq \frac{1}{a} \sup_{i \in I} \int |X_i| dP.$$

La famille étant bornée dans L^1 , il vient alors :

$$\lim_{a \rightarrow +\infty} \sup_{i \in I} P(|X_i| > a) = 0. \quad (10.1)$$

Soit $\varepsilon > 0$ quelconque ; la famille étant équi-continu, on choisit $\eta > 0$ tel que l'on ait l'implication

$$P(A) \leq \eta \implies \sup_{i \in I} \int_A |X_i| dP \leq \varepsilon. \quad (10.2)$$

On choisit alors $M > 0$ tel que l'on ait, pour tout $a \geq M$, $\sup_{i \in I} P(|X_i| > a) \leq \eta$, ce qui est possible d'après (10.1). Il résulte alors de (10.2) que l'on a

$$\forall a \geq M \quad \text{et} \quad \forall i \in I \quad \int_{\{|X_i| > a\}} |X_i| dP \leq \varepsilon,$$

et donc

$$\forall a \geq M \quad \sup_{i \in I} \int_{\{|X_i| > a\}} |X_i| dP \leq \varepsilon,$$

ce qui démontre l'équi-intégrabilité de la famille. \square

Définition 10.10. Soit $p \geq 1$. Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires admettant un moment d'ordre p **converge dans \mathcal{L}^p** vers une variable aléatoire X si $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ et si

$$\lim_n E[|X_n - X|^p] = 0.$$

On note :

$$X_n \xrightarrow{\mathcal{L}^p} X.$$

La suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires **converge dans \mathcal{L}^p** s'il existe une variable aléatoire $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ telle que cette suite **converge dans \mathcal{L}^p** vers X .

Remarque. Si $p = 1$ (resp. $p = 2$) on dit que la suite **converge en moyenne** (resp. **en moyenne quadratique**). Si $p \geq 1$, comme il résulte de l'inégalité de Minkowski (cf. chapitre 8), l'application $X \mapsto [E|X|^p]^{1/p}$ est une **semi-norme** sur $\mathcal{L}^p(\Omega, \mathcal{A}, P)$: les notions de convergence \mathcal{L}^p sont les notions de convergence relativement à cette semi-norme. En particulier, si une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires **converge dans \mathcal{L}^p** , sa limite est **P-p.s. unique**. L'ensemble quotient de $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ par la relation d'équivalence d'égalité P-p.s. est noté $L^p(\Omega, \mathcal{A}, P)$; c'est, en vertu de l'inégalité de Minkowski, un espace vectoriel normé, dont la **norme** est obtenue par passage

au quotient de la semi-norme $X \mapsto [E|X|^p]^{1/p}$ (on parle de la norme p de X et on note usuellement $\|X\|_p = [E|X|^p]^{1/p}$) : dans cet espace de classes de variables aléatoires, la limite d'une suite est alors unique. Il est d'usage de noter de la même façon une variable aléatoire et sa classe ; on en fera de même pour la semi-norme et la norme associée et on parlera indifféremment de convergence \mathcal{L}^p ou L^p .

Le théorème suivant établit les relations entre convergence en probabilité et convergence L^p et démontre que si $p \geq 1$, l'ensemble $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ est complet (non séparé). L'espace $L^p(\Omega, \mathcal{A}, P)$ est alors un espace de Banach. Pour démontrer ce théorème, nous utiliserons l'inégalité de convexité suivante :

Lemme 10.11. Soit $p \geq 1$. Pour tous réels a, b, c on a

$$|a - b|^p \leq 2^{p-1} [|a - c|^p + |c - b|^p] \quad (10.3)$$

Démonstration. La fonction $x \mapsto x^p$ étant convexe sur \mathbb{R}^+ , on a, pour tous u, v positifs

$$\left[\frac{1}{2}(u + v) \right]^p \leq \frac{1}{2}(u^p + v^p),$$

soit

$$(u + v)^p \leq 2^{p-1} (u^p + v^p);$$

tenant compte de l'inégalité triangulaire et de la croissance de la fonction $x \mapsto x^p$, il vient alors

$$|a - b|^p \leq (|a - c| + |c - b|)^p \leq 2^{p-1} [|a - c|^p + |c - b|^p]. \quad \square$$

Théorème 10.12. Soient $p \geq 1$ et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires admettant un moment d'ordre p . Les assertions suivantes sont équivalentes :

- (i) la suite $(X_n)_{n \in \mathbb{N}}$ converge dans L^p ;
- (ii) la suite $(X_n)_{n \in \mathbb{N}}$ est de Cauchy dans L^p , c'est à dire que

$$\lim_{m, n} E|X_n - X_m|^p = 0 :$$

(iii) la suite $(|X_n|^p)_{n \in \mathbb{N}}$ est équi-intégrable et il existe $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ telle que $X_n \xrightarrow{P} X$.

Démonstration. (i) \Rightarrow (ii) : si la suite $(X_n)_{n \in \mathbb{N}}$ converge dans L^p , il existe $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ telle que

$$\lim_n E\{|X_n - X|^p\} = 0 ;$$

l'inégalité de Minkowski assure que, pour tous m et n ,

$$\|X_n - X_m\|_p \leq \|X_n - X\|_p + \|X - X_m\|_p ,$$

ce qui démontre que la suite est de Cauchy (on n'a fait que rétablir dans ce cas particulier le fait général que toute suite convergente pour une semi-norme est de Cauchy relativement à cette semi-norme).

(ii) \Rightarrow (iii) : soit $\varepsilon > 0$ et un entier N tel que l'on ait, pour tout $n, m \geq N$, $E[|X_n - X_m|^p] \leq \varepsilon/2^p$. Il résulte alors de l'inégalité (10.3) que, pour tout $A \in \mathcal{A}$, on a, pour tout $n \geq N$,

$$\int_A |X_n|^p dP \leq 2^{p-1} \left[\int_A |X_N|^p dP + \int_A |X_n - X_N|^p dP \right] \leq 2^{p-1} \int_A |X_N|^p dP + \frac{\varepsilon}{2};$$

on a alors, pour tout $A \in \mathcal{A}$,

$$\sup_{n \geq N} \int_A |X_n|^p dP \leq 2^{p-1} \int_A |X_N|^p dP + \frac{\varepsilon}{2},$$

et donc,

$$\sup_{n \in \mathbb{N}} \int_A |X_n|^p dP \leq \sup_{n \leq N} \int_A |X_n|^p dP + 2^{p-1} \int_A |X_N|^p dP + \frac{\varepsilon}{2}. \quad (10.4)$$

Il en résulte déjà que la famille $\{|X_n|^p \mid n \in \mathbb{N}\}$ est bornée dans L^1 . De plus, la famille finie $\{|X_n|^p \mid n \leq N\}$ étant équi-intégrable, est en particulier équi-continue. La majoration (10.4) montre alors que la famille $\{|X_n|^p \mid n \in \mathbb{N}\}$ est aussi équi-continue, donc équi-intégrable, puisque bornée dans L^1 . Enfin, il résulte de la croissance de la fonction $x \mapsto x^p$ et de l'inégalité de Markov que l'on a, pour tout $\varepsilon > 0$, pour tous n et m ,

$$P(|X_n - X_m| > \varepsilon) \leq \varepsilon^{-p} E|X_n - X_m|^p.$$

Il en résulte que la suite $(X_n)_{n \in \mathbb{N}}$ est de Cauchy en probabilité et converge donc en probabilité vers une variable aléatoire X . La suite $(|X_n|^p)_{n \in \mathbb{N}}$ étant bornée dans L^1 , il résulte du lemme de Fatou que

$$\int_{\Omega} |X|^p dP \leq \liminf_n E|X_n|^p \leq \sup_{n \in \mathbb{N}} E|X_n|^p < +\infty,$$

ce qui démontre que $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$.

(iii) \Rightarrow (i) : pour tout $\varepsilon > 0$, on a, d'après l'inégalité (10.3),

$$E|X_n - X|^p \leq \int_{(|X_n - X| \leq \varepsilon^{1/p})} |X_n - X|^p dP + 2^{p-1} \left[\int_{(|X_n - X| > \varepsilon^{1/p})} (|X_n|^p + |X|^p) dP \right],$$

soit

$$E|X_n - X|^p \leq \varepsilon + 2^{p-1} \left[\int_{(|X_n - X| \leq \varepsilon^{1/p})} |X_n|^p dP + \int_{(|X_n - X| > \varepsilon^{1/p})} |X|^p dP \right]. \quad (10.5)$$

L'équi-continuité de la famille $\{|X_n|^p, n \in \mathbb{N}; |X|^p\}$ permet de trouver $\eta > 0$ tel que l'on ait

$$\sup_{n \in \mathbb{N}} \int_A |X_n|^p dP + \int_A |X|^p dP \leq \frac{\varepsilon}{2^{p-1}} \quad \text{dès que } P(A) \leq \eta;$$

de plus, la convergence en probabilité de la suite $(X_n)_{n \in \mathbb{N}}$ vers X implique qu'il existe N tel que l'on ait, pour tout $n \geq N$,

$$P(|X_n - X| > \varepsilon^{1/p}) \leq \eta.$$

Il résulte alors de l'inégalité (10.5) que l'on a

$$\forall \varepsilon > 0 \quad \limsup_n E|X_n - X|^p \leq 2\varepsilon,$$

ce qui démontre que la suite $(X_n)_{n \in \mathbb{N}}$ converge dans L^p vers X . \square

Un contre-exemple : si, pour tout $n \in \mathbb{N}^*$, X_n est de loi $\frac{1}{n}\delta_{n^2} + (1 - \frac{1}{n})\delta_0$, on a, pour tout $\varepsilon > 0$, $P(|X_n| > \varepsilon) = \frac{1}{n}$ et $EX_n = n$; **la suite $(X_n)_{n \in \mathbb{N}}$ converge donc en probabilité vers 0 mais ne converge pas dans \mathcal{L}^1 .**

10.3. Séries de variables aléatoires indépendantes

On étudie une condition suffisante de convergence P-p.s. et L^2 de série de variables aléatoires indépendantes admettant un moment d'ordre deux. On donne d'abord l'inégalité de Kolmogorov qui généralise l'inégalité de Tchebitchev.

Théorème 10.13 (Inégalité de Kolmogorov). *Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes, admettant un moment d'ordre deux et centrées. Pour tout $\varepsilon > 0$, on a l'inégalité*

$$P\left(\max_{1 \leq k \leq n} \left| \sum_{i=1}^k X_i \right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \left(\sum_{i=1}^n \sigma_{X_i}^2 \right).$$

Démonstration. On note, pour tout k tel que $1 \leq k \leq n$,

$$S_k = \sum_{i=1}^k X_i \quad \text{et} \quad M_k = \max_{1 \leq i \leq k} |S_i|.$$

Il s'agit de majorer la probabilité de l'ensemble $E = (M_n \geq \varepsilon)$; s'il est vide, l'inégalité est triviale et on se place dans le cas où il ne l'est pas. Faisons apparaître l'indice k pour lequel $|S_k|$ dépasse pour la première fois le seuil ε ; on introduit pour cela les ensembles

$$E_1 = (|S_1| \geq \varepsilon) \quad \text{et, si } 2 \leq k \leq n, \quad E_k = (|S_k| \geq \varepsilon) \cap \left[\bigcap_{i=1}^{k-1} (|S_i| < \varepsilon) \right].$$

Ces ensembles forment une partition de E et par conséquent, on a

$$P(E) = \sum_{k=1}^n P(E_k).$$

Il résulte de la définition de E_k que

$$P(E_k) \leq \frac{1}{\varepsilon^2} E(\mathbf{1}_{E_k} S_k^2),$$

et donc que

$$P(E) \leq \frac{1}{\varepsilon^2} \sum_{k=1}^n E(\mathbf{1}_{E_k} S_k^2). \quad (10.6)$$

Démontrons alors que, si $1 \leq k \leq n$, on a

$$E(\mathbf{1}_{E_k} S_n^2) = E(\mathbf{1}_{E_k} S_n^2); \quad (10.7)$$

en effet, on a

$$E(\mathbf{1}_{E_k} S_n^2) = E\left[\mathbf{1}_{E_k} \left(S_k + \sum_{j=k+1}^n X_j\right)^2\right]$$

soit

$$E(\mathbf{1}_{E_k} S_n^2) = E(\mathbf{1}_{E_k} S_k^2) + 2E\left[\mathbf{1}_{E_k} S_k \left(\sum_{j=k+1}^n X_j\right)\right] + E\left[\mathbf{1}_{E_k} \left(\sum_{j=k+1}^n X_j\right)^2\right].$$

Les variables aléatoires $\mathbf{1}_{E_k} S_k$ et $\sum_{j=k+1}^n X_j$ sont indépendantes et la variable aléatoire $\sum_{j=k+1}^n X_j$ est centrée; le terme médian du second membre est donc nul et, le troisième terme étant positif, on obtient l'inégalité (10.7). En reportant dans l'inégalité (10.6) le majorant ainsi obtenu, et en tenant compte de ce que les ensembles E_k forment une partition de E , on obtient :

$$P(E) \leq \frac{1}{\varepsilon^2} \sum_{k=1}^n E(\mathbf{1}_{E_k} S_n^2) = \frac{1}{\varepsilon^2} E(\mathbf{1}_E S_n^2) \leq \frac{1}{\varepsilon^2} E(S_n^2).$$

Les variables aléatoires X_i étant indépendantes et centrées, on a $E(S_n^2) = \sum_{i=1}^n \sigma_{X_i}^2$, ce qui achève la démonstration. \square

On en déduit une condition suffisante de convergence P-p.s. d'une série de variables aléatoires indépendantes.

Proposition 10.14. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes, centrées et admettant un moment d'ordre deux. Si $\sum_{n=1}^{+\infty} \sigma_{X_n}^2 < +\infty$, la série $\sum X_n$ de terme général X_n converge P-p.s. et dans L^2 .

Démonstration. On démontre d'abord la convergence P-p.s. Pour $m \in \mathbb{N}^*$, notons

$$S_m = \sum_{j=1}^m X_j \quad A_m = \sup_{k \in \mathbb{N}^*} |S_{m+k} - S_m| \quad \text{et} \quad A = \inf_{m \in \mathbb{N}^*} A_m.$$

Il résulte du critère de Cauchy pour les séries numériques que l'on a

$$\left\{ \sum X_n \text{ converge} \right\} = \{A = 0\}.$$

Mais on a $\{A \neq 0\} = \bigcup_{n \in \mathbb{N}^*} \{A > \frac{1}{n}\}$ et, pour tout $n \in \mathbb{N}^*$, $\{A > \frac{1}{n}\} \subset \bigcap_{m \in \mathbb{N}^*} \{A_m > \frac{1}{n}\}$, ce qui donne l'inclusion :

$$\{A \neq 0\} \subset \bigcup_{n \in \mathbb{N}^*} \bigcap_{m \in \mathbb{N}^*} \left\{ A_m > \frac{1}{n} \right\}. \quad (10.8)$$

Puisque $\sup_{k \in \mathbb{N}^*} |S_{m+k} - S_m| = \lim_r \nearrow \sup_{1 \leq k \leq r} |S_{m+k} - S_m|$, la suite d'ensembles $\left\{ \sup_{1 \leq k \leq r} |S_{m+k} - S_m| > \frac{1}{n} \right\}$ est croissante, et on a

$$\left\{ A_m > \frac{1}{n} \right\} = \bigcup_{r \in \mathbb{N}^*} \left\{ \sup_{1 \leq k \leq r} |S_{m+k} - S_m| > \frac{1}{n} \right\}. \quad (10.9)$$

Il résulte de l'inégalité de Kolmogorov que

$$P\left(\sup_{1 \leq k \leq r} |S_{m+k} - S_m| > \frac{1}{n} \right) \leq n^2 \sum_{j=m+1}^r \sigma_{X_j}^2;$$

l'égalité (10.9) faisant intervenir une suite croissante d'ensembles, il vient

$$P\left(A_m > \frac{1}{n} \right) = \lim_r P\left(\sup_{1 \leq k \leq r} |S_{m+k} - S_m| > \frac{1}{n} \right) \leq n^2 \sum_{j=m+1}^{+\infty} \sigma_{X_j}^2.$$

Il en résulte que, pour tout $m \in \mathbb{N}^*$, on a

$$0 \leq P\left[\bigcap_{p \in \mathbb{N}^*} \left(A_p > \frac{1}{n} \right) \right] \leq P\left(A_m > \frac{1}{n} \right) \leq n^2 \sum_{j=m+1}^{+\infty} \sigma_{X_j}^2;$$

le membre de droite convergeant vers 0 quand m tend vers l'infini (reste d'une série convergente), il vient que, pour tout $n \in \mathbb{N}^*$, $P\left[\bigcap_{p \in \mathbb{N}^*} (A_p > 1/n) \right] = 0$. Il résulte alors de l'inclusion (10.8) que $P(A \neq 0) = 0$, c'est à dire que la série de terme général X_n converge P-p.s.

Il y a aussi convergence dans L^2 puisque la suite des sommes partielles est de Cauchy pour la norme 2: en effet, les variables aléatoires X_n étant centrées et indépendantes, on a, si $m < n$,

$$E[(S_n - S_m)^2] = \sum_{j=m+1}^n \sigma_{X_j}^2.$$

ce qui démontre le résultat, la série des variances étant convergente. \square

10.4. Lois des grands nombres

L'étude de phénomènes aléatoires conduit fréquemment à poser le problème de convergence de la suite des **moyennes arithmétiques** d'une suite de variables aléatoires **indépendantes** de **même loi**. C'est le cas par exemple, en statistique, en **théorie de l'estimation** : si X est une variable aléatoire qui modélise une grandeur liée à un phénomène aléatoire, se pose le problème d'estimer sa loi, ou certains paramètres de cette loi, au vu d'une suite de réalisations de ce phénomène, réalisations obtenues au cours d'expériences **indépendantes**. On est alors conduit à introduire une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes de même loi que X , et à étudier, pour une certaine fonction f , la suite de terme général de la forme $\frac{1}{n} \sum_{j=1}^n f(X_j)$.

Si $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires réelles, on note, pour tout $n \in \mathbb{N}^*$,

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

En termes statistiques, \bar{X}_n est appelée **moyenne empirique** de l'échantillon (X_1, X_2, \dots, X_n) . On donne le nom de **loi des grands nombres** à deux théorèmes principaux qui affirment la convergence de la suite de terme général \bar{X}_n sous certaines hypothèses. Pour la **loi faible**, il s'agit de **convergence en probabilité**. Pour la **loi forte** il s'agit de **convergence presque sûre**. On donne aussi le nom de loi des grands nombres (faible ou forte, suivant le cas) à de nombreuses variantes de ces deux énoncés, obtenues avec des hypothèses plus ou moins fortes.

Noter que pour les **lois faibles**, l'hypothèse d'indépendance des variables aléatoires X_n n'est pas nécessaire¹ : on demande seulement la **non corrélation** ou l'**indépendance deux à deux** des variables aléatoires, alors que pour les **lois fortes**, on demande l'**indépendance globale**² des variables aléatoires. On laisse le lecteur établir, à titre d'exercice, une version de la loi forte pour des variables aléatoires deux à deux indépendantes.

Rappelons pour commencer deux lemmes élémentaires d'analyse qui seront plusieurs fois utilisés pour établir des « lois des grands nombres ».

1. On a fait l'hypothèse d'indépendance globale dans l'énoncé du théorème 7.9 (tome 1) par souci de simplicité.

2. Les lois des grands nombres apparaissent dans de multiples problématiques et font l'objet d'une abondante littérature ; en particulier, dans le cadre de notre étude, l'hypothèse d'indépendance peut être levée grâce à la théorie des martingales.

Lemme 10.15 (Lemme de Cesàro³). Soit une suite $(x_n)_{n \in \mathbb{N}^*}$ de réels convergeant vers x quand n tend vers l'infini. La suite de terme général $\frac{1}{n} \sum_{j=1}^n x_j$ est convergente de limite x .

Démonstration. Soit $\varepsilon > 0$ quelconque puis N tel qu'on ait, pour tout $n \geq N$, $|x_n - x| \leq \varepsilon$; puisque

$$\left| \frac{1}{n} \sum_{j=1}^n x_j - x \right| \leq \frac{1}{n} \sum_{j=1}^N |x_j - x| + \frac{1}{n} \sum_{j=N+1}^n |x_j - x| \leq \frac{1}{n} \sum_{j=1}^N |x_j - x| + \varepsilon.$$

on a $\limsup_n \left| \frac{1}{n} \sum_{j=1}^n x_j - x \right| \leq \varepsilon$, ce qui démontre le résultat, vu l'arbitraire de ε . \square

Lemme 10.16 (Lemme de Kronecker⁴). Soient une série de terme général réel x_n convergente et une suite croissante $(b_n)_{n \in \mathbb{N}^*}$ de réels tendant vers l'infini avec n . On a alors

$$\lim_n \frac{1}{b_n} \sum_{j=1}^n b_j x_j = 0.$$

Démonstration. On note $S = \sum_{j=1}^{+\infty} x_j$ et, pour tout $n \in \mathbb{N}^*$, $S_n = -S + \sum_{j=1}^n x_j$, si bien que la suite $(S_n)_{n \in \mathbb{N}^*}$ tend vers 0 quand n tend vers l'infini. On a $x_n = S_n - S_{n-1}$ et donc, par transformation (ou sommation) d'Abel, pour tous entiers n et N tels que $n > N \geq 2$,

$$\sum_{j=N}^n b_j x_j = \sum_{j=N}^n b_j (S_j - S_{j-1}) = S_n b_n - b_N S_{N-1} - \sum_{j=N}^{n-1} S_j (b_{j+1} - b_j),$$

soit, dès que $b_n \neq 0$,

$$\frac{1}{b_n} \sum_{j=1}^n b_j x_j = \frac{1}{b_n} \sum_{j=1}^N b_j x_j - \frac{b_N S_{N-1}}{b_n} + S_n - \frac{1}{b_n} \left[\sum_{j=N}^{n-1} S_j (b_{j+1} - b_j) \right].$$

La suite de terme général $\frac{1}{b_n} \sum_{j=1}^N b_j x_j - \frac{b_N S_{N-1}}{b_n} + S_n$ tend vers 0 quand n tend vers l'infini. Soit alors $\varepsilon > 0$ et N tel que, pour tout $n \geq N$, on ait

3. Ernesto Cesàro (1859-1906), né à Naples, devient professeur à l'université de cette ville en 1883. Ses activités mathématiques sont variées; il s'est en particulier intéressé aux liens entre l'arithmétique et le calcul intégral, ainsi qu'au comportement des séries entières sur le cercle de convergence.

4. Leopold Kronecker (1823-1891) est né à Liegnitz, en Pologne. Après des études à Berlin et Bonn, il s'enrichit dans les finances, ce qui lui permet ensuite de se consacrer aux mathématiques. Il enseigne à Berlin à partir de 1861. Ses travaux portent sur la théorie des équations, sur les fonctions elliptiques et la théorie algébrique des nombres. Il s'est farouchement opposé à la théorie des ensembles de Cantor, et à la construction des nombres réels proposée par Weierstrass.

$$\left| \frac{1}{b_n} \sum_{j=1}^N b_j x_j - \frac{b_N S_{N-1}}{b_n} + S_n \right| \leq \frac{\varepsilon}{2} \quad \text{et} \quad |S_n| \leq \frac{\varepsilon}{2}.$$

Puisque la suite $(b_n)_{n \in \mathbb{N}^*}$ est croissante, on a

$$\left| \sum_{j=N}^{n-1} S_j (b_{j+1} - b_j) \right| \leq \sum_{j=N}^{n-1} |S_j| (b_{j+1} - b_j) \leq \frac{\varepsilon}{2} \sum_{j=N}^{n-1} (b_{j+1} - b_j) = \frac{\varepsilon}{2} (b_n - b_N).$$

Il en résulte que

$$\limsup_n \frac{1}{b_n} \left| \sum_{j=N}^{n-1} S_j (b_{j+1} - b_j) \right| \leq \frac{\varepsilon}{2},$$

et donc que $\limsup_n \left| \frac{1}{b_n} \sum_{j=1}^n b_j x_j \right| \leq \varepsilon$, ce qui démontre le résultat, vu l'arbitraire de ε . \square

Théorème 10.17 (Loi faible des grands nombres). Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, admettant un moment d'ordre deux, et deux à deux non corrélées⁵. On suppose la convergence des suites :

$$\frac{1}{n} \sum_{j=1}^n \mathbb{E} X_j \xrightarrow{n \rightarrow +\infty} m \quad \text{et} \quad \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2 \xrightarrow{n \rightarrow +\infty} 0.$$

Alors, la suite des variables aléatoires $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ converge en probabilité vers m .

Démonstration. On a

$$\mathbb{E} \bar{X}_n = \frac{1}{n} \sum_{j=1}^n \mathbb{E} X_j.$$

Les variables aléatoires X_n étant deux à deux non corrélées, on a aussi :

$$\sigma_{\bar{X}_n}^2 = \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2.$$

L'inégalité triangulaire conduit à l'inégalité :

$$|\bar{X}_n - m| \leq \left| \bar{X}_n - \frac{1}{n} \sum_{j=1}^n \mathbb{E} X_j \right| + \left| \frac{1}{n} \sum_{j=1}^n \mathbb{E} X_j - m \right|.$$

5. Deux variables aléatoires réelles admettant un moment d'ordre deux sont dites **non corrélées** si leur coefficient de corrélation est nul (ce qui est équivalent à dire que leur covariance est nulle).

Mais, pour tout $\varepsilon > 0$, il existe $N(\varepsilon) \in \mathbb{N}^*$ tel que, pour tout $n \geq N(\varepsilon)$, on ait $\left| \frac{1}{n} \sum_{j=1}^n EX_j - m \right| \leq \frac{\varepsilon}{2}$. Pour tout $n \geq N(\varepsilon)$, on a donc l'inclusion des ensembles

$$\left(\left| \bar{X}_n - \frac{1}{n} \sum_{j=1}^n EX_j \right| \leq \frac{\varepsilon}{2} \right) \subset \left(|\bar{X}_n - m| \leq \varepsilon \right),$$

ou encore, pour les complémentaires, l'inclusion :

$$\left(|\bar{X}_n - m| > \varepsilon \right) \subset \left(\left| \bar{X}_n - \frac{1}{n} \sum_{j=1}^n EX_j \right| > \frac{\varepsilon}{2} \right).$$

L'inégalité de Bienaymé-Tchebitchev permet d'écrire :

$$P\left(\left| \bar{X}_n - \frac{1}{n} \sum_{j=1}^n EX_j \right| > \frac{\varepsilon}{2} \right) \leq \frac{4}{\varepsilon^2} \sigma_{\bar{X}_n}^2 = \frac{4}{\varepsilon^2} \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2.$$

Il en résulte que, pour tout $n \geq N(\varepsilon)$, on a :

$$P(|\bar{X}_n - m| > \varepsilon) \leq \frac{4}{\varepsilon^2} \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2,$$

ce qui, en utilisant la seconde hypothèse, démontre le résultat. \square

Remarque. L'hypothèse sur les variances contraint les variables aléatoires à ne pas être « trop dispersées » autour de leur moyenne. Les hypothèses du théorème précédent sont toutes satisfaites, en particulier, si les variables aléatoires X_n sont **indépendantes et de même loi** et si X_1 admet un moment d'ordre **deux**. En fait, si les variables aléatoires sont **indépendantes et de même loi**, il suffit de l'existence d'un moment d'ordre **un**, comme le montre le théorème de **Khintchine**⁶ énoncé ci-dessous.

Avant d'étudier ce théorème, on rappelle un cas particulier du théorème précédent (il faut noter toutefois qu'il lui soit historiquement antérieur); il s'agit du théorème de Bernoulli étudié au tome 1, chapitre 7, p. 236.

Théorème 10.18 (Théorème de Bernoulli). Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements **indépendants de même probabilité** p . La suite des variables aléatoires $\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{A_n}$ converge en probabilité vers p .

Démonstration. Les variables aléatoires $\mathbf{1}_{A_n}$, $n \in \mathbb{N}^*$, sont indépendantes de même loi de **Bernoulli**; elles admettent un moment d'ordre deux, et on se trouve dans la situation de la remarque. \square

6. Alexandre Khintchine (1894-1959) a été professeur à l'université de Moscou à partir de 1922. Ses travaux concernent l'analyse réelle, la théorie des nombres, et les probabilités. Il a introduit, parallèlement à P. Lévy, la notion de variable aléatoire. On lui doit la définition de processus aléatoires stationnaires.

Remarque. Ce théorème assure que, si on fait une suite d'expériences aléatoires répétées de manière indépendante, la suite des **fréquences relatives** d'apparition d'une certaine propriété liée à cette expérience converge en probabilité (au sens de la probabilité P du modèle adopté) vers la probabilité de l'événement lié à cette propriété. C'est donc un théorème de cohérence du modèle probabiliste avec l'approche fréquentiste et intuitive de la notion de probabilité qui est à l'origine du calcul des probabilités.

Théorème 10.19 (Théorème de Khintchine ; loi faible des grands nombres).

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , deux à deux indépendantes de même loi μ et admettant une moyenne. Alors, la suite des variables aléatoires $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ converge en probabilité vers la moyenne commune EX_1 .

Démonstration. On se ramène au théorème 10.17 par un procédé de troncature. Soit

$$Y_n = \mathbf{1}_{\{|X_n| \leq n\}} X_n \quad \text{et} \quad \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Les variables aléatoires Y_n sont indépendantes deux à deux et bornées ; on va démontrer qu'elles satisfont les hypothèses du théorème 10.17. On a

$$\begin{aligned} \sum_{i=1}^n EY_j &= \sum_{j=1}^n \int_{\{|x| \leq j\}} x \, d\mu(x) = \sum_{i=1}^n \sum_{k=0}^{i-1} \int_{\{k < |x| \leq k+1\}} x \, d\mu(x) \\ &= \sum_{k=0}^{n-1} \sum_{j=k+1}^n \int_{\{k < |x| \leq k+1\}} x \, d\mu(x), \end{aligned}$$

soit

$$\sum_{j=1}^n EY_j = \sum_{k=0}^{n-1} (n-k) \int_{\{k < |x| \leq k+1\}} x \, d\mu(x),$$

ce qui donne

$$\frac{1}{n} \sum_{i=1}^n EY_i = \int_{\{|x| \leq n\}} x \, d\mu(x) - \frac{1}{n} \sum_{k=0}^{n-1} k \int_{\{k < |x| \leq k+1\}} x \, d\mu(x).$$

L'existence de la moyenne de X_1 , implique que l'on a

$$\lim_n \int_{\{|x| \leq n\}} x \, d\mu(x) = EX_1$$

et que la série de terme général $\int_{\{k < |x| \leq k+1\}} x \, d\mu(x)$ est convergente : le lemme de Kronecker assure alors que

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} k \int_{(k < |x| \leq k+1)} x \, d\mu(x) = 0,$$

ce qui donne

$$\lim_n \frac{1}{n} \sum_{j=1}^n \text{E}Y_j = \text{E}X_1.$$

Par ailleurs, on a

$$\sigma_{\bar{Y}_j}^2 \leq \text{E}Y_j^2 = \int_{(|x| \leq j)} x^2 \, d\mu(x) \leq \int_{(|x| \leq n)} x^2 \, d\mu(x),$$

et donc

$$\begin{aligned} 0 \leq \frac{1}{n^2} \sum_{j=1}^n \sigma_{\bar{Y}_j}^2 &\leq \frac{1}{n} \int_{(|x| \leq n)} x^2 \, d\mu(x) \\ &\leq \frac{1}{n} \left[\int_{(|x| \leq \sqrt{n})} x^2 \, d\mu(x) + \int_{(\sqrt{n} < |x| \leq n)} x^2 \, d\mu(x) \right]. \end{aligned}$$

Il en résulte que l'on a

$$0 \leq \frac{1}{n^2} \sum_{j=1}^n \sigma_{\bar{Y}_j}^2 \leq \frac{1}{\sqrt{n}} \int_{(|x| \leq \sqrt{n})} |x| \, d\mu(x) + \int_{(\sqrt{n} < |x|)} |x| \, d\mu(x);$$

puisque $\int |x| \, d\mu(x) < +\infty$, le membre de droite tend vers 0 et on a

$$\lim_n \frac{1}{n^2} \sum_{j=1}^n \sigma_{\bar{Y}_j}^2 = 0.$$

Le théorème 10.17 assure alors que

$$\bar{Y}_n \xrightarrow{P} \text{E}X_1. \quad (10.10)$$

Si $n > r$, on note

$$\bar{Y}_{n,r} = \frac{1}{n} \left(\sum_{j=1}^r X_j + \sum_{j=r+1}^n Y_j \right).$$

On a

$$\begin{aligned} \text{P}(\bar{Y}_{n,r} \neq \bar{X}_n) &\leq \text{P}\left(\bigcup_{j=r+1}^n (Y_j \neq X_j) \right) \\ &\leq \sum_{j=r+1}^n \text{P}(Y_j \neq X_j) = \sum_{j=r+1}^n \int \mathbf{1}_{[0,|x|](j)} \, d\mu(x), \end{aligned}$$

et, puisque

$$\sum_{j=r+1}^n \mathbf{1}_{\{0, |x| \leq j\}}(j) \leq \sum_{j=r+1}^{+\infty} \mathbf{1}_{\{0, |x| \leq j\}}(j) \leq \mathbf{1}_{\{|x| > r\}} |x|,$$

il vient

$$P(\bar{Y}_{n,r} \neq \bar{X}_n) \leq \int_{\{|x| > r\}} |x| d\mu(x).$$

Pour tout $\varepsilon > 0$, il existe donc $r > 0$ tel que, dès que $n > r$, on ait $P(\bar{Y}_{n,r} \neq \bar{X}_n) \leq \frac{\varepsilon}{2}$. Alors, pour tout $\delta > 0$, on a

$$\begin{aligned} P(|\bar{X}_n - EX_1| > \delta) &= \\ P[(|\bar{X}_n - EX_1| > \delta) \cap (\bar{Y}_{n,r} \neq \bar{X}_n)] &+ P[(|\bar{X}_n - EX_1| > \delta) \cap (\bar{Y}_{n,r} = \bar{X}_n)]. \end{aligned}$$

et donc :

$$\begin{aligned} P(|\bar{X}_n - EX_1| > \delta) &\leq P(\bar{Y}_{n,r} \neq \bar{X}_n) + P(|\bar{Y}_{n,r} - EX_1| > \delta) \\ &\leq \frac{\varepsilon}{2} + P(|\bar{Y}_{n,r} - EX_1| > \delta). \end{aligned}$$

Mais d'après (10.10), on a $\lim_n P(|\bar{Y}_{n,r} - EX_1| > \delta) = 0$; il existe donc $N > r$ tel que, pour tout $n \geq N$ on ait

$$P(|\bar{Y}_{n,r} - EX_1| > \delta) \leq \frac{\varepsilon}{2};$$

en résumé, on a montré que, pour tout $\varepsilon > 0$ et tout $\delta > 0$, il existe N tel que $P(|\bar{X}_n - EX_1| > \delta) \leq \varepsilon$ dès que $n \geq N$, c'est à dire que la suite $(\bar{X}_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers m . \square

Théorème 10.20 (Loi forte des grands nombres). Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , indépendantes et admettant un moment d'ordre deux. On suppose que :

$$EX_n \xrightarrow{n \rightarrow +\infty} m \quad \text{et} \quad \sum_{j=1}^{+\infty} \frac{1}{j^2} \sigma_{X_j}^2 < +\infty.$$

Alors, la suite des variables aléatoires $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ converge P-p.s. et dans L^2 vers m .

Démonstration. Il résulte du lemme de Cesàro que

$$E\bar{X}_n = \frac{1}{n} \sum_{j=1}^n EX_j \xrightarrow{n \rightarrow +\infty} m. \quad (10.11)$$

Les variables aléatoires $Y_n = \frac{X_n - EX_n}{n}$ sont indépendantes, centrées, et de variance $\frac{1}{n^2} \sigma_{X_n}^2$; on a donc

$$\sum_{j=1}^{+\infty} \sigma_{Y_j}^2 = \sum_{j=1}^{+\infty} \frac{1}{j^2} \sigma_{X_j}^2 < +\infty,$$

ce qui, en vertu du théorème 10.14 démontre la convergence P-p.s. de la série de terme général Y_n . Le lemme de Kronecker assure alors que la suite des moyennes arithmétiques des variables aléatoires nY_n converge P-p.s vers 0, et donc que **la suite des variables aléatoires \bar{X}_n converge P-p.s vers m .**

Pour la convergence L^2 , remarquons que, par indépendance des variables aléatoires $X_n - EX_n$, on a

$$\left\| \frac{1}{n} \sum_{j=1}^n (X_j - EX_j) \right\|_2^2 = \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2 ;$$

le lemme de Kronecker, conjointement à l'hypothèse, assure que

$$\lim_n \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2 = 0.$$

Puisque l'on a, pour tout $n \in \mathbb{N}^*$,

$$\bar{X}_n - m = \frac{1}{n} \sum_{j=1}^n (X_j - EX_j) + \left[\frac{1}{n} \sum_{j=1}^n EX_j - m \right],$$

la relation (10.11) et l'inégalité triangulaire conduisent à la **convergence L^2 vers m de la suite de terme général \bar{X}_n .** \square

Comme le montre l'exemple suivant, **une suite de variables aléatoires peut suivre la loi faible des grands nombres sans pour autant suivre la loi forte.**

Exemple 10.2. Soit $(X_n)_{n \geq 2}$ une suite de variables aléatoires indépendantes de lois données par

$$P_{X_n} = \frac{1}{2n \ln n} (\delta_n + \delta_{-n}) + \left(1 - \frac{1}{2n \ln n}\right) \delta_0.$$

Les X_n sont centrées : s'il y a convergence P-p.s. de $\frac{S_n}{n}$, c'est vers 0. Mais on a

$$\sum_{n=1}^{+\infty} P(|X_n| \geq n) = \sum_{n=1}^{+\infty} \frac{1}{n \ln n} = +\infty ;$$

les X_n étant indépendantes, le lemme de Borel-Cantelli assure que $P\left[\limsup_n (|X_n| > n)\right] = 1$; puisque l'on a $\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1}$, il en résulte que, P-p.s. **la suite de terme général $\frac{S_n}{n}$ ne converge pas vers 0.**

Par contre, la loi faible s'applique. En effet, on a $\sigma_{X_n}^2 = \frac{n}{\ln n}$. La fonction $x \mapsto \frac{x}{\ln x}$ étant croissante sur $[e, +\infty[$, on a la majoration

$$\sum_{k=3}^n \frac{k}{\ln k} \leq \sum_{k=3}^n \frac{n}{\ln n} = \frac{(n-2)n}{\ln n}.$$

et donc

$$0 \leq \frac{1}{n^2} \sum_{k=2}^n \sigma_{X_k}^2 \neq \frac{2}{n^2 \ln 2} + \frac{(n-2)}{n \ln n},$$

ce qui démontre que $\lim_n \frac{1}{n^2} \sum_{k=2}^n \sigma_{X_k}^2 = 0$. Par le théorème 10.17, la suite de terme général $\frac{S_n}{n}$ converge vers 0 en probabilité.

Si les variables aléatoires X_n sont seulement intégrables, on a encore une loi forte des grands nombres, à condition de rajouter une hypothèse, à savoir que les X_n sont équidistribuées. C'est l'objet du théorème suivant.

Théorème 10.21 (Théorème de Kolmogorov-Khinchine). Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , indépendantes, de même loi. Les deux assertions suivantes sont équivalentes :

(i) il existe un réel c tel que la suite des variables aléatoires $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ converge P-p.s. vers c ;

(ii) $X_1 \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$;

Si l'assertion (i) est vraie, on a $c = EX_1$.

Démonstration. On a démontré en exercice (exercice 12, chapitre 9) que, si les X_n sont indépendantes et de même loi, pour que la suite de terme général X_n/n converge P-p.s. vers 0, il faut et il suffit que X_1 soit intégrable.

• Supposons que la suite $(\bar{X}_n)_{n \in \mathbb{N}^*}$ converge P-p.s. vers c ; puisque, pour tout $n \in \mathbb{N}^*$, on a

$$\frac{X_n}{n} = \bar{X}_n - \frac{n-1}{n} \bar{X}_{n-1},$$

la suite de terme général X_n/n converge P-p.s. vers 0, et en conséquence X_1 est intégrable.

• Inversement, supposons que X_1 soit intégrable ; la suite de terme général X_n/n converge alors P-p.s. vers 0 et on a donc $P(\limsup_n (|X_n| > n)) = 0$. En introduisant pour tout n la variable aléatoire $\tilde{X}_n = \mathbf{1}_{(|X_n| \leq n)} X_n$, il en résulte que l'on a $P(\liminf_n (X_n = \tilde{X}_n)) = 1$. Si on note

$$S_n = \sum_{j=1}^n X_j \quad \text{et} \quad \tilde{S}_n = \sum_{j=1}^n \tilde{X}_j,$$

on a alors

$$\begin{aligned} P\left[\left(\frac{S_n}{n}\right)_{n \in \mathbb{N}^*} \text{ converge}\right] &= P\left[\left(\left(\frac{S_n}{n}\right)_{n \in \mathbb{N}^*} \text{ converge}\right) \cap \liminf_n (X_n = \tilde{X}_n)\right] \\ &= P\left[\left(\left(\frac{\tilde{S}_n}{n}\right)_{n \in \mathbb{N}^*} \text{ converge}\right) \cap \liminf_n (X_n = \tilde{X}_n)\right]. \end{aligned}$$

soit

$$P\left[\left(\frac{S_n}{n}\right)_{n \in \mathbb{N}^*} \text{ converge}\right] = P\left[\left(\frac{\tilde{S}_n}{n}\right)_{n \in \mathbb{N}^*} \text{ converge}\right].$$

Il suffit donc de démontrer la convergence P.-p.s. de la suite $(\frac{\tilde{S}_n}{n})_{n \in \mathbb{N}^+}$, ce qui se fait à l'aide du théorème 10.20, puisque les \tilde{X}_n sont dans \mathcal{L}^2 . Vérifions les deux conditions sur les moments :

- les X_n ayant même loi, il en est de même des \tilde{X}_n ; ces dernières ont donc même moyenne, et, X_1 étant intégrable, on a

$$\lim_n E \tilde{X}_n = \lim_n E [X_1 \mathbf{1}_{(|X_1| \leq n)}] = E X_1;$$

- on a

$$0 \leq \sigma_{\tilde{X}_n}^2 \leq E(\tilde{X}_n^2) = E[X_1^2 \mathbf{1}_{(|X_1| \leq n)}],$$

et, par la propriété de Beppo Levi,

$$0 \leq \sum_{n=1}^{+\infty} \frac{1}{n^2} \sigma_{\tilde{X}_n}^2 \leq E \left[\sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{(|X_1| \leq n)} \right]. \quad (10.12)$$

Mais, puisque $X_1^2 \mathbf{1}_{(|X_1| \leq n)} = X_1^2 \sum_{m=1}^n \mathbf{1}_{(m-1 < |X_1| \leq m)}$, on a, dans $\overline{\mathbb{R}}^+$,

$$\begin{aligned} \sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{(|X_1| \leq n)} &= \sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \left[\sum_{m=1}^n \mathbf{1}_{(m-1 < |X_1| \leq m)} \right] \\ &= \sum_{m=1}^{+\infty} |X_1| \left[\sum_{n=m}^{+\infty} |X_1| \frac{1}{n^2} \mathbf{1}_{(m-1 < |X_1| \leq m)} \right], \end{aligned}$$

ce qui donne l'inégalité :

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{(|X_1| \leq n)} \leq \sum_{m=1}^{+\infty} |X_1| \mathbf{1}_{(m-1 < |X_1| \leq m)} \left[m \sum_{n=m}^{+\infty} \frac{1}{n^2} \right].$$

De plus, on a la majoration

$$\begin{aligned} \sum_{n=m}^{+\infty} \frac{1}{n^2} &= \frac{1}{m^2} + \sum_{n=m}^{+\infty} \frac{1}{(n+1)^2} \\ &\leq \frac{1}{m^2} + \sum_{n=m}^{+\infty} \int_n^{n+1} \frac{1}{x^2} dx = \frac{1}{m^2} + \int_m^{+\infty} \frac{1}{x^2} dx. \end{aligned}$$

ce qui donne $m \sum_{n=m}^{+\infty} \frac{1}{n^2} \leq \frac{1}{m} + 1 \leq 2$. Il en résulte l'inégalité

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{(|X_1| \leq n)} \leq 2 \sum_{m=1}^{+\infty} |X_1| \mathbf{1}_{(m-1 < |X_1| \leq m)} = 2 |X_1|;$$

on a alors, d'après l'inégalité (10.12),

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} \sigma_{\tilde{X}_n}^2 \leq E \left[\sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{(|X_1| \leq n)} \right] \leq 2 E |X_1| < +\infty.$$

On a donc démontré la convergence P-p.s. de la suite $(\frac{\tilde{S}_n}{n})_{n \in \mathbb{N}^*}$ et donc aussi de la suite $(\bar{X}_n)_{n \in \mathbb{N}^*}$ vers EX_1 . \square

Une application importante du théorème précédent est le **théorème fondamental de la statistique** sur la **convergence des fonctions de répartition empiriques**.

Soit X une variable aléatoire réelle de loi μ ; soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles, indépendantes de même loi μ et de fonction de répartition F .

Définition 10.22. Le vecteur aléatoire (X_1, X_2, \dots, X_n) est appelé **échantillon de taille n de X** . La fonction F_n de $\mathbb{R} \times \Omega$ dans $[0, 1]$ définie par

$$\forall (x, \omega) \in \mathbb{R} \times \Omega \quad F_n(x, \omega) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(X_j \leq x)}(\omega)$$

est appelée **fonction de répartition empirique** (associée à X) basée sur l'échantillon (X_1, X_2, \dots, X_n) .

Remarque. Pour une réalisation ω , le vecteur $(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$ est appelé **échantillon empirique**; pour tout réel x , le nombre $nF_n(x, \omega)$ est le nombre d'indices k pour lesquels on a $X_k(\omega) \leq x$.

Théorème 10.23 (Théorème fondamental de la statistique ou théorème de Glivenko-Cantelli). Avec les notations ci-dessus, pour P-presque tout ω , la suite des fonction de répartition $F_n(\cdot, \omega)$ converge uniformément vers F , autrement dit, on a

$$\text{P-p.s.} \quad \lim_n \sup_{x \in \mathbb{R}} |F_n(x, \cdot) - F(x)| = 0.$$

Démonstration. Il faut remarquer que, pour tout ω , $F_n(\cdot, \omega)$ et F étant continues à droite, on a

$$\sup_{x \in \mathbb{R}} |F_n(x, \cdot) - F(x)| = \sup_{x \in \mathbb{Q}} |F_n(x, \cdot) - F(x)|,$$

ce qui montre que $\sup_{x \in \mathbb{R}} |F_n(x, \cdot) - F(x)|$ est bien une variable aléatoire.

Pour tout réel x , la suite $(\mathbf{1}_{(X_j \leq x)})_{j \in \mathbb{N}^*}$ (resp. $(\mathbf{1}_{(X_j < x)})_{j \in \mathbb{N}^*}$) est une suite de variables aléatoires indépendantes de même loi et intégrables. De plus, on a

$$E\mathbf{1}_{(X_j \leq x)} = P(X_j \leq x) = F(x) \quad \text{et} \quad E\mathbf{1}_{(X_j < x)} = P(X_j < x) = F(x-);$$

il résulte alors du théorème 10.21 que

$$\text{P-p.s.} \quad \lim_n F_n(x, \cdot) = F(x) \quad \text{et} \quad \lim_n F_n(x-, \cdot) = F(x-). \quad (10.13)$$

Soit alors D la réunion de l'ensemble des rationnels \mathbb{Q} et de l'ensemble, dénombrable et éventuellement vide, des points de discontinuité de F ; cet ensemble D est dénombrable et dense. De l'assertion (10.13) résulte l'existence, pour tout $x \in D$, de deux ensembles de probabilité nulle, N_x^1 et N_x^2 , tels que l'on ait

$$\forall \omega \notin N_x^1 \quad \lim_n F_n(x, \omega) = F(x) \quad \text{et} \quad \forall \omega \notin N_x^2 \quad \lim_n F_n(x-, \omega) = F(x-).$$

L'ensemble $N = [\bigcup_{x \in D} N_x^1] \cup [\bigcup_{x \in D} N_x^2]$ est encore de probabilité nulle et on a

$$\forall \omega \notin N, \quad \forall x \in D \quad \lim_n F_n(x, \omega) = F(x) \quad \text{et} \quad \lim_n F_n(x-, \omega) = F(x-).$$

En appliquant le lemme 10.24 ci-dessous, on obtient que

$$\forall \omega \notin N, \quad \lim_n \sup_{x \in \mathbb{R}} |F_n(x, \omega) - F(x)| = 0,$$

ce qui est le résultat annoncé. \square

Pour être complet, il ne reste plus qu'à énoncer et démontrer ce lemme.

Lemme 10.24. *Soient f et f_n , $n \in \mathbb{N}^*$, des fonctions définies sur \mathbb{R} positives, croissantes et bornées par 1.*

(a) *Si la suite $(f_n)_{n \in \mathbb{N}^*}$ converge simplement sur un ensemble D dénombrable et dense de \mathbb{R} , elle converge simplement sur l'ensemble des points de continuité de f .*

(b) *Supposons de plus que les fonctions f et f_n sont des **fonctions de répartition**. Soit D l'ensemble, dénombrable et dense, réunion de \mathbb{Q} et de l'ensemble des points de discontinuité de f . Si*

$$\forall x \in D \quad \lim_n f_n(x) = f(x) \quad \text{et} \quad \lim_n f_n(x-) = f(x-),$$

la suite $(f_n)_{n \in \mathbb{N}^}$ converge vers f uniformément sur \mathbb{R} .*

Démonstration. (a) Soit x un point de continuité de f . Soient $\varepsilon > 0$ quelconque et $\eta > 0$ tel que l'on ait

$$x' \in [x - \eta, x + \eta] \implies |f(x) - f(x')| \leq \varepsilon.$$

Soient y et y' appartenant à D tels que l'on ait $x - \eta < y < x < y' < x + \eta$; la croissance des fonctions f et f_n et l'hypothèse de convergence de la suite permettent d'écrire :

$$f(y) = \lim_n f_n(y) \leq \lim_n \inf_n f_n(x) \leq \lim_n \sup_n f_n(x) \leq \lim_n f_n(y') = f(y').$$

Il en résulte que, pour tout $\varepsilon > 0$, on a

$$\left| \lim_n \inf_n f_n(x) - \lim_n \sup_n f_n(x) \right| \leq \varepsilon \quad \text{et} \quad \left| \lim_n \sup_n f_n(x) - f(x) \right| \leq \varepsilon,$$

ce qui démontre, ε étant quelconque, que l'on a

$$\liminf_n f_n(x) = \limsup_n f_n(x) = f(x);$$

autrement dit, la suite $(f_n(x))_{n \in \mathbb{N}^*}$ converge vers $f(x)$.

(b) Pour tous entiers j, k tels que $1 \leq j \leq k$, posons

$$x_{j,k} = \sup \left\{ x \in \mathbb{R} \mid f(x-) \leq \frac{j}{k} \leq f(x) \right\}$$

(on convient que $\sup \emptyset = +\infty$) et $x_{0,k} = -\infty$. Puisque f est une fonction de répartition, on a $\lim_{x \rightarrow -\infty} f(x) = 0$ et $\lim_{x \rightarrow +\infty} f(x) = 1$. Il en résulte que $x_{j,k} < x_{j+1,k}$ dès que $k \in \mathbb{N}^*$ et $0 \leq j \leq k-1$. Les intervalles $[x_{j,k}, x_{j+1,k}]$ forment donc une partition de \mathbb{R} . Posons

$$\Delta_n^1(k) = \max_{0 \leq j \leq k-1} |f_n(x_{j,k}) - f(x_{j,k})|,$$

$$\Delta_n^2(k) = \max_{0 \leq j \leq k-1} |f_n(x_{j,k}-) - f(x_{j,k}-)| \text{ et } \Delta_n = \sup_{x \in \mathbb{R}} |f_n(x) - f(x)|.$$

On a, pour tout $k \in \mathbb{N}^*$,

$$\Delta_n \leq \max(\Delta_n^1(k), \Delta_n^2(k)) + \frac{1}{k}. \quad (10.14)$$

En effet, si $x \in]x_{j,k}, x_{j+1,k}[$, on a

$$\begin{aligned} f(x_{j,k}) \leq f(x) \leq f(x_{j+1,k}-) & \quad f_n(x_{j,k}) \leq f_n(x) \leq f_n(x_{j+1,k}-) \\ \text{et} \quad 0 \leq f(x_{j+1,k}-) - f(x_{j,k}) & \leq \frac{1}{k}, \end{aligned}$$

ce qui donne

$$f_n(x) - f(x) \leq f_n(x_{j+1,k}-) - f(x_{j,k}) \leq f_n(x_{j+1,k}-) - f(x_{j+1,k}-) + \frac{1}{k}$$

et

$$f_n(x) - f(x) \geq f_n(x_{j,k}) - f(x_{j+1,k}-) \geq f_n(x_{j,k}) - f(x_{j,k}) - \frac{1}{k},$$

ce qui démontre l'inégalité (10.14), puisque l'on a, pour tout $k \in \mathbb{N}^*$,

$$\Delta_n = \max_{1 \leq j \leq k} \sup_{x \in]x_{j,k}, x_{j+1,k}[} |f_n(x) - f(x)|.$$

Par ailleurs, pour tout $k \in \mathbb{N}^*$, on a $\lim_n \Delta_n^1(k) = \lim_n \Delta_n^2(k) = 0$; en effet on a

$$\lim_n f_n(x_{j,k}) = f(x_{j,k}) \quad \text{et} \quad \lim_n f_n(x_{j,k}-) = f(x_{j,k}-) = f(x_{j,k}).$$

car soit $x_{j,k}$ est point de continuité de f , et cela résulte de la première partie du lemme, soit $x_{j,k}$ est point de discontinuité de f , et c'est l'hypothèse; il

suffit alors de remarquer que dans $\Delta_n^1(k)$ et $\Delta_n^2(k)$ ne figurent qu'un nombre fini de quantités $|f_n(x_{j,k}) - f(x_{j,k})|$ ou $|f_n(x_{j,k-}) - f(x_{j,k-})|$. Il en résulte que, pour tout $k \in \mathbb{N}^*$, on a

$$0 \leq \limsup_n \Delta_n = 0,$$

ce qui démontre que la suite de terme général converge vers 0. \square

Le théorème de Glivenko-Cantelli 10.23 suggère l'idée du **test de Kolmogorov-Smirnov**; avec les notations employées dans ce théorème, il s'agit de tester, au vu d'un échantillon de taille n , l'hypothèse que la variable aléatoire X a pour fonction de répartition la fonction **continue** F . La méthode de test consiste à donner une **région d'acceptation** de l'hypothèse avec une probabilité d'erreur de α , α étant appelé **seuil** ou **niveau** du test. Ce test est **non paramétrique**, au sens où l'hypothèse consiste à dire que F appartient à une classe de fonctions, à savoir les fonctions continues, par opposition à un **test paramétrique** où on suppose que F appartient à une famille de fonctions déterminées par des paramètres (par exemple la famille de toutes les lois gaussiennes de paramètres m et σ^2) et où l'hypothèse porte sur les valeurs de ces paramètres.

Le test est basé sur la remarque que la variable aléatoire D_n , appelée statistique de Kolmogorov-Smirnov, et définie par

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x, \cdot) - F(x)|,$$

a une loi indépendante de F . Démontrons ceci : soit G la pseudo-inverse de F , à savoir la fonction définie par

$$G(y) = \inf\{x \mid F(x) \geq y\}.$$

On rappelle (voir exercice 1 du chapitre 8) que G est définie sur $[0, 1]$ et que, F étant continu, on a, pour tout $y \in [0, 1]$, $F(G(y)) = y$; de plus la loi de $F(X)$ est la loi uniforme sur l'intervalle $[0, 1]$. Enfin, puisque F est continue, en tout point x de croissance stricte⁷ de F , on a l'équivalence $F(x) \leq F(y)$ si et seulement si $x \leq y$. Si on note C l'ensemble des points de croissance stricte de F , l'ensemble C^c des paliers de F est réunion dénombrable d'intervalles $]a_i, b_i[$, $i \in I$, les paliers de F correspondant aux sauts de la fonction croissante G ; on a alors, si $1 \leq j \leq n$,

$$P(X_j \in C^c) \leq \sum_{i \in I} P[X_j \in (a_i, b_i)] = \sum_{i \in I} [F(b_i) - F(a_i)] = 0;$$

ainsi, puisque les variables aléatoires X_j sont indépendantes, on a P-p.s., $(X_1, X_2, \dots, X_n) \in C^n$. Pour des réels x_1, x_2, \dots, x_n, x , notons

7. Un point x est un **point de croissance stricte** pour la fonction F , s'il existe un intervalle ouvert contenant x sur lequel F est strictement croissante.

$$v_n(x_1, x_2, \dots, x_n, x) = \sum_{j=1}^n \mathbf{1}_{(x_j \leq x)}$$

le nombre de x_j inférieurs ou égaux à x . On a P-p.s.

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x, \cdot) - F(x)| = \sup_{x \leq G(1)} |F_n(x, \cdot) - F(x)|,$$

puisque, si $x > G(1)$, on a P-p.s. $F_n(x, \cdot) = F(x) = 1$; ainsi on a P-p.s.

$$\begin{aligned} D_n &= \sup_{x \leq G(1)} \left| \frac{1}{n} v_n(X_1, X_2, \dots, X_n, x) - F(x) \right| \\ &= \sup_{y \in [0,1]} \left| \frac{1}{n} v_n(X_1, X_2, \dots, X_n, G(y)) - F(G(y)) \right| \end{aligned}$$

c'est à dire,

$$\text{P-p.s. } D_n = \sup_{y \in [0,1]} \left| \frac{1}{n} v_n(X_1, X_2, \dots, X_n, G(y)) - y \right|;$$

mais, puisque P-p.s., $(X_1, X_2, \dots, X_n) \in C^n$, on a P-p.s.

$$\begin{aligned} v_n(X_1, X_2, \dots, X_n, G(y)) &= v_n(F(X_1), F(X_2), \dots, F(X_n), F(G(y))) \\ &= v_n(F(X_1), F(X_2), \dots, F(X_n), y). \end{aligned}$$

autrement dit,

$$\boxed{\text{P-p.s. } D_n = \sup_{y \in [0,1]} \left| \frac{1}{n} v_n(F(X_1), F(X_2), \dots, F(X_n), y) - y \right| .}$$

En remarquant que les variables aléatoires $F(X_1), F(X_2), \dots, F(X_n)$, sont indépendantes de même loi uniforme sur $[0, 1]$, on a bien montré que D_n a une loi indépendante de F . Cette loi est tabulée⁸. Le test consiste alors, pour un niveau α donné à déterminer dans cette table la valeur d_α pour laquelle on a $P(D_n \leq d_\alpha) = 1 - \alpha$. On accepte l'hypothèse que X a pour fonction de répartition F si on a, pour tout $x \in \mathbb{R}$, $|F_n(x, \cdot) - F(x)| \leq d_\alpha$, c'est à dire si le graphe de F est situé dans la bande déterminée par les graphes translatés de $\pm d_\alpha$ de la fonction de répartition empirique construite à partir de l'échantillon (x_1, x_2, \dots, x_n) .

Exemple numérique. On se demande si, au niveau 0.05, on peut accepter l'hypothèse que l'échantillon suivant, de taille 15, issu d'un générateur aléatoire, soit bien celui d'une variable aléatoire de loi uniforme sur $[0, 1]$:

0.8 0.4 0.25 0.7 0.6 0.2 0.5 0.3 0.15 0.1 0.65 0.9 0.45 0.85 0.55

8. On peut par exemple trouver une table de la loi de D_n dans le livre de Kishor S.Trivedi (1982), *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*, Prentice-Hall, Inc., Englewood Cliffs, N.J., p. 599.

La fonction de répartition empirique croît par saut de $\frac{1}{13}$ aux points d'abscisse x_i . La table donne $d_{0,05} = 0.34$. Une représentation graphique montre rapidement que la « première bissectrice est intérieure à la bande délimitée par les fonctions $F_n \pm 0.34$ » (un calcul peut d'ailleurs aussi bien le montrer). Ainsi, avec une probabilité 0.05 de se tromper, on accepte l'hypothèse que l'échantillon est bien celui d'une variable aléatoire de loi uniforme sur $[0, 1]$.

Exercices

Sauf mention du contraire, toutes les variables aléatoires sont définies sur un même espace probabilisé (Ω, \mathcal{A}, P) .

Exercice 10.1. Métriques et convergence en probabilité. Soit \mathcal{L}^0 (resp. L^0), l'ensemble des variables aléatoires (resp. des classes de variables aléatoires) définies P-p.s., et, dans le cas de \mathbb{R} , P-p.s. finies. On pose, pour tout X et Y de \mathcal{L}^0 (resp. de L^0),

$$d(X, Y) = E \left[\frac{|X - Y|}{1 + |X - Y|} \right] \quad \text{et} \quad \delta(X, Y) = E [\min(1, |X - Y|)] .$$

Démontrer que d et δ définissent sur l'espace vectoriel \mathcal{L}^0 (resp. sur L^0) deux pseudo-métriques⁹ (resp. métriques) équivalentes et telles que la convergence des suites au sens de ces métriques soit équivalente à la convergence en probabilité. En déduire que ces espaces « métriques » sont complets.

Solution. La fonction $x \mapsto \frac{x}{1+x}$ étant croissante sur \mathbb{R}^+ , il résulte de l'inégalité triangulaire que, pour tous réels positifs x, y et z , on a

$$\frac{|x - y|}{1 + |x - y|} \leq \frac{|x - z| + |z - y|}{1 + |x - z| + |z - y|} \leq \frac{|x - z|}{1 + |x - z|} + \frac{|z - y|}{1 + |z - y|} ;$$

l'inégalité triangulaire pour d en résulte par croissance de l'intégrale. De plus $d(X, Y) = 0$ si et seulement si $X = Y$ P-p.s., si bien que d , étant de plus symétrique, est un écart sur \mathcal{L}^0 .

Il résulte de l'inégalité triangulaire que, pour tous réels positifs x, y et z , on a

$$\min(1, |x - y|) \leq \min(1, |x - z| + |z - y|) \leq \min(1, |x - z|) + \min(1, |z - y|) ;$$

l'inégalité triangulaire pour δ en résulte par croissance de l'intégrale. De plus $\delta(X, Y) = 0$ si et seulement si $X = Y$ P-p.s., si bien que δ , étant de plus symétrique, est un écart sur \mathcal{L}^0 .

On vérifie facilement que l'on a, pour tout $x \geq 0$,

$$\frac{1}{2} \min(1, x) \leq \frac{x}{1+x} \leq \min(1, x) ,$$

9. Une pseudo-métrique est encore appelée **écart**.

ce qui conduit à l'encadrement

$$\frac{1}{2}\delta(X, Y) \leq d(X, Y) \leq \delta(X, Y);$$

autrement dit, **les écarts d et δ sont équivalents.**

La fonction $x \mapsto \frac{x}{1+x}$ étant croissante sur \mathbb{R}^+ et bornée par 1, on a pour tout $\varepsilon > 0$:

$$\frac{\varepsilon}{1+\varepsilon} \mathbf{P}(|X-Y| > \varepsilon) \leq d(X, Y) \leq \frac{\varepsilon}{1+\varepsilon} \mathbf{E}[\mathbf{1}_{(|X-Y| \leq \varepsilon)}] + \mathbf{E}[\mathbf{1}_{(|X-Y| > \varepsilon)}].$$

ce qui donne

$$\frac{\varepsilon}{1+\varepsilon} \mathbf{P}(|X-Y| > \varepsilon) \leq d(X, Y) \leq \varepsilon + \mathbf{P}(|X-Y| > \varepsilon);$$

il est alors clair que la convergence d'une suite au sens de la métrique d soit équivalente à la convergence en probabilité. Puisque pour qu'une suite converge en probabilité, il faut et il suffit qu'elle soit de Cauchy pour la convergence en probabilité, il en résulte que **l'espace pseudo-métrique (\mathcal{L}^0, d) est donc aussi (\mathcal{L}^0, δ) , est complet.**

Remarque. On aurait pu procéder de manière inverse, à savoir, montrer préalablement que (\mathcal{L}^0, d) est complet, puis en déduire que pour qu'une suite converge en probabilité il faut et il suffit qu'elle soit de Cauchy pour la convergence en probabilité. Montrons directement que (\mathcal{L}^0, d) est complet: si $(X_n)_{n \in \mathbb{N}}$ est une suite de Cauchy pour d , on extrait une sous-suite $(X_{n_k})_{k \in \mathbb{N}}$ telle que $\sum_{k=0}^{+\infty} d(X_{n_k}, X_{n_{k+1}}) < +\infty$, c'est à dire, par la propriété de Beppo Levi, que

$$\mathbf{E}\left[\sum_{k=0}^{+\infty} \frac{|X_{n_k} - X_{n_{k+1}}|}{1 + |X_{n_k} - X_{n_{k+1}}|}\right] < +\infty.$$

Il en résulte que P-p.s. $\sum_{k=0}^{+\infty} \frac{|X_{n_k} - X_{n_{k+1}}|}{1 + |X_{n_k} - X_{n_{k+1}}|} < +\infty$; mais, puisqu'alors on a

P-p.s. $\lim_k |X_{n_k} - X_{n_{k+1}}| = 0$, on a encore P-p.s. $\sum_{k=0}^{+\infty} |X_{n_k} - X_{n_{k+1}}| < +\infty$. Par conséquent, la sous-suite $(X_{n_k})_{k \in \mathbb{N}}$ converge P-p.s. et, par le théorème de convergence dominée, il y a aussi convergence pour la métrique d ; mais toute suite de Cauchy qui admet une sous-suite convergente est elle-même convergente.

Exercice 10.2. Une réciproque partielle au théorème 10.3. Démontrer que si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes qui converge P-p.s. vers 0, on a, pour tout $\varepsilon > 0$,

$$\sum_{n=0}^{+\infty} \mathbf{P}[(|X_n| > \varepsilon)] < +\infty.$$

Solution. L'ensemble $C = \bigcap_{\varepsilon > 0} \liminf_n (|X_n| \leq \varepsilon)$ est l'ensemble des ω pour lesquels la suite de terme général $X_n(\omega)$ converge vers 0; il est de probabilité 1. Par conséquent, pour tout $\varepsilon > 0$, on a $\mathbf{P}[\liminf_n (|X_n| \leq \varepsilon)] = 1$, soit encore,

$P[\limsup_n (|X_n| > \varepsilon)] = 0$. Les X_n étant indépendantes, le lemme de Borel-Cantelli assure que pour tout $\varepsilon > 0$,

$$\sum_{n=0}^{+\infty} P[(|X_n| > \varepsilon)] < +\infty.$$

Exercice 10.3. Équi-intégrabilité et convergence \mathcal{L}^p de suites de variables aléatoires gaussiennes. Soit X une variable aléatoire réelle de loi gaussienne $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$.

1. Démontrer l'inégalité

$$E[\exp |X|] \leq 2 \operatorname{ch} m \exp\left(\frac{\sigma^2}{2}\right). \quad (10.15)$$

2. Soient I un ensemble quelconque et, pour tout $i \in I$, une variable aléatoire X_i de loi gaussienne $\mathcal{N}_{\mathbb{R}}(m_i, \sigma_i^2)$. Démontrer que si les familles de réels $(m_i)_{i \in I}$ et $(\sigma_i^2)_{i \in I}$ sont bornées, les familles de variables aléatoires $(|X_i|^p)_{i \in I}$ sont équi-intégrables pour tout $p \geq 1$.

3. Si de plus $I = \mathbb{N}$ et si la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers une variable aléatoire X , démontrer que X est gaussienne et que la convergence a lieu dans tout \mathcal{L}^p , $p \geq 1$.

Solution.

1. On a

$$\begin{aligned} E[\exp X] &= \int_{\mathbb{R}} \exp(x) \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \\ &= \exp\left(m + \frac{\sigma^2}{2}\right) \int_{\mathbb{R}} \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x - (m + \sigma^2))^2\right] dx. \end{aligned}$$

soit

$$E[\exp X] = \exp\left(m + \frac{\sigma^2}{2}\right).$$

Puisque $-X$ est de loi $\mathcal{N}_{\mathbb{R}}(-m, \sigma^2)$, il en résulte que l'on a

$$E[\exp |X|] \leq E[\exp X] + E[\exp -X] \leq \exp\left(m + \frac{\sigma^2}{2}\right) + \exp\left(-m + \frac{\sigma^2}{2}\right),$$

ce qui donne la formule (10.15).

2. Soit $p \geq 1$. Il existe $M > 0$ tel que l'on ait, pour tout $x \geq M$, $|x|^p \leq \exp\left(\frac{x}{2}\right)$, si bien que, pour tout i , on a la majoration

$$|X_i|^p \leq M^p \mathbf{1}_{(|X_i| \leq M)} + \mathbf{1}_{(|X_i| > M)} \exp\left(\frac{|X_i|}{2}\right).$$

Il en résulte que, pour tout $A \in \mathcal{A}$, on a, par l'inégalité de Schwarz,

$$\int_A |X_i|^p dP \leq M^p P(A) + [P(A)]^{\frac{1}{2}} E \exp(|X_i|).$$

Il résulte de (10.15) et du fait que les familles de réels $(m_i)_{i \in I}$ et $(\sigma_i^2)_{i \in I}$ sont bornées que $\sup_{i \in I} E \exp(|X_i|) < +\infty$, ce qui permet de conclure que la famille $(|X_i|^p)_{i \in I}$ est équi-intégrable.

3. La suite $(X_n^p)_{n \in \mathbb{N}}$ est alors équi-intégrable pour tout $p \geq 1$. Ainsi, puisque la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X , elle converge dans tout \mathcal{L}^p ; en particulier, en prenant $p = 1, 2$, il en résulte que les suites de terme général m_n et σ_n sont convergentes et que

$$\lim_n m_n = EX \quad \text{et} \quad \lim_n \sigma_n^2 = \sigma_X^2.$$

Enfin, soit $f \in \mathcal{C}_K^+(\mathbb{R})$ quelconque; on a

$$Ef(X_n) = \int_{\mathbb{R}} f(x) \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left(-\frac{(x-m_n)^2}{2\sigma_n^2}\right) dx,$$

soit, par le changement de variables $y = \frac{x-m_n}{\sigma_n}$,

$$Ef(X_n) = \int_{\mathbb{R}} f(\sigma_n y + m_n) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy.$$

La fonction f étant bornée, le théorème de convergence dominée assure que

$$\lim_n Ef(X_n) = \int_{\mathbb{R}} f(\sigma y + m) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy,$$

soit encore, par le changement de variables $y = \frac{x-m}{\sigma}$,

$$\lim_n Ef(X_n) = \int_{\mathbb{R}} f(x) \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

Puisque f est continue et bornée, la suite de terme général $f(X_n)$, d'une part converge en probabilité vers $f(X)$, et d'autre part, est équi-intégrable; elle est donc aussi convergente dans \mathcal{L}^1 , ce qui démontre que $\lim_n Ef(X_n) = Ef(X)$, et donc que:

$$Ef(X) = \int_{\mathbb{R}} f(x) \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

Il en résulte que, f étant quelconque, la loi de X est la loi $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$.

Exercice 10.4. Une condition nécessaire à la convergence P-p.s. de séries de variables aléatoires uniformément bornées. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires centrées et indépendantes, P-p.s. uniformément bornée par $c > 0$. On note $S_n = \sum_{k=0}^n X_k$; pour un entier $l \geq 1$ fixé, on introduit les ensembles

$$A = \left\{ \sup_{n \in \mathbb{N}} |S_n| \leq l \right\} \quad \text{et} \quad A_p = \left\{ \sup_{0 \leq n \leq p} |S_n| \leq l \right\}.$$

1. Démontrer que l'on a l'inégalité

$$E \left[\mathbf{1}_{A_p} S_{p+1}^2 \right] \geq E \left[\mathbf{1}_{A_p} S_p^2 \right] + P(A) \sigma_{X_{p+1}}^2. \quad (10.16)$$

2. En déduire que la condition $P(\sup_n |S_n| < +\infty) > 0$ implique que l'on a $\sum_{n=0}^{+\infty} \sigma_{X_n}^2 < +\infty$: en particulier il en est ainsi si la série de terme général X_n converge P.-s.

Remarque. C'est une réciproque partielle du théorème 10.14.

Solution.

1. On a

$$\begin{aligned} E[\mathbf{1}_{A_p} S_{p+1}^2] &= E[\mathbf{1}_{A_p} (S_p + X_{p+1})^2] \\ &= E[\mathbf{1}_{A_p} S_p^2] + 2E[\mathbf{1}_{A_p} S_p X_{p+1}] + E[\mathbf{1}_{A_p} X_{p+1}^2]. \end{aligned}$$

et, les variables aléatoires $\mathbf{1}_{A_p} S_p$ et X_{p+1} étant indépendantes,

$$E[\mathbf{1}_{A_p} S_{p+1}^2] = E[\mathbf{1}_{A_p} S_p^2] + 2E[\mathbf{1}_{A_p} S_p]E[X_{p+1}] + E[\mathbf{1}_{A_p} X_{p+1}^2].$$

Puisque X_{p+1} est centrée et que $\mathbf{1}_{A_p}$ et X_{p+1} sont indépendantes, on a alors

$$E[\mathbf{1}_{A_p} S_{p+1}^2] = E[\mathbf{1}_{A_p} S_p^2] + E[\mathbf{1}_{A_p}] E[X_{p+1}^2];$$

il ne reste plus qu'à remarquer que $A \subset A_p$ pour obtenir l'inégalité (10.16).

2. Puisque l'ensemble $\{\sup_n |S_n| < +\infty\}$ est réunion croissante de la suite des ensembles $\{\sup_n |S_n| \leq l\}$, $l \in \mathbb{N}^*$, on a

$$P(\sup_n |S_n| < +\infty) = \lim_l P(\sup_n |S_n| \leq l);$$

on peut donc choisir un entier l tel que $P(\sup_n |S_n| \leq l) > 0$, c'est à dire, avec les notations ci-dessus tel que $P(A) > 0$. Il résulte alors de l'inégalité (10.16) et de l'inclusion $A_p \supset A_{p+1}$ que

$$P(A) \sigma_{X_{p+1}}^2 \leq E[\mathbf{1}_{A_p \setminus A_{p+1}} S_{p+1}^2] + E[\mathbf{1}_{A_{p+1}} S_{p+1}^2] - E[\mathbf{1}_{A_p} S_p^2].$$

Mais, la suite $(X_n)_{n \in \mathbb{N}}$ étant P.-s. uniformément bornée par $c > 0$, on a, sur $A_p \setminus A_{p+1}$,

$$|S_{p+1}| \leq |S_p| + |X_{p+1}| \leq l + c,$$

ce qui conduit à l'inégalité

$$P(A) \sigma_{X_{p+1}}^2 \leq (l + c)^2 P(A_p \setminus A_{p+1}) + E[\mathbf{1}_{A_{p+1}} S_{p+1}^2] - E[\mathbf{1}_{A_p} S_p^2].$$

En sommant membre à membre, il vient, pour tout $n \geq 2$,

$$P(A) \sum_{p=1}^{n-1} \sigma_{X_{p+1}}^2 \leq (l + c)^2 + E[\mathbf{1}_{A_n} S_n^2],$$

et donc, par définition de A_n ,

$$P(A) \sum_{p=1}^{n-1} \sigma_{X_{p+1}}^2 \leq (l + c)^2 + l^2.$$

Puisque $P(A) > 0$, il en résulte que $\sum_{n=0}^{+\infty} \sigma_{X_n}^2 < +\infty$. En particulier, si la série de terme général X_n converge P-p.s. on a $P(\sup_n |S_n| < +\infty) > 0$ et donc $\sum_{n=0}^{+\infty} \sigma_{X_n}^2 < +\infty$.

Remarque. Comme le montre l'exemple ci-après, l'hypothèse de bornitude est nécessaire. Si les $X_n, n \in \mathbb{N}^*$, sont indépendantes et de loi $\frac{1}{2n^3}(\delta_n + \delta_{-n}) + (1 - \frac{1}{n^3})\delta_0$, elles sont centrées. On a $E|X_n| = \frac{1}{n^2}$, si bien que l'on a $E[\sum_{n=1}^{+\infty} |X_n|] < +\infty$ et donc $P[\sum_{n=1}^{+\infty} |X_n| < +\infty] = 1$. De plus on a $\sigma_{X_n}^2 = EX_n^2 = \frac{1}{n}$, si bien que $\sum_{n=0}^{+\infty} \sigma_{X_n}^2 = +\infty$; mais la suite n'est pas P-p.s. uniformément bornée par une constante $c > 0$!

Exercice 10.5. Théorème des trois séries de Kolmogorov. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes. On note $S_n = \sum_{k=0}^n X_k$ et $Y_n = \mathbf{1}_{(|X_n| \leq c)} X_n$, où c est un réel strictement positif quelconque. Démontrer que pour la série $\sum X_n$ converge P-p.s., il faut et il suffit que les trois séries $\sum EY_n, \sum \sigma_{Y_n}^2$ et $\sum P(|X_n| > c)$ convergent.

Pour la condition nécessaire, on se ramènera au cas de variables aléatoires centrées par le procédé de symétrisation (cf. exercice 13, chapitre 9) et on utilisera l'exercice 10.4 ci-dessus.

Solution. Si la série $\sum X_n$ converge P-p.s., la suite de terme général X_n converge P-p.s. vers 0; on a donc

$$P(\liminf_n (|X_n| \leq c)) = 1, \quad (10.17)$$

ou encore $P(\limsup_n (|X_n| > c)) = 0$. Les événements $(|X_n| > c)$ étant indépendants, il résulte alors du lemme de Borel-Cantelli que l'on a $\sum_{n=0}^{+\infty} P(|X_n| > c) < +\infty$.

Par ailleurs, l'égalité (10.17) s'écrit encore $P(\liminf_n (X_n = Y_n)) = 1$. Il en résulte que la série $\sum Y_n$ converge P-p.s. Soit, sur l'espace probabilisé produit, la symétrisée Y_n^s de Y_n (cf. exercice 13 chapitre 9). Les Y_n^s sont indépendantes et centrées; puisque la série $\sum Y_n$ converge P-p.s., il résulte du théorème de Fubini que la série $\sum Y_n^s$ converge P \otimes P-p.s. Les variables Y_n^s étant de plus bornées par $2c$, il résulte de l'exercice 10.4 que $\sum_{n=0}^{+\infty} \sigma_{Y_n^s}^2 < +\infty$. Comme $\sigma_{Y_n^s}^2 = 2\sigma_{Y_n}^2$, on a aussi $\sum_{n=0}^{+\infty} \sigma_{Y_n}^2 < +\infty$.

Enfin, les variables aléatoires indépendantes et centrées $\overset{0}{Y}_n = Y_n - EY_n$ vérifient $\sum_{n=0}^{+\infty} \sigma_{\overset{0}{Y}_n}^2 < +\infty$. Il en résulte que la série $\sum \overset{0}{Y}_n$ converge P-p.s.; comme de plus la série $\sum Y_n$ converge P-p.s., la série $\sum EY_n$ converge.

Inversement, si les trois séries $\sum EY_n, \sum \sigma_{Y_n}^2$ et $\sum P(|X_n| > c)$ convergent, on a $\sum_{n=0}^{+\infty} \sigma_{\overset{0}{Y}_n}^2 < +\infty$, et la série $\sum \overset{0}{Y}_n$ converge donc P-p.s.; il en est alors de même de la série $\sum Y_n$. De plus, puisque l'on a

$$\sum_{n=0}^{+\infty} P(X_n \neq Y_n) = \sum_{n=0}^{+\infty} P(|X_n| > c) < +\infty,$$

il résulte du lemme de Borel-Cantelli que $P(\limsup_n (X_n \neq Y_n)) = 0$, ce qui s'écrit encore $P(\liminf_n (X_n = Y_n)) = 1$. **La convergence P-p.s. de la série $\sum X_n$ en résulte.**

Exercice 10.6. Il n'y a pas de lemme de Césaro pour la convergence en probabilité.

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes, X_n étant de fonction de répartition F_n définie par

$$F_n(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - \frac{1}{x+n} & \text{si } x > 0 \end{cases}$$

On note $S_n = \sum_{k=1}^n X_k$ et $Y_n = \frac{S_n}{n}$. Démontrer que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge vers 0 en probabilité, mais que la suite $(Y_n)_{n \in \mathbb{N}^*}$ ne converge pas vers 0 en probabilité.

Solution. Les variables aléatoires X_n sont P-p.s. positives. On a, pour tout $\varepsilon > 0$,

$$P(|X_n| > \varepsilon) = 1 - F_n(\varepsilon) = \frac{1}{\varepsilon + n},$$

et donc $\lim_n P(|X_n| > \varepsilon) = 0$, c'est à dire que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge vers 0 en probabilité.

Par ailleurs, si $M_n = \max_{1 \leq k \leq n} X_k$, les X_n étant P-p.s. positives, on a $\frac{M_n}{n} \leq Y_n$ P-p.s.; on a donc, pour tout $\varepsilon > 0$,

$$P\left(\varepsilon < \frac{M_n}{n}\right) \leq P(\varepsilon < Y_n).$$

Les X_n étant indépendantes, on a, pour tout $x > 0$,

$$\begin{aligned} P(M_n \leq x) &= P\left[\bigcap_{k=1}^n (X_k \leq x)\right] = \prod_{k=1}^n P[(X_k \leq x)] \\ &= \prod_{k=1}^n \left(1 - \frac{1}{x+k}\right) < \left(1 - \frac{1}{x+n}\right)^n. \end{aligned}$$

Il en résulte que

$$1 - \left(1 - \frac{1}{n(\varepsilon+1)}\right)^n \leq P\left(\varepsilon < \frac{M_n}{n}\right) \leq P(\varepsilon < Y_n),$$

et donc, en passant à la limite inférieure,

$$0 < 1 - \exp\left(-\frac{1}{(\varepsilon+1)}\right) \leq \liminf_n P(\varepsilon < Y_n),$$

ce qui prouve que la suite $(Y_n)_{n \in \mathbb{N}^*}$ ne converge pas vers 0 en probabilité.

Exercice 10.7. Le théorème 10.20 donne une condition suffisante mais non nécessaire à la loi forte des grands nombres. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes, X_n étant de loi $P_{X_n} = \frac{1}{2}\left(1 - \frac{1}{2^n}\right)(\delta_1 + \delta_{-1}) +$

$\frac{1}{2^{n+1}}(\delta_{2^n} + \delta_{-2^n})$. Démontrer que $\sum_{n=1}^{+\infty} \frac{1}{n^2} \sigma_{X_n}^2 < +\infty$ et que, toutefois, la suite $(X_n)_{n \in \mathbb{N}^*}$ satisfait à la loi des grands nombres.

Solution. La loi de X_n étant symétrique, on a $EX_n = 0$. On a donc

$$\sigma_{X_n}^2 = EX_n^2 = 1 - \frac{1}{2^n} + (2^n)^2 \frac{1}{2^n} = 1 - \frac{1}{2^n} + 2^n.$$

ce qui prouve que $\sum_{n=1}^{+\infty} \frac{1}{n^2} \sigma_{X_n}^2 < +\infty$.

Soit par ailleurs $Y_n = \mathbf{1}_{(|X_n| \leq 1)} X_n$; Y_n est à valeurs 0 ou ± 1 et on a $P(Y_n = +1) = P(X_n = \pm 1) = 1 - \frac{1}{2^n}$ et $P(Y_n = 0) = P(|X_n| = 2^n) = \frac{1}{2^n}$, et donc

$$P(X_n \neq Y_n) = P(|X_n| = 2^n) = \frac{1}{2^n},$$

ce qui prouve que

$$\sum_{n=1}^{+\infty} P(X_n \neq Y_n) < +\infty.$$

Le lemme de Borel-Cantelli assure alors que

$$P\left[\limsup_n (X_n \neq Y_n)\right] = 0,$$

soit encore

$$P\left[\liminf_n (X_n = Y_n)\right] = 1. \quad (10.18)$$

De plus, on a

$$\sigma_{Y_n}^2 = EY_n^2 = 1 - \frac{1}{2^n}, \text{ et donc } \sum_{n=1}^{+\infty} \frac{1}{n^2} \sigma_{Y_n}^2 < +\infty.$$

Il résulte alors du théorème 10.20 que la suite de terme général $\frac{1}{n} \sum_{j=1}^n Y_j$ converge P.-p.s. vers 0. D'après (10.18), la suite de terme général $\frac{1}{n} \sum_{j=1}^n X_j$ converge aussi P.-p.s. vers 0.

Exercice 10.8. Une application de la loi forte des grands nombres : la méthode de Monte-Carlo pour le calcul d'intégrales. Soient D un domaine de \mathbb{R}^d et f une fonction réelle définie sur \mathbb{R}^d mesurable, tels que $\mathbf{1}_D \cdot f$ soit Lebesgue-intégrable. Soit $(U_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes, de loi uniforme sur $[0, 1]$. On définit, pour tout n , la variable aléatoire \underline{U}_n à valeurs dans \mathbb{R}^d par $\underline{U}_n = (U_{nd+1}, U_{nd+2}, \dots, U_{(n+1)d})$ et la variable aléatoire réelle $X_n = (\mathbf{1}_D \cdot f) \circ \underline{U}_n$. Démontrer que la suite de terme général $S_n = \frac{1}{n} \sum_{j=1}^n X_j$ converge P.-p.s. vers l'intégrale $\mathbf{1} = \int_D \mathbf{1}_{[0,1]^d} f(x) dx$ et que, si f est bornée par $c > 0$, on a, pour tout $\varepsilon > 0$,

$$P(|S_n - \mathbf{1}| > \varepsilon) \leq \frac{c^2}{n\varepsilon^2}. \quad (10.19)$$

Solution. Les variables aléatoires \underline{U}_n sont indépendantes; il en est donc de même des X_n . De plus, les X_n sont de même loi et, par le théorème de transfert et l'hypothèse que $\mathbf{1}_D \cdot f$ est Lebesgue-intégrable, elles admettent une moyenne. Le théorème de loi forte 10.21 s'applique. Il reste à calculer la moyenne de X_1 . Le théorème de transfert donne

$$EX_1 = \int_{E^d} (\mathbf{1}_D \cdot f)(x) dP_{\underline{U}_n}(x),$$

et, puisque \underline{U}_n est de loi uniforme sur $[0, 1]^d$, il vient :

$$EX_1 = \int_{E^d} (\mathbf{1}_D \cdot f)(x) \mathbf{1}_{[0,1]^d}(x) dx = 1;$$

ainsi

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{\text{P.p.s.}} \int_{D \cap [0,1]^d} f(x) dx.$$

Si f est bornée par $c > 0$, puisque $D \cap [0, 1]^d$ est un domaine borné, les variables aléatoires X_n sont dans \mathcal{L}^2 ; puisque elles ont même loi et sont indépendantes, on a

$$ES_n = EX_1 = 1 \quad \text{et} \quad \sigma_{S_n}^2 = \frac{\sigma_{X_1}^2}{n}.$$

L'inégalité de Tchebitchev appliquée à S_n et la majoration

$$\sigma_{X_1}^2 = E[X_1^2] - [EX_1]^2 \leq E[X_1^2] \leq c^2$$

donnent l'inégalité (10.19).

Remarque. En dimension 1 et pour des fonctions bien régulières, cette méthode ne peut rivaliser avec les méthodes classiques d'analyse numérique; par contre elle devient utile si la fonction est très irrégulière (on n'a demandé que la mesurabilité) ou si $d \geq 2$. On peut aussi améliorer la majoration (10.19) avec une inégalité du type Bernstein.

Exercice 10.9. Inégalité d'Ottaviani. Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes. On note, pour tout k tel que $1 \leq k \leq n$,

$$S_k = \sum_{i=1}^k X_i \quad \text{et} \quad M_k = \max_{1 \leq i \leq k} |S_i|,$$

et si $0 \leq k \leq n-1$,

$$S_{k,n} = \sum_{i=k+1}^n X_i.$$

Pour tout $\varepsilon > 0$, on introduit les ensembles $E = (M_n > 2\varepsilon)$, $F_k = (|S_k| > 2\varepsilon)$ et, si $2 \leq k \leq n$,

$$E_k = (|S_k| > 2\varepsilon) \cap \left[\bigcap_{i=1}^{k-1} (|S_i| \leq 2\varepsilon) \right].$$

Démontrer l'inégalité

$$\mathbf{P}(|S_n| > \varepsilon) \geq \sum_{k=1}^n \mathbf{P}(|S_{k,n}| \leq \varepsilon) \cap E_k$$

et en déduire l'**inégalité d'Ottaviani** :

$$\min_{1 \leq k \leq n} \mathbf{P}(|S_{k,n}| \leq \varepsilon) \mathbf{P}(|M_n| > 2\varepsilon) \leq \mathbf{P}(|S_n| > \varepsilon).$$

Solution. Les ensembles E_k formant une partition de Ω , on a

$$(|S_n| > \varepsilon) \supset (|S_n| > \varepsilon) \cap \Omega = \bigcup_{k=1}^n [(|S_n| > \varepsilon) \cap E_k]. \quad (10.20)$$

Puisque $S_n = S_k + S_{k,n}$, on a

$$|S_k| > 2\varepsilon \text{ et } |S_{k,n}| \leq \varepsilon \implies |S_n| > \varepsilon,$$

car sinon on aurait $|S_k| \leq |S_n| + |S_{k,n}| \leq 2\varepsilon$ et il y aurait contradiction avec $|S_k| > 2\varepsilon$. On a donc

$$(|S_n| > \varepsilon) \cap E_k \supset (|S_{k,n}| \leq \varepsilon) \cap E_k,$$

et par conséquent, d'après (10.20),

$$\mathbf{P}(|S_n| > \varepsilon) \geq \sum_{k=1}^n \mathbf{P}(|S_{k,n}| \leq \varepsilon) \cap E_k.$$

Les ensembles $(|S_{k,n}| \leq \varepsilon)$ et E_k étant indépendants, on a alors

$$\mathbf{P}(|S_n| > \varepsilon) \geq \sum_{k=1}^n \mathbf{P}(|S_{k,n}| \leq \varepsilon) \mathbf{P}(E_k) \geq \min_{1 \leq k \leq n} \mathbf{P}(|S_{k,n}| \leq \varepsilon) \sum_{k=1}^n \mathbf{P}(E_k);$$

en tenant compte de l'égalité $\sum_{k=1}^n \mathbf{P}(E_k) = \mathbf{P}(\Omega)$, on obtient l'inégalité d'Ottaviani.

Remarque. Contrairement à l'inégalité de Kolmogorov, cette inégalité ne nécessite l'existence d'aucun moment pour les variables aléatoires.

Exercice 10.10. Équivalence des convergences en probabilité et P.-p.s. de séries de variables aléatoires indépendantes (théorème de Lévy). Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes. Démontrer que si la série $\sum X_n$ de terme général X_n converge en probabilité, elle converge P.-p.s. (on utilisera l'inégalité d'Ottaviani démontrée à l'exercice 10.9, chapitre 10).

Solution. Pour $m \in \mathbb{N}^*$, notons

$$S_m = \sum_{i=1}^m X_i, \quad A_m = \sup_{k \in \mathbb{N}^*} |S_{m+k} - S_m| \quad \text{et} \quad A = \inf_{m \in \mathbb{N}^*} A_m.$$

Il résulte du critère de Cauchy pour les séries numériques que l'on a

$$\left\{ \sum X_n \text{ converge} \right\} = \{A = 0\}.$$

Mais on a

$$\{A \neq 0\} = \bigcup_{\varepsilon \in \mathbb{Q}^{+*}} \{A > \varepsilon\}$$

et, pour tout $n \in \mathbb{N}^*$,

$$\{A > \varepsilon\} \subset \bigcap_{m \in \mathbb{N}^*} \{A_m > \varepsilon\},$$

ce qui donne l'inclusion :

$$\{A \neq 0\} \subset \bigcup_{\varepsilon \in \mathbb{Q}^{+*}} \bigcap_{m \in \mathbb{N}^*} \{A_m > \varepsilon\}. \quad (10.21)$$

Puisque $\sup_{k \in \mathbb{N}^*} |S_{m+k} - S_m| = \lim_r \nearrow \sup_{1 \leq k \leq r} |S_{m+k} - S_m|$, la suite d'ensembles $\{\sup_{1 \leq k \leq r} |S_{m+k} - S_m| > \varepsilon\}$ est croissante en r , et on a

$$\{A_m > \varepsilon\} = \bigcup_{r \in \mathbb{N}^*} \left\{ \sup_{1 \leq k \leq r} |S_{m+k} - S_m| > \varepsilon \right\}. \quad (10.22)$$

Remarquant que $S_{m+k} - S_m = \sum_{j=1}^k X_{j+m}$, on applique l'inégalité d'Ottaviani à la suite $(X_{j+m})_{j \in \mathbb{N}^*}$, ce qui donne l'inégalité :

$$\begin{aligned} \min_{1 \leq k \leq r} \mathbb{P}(|S_{m+r} - S_{m+k}| \leq \varepsilon) \mathbb{P}\left(\max_{1 \leq k \leq r} |S_{m+k} - S_m| > 2\varepsilon\right) \\ \leq \mathbb{P}(|S_{r+m} - S_m| > \varepsilon). \end{aligned} \quad (10.23)$$

La série $\sum X_n$ convergeant en probabilité, la suite de terme général S_m est de Cauchy en probabilité. Pour $\eta > 0$ donné, on peut alors choisir un entier $N_{\varepsilon, \eta}$ tel que l'on ait, pour tout $m \geq N_{\varepsilon, \eta}$,

$$\mathbb{P}(|S_{m+r} - S_{m+k}| > \varepsilon) \leq \eta \text{ dès que } 0 \leq k \leq r, \quad (10.24)$$

soit encore :

$$1 - \eta \leq \mathbb{P}(|S_{m+r} - S_{m+k}| \leq \varepsilon) \text{ dès que } 0 \leq k \leq r.$$

Pour un tel choix, on a donc

$$1 - \eta \leq \min_{1 \leq k \leq r} \mathbb{P}(|S_{m+r} - S_{m+k}| \leq \varepsilon),$$

et, par les inégalités (10.23) puis (10.24),

$$\mathbb{P}\left(\max_{1 \leq k \leq r} |S_{m+k} - S_{m+r}| > 2\varepsilon\right) \leq \frac{1}{1 - \eta} \mathbb{P}(|S_{m+r} - S_m| > \varepsilon) \leq \frac{\eta}{1 - \eta}.$$

L'égalité (10.22) faisant intervenir une suite croissante d'ensembles, il vient

$$\mathbb{P}(A_m > \varepsilon) = \lim_r \mathbb{P}\left(\sup_{1 \leq k \leq r} |S_{m+k} - S_m| > \varepsilon\right) \leq \frac{\eta}{1 - \eta}.$$

Il en résulte que, pour tout $m \geq N_{\varepsilon, \eta}$, on a

$$0 \leq \mathbb{P}\left[\bigcap_{p \in \mathbb{N}^*} (A_p > \varepsilon)\right] \leq \mathbb{P}(A_m > \varepsilon) \leq \frac{\eta}{1 - \eta}.$$

En conclusion, on a montré que, pour tout $\eta > 0$, on a

$$0 \leq \mathbb{P}\left[\bigcap_{p \in \mathbb{N}^*} (A_p > \varepsilon)\right] \leq \frac{\eta}{1 - \eta}.$$

ce qui démontre que, pour tout $\varepsilon > 0$, on a $\mathbb{P}\left[\bigcap_{p \in \mathbb{N}^*} (A_p > \varepsilon)\right] = 0$. Il résulte alors de l'inclusion (10.21) que $\mathbb{P}(A \neq 0) = 0$, c'est à dire que **la série de terme général X_n converge P-p.s.**

Remarque. La réciproque étant toujours vraie, on a bien **équivalence** des convergences en probabilité et P-p.s. de **séries** de variables aléatoires **indépendantes**.

Exercice 10.11. Inégalité de type grandes déviations : inégalité de Hoeffding.

1. Soit X une variable aléatoire réelle P-p.s. bornée par 1, c'est à dire telle que $|X| \leq 1$ P-p.s. : on suppose que X est centrée.

(a) Soit un réel t quelconque. Justifier l'inégalité de convexité suivante :

$$\forall x \in [-1, 1] \quad \exp(tx) \leq \frac{1}{2}(1 - x) \exp(-t) + \frac{1}{2}(1 + x) \exp(t)$$

(b) Après avoir justifié l'existence de la moyenne de la variable aléatoire $\exp(tX)$, en déduire l'inégalité¹⁰

$$\mathbb{E} \exp(tX) \leq \frac{1}{2}(\exp(-t) + \exp(t)).$$

Démontrer alors l'inégalité

$$\mathbb{E} \exp(tX) \leq \exp\left(\frac{t^2}{2}\right); \quad (10.25)$$

(ou pourra comparer les termes généraux des développements en série entière des fonctions concernées).

2. On considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires réelles indépendantes, bornées P-p.s. et centrées : on suppose que $|X_n| \leq c_n$ P-p.s., avec $c_n > 0$. On note, pour tout $n \in \mathbb{N}^*$, $S_n = \sum_{j=1}^n X_j$.

(a) Démontrer que, pour tout t , on a

$$\mathbb{E} \exp(tS_n) \leq \exp\left(\frac{t^2}{2} \sum_{j=1}^n c_j^2\right). \quad (10.26)$$

¹⁰ La fonction $t \mapsto \mathbb{E} \exp(tX)$ est appelée **transformée de Laplace** ou **fonction génératrice** de la variable aléatoire X .

(b) Dédurre alors de l'inégalité de Markov que, pour tout $t > 0$ et tout $\varepsilon > 0$, on a

$$P(S_n > \varepsilon) \leq \exp\left(-t\varepsilon + \frac{t^2}{2} \sum_{j=1}^n c_j^2\right). \quad (10.27)$$

(c) En minimisant en t le second membre de l'inégalité (10.32), en déduire que, pour tout $\varepsilon > 0$, on a l'inégalité

$$P(S_n > \varepsilon) \leq \exp\left(-\frac{\varepsilon^2}{2 \sum_{j=1}^n c_j^2}\right). \quad (10.28)$$

(d) Démontrer alors que, pour tout $\varepsilon > 0$, on a l'inégalité de **Hoeffding** :

$$P(|S_n| > \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{2 \sum_{j=1}^n c_j^2}\right). \quad (10.29)$$

(e) Soit $\alpha > 0$. On suppose que la suite de terme général c_n est telle que $\sum_{j=1}^n c_j^2 \leq n^{2\alpha-\beta}$, où $\beta > 0$.

Démontrer que, pour tout $\varepsilon > 0$, la série de terme général $P(|S_n| > n^\alpha \varepsilon)$ est convergente. En déduire que

$$P\left[\bigcup_{\varepsilon \in \mathbb{Q}^{+*}} \limsup_n (|S_n| > n^\alpha \varepsilon)\right] = 0;$$

que dire alors de la convergence P.-p.s. de la suite de terme général $n^{-\alpha} S_n$?

3. On suppose de plus que les X_n ont la même loi triangulaire, ou plus précisément qu'elles admettent la densité g définie par

$$g(x) = \mathbf{1}_{[0,1]}(|x|) (1 - |x|).$$

(a) Calculer, pour tout réel t , $E \exp(tX_1)$.

(b) Démontrer que l'application $\Phi : t \mapsto E \exp(tX_1)$ est indéfiniment dérivable et que l'on a, pour tout $k \in \mathbb{N}^*$,

$$\Phi^{(k)}(0) = E(X_1^k).$$

En déduire la variance de S_n .

(c) En utilisant les résultats de la deuxième question, démontrer que, pour tout $\alpha > \frac{1}{2}$, on a $\lim_n n^{-\alpha} S_n = 0$ P.-p.s.

Solution.

1. (a) Soit t un réel quelconque. On remarque que, pour tout x tel que $|x| \leq 1$, on a

$$0 \leq \frac{1}{2}(1-x) \leq 1, 0 \leq \frac{1}{2}(1+x) \leq 1. \text{ et } \frac{1}{2}(1-x) + \frac{1}{2}(1+x) = 1:$$

puisque, de plus, on a l'égalité $tx = \frac{1}{2}(1-x)(-t) + \frac{1}{2}(1+x)t$, la fonction $x \mapsto \exp(tx)$ étant **convexe** (sa dérivée seconde est strictement positive), on a, pour tout $x \in [-1, 1]$,

$$\exp(tx) \leq \frac{1}{2}(1-x)\exp(-t) + \frac{1}{2}(1+x)\exp(t).$$

(b) La variable aléatoire X étant P-p.s. bornée par 1, la variable aléatoire $\exp(tX)$ est bornée P-p.s. et admet donc une moyenne. De plus, d'après la question précédente, on a P-p.s.

$$\exp(tX) \leq \frac{1}{2}(1-X)\exp(-t) + \frac{1}{2}(1+X)\exp(t).$$

Il en résulte que

$$E \exp(tX) \leq \frac{1}{2}E(1-X)\exp(-t) + \frac{1}{2}E(1+X)\exp(t):$$

la variable aléatoire X étant centrée, il vient

$$E \exp(tX) \leq \frac{1}{2}(\exp(-t) + \exp(t)) = \operatorname{ch}(t).$$

On a

$$\operatorname{ch}(t) = \sum_{n=0}^{+\infty} \frac{t^{2n}}{(2n)!} \quad \text{et} \quad \exp\left(\frac{t^2}{2}\right) = \sum_{n=0}^{+\infty} \frac{t^{2n}}{n!2^n};$$

mais, pour tout $n \in \mathbb{N}$, on a

$$n! 2^n = 2 \cdot 4 \cdots 2n \leq (2n)!,$$

ce qui démontre l'inégalité $\operatorname{ch}(t) \leq \exp\left(\frac{t^2}{2}\right)$. Il en résulte que

$$\boxed{E \exp(tX) \leq \exp\left(\frac{t^2}{2}\right)}. \quad (10.30)$$

2. (a) Soit t quelconque. En appliquant l'inégalité (10.30) à la variable aléatoire $\frac{X_n}{c_n}$, on a, pour tout t' ,

$$E \exp\left(t' \frac{X_n}{c_n}\right) \leq \exp\left(\frac{t'^2}{2}\right),$$

et, en prenant $t' = tc_n$, il vient

$$E \exp(tX_n) \leq \exp\left(\frac{t^2}{2}c_n^2\right).$$

Par ailleurs, les variables aléatoires $\exp(tX_n)$ étant indépendantes, on a

$$E \exp(tS_n) = \prod_{j=1}^n E \exp(tX_j):$$

on a ainsi démontré que, pour tout t , on a

$$\boxed{\mathbb{E} \exp(tS_n) \leq \exp\left(\frac{t^2}{2} \sum_{j=1}^n c_j^2\right)}. \quad (10.31)$$

(b) Soient $t > 0$ et $\varepsilon > 0$ quelconques. La fonction $x \mapsto \exp(tx)$ étant **croissante**, on a

$$(S_n > \varepsilon) \subset (\exp(tS_n) > \exp(t\varepsilon));$$

il résulte alors de l'inégalité de Markov que

$$\mathbb{P}(S_n > \varepsilon) \leq \mathbb{P}(\exp(tS_n) > \exp(t\varepsilon)) \leq \frac{\mathbb{E} \exp(tS_n)}{\exp(t\varepsilon)},$$

et, d'après l'inégalité (10.31), que

$$\mathbb{P}(S_n > \varepsilon) \leq \exp\left(-t\varepsilon + \frac{t^2}{2} \sum_{j=1}^n c_j^2\right). \quad (10.32)$$

(c) Soit $\varepsilon > 0$ quelconque et soit $a = \sum_{j=1}^n c_j^2$; la fonction $t \mapsto at^2/2 - t\varepsilon$ atteint son minimum pour $t = \varepsilon/a > 0$; ce minimum vaut $-\varepsilon^2/2a$. L'exponentielle étant (strictement) croissante, il vient

$$\mathbb{P}(S_n > \varepsilon) \leq \exp\left[\min_{t>0}\left(-t\varepsilon + \frac{t^2}{2} \sum_{j=1}^n c_j^2\right)\right] = \exp\left(-\frac{\varepsilon^2}{2 \sum_{j=1}^n c_j^2}\right). \quad (10.33)$$

(d) Soit $\varepsilon > 0$ quelconque. On a les égalités

$$(|S_n| > \varepsilon) = (S_n > \varepsilon) \cup (S_n < -\varepsilon) = (S_n > \varepsilon) \cup (-S_n > \varepsilon),$$

et donc l'inégalité

$$\mathbb{P}(|S_n| > \varepsilon) \leq \mathbb{P}(S_n > \varepsilon) + \mathbb{P}(-S_n > \varepsilon).$$

Appliquant l'inégalité (10.33) aux variables aléatoires $-X_n$, il vient

$$\mathbb{P}(-S_n > \varepsilon) \leq \exp\left(-\frac{\varepsilon^2}{2 \sum_{j=1}^n c_j^2}\right).$$

Il en résulte que

$$\boxed{\mathbb{P}(|S_n| > \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{2 \sum_{j=1}^n c_j^2}\right)}. \quad (10.34)$$

(c) Soit $\varepsilon' > 0$ quelconque ; prenant $\varepsilon = n^\alpha \varepsilon'$ dans l'inégalité (10.34), il vient

$$P(|S_n| > n^\alpha \varepsilon') \leq 2 \exp\left(-\frac{n^{2\alpha} \varepsilon'^2}{2 \sum_{j=1}^n c_j^2}\right),$$

soit, puisque $\sum_{j=1}^n c_j^2 \leq n^{2\alpha-\beta}$, où $\beta > 0$,

$$P(|S_n| > n^\alpha \varepsilon') \leq 2 \exp(-\varepsilon'^2 n^\beta). \quad (10.35)$$

La série de terme général $\exp(-\varepsilon'^2 n^\beta)$ est convergente : en effet, à partir d'un certain rang, on a

$$\varepsilon'^2 n^\beta \geq 2 \ln n, \text{ et donc, } 0 \leq \exp(-\varepsilon'^2 n^\beta) \leq n^{-2}.$$

De l'inégalité (10.35) résulte alors la convergence de la série de terme général $P(|S_n| > n^\alpha \varepsilon')$, et le lemme de Borel Cantelli assure alors que

$$P\left[\limsup_n (|S_n| > n^\alpha \varepsilon')\right] = 0.$$

Puisque \mathbb{Q}^{+*} est dénombrable, il en résulte que

$$P\left[\bigcup_{\varepsilon \in \mathbb{Q}^{+*}} \limsup_n (|S_n| > n^\alpha \varepsilon)\right] = 0.$$

On a donc, en passant au complémentaire,

$$P\left[\bigcap_{\varepsilon \in \mathbb{Q}^{+*}} \liminf_n (|S_n| \leq n^\alpha \varepsilon)\right] = 1.$$

ce qui veut dire que **la suite de terme général $n^{-\alpha} S_n$ converge P-p.s. vers 0**.

3. (a) Soit t quelconque. On a, par le théorème de transfert et le théorème d'intégration par rapport à une mesure à densité,

$$E \exp(tX_1) = \int_{\mathbb{R}} \exp(tx) dP_X(x) = \int_{\mathbb{R}} \exp(tx) \mathbf{1}_{[0,1]}(|x|) (1 - |x|) dx,$$

soit

$$E \exp(tX_1) = \int_{-1}^0 \exp(tx) (1 + x) dx + \int_0^1 \exp(tx) (1 - x) dx;$$

en faisant le changement de variables $y = -x$ dans la première intégrale, et en regroupant, on obtient

$$E \exp(tX_1) = \int_0^1 [\exp(-tx) + \exp(tx)] (1 - x) dx,$$

soit

$$E \exp(tX_1) = 2 \int_0^1 \operatorname{ch}(tx) (1 - x) dx.$$

Si $t = 0$, $E \exp(tX_1) = 1$; si $t \neq 0$, une intégration par parties donne alors

$$E \exp(tX_1) = 2 \left[\left. (1-x) \frac{\text{sh}(tx)}{t} \right|_0^1 + \int_0^1 \frac{\text{sh}(tx)}{t} dx \right],$$

soit

$$E \exp(tX_1) = \begin{cases} 2 \frac{\text{ch}(t) - 1}{t^2} & \text{si } t \neq 0, \\ 1 & \text{si } t = 0. \end{cases}$$

(b) L'application $t \mapsto \exp(tX_1)$ est indéfiniment dérivable et on a, pour tout $k \in \mathbb{N}$ et tout $t \in \mathbb{R}$,

$$|X_1^k \exp(tX_1)| \leq \exp |t| \quad \text{P-p.s. ;}$$

le théorème de dérivation d'une intégrale dépendant d'un paramètre (corollaire du théorème de convergence dominée) assure que Φ est indéfiniment dérivable et que l'on a, pour tout $k \in \mathbb{N}^*$,

$$\Phi^{(k)}(t) = E(X_1^k \exp(tX_1));$$

en particulier on a

$$\Phi^{(k)}(0) = E(X_1^k).$$

Il en résulte (formule de Taylor-Young) que, pour tout $n \in \mathbb{N}^*$,

$$\Phi(t) = 1 + \sum_{k=1}^n \frac{\Phi^{(k)}(0)}{k!} t^k + o(t^n) = 1 + \sum_{k=1}^n \frac{E(X_1^k)}{k!} t^k + o(t^n);$$

en particulier, pour $n = 2$,

$$\Phi(t) = 1 + tE(X_1) + \frac{t^2}{2} E(X_1^2) + o(t^2).$$

Puisque par ailleurs on a, d'après l'expression de Φ ,

$$\Phi(t) = 1 + 2 \frac{t^2}{4!} + o(t^2),$$

par unicité du développement limité, on retrouve que $E(X_1) = 0$, et on obtient que

$$\sigma_{X_1}^2 = E(X_1^2) = 2 \cdot 2 \frac{1}{4!} = \frac{1}{6};$$

les variables aléatoires X_n étant indépendantes de même loi, on a alors

$$\sigma_{S_n}^2 = nE(X_1^2) = \frac{n}{6}.$$

- (c) On peut prendre, dans ce cas, $c_n = 1$ pour tout $n \in \mathbb{N}^*$. On a alors $\sum_{j=1}^n c_j^2 = n$ et, pour réaliser la condition de la question 2.e., il suffit de trouver $\beta > 0$ tel que $1 = 2\alpha - \beta > 0$; si $\alpha > \frac{1}{2}$, $\beta = 2\alpha - 1 > 0$ convient. Il en résulte que, pour tout $\alpha > \frac{1}{2}$, on a $\lim_n n^{-\alpha} S_n = 0$ P.p.s.

Remarque. Comme on le verra au chapitre 14, il résulte du théorème limite central que, pour tout $x \in \mathbb{R}$,

$$\lim_n \mathbb{P}(n^{-\frac{1}{2}} S_n \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du;$$

autrement dit la suite de terme général $n^{-\frac{1}{2}} S_n$ ne converge pas P.p.s. vers 0.

Chapitre 11

Probabilités et espérances conditionnelles

11.1. Noyaux et lois conditionnelles

Nous avons vu (chapitre 4 du tome 1) comment définir la loi conditionnelle d'une variable aléatoire Y par rapport à une variable aléatoire discrète X ; il est évident que ce procédé est impossible à mettre en œuvre dès que X n'est plus discrète, en raison de l'impossibilité de diviser par zéro. Dans ce chapitre, on introduit la notion de **noyau** ou **probabilité de transition** pour définir les lois conditionnelles dans un contexte général.

Dans la suite, (E, \mathcal{E}) et (F, \mathcal{F}) désignent deux espaces probabilisables quelconques.

Définition 11.1. Une application ν de $E \times \mathcal{F}$ dans $[0, 1]$ est appelée **noyau (de probabilité)** sur $E \times \mathcal{F}$ ou **probabilité de transition** de (E, \mathcal{E}) vers (F, \mathcal{F}) , si elle satisfait les deux propriétés :

- (i) pour tout $x \in E$, l'application $\nu(x, \cdot)$ est une probabilité sur (F, \mathcal{F}) ;
- (ii) pour tout $B \in \mathcal{F}$, l'application $\nu(\cdot, B)$ est une fonction \mathcal{E} -mesurable.

Exemple 11.1. Soit P une probabilité sur (F, \mathcal{F}) ; l'application ν de $E \times \mathcal{F}$ dans $[0, 1]$ définie par

$$\forall B \in \mathcal{F} \quad \nu(\cdot, B) = P(B)$$

est un noyau sur $E \times \mathcal{F}$.

Exemple 11.2. Soit p une application mesurable de (E, \mathcal{E}) dans (F, \mathcal{F}) . Si, pour tout $y \in F$, on note δ_y la mesure de Dirac en y , l'application ν de $E \times \mathcal{F}$ dans $[0, 1]$ définie par

$$\forall x \in E \quad \nu(x, \cdot) = \delta_{p(x)}$$

est un noyau sur $E \times \mathcal{F}$; la mesurabilité, pour tout $B \in \mathcal{F}$, de $\nu(\cdot, B)$ résulte de la relation $\nu(\cdot, B) = \mathbf{1}_B \circ p$.

Exemple 11.3. Soit f une application de $E \times F$ dans \mathbb{R}^+ , $\mathcal{E} \otimes \mathcal{F}$ -mesurable et soit μ une mesure positive sur (F, \mathcal{F}) . Soit de plus une probabilité ρ sur

$(\mathcal{F}, \mathcal{F})$. L'application ν de $E \times \mathcal{F}$ dans $[0, 1]$ définie en tout $(x, B) \in E \times \mathcal{F}$ par

$$\nu(x, B) = \begin{cases} \frac{\int_B f(x, y) d\mu(y)}{\int_{\mathcal{F}} f(x, y) d\mu(y)} & \text{si } \int_{\mathcal{F}} f(x, y) d\mu(y) \in]0, +\infty[, \\ \rho(B) & \text{sinon,} \end{cases}$$

est un noyau sur $E \times \mathcal{F}$.

La première propriété est évidente. La mesurabilité pour tout $B \in \mathcal{F}$ de $\nu(\cdot, B)$ s'obtient en utilisant un résultat intermédiaire du théorème de Fubini qui assure la mesurabilité des applications $x \mapsto \int_B f(x, y) d\mu(y)$ et $x \mapsto \int_{\mathcal{F}} f(x, y) d\mu(y)$.

Remarque. Si $\int_{\mathcal{F}} f(x, y) d\mu(y) \in]0, +\infty[$, l'application $B \mapsto \nu(x, B)$ est la mesure de densité $f(x, \cdot) / \int_{\mathcal{F}} f(x, y) d\mu(y)$ par rapport à μ . Un cas particulièrement utilisé est celui où $E = \mathbb{R}^n$ et $\mathcal{F} = \mathbb{R}^m$, ces espaces étant munis de leur tribu borélienne et la mesure de référence μ étant la mesure de Lebesgue sur \mathbb{R}^m .

La donnée d'un noyau et d'une probabilité permet de définir une probabilité sur l'espace produit :

Théorème 11.2. *Soit λ une probabilité sur (E, \mathcal{E}) et ν un noyau sur $E \times \mathcal{F}$. Soit $\lambda \cdot \nu$ l'application définie sur la semi-algèbre des pavés, notée abusivement $\mathcal{E} \times \mathcal{F}$, par*

$$\forall A \times B \in \mathcal{E} \times \mathcal{F} \quad \lambda \cdot \nu(A \times B) = \int_A \nu(x, B) d\lambda(x). \quad (11.1)$$

L'application $\lambda \cdot \nu$ est σ -additive sur $\mathcal{E} \times \mathcal{F}$; il existe un prolongement unique en une probabilité sur l'espace probabilisable produit $(E \times F, \mathcal{E} \otimes \mathcal{F})$ encore notée $\lambda \cdot \nu$.

Démonstration. On applique le théorème d'existence et d'unicité d'un prolongement d'une fonction σ -additive sur une semi-algèbre en une probabilité sur la tribu engendrée. Démontrons la σ -additivité de $\lambda \cdot \nu$ sur la semi-algèbre $\mathcal{E} \times \mathcal{F}$. Soient $A \times B \in \mathcal{E} \times \mathcal{F}$ et $(A_n \times B_n)_{n \in \mathbb{N}} \subset \mathcal{E} \times \mathcal{F}$ une suite d'ensembles *disjoints* telle que

$$A \times B = \bigsqcup_{n \in \mathbb{N}} (A_n \times B_n),$$

ce qui est équivalent à

$$\forall (x, y) \in E \times F \quad \mathbf{1}_{(A \times B)}(x, y) = \sum_{n=0}^{+\infty} \mathbf{1}_{A_n}(x) \mathbf{1}_{B_n}(y).$$

Intégrant, à x fixé, en y par rapport à la probabilité $\nu(x, \cdot)$, il vient :

$$\forall x \in E \quad \mathbf{1}_A(x)\nu(x, B) = \sum_{n=0}^{+\infty} \mathbf{1}_{A_n}(x)\nu(x, B_n).$$

Reste à intégrer par rapport à la probabilité λ et à utiliser la σ -additivité de l'intégrale :

$$\int_E \mathbf{1}_A(x)\nu(x, B) d\lambda(x) = \sum_{n=0}^{+\infty} \int_E \mathbf{1}_{A_n}(x)\nu(x, B_n) d\lambda(x),$$

c'est-à-dire :

$$\lambda \cdot \nu(A \times B) = \sum_{n=0}^{+\infty} \lambda \cdot \nu(A_n \times B_n). \quad \square$$

Remarque. 1. Dans l'exemple 11.1, la mesure $\lambda \cdot \nu$ est alors la probabilité produit $\lambda \otimes P$.

2. Dans l'exemple 11.2, la mesure $\lambda \cdot \nu$ est définie par

$$\forall A \times B \in \mathcal{E} \times \mathcal{F} \quad \lambda \cdot \nu(A \times B) = \int_{\Lambda} \mathbf{1}_B \circ p d\lambda.$$

3. Dans l'exemple 11.3, si μ est une mesure σ -finie, la mesure $\lambda \cdot \nu$ est définie en tout $A \times B \in \mathcal{E} \times \mathcal{F}$ par

$$\lambda \cdot \nu(A \times B) = \int_{(A \cap C) \times B} \frac{f(x, y)}{g(x)} d(\lambda \otimes \mu)(x, y) + \lambda(A \cap C^c)p(B),$$

où g est définie par

$$\forall x \in E \quad g(x) = \int_F f(x, y) d\mu(y).$$

et

$$C = \{x \in E \mid g(x) \in]0, +\infty[\}.$$

Si $\lambda(C) = 1$, on a $\lambda \otimes \mu[(A \cap C) \times B] = \lambda \otimes \mu[A \times B]$ et la mesure $\lambda \cdot \nu$ est définie par

$$\forall A \times B \in \mathcal{E} \times \mathcal{F} \quad \lambda \cdot \nu(A \times B) = \int_{A \times B} \frac{f(x, y)}{g(x)} d\lambda \otimes \mu(x, y).$$

C'est alors une **mesure à densité** par rapport à la mesure produit $\lambda \otimes \mu$.

Notation. Selon le contexte, si f est une fonction sur F intégrable par rapport à la probabilité $\nu(x, \cdot)$, son intégrale est notée indifféremment $\nu(x, f)$ ou $\int_F f(y) \nu(x, dy)$ ou encore $\nu f(x)$.

On donne maintenant un théorème d'intégration par rapport à la mesure $\lambda \cdot \nu$ qui généralise le théorème de Fubini (lequel correspond au cas de l'exemple 11.1).

Théorème 11.3 (Théorème de Fubini généralisé). Soit f une application mesurable de l'espace probabilisable produit $(E \times F, \mathcal{E} \otimes \mathcal{F})$ dans $(\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$.

(a) Si f est positive, l'application $x \mapsto \int_F f(x, y) \nu(x, dy)$ est \mathcal{E} -mesurable et on a

$$\int_{E \times F} f \, d\lambda \cdot \nu = \int_E \left[\int_F f(x, y) \nu(x, dy) \right] d\lambda(x). \quad (11.2)$$

(b) Si f est $\lambda \cdot \nu$ -intégrable, alors pour λ -presque tout x , l'application partielle $f(x, \cdot)$ est $\nu(x, \cdot)$ -intégrable, et l'application définie pour λ -presque tout x par $\mapsto \int_F f(x, y) \nu(x, dy)$ est λ -intégrable et l'égalité (11.2) est encore vraie.

Démonstration. La démonstration est standard :

(a) La famille de parties

$$\mathcal{S} = \{A \in \mathcal{E} \otimes \mathcal{F} \mid x \mapsto \nu(x, A_x^2) \text{ est } \mathcal{E}\text{-mesurable}\},$$

où A_x^2 est la **section**¹ en x de A , est un λ -système ; en effet,

– pour tous A, B de \mathcal{S} tels que $A \subset B$, on a $A_x^2 \subset B_x^2$ et $(B \setminus A)_x^2 = B_x^2 \setminus A_x^2$ et donc, $\nu(x, \cdot)$ étant une probabilité,

$$\nu(x, (B \setminus A)_x^2) = \nu(x, B_x^2) - \nu(x, A_x^2);$$

l'application $x \mapsto \nu(x, (B \setminus A)_x^2)$ est alors \mathcal{E} -mesurable et on a ainsi $B \setminus A \in \mathcal{S}$.

– Pour toute suite croissante $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{S} , on a, pour tout $n \in \mathbb{N}$, $(A_n)_x^2 \subset (A_{n+1})_x^2$ et $[\bigcup_{n \in \mathbb{N}} A_n]_x^2 = \bigcup_{n \in \mathbb{N}} (A_n)_x^2$; $\nu(x, \cdot)$ étant une probabilité, il vient :

$$\nu\left(x, \left[\bigcup_{n \in \mathbb{N}} A_n\right]_x^2\right) = \lim_n \nearrow \nu(x, (A_n)_x^2);$$

l'application $x \mapsto \nu(x, [\bigcup_{n \in \mathbb{N}} A_n]_x^2)$ est alors \mathcal{E} -mesurable et $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{S}$.

Il est évident que \mathcal{S} contient le π -système des pavés $\mathcal{E} \times \mathcal{F}$ et donc la tribu $\mathcal{E} \otimes \mathcal{F}$ engendrée par $\mathcal{E} \times \mathcal{F}$. On vient de prouver que $\mathcal{S} = \mathcal{E} \otimes \mathcal{F}$, c'est-à-dire encore que l'application $x \mapsto \int_F f(x, y) \nu(x, dy)$ est \mathcal{E} -mesurable pour tout f fonction indicatrice d'un ensemble $A \in \mathcal{E} \otimes \mathcal{F}$, et aussi, par linéarité, pour toute fonction mesurable étagée. On obtient le résultat, pour toute fonction mesurable positive, en prenant une suite croissante de fonctions étagées positives convergeant simplement vers f . On peut alors définir, pour toute fonction mesurable positive f , l'élément de $\overline{\mathbb{R}}^+$:

1. Si $A \in \mathcal{P}(E \times F)$, on définit les **sections**, éventuellement vides : $A_x^1 = \{x \in E \mid (x, y) \in A\}$, si $x \in F$, et $A_x^2 = \{y \in F \mid (x, y) \in A\}$, si $x \in E$.

$$\Phi(f) = \int_E \left[\int_F f(x, y) \nu(x, dy) \right] d\lambda(x).$$

Φ est une intégrale qui coïncide, par définition de $\lambda \cdot \nu$, avec l'intégrale $f \mapsto \int_{E \times F} f d(\lambda \cdot \nu)$ sur l'ensemble des fonctions indicatrices des pavés mesurables; elles sont égales.

(b) Si f est $\lambda \cdot \nu$ -intégrable, l'égalité (11.2) est encore vraie pour $|f|$, ce qui démontre que pour λ -presque tout x , l'application partielle $f(x, \cdot)$ est $\nu(x, \cdot)$ -intégrable, et que l'application, définie λ -presque partout, $x \mapsto \int_F f(x, y) \nu(x, dy)$ est λ -intégrable. L'égalité (11.2) est aussi vraie pour f^+ et f^- ; alors, par définition de l'intégrale de f par rapport à la mesure $\lambda \cdot \nu$, on a

$$\begin{aligned} \int_{E \times F} f d(\lambda \cdot \nu) &= \int_{E \times F} f^+ d(\lambda \cdot \nu) - \int_{E \times F} f^- d(\lambda \cdot \nu) \\ &= \int_E \left[\int_F f^+(x, y) \nu(x, dy) \right] d\lambda(x) - \int_E \left[\int_F f^-(x, y) \nu(x, dy) \right] d\lambda(x) \\ &= \int_E \left[\int_F f(x, y) \nu(x, dy) \right] d\lambda(x). \quad \square \end{aligned}$$

Corollaire 11.4. Avec les notations du théorème 11.3, l'application définie sur \mathcal{F} par

$$\forall B \in \mathcal{F} \quad \mu(B) = \int_E \nu(x, B) d\lambda(x)$$

est une probabilité sur l'espace probabilisable (F, \mathcal{F}) . Soit g une application mesurable de l'espace probabilisable (F, \mathcal{F}) dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

(a) Si g est positive, on a

$$\int_F g d\mu = \int_E \nu(x, g) d\lambda(x). \quad (11.3)$$

(b) Si g est de signe quelconque et μ -intégrable, l'application $\nu(\cdot, g)$ est définie λ -presque partout, égale λ -presque partout à une fonction \mathcal{E} -mesurable et l'égalité (11.3) est encore vraie.

Remarque. On peut obtenir sans difficulté une version du théorème 11.2 et de son corollaire dans le cadre des fonctions à valeurs dans \mathbb{R}^d ou un espace euclidien.

Dans ce qui suit, X et Y désignent deux variables aléatoires à valeurs respectivement dans les espaces probabilisables quelconques (E, \mathcal{E}) et (F, \mathcal{F}) .

Si X est discrète et si \mathcal{E} contient les points, c'est-à-dire si tout singleton est un élément de \mathcal{E} , définissons le noyau ν sur $E \times \mathcal{F}$ par

$$\forall (x, B) \in E \times \mathcal{F} \quad \nu(x, B) = \begin{cases} P^{(X=x)}(Y \in B) & \text{si } x \in \text{val}(X) \\ \rho(B) & \text{sinon,} \end{cases}$$

où ρ est une probabilité quelconque sur (F, \mathcal{F}) . Remarquons que $\text{val}(X) = \{x \in E \mid P(X = x) \neq 0\}$, réunion dénombrable de singletons, appartient alors à \mathcal{E} et que $P_X[\text{val}(X)] = 1$. Pour tout $A \in \mathcal{E}$ et $B \in \mathcal{F}$, on a

$$\begin{aligned} P[(X \in A) \cap (Y \in B)] &= \sum_{x \in \text{val}(X) \cap A} P[(X = x) \cap (Y \in B)] \\ &= \sum_{x \in \text{val}(X) \cap A} \nu(x, B) P_X(\{x\}); \end{aligned}$$

puisque la probabilité P_X s'écrit

$$P_X = \sum_{x \in \text{val}(X)} P(X = x) \delta_x.$$

il vient

$$P_{(X,Y)}(A \times B) = \int_A \nu(x, B) dP_X(x),$$

ce qui est équivalent à $P_{(X,Y)} = P_X \cdot \nu$. Cette relation sera le point de départ pour définir une loi conditionnelle dans le cas général.

Définition 11.5. On appelle **loi conditionnelle de Y sachant X** un noyau ν sur $E \times \mathcal{F}$ tel que

$$P_{(X,Y)} = P_X \cdot \nu.$$

On la note souvent $P_Y^{X=\cdot}$ et la formule de définition s'écrit ainsi :

$$P_{(X,Y)} = P_X \cdot P_Y^{X=\cdot}. \quad (11.4)$$

Exemple 11.4. Si λ et μ sont des mesures σ -finies respectivement sur (E, \mathcal{E}) et (F, \mathcal{F}) , et si $P_{(X,Y)} = f \cdot (\lambda \otimes \mu)$, où f est une fonction mesurable positive sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$ de $\lambda \otimes \mu$ -intégrale 1, le noyau ν défini dans l'exemple 11.3 est une loi conditionnelle de Y sachant X. L'application $x \mapsto \int_F f(x, y) d\mu(y)$ n'est autre que la densité de P_X par rapport à λ .

Définition 11.6. Dans le cas où $E = \mathbb{R}^n$ et $F = \mathbb{R}^m$, munis de leur tribu borélienne, si ν est une loi conditionnelle de Y sachant X telle que, pour P_X -presque tout $x \in \mathbb{R}^n$, $\nu(x, \cdot)$ est une mesure de densité $f_Y^{X=x}$ par rapport à la mesure de Lebesgue de \mathbb{R}^m , on dit que $f_Y^{X=x}$ est une **densité conditionnelle de Y sachant X = x**.

Exemple 11.5. Cas particulier usuel : $E = \mathbb{R}^n$ et $F = \mathbb{R}^m$ munis de leur tribu borélienne, les mesures de références étant les mesures de Lebesgue sur ces espaces.

Supposons que (X, Y) admette une densité $f_{(X,Y)}$; X admet alors une densité f_X et pour toute probabilité ρ sur $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$, le noyau ν défini pour $(x, B) \in \mathbb{R}^n \times \mathcal{B}_{\mathbb{R}^m}$ par

$$\nu(x, B) = \begin{cases} \int_B \frac{f_{(X,Y)}(x, y)}{f_X(x)} d\lambda_m(y) & \text{si } f_X(x) > 0, \\ \rho(B) & \text{si } f_X(x) = 0. \end{cases}$$

est une *loi conditionnelle de Y sachant X* (c'est un cas particulier de l'exemple 11.4). Si $f_X(x) > 0$, la mesure $\nu(x, \cdot)$ admet la densité $f_{(X,Y)}(x, \cdot)/f_X(x)$ par rapport à la mesure de Lebesgue λ_m . La marginale Y admet donc une *densité conditionnelle* par rapport à X (ou sachant X) notée $f_Y^{X=x}$ qui vérifie, pour tout $(x, y) \in (\mathbb{R}^n \times \mathbb{R}^m)$ tel que $f_X(x) > 0$:

$$f_Y^{X=x}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}. \quad (11.5)$$

Inversement, si, pour P_X -presque tout $x \in \mathbb{R}^n$, il existe une densité conditionnelle de Y sachant $X = x$, notée $f_Y^{X=x}$, et si X admet une densité f_X , la variable aléatoire (X, Y) admet une densité $f_{(X,Y)}$ qui vérifie, pour $P_{(X,Y)}$ -presque tout $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$:

$$f_{(X,Y)}(x, y) = f_X(x) f_Y^{X=x}(y). \quad (11.6)$$

En effet, par définition de la loi conditionnelle, on a alors, pour tout $A \in \mathcal{B}_{\mathbb{R}^n}$ et $B \in \mathcal{B}_{\mathbb{R}^m}$:

$$P_{(X,Y)}(A \times B) = \int_A \left[\int_B f_Y^{X=x}(y) d\lambda_m(y) \right] f_X(x) d\lambda_n(x)$$

ce qui donne le résultat après application du théorème de Fubini.

Montrons comment ces deux situations peuvent apparaître simultanément. On considère deux variables aléatoires réelles X et Y *indépendantes*, de même loi exponentielle $\exp(-\lambda)$; on note $S = X + Y$ et on cherche une loi conditionnelle de X sachant S (X et Y représentent par exemple les **temps d'attente** de deux clients arrivant indépendamment à un guichet).

En raison de l'indépendance de X et Y, la variable aléatoire (X, Y) admet une densité produit direct des densités des marginales, et par le changement de variables (de jacobien 1) défini sur \mathbb{R}^2 par

$$\begin{cases} x = t \\ y = s - t, \end{cases}$$

la variable aléatoire (X, S) admet une densité $f_{(X,S)}$ donnée par

$$\forall (t, s) \in \mathbb{R}^2 \quad f_{(X,S)}(t, s) = f_X(t) f_Y(t, s - t).$$

Il en résulte que (résultat déjà vu par ailleurs) S admet une densité f_S donnée par

$$\forall s \in \mathbb{R} \quad f_S(s) = \mathbf{1}_{\mathbb{R}^+}(s) \lambda^2 s \exp(-\lambda s),$$

et, que, pour tout $s > 0$, X admet une *densité conditionnelle sachant* $S = s$ donnée, après réduction, par

$$f_X^{S=s}(x) = \frac{1}{s} \mathbf{1}_{[0,s]}(x).$$

Ainsi, pour tout s appartenant à l'intérieur du support de la loi de S , la loi conditionnelle de X sachant S est la loi uniforme sur l'intervalle $[0, s]$.

Un exemple concret va montrer que l'on a bien défini une notion de loi conditionnelle qui correspond à l'intuition.

Exemple 11.6. La variable aléatoire (X, Y) représente un point aléatoire tiré au hasard dans le carré $[0, 1]^2$, autrement dit, (X, Y) est une variable aléatoire de loi uniforme sur $[0, 1]^2$ et a pour densité $\mathbf{1}_{[0,1]^2}$. On note $S = X + Y$ et l'on cherche une loi conditionnelle de X sachant S (noter que X et Y sont alors indépendantes de même loi uniforme sur $[0, 1]$).

On est dans la même situation que dans l'exemple précédent, mais avec une loi différente. On développe le calcul plus en détail; soit T le difféomorphisme de \mathbb{R}^2 sur lui-même défini par

$$\forall (x, y) \in \mathbb{R}^2 \quad T(x, y) = (x, x + y).$$

Son inverse est donné par

$$\forall (u, s) \in \mathbb{R}^2 \quad T^{-1}(u, s) = (u, s - u);$$

le jacobien du difféomorphisme est de valeur absolue 1 et l'on a $(X, S) = T \circ (X, Y)$. La variable aléatoire (X, S) admet donc une densité $f_{(X,S)}$ donnée par

$$\forall (u, s) \in \mathbb{R}^2 \quad f_{(X,S)}(u, s) = f_{(X,Y)}(u, s - u),$$

soit

$$\forall (u, s) \in \mathbb{R}^2 \quad f_{(X,S)}(u, s) = \mathbf{1}_{[0,1]}(u) \mathbf{1}_{[0,1]}(s - u).$$

La marginale S admet alors la densité donnée par

$$\forall s \in \mathbb{R} \quad f_S(s) = \int_{\mathbb{R}} f_{(X,S)}(u, s) du = \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(u) \mathbf{1}_{[0,1]}(s - u) du.$$

En décomposant le produit d'indicatrices sous la forme

$$\mathbf{1}_{[0,1]}(u) \mathbf{1}_{[0,1]}(s - u) = \mathbf{1}_{[0,1]}(s) \mathbf{1}_{[0,s]}(u) + \mathbf{1}_{[1,2]}(s) \mathbf{1}_{[s-1,1]}(u),$$

il vient :

$$\forall s \in \mathbb{R} \quad f_S(s) = s \mathbf{1}_{[0,1]}(s) + (2 - s) \mathbf{1}_{[1,2]}(s).$$

La loi de S est la **loi triangulaire**. Pour tout s de l'intérieur du support de f_S , la variable aléatoire X admet une *densité conditionnelle sachant* $S = s$, $f_X^{S=s}$, donnée par

$$\forall x \in \mathbb{R} \quad f_X^{S=s}(x) = \begin{cases} \frac{1}{s} \mathbf{1}_{[0,s]}(x) & \text{si } 0 < s \leq 1, \\ \frac{1}{2-s} \mathbf{1}_{[s-1,1]}(x) & \text{si } 1 < s < 2. \end{cases}$$

La variable aléatoire X admet donc comme loi conditionnelle sachant $S = s$ la loi uniforme sur $[0, s]$ si $0 < s \leq 1$, sur $[s-1, 1]$ si $1 < s < 2$, résultat bien naturel.

Exemple 11.7. Si X et Y sont deux variables aléatoires *indépendantes* à valeurs respectivement dans les espaces probabilisables quelconques (E, \mathcal{E}) et (F, \mathcal{F}) , le noyau « constant » ν défini par

$$\forall x \in E \quad \nu(x, \cdot) = P_Y$$

est une loi conditionnelle de Y sachant X .

En effet, X et Y étant indépendantes, on a $P_{(X,Y)} = P_X \otimes P_Y$, ce qui peut s'écrire, pour tout $A \in \mathcal{E}$ et $B \in \mathcal{F}$:

$$P_{(X,Y)}(A \times B) = \int_A \nu(x, B) dP_X(x).$$

Remarque. Il est évident que tout autre noyau ν' sur $E \times \mathcal{F}$ qui vérifie, pour tout $B \in \mathcal{F}$, $\nu'(\cdot, B) = \nu(\cdot, B) P_X$ -p.s. est encore une loi conditionnelle de Y sachant X . Il n'y a donc pas unicité de la loi conditionnelle. Se pose maintenant le problème de l'existence. Celui-ci est partiellement résolu dans les exemples ci-dessus. Nous donnons à titre d'information un théorème assez général d'existence ; sa démonstration sort du cadre de ce livre.

Théorème 11.7 (Théorème de Jirina). *Soient E et F deux espaces métriques séparables complets (en particulier des espaces euclidiens) munis de leur tribu borélienne et X et Y deux variables aléatoires à valeurs respectivement dans E et F ; il existe une loi conditionnelle de Y sachant X .*

Donnons une version du théorème de Fubini généralisé en termes de lois conditionnelles, aucune démonstration n'étant nécessaire.

Théorème 11.8. *Soit (X, Y) une variable aléatoire à valeurs dans un espace probabilisable quelconque $(E \times F, \mathcal{E} \otimes \mathcal{F})$ telle qu'existe une loi conditionnelle $P_Y^{X=x}$ de Y sachant X . Soit f une application mesurable de l'espace probabilisable $(E \times F, \mathcal{E} \otimes \mathcal{F})$ dans $(\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$.*

(a) *Si f est positive, l'application $x \mapsto \int_F f(x, y) dP_Y^{X=x}(y)$ est \mathcal{E} -mesurable et on a :*

$$\int_{E \times F} f dP_{(X,Y)} = \int_E \left[\int_F f(x, y) dP_Y^{X=x}(y) \right] dP_X(x). \quad (11.7)$$

(b) Si f est de signe quelconque et $P_{(X,Y)}$ -intégrable, pour P_X -presque tout x , l'application partielle $f(x, \cdot)$ est $P_Y^{X=x}$ -intégrable, et l'application définie P_X -presque sûrement par $x \mapsto \int_F f(x, y) dP_Y^{X=x}(y)$ est P_X -intégrable et l'égalité (11.7) est encore vraie.

Il en résulte un théorème de transfert « conditionnel » qui est souvent utilisé dans les calculs de lois conditionnelles.

Théorème 11.9 (Théorème de transfert conditionnel). Soit (X, Y) une variable aléatoire à valeurs dans l'espace probabilisable $(E \times F, \mathcal{E} \otimes \mathcal{F})$ telle qu'existe une loi conditionnelle $\nu = P_Y^{X=\cdot}$ de Y sachant X . Soit f une application mesurable de $(E \times F, \mathcal{E} \otimes \mathcal{F})$ dans un autre espace probabilisable (G, \mathcal{G}) . Une loi conditionnelle de $f(X, Y)$ sachant X est donnée par le noyau μ en termes de mesure image par

$$\forall x \in E \quad \mu(x, \cdot) = f(x, \cdot) [\nu(x, \cdot)] ,$$

ce qui peut s'écrire de manière plus suggestive :

$$\boxed{\forall x \in E \quad P_{f(X,Y)}^{X=x} = P_{f(x,Y)}^{X=x}} \quad (11.8)$$

En particulier, si X et Y sont indépendantes, on a :

$$\boxed{\forall x \in E \quad P_{f(X,Y)}^{X=x} = P_{f(x,Y)}} \quad (11.9)$$

Démonstration. Pour tout $A \in \mathcal{E}$ et tout $B \in \mathcal{G}$, on a, avec les notations habituelles, et en utilisant le théorème 11.8 :

$$\begin{aligned} P_{(X,f(X,Y))}(A \times B) &= P_{(X,Y)}[(A \times F) \cap f^{-1}(B)] \\ &= \int_E \left[\int_F \mathbf{1}_{A \times F}(x, y) \mathbf{1}_{f^{-1}(B)}(x, y) \nu(x, dy) \right] dP_X(x) . \end{aligned}$$

Mais on a

$$\mathbf{1}_{A \times F}(x, y) \mathbf{1}_{f^{-1}(B)}(x, y) = \mathbf{1}_A(x) \mathbf{1}_{[f(x, \cdot)]^{-1}(B)}(y) ,$$

ce qui donne, par définition de la mesure image $\mu(x, \cdot)$ de $\nu(x, \cdot)$ par $f(x, \cdot)$:

$$P_{(X,f(X,Y))}(A \times B) = \int_A \mu(x, B) dP_X(x) . \quad \square$$

Les deux exemples suivants illustrent les différentes notions et théorèmes introduits jusqu'à maintenant dans des situations non standards où les lois et lois conditionnelles sont des **mélanges de lois à densité et de mesures ponctuelles** (pondérations de mesures de Dirac).

Exemple 11.8. Soient (X, Y) une variable aléatoire à valeurs dans $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$ et $\varepsilon \in]0, 1[$. On suppose que X est de **loi uniforme** sur l'intervalle $[0, 1]$. On note λ la mesure de Lebesgue sur \mathbb{R} . On considère les deux cas suivants :

- **Cas 1.** Une loi conditionnelle $P_Y^{X=y}$ de Y sachant X est donnée par

$$\forall x \in [0, 1] \quad P_Y^{X=x} = \mathbf{1}_{[\varepsilon, 1]} \cdot \lambda + \varepsilon \delta_x,$$

c'est-à-dire que la probabilité $P_Y^{X=x}$ est **mélange de la probabilité uniforme sur l'intervalle $[\varepsilon, 1]$ et de la masse de Dirac en x** .

- **Cas 2.** Une loi conditionnelle $P_Y^{X=x}$ de Y sachant X est donnée par

$$P_Y^{X=x} = \begin{cases} \mathbf{1}_{[\varepsilon, 1]} \cdot \lambda + \varepsilon \delta_x & \forall x \in [0, \varepsilon[\\ \mathcal{U}([0, 1]) & \forall x \in [\varepsilon, 1[, \end{cases}$$

c'est-à-dire que, si $0 \leq x < \varepsilon$, la probabilité $P_Y^{X=x}$ est encore mélange de la probabilité uniforme sur l'intervalle $[\varepsilon, 1]$ et de la masse de Dirac en x et si $\varepsilon \leq x \leq 1$, $P_Y^{X=x}$ est la loi uniforme sur $[0, 1]$.

On étudie dans ces deux cas la loi de la variable aléatoire Y . Par définition d'une loi conditionnelle, on a, pour tous boréliens A et B de \mathbb{R} ,

$$P_{(X,Y)}(A \times B) = \int_A P_Y^{X=x}(B) dP_X(x),$$

soit :

- **Cas 1.**

$$P_{(X,Y)}(A \times B) = \int_A \mathbf{1}_{[0,1]}(x) [\lambda(B \cap [\varepsilon, 1]) + \varepsilon \mathbf{1}_B(x)] d\lambda(x),$$

ce qui peut s'écrire

$$P_{(X,Y)}(A \times B) = \lambda(A \cap [0, 1]) \lambda(B \cap [\varepsilon, 1]) + \varepsilon \lambda(A \cap B \cap [0, 1]).$$

La loi de Y est obtenue en prenant $A = \mathbb{R}$, soit, $\forall B \in \mathcal{B}_{\mathbb{R}}$:

$$P_Y(B) = \lambda(B \cap [\varepsilon, 1]) + \varepsilon \lambda(B \cap [0, 1]) = \int_B [\mathbf{1}_{[\varepsilon, 1]} + \varepsilon \mathbf{1}_{[0, 1]}] d\lambda.$$

La variable aléatoire Y admet donc une densité f_Y donnée par

$$f_Y = \mathbf{1}_{[\varepsilon, 1]} + \varepsilon \mathbf{1}_{[0, 1]}.$$

ou encore :

$$f_Y = \varepsilon \mathbf{1}_{[0, \varepsilon[} + (1 + \varepsilon) \mathbf{1}_{[\varepsilon, 1]}.$$

- **Cas 2.**

$$P_{(X,Y)}(A \times B) = \int_{A \cap [0, \varepsilon[} \mathbf{1}_{[0, 1]}(x) [\lambda(B \cap [\varepsilon, 1]) + \varepsilon \mathbf{1}_B(x)] d\lambda(x) + \int_{A \cap [\varepsilon, 1[} \mathbf{1}_{[0, 1]}(x) \lambda(B \cap [0, 1]) d\lambda(x).$$

ce qui peut s'écrire :

$$P_{(X,Y)}(A \times B) = \lambda(A \cap [0, \varepsilon]) \lambda(B \cap [\varepsilon, 1]) + \varepsilon \lambda(A \cap B \cap [0, \varepsilon]) \\ + \lambda(A \cap [\varepsilon, 1]) \lambda(B \cap [0, 1]).$$

La loi de Y est obtenue en prenant $A = \mathbb{R}$, soit, après simplifications,

$$\forall B \in \mathcal{B}_{\mathbb{R}} \quad P_Y(B) = \lambda(B \cap [0, 1]).$$

La variable aléatoire Y est donc de *loi uniforme* sur l'intervalle $[0, 1]$. Ceci est donc un **exemple de variables aléatoires X et Y , chacune de loi uniforme sur l'intervalle $[0, 1]$, et telle que la loi du couple (X, Y) n'est pas la loi uniforme sur le carré $[0, 1]^2$.**

On calcule dans ces deux cas la covariance des variables aléatoires X et Y .

Les variables aléatoires X et Y sont bornées par 1 ; par le théorème de transfert, l'application $(x, y) \mapsto xy$ est donc $P_{(X,Y)}$ -intégrable et l'on peut appliquer le théorème 11.8. ce qui donne

$$E(XY) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} xy \, dP_Y^{X=x}(y) \right] dP_X(x),$$

soit :

- **Cas 1.**

$$E(XY) = \int_{\mathbb{R}} x \left[\int_{[\varepsilon, 1]} y \, d\lambda(y) + \varepsilon x \right] \mathbf{1}_{[0,1]}(x) \, d\lambda(x),$$

et, en identifiant intégrales de Lebesgue et Riemann,

$$E(XY) = \int_0^1 x \left[\frac{1}{2}(1 - \varepsilon^2) + \varepsilon x \right] dx,$$

ce qui donne $E(XY) = \frac{1}{4}(1 - \varepsilon^2) + \frac{\varepsilon}{3}$. Un calcul élémentaire donne

$$EX = \frac{1}{2} \quad \text{et} \quad EY = \frac{1}{2}(1 - \varepsilon^2) + \frac{\varepsilon}{2};$$

en tenant compte de l'égalité $\text{cov}(X, Y) = E(XY) - (EX)(EY)$ il vient alors

$$\boxed{\text{cov}(X, Y) = \frac{\varepsilon}{12}.}$$

- **Cas 2.** De même

$$E(XY) = \int_{[0, \varepsilon]} x \left[\int_{[\varepsilon, 1]} y \, d\lambda(y) + \varepsilon x \right] d\lambda(x) \\ + \int_{[\varepsilon, 1]} x \left[\int_{[0, 1]} y \, d\lambda(y) \right] d\lambda(x),$$

soit, en identifiant intégrales de Lebesgue et Riemann,

$$E(XY) = \int_0^{\varepsilon} x \left[\frac{1}{2}(1 - \varepsilon^2) + \varepsilon x \right] dx + \frac{1}{2} \int_{\varepsilon}^1 x \, dx,$$

ce qui donne après calcul : $E(XY) = \frac{1}{4} + \frac{\varepsilon^4}{12}$. Les variables aléatoires X et Y étant, dans ce cas, de loi uniforme sur $[0, 1]$, on a $EX = EY = \frac{1}{2}$, ce qui conduit à l'égalité :

$$\text{cov}(X, Y) = \frac{\varepsilon^4}{12}.$$

11.2. Moments conditionnels

On définit, lorsqu'ils existent, les **moments conditionnels**.

Proposition 11.10 (Proposition et définition). Soient X une variable aléatoire à valeurs dans un espace probablisable quelconque (E, \mathcal{E}) et Y une variable aléatoire réelle telles qu'existe une loi conditionnelle $P_Y^{X=\cdot}$ de Y sachant X . Si, pour un $p \in \mathbb{N}^*$, Y admet un moment d'ordre p , alors, P_X -presque sûrement :

$$\int_{\mathbb{R}} |y|^p dP_Y^{X=\cdot}(y) < +\infty.$$

Si $p = 1$, on appelle **moyenne conditionnelle** une fonction \mathcal{E} -mesurable égale P_X -presque sûrement à $m_Y^{X=\cdot} \equiv \int_{\mathbb{R}} y dP_Y^{X=\cdot}(y)$.

Si $p = 2$, on appelle **variance conditionnelle** une fonction \mathcal{E} -mesurable égale P_X -presque sûrement à $\int_{\mathbb{R}} [y - \int_{\mathbb{R}} y dP_Y^{X=\cdot}(y)]^2 dP_Y^{X=\cdot}(y)$.

Démonstration. Puisque, par hypothèse

$$E|Y|^p = \int_{E \times \mathbb{R}} |y|^p dP_{(X,Y)}(x, y) < +\infty,$$

le théorème de Fubini généralisé assure que :

$$\int_E \left[\int_{\mathbb{R}} |y|^p dP_Y^{X=x}(y) \right] dP_X(x) < +\infty.$$

On obtient le résultat. □

Remarque. Il résulte du théorème de Fubini généralisé, puis du théorème de transfert, que l'on a, pour tout $C \in \mathcal{E}$,

$$\begin{aligned} \int_C m_Y^{X=x} dP_X(x) &= \int_E \mathbf{1}_C(x) \left[\int_{\mathbb{R}} y dP_Y^{X=x}(y) \right] dP_X(x) \\ &= \int_{E \times \mathbb{R}} \mathbf{1}_C(x) y dP_{(X,Y)}(x, y) \\ &= \int_{\Omega} \mathbf{1}_C(X) Y dP. \end{aligned}$$

soit :

$$\forall C \in \mathcal{E} \quad \int_C m_Y^{X=\lambda} dP_X(\lambda) = \int_{X^{-1}(C)} Y dP. \quad (11.10)$$

Autrement dit, la moyenne conditionnelle $m_Y^{X=\lambda}$ et la loi de X suffisent pour calculer la moyenne de Y sur un élément quelconque $X^{-1}(C)$ de la tribu engendrée par X .

Nous donnons un exemple de calcul de moyenne conditionnelle.

Exemple 11.9. Reprenons l'exemple 11.8 et calculons la moyenne conditionnelle de Y sachant X dans les deux cas envisagés.

– **Cas 1.** On a, pour tout $x \in [0, 1]$, $m_Y^{X=x} = \int_{[\varepsilon, 1]} y d\lambda(y) + \varepsilon x$, soit :

$$m_Y^{X=x} = \frac{1}{2}(1 - \varepsilon^2) + \varepsilon x.$$

La moyenne conditionnelle de Y sachant X est *affine* sur $[0, 1]$.

– **Cas 2.** Utilisant le calcul précédent, on a, pour tout $x \in [0, \varepsilon]$:

$$m_Y^{X=x} = \frac{1}{2}(1 - \varepsilon^2) + \varepsilon x.$$

Puisque, pour tout $x \in [\varepsilon, 1]$, on a

$$m_Y^{X=x} = \frac{1}{2}.$$

la moyenne conditionnelle de Y sachant X est *affine par morceaux* sur $[0, 1]$, avec une discontinuité en ε ; elle s'écrit :

$$m_Y^{X=x} = \mathbf{1}_{[0, \varepsilon]}(x) \left[\frac{1}{2}(1 - \varepsilon^2) + \varepsilon x \right] + \mathbf{1}_{[\varepsilon, 1]}(x) \frac{1}{2}.$$

Nous reprenons la situation du théorème de transfert conditionnel 11.9 et donnons une **formule de calcul de la moyenne conditionnelle** très utile.

Lemme 11.11. Soit (X, Y) une variable aléatoire à valeurs dans l'espace probabilisable $(E \times F, \mathcal{E} \otimes \mathcal{F})$ telle qu'existe une loi conditionnelle $\nu = P_X^{X=\cdot}$ de Y sachant X . Soit f une application mesurable de $(E \times F, \mathcal{E} \otimes \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. On suppose que $f(X, Y) \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. On a alors, pour P_X -presque tout $x \in E$:

$$m_{f(X, Y)}^{X=x} = m_{f(\cdot, Y)}^{X=x}.$$

En particulier, si X et Y sont **indépendantes**, on a, pour P_X -presque tout $x \in E$:

$$m_{f(X, Y)}^{X=x} = E[f(x, Y)].$$

Démonstration. La définition de la moyenne conditionnelle $m_{f(X,Y)}^{X=x}$ et le théorème de transfert conditionnel 11.9 permettent d'écrire, pour P_X -presque tout $x \in E$:

$$m_{f(X,Y)}^{X=x} = \int_{\mathbb{R}} z \, dP_{f(X,Y)}^{X=x}(z).$$

ce qui donne le résultat. Dans le cas d'indépendance, il suffit alors d'appliquer la relation (11.9). \square

Nous terminons ce paragraphe en donnant une **application de la notion de moyenne conditionnelle au problème de régression**. C'est un problème de moindres carrés qui généralise celui de régression linéaire que nous avons étudié au premier tome.

Le problème général : soit (X, Y) une variable aléatoire à valeurs dans l'espace probabilisable $(E \times F, \mathcal{E} \otimes \mathcal{F})$ telle qu'existe une loi conditionnelle $\nu = P_Y^{X=x}$ de Y sachant X . On veut estimer dans quelle mesure Y est « voisine » d'une fonctionnelle de X . Cette formulation heuristique donne naissance au problème de minimisation précis suivant auquel nous nous limiterons :

On suppose que F est l'ensemble \mathbb{R} muni de sa tribu borélienne (une généralisation possible et simple est de prendre F euclidien) et que la variable aléatoire Y admet un moment d'ordre 2. On cherche à résoudre le problème de minimisation :

$$\min \{E[Y - f \circ X]^2 \mid f \in \mathcal{L}^2(E, \mathcal{E}, P_X)\}. \quad (11.11)$$

Remarque. Pour interpréter géométriquement ce problème, transformons-le en un problème de projection dans l'espace hilbertien $L^2(\Omega, \mathcal{A}, P)$: admettons provisoirement que le sous-espace

$$\Pi_X \equiv \{\widehat{f \circ X} \mid f \in \mathcal{L}^2(E, \mathcal{E}, P_X)\}$$

est un sous-espace fermé de $L^2(\Omega, \mathcal{A}, P)$ ($\widehat{f \circ X}$ désignant la classe de $f \circ X$). Les solutions du problème (11.11) sont alors les représentants de la projection orthogonale de la classe de Y sur Π_X .

Lemme 11.12. *Le sous-espace Π_X est fermé dans $L^2(\Omega, \mathcal{A}, P)$.*

Démonstration. Soit une suite $(f_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{L}^2(E, \mathcal{E}, P_X)$ telle que la suite $(\widehat{f_n \circ X})_{n \in \mathbb{N}}$ converge vers $Z \in L^2(\Omega, \mathcal{A}, P)$. La suite $(f_n \circ X)_{n \in \mathbb{N}}$ est bornée dans $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ par un réel $c > 0$ et il existe une sous-suite $(f_{n_k})_{k \in \mathbb{N}}$ telle que la suite $(f_{n_k} \circ X)_{k \in \mathbb{N}}$ converge P.-p.s. (vers un représentant de Z) ; en particulier, si $f = \limsup_k f_{n_k}$, la suite $(f_{n_k} \circ X)_{k \in \mathbb{N}}$ converge

P-p.s. vers $f \circ X$. On a, d'après le théorème de transfert et le lemme de Fatou :

$$\begin{aligned} \int_E [f(x)]^2 dP_X(x) &= \int_\Omega [f \circ X]^2 dP = \int_\Omega \lim_k [f_{n_k} \circ X]^2 dP \\ &\leq \lim_k \int_\Omega [f_{n_k} \circ X]^2 dP \leq c, \end{aligned}$$

ce qui démontre que $f \in \mathcal{L}^2(E, \mathcal{E}, P_X)$. Puisque de plus $f \circ \widetilde{X} = Z$, le lemme est démontré. \square

Proposition 11.13. *La moyenne conditionnelle $m_Y^{X=}$ est une solution du problème de régression (11.11).*

Démonstration. Il résulte du théorème de Fubini généralisé que, pour tout $f \in \mathcal{L}^2(E, \mathcal{E}, P_X)$:

$$E\{Y - f \circ X\}^2 = \int_E \left[\int_F [y - f(x)]^2 dP_Y^{X=x}(y) \right] dP_X(x).$$

Toute solution f_0 au problème (11.11) vérifie, pour P_X -presque tout x :

$$\int_F [y - f_0(x)]^2 dP_Y^{X=x}(y) = \min \left\{ \int_F [y - f(x)]^2 dP_Y^{X=x}(y) \mid f \in \mathcal{L}^2(E, \mathcal{E}, P_X) \right\}.$$

Il s'en suit que, pour P_X -presque tout x , $f_0(x)$ doit être un point stationnaire du polynôme Q , du second degré en z :

$$Q(z) = z^2 - 2z \int_F y dP_Y^{X=x}(y),$$

soit :

$$f_0(x) = \int_F y dP_Y^{X=x}(y),$$

et ce point correspond bien à un minimum. \square

11.3. Espérance conditionnelle

L'étude d'un phénomène aléatoire conduit, pour une certaine information, à adopter comme modèle de base un espace probabilisé (Ω, \mathcal{A}, P) . Si l'information est « moins grande », on peut être conduit à travailler avec un espace probabilisé (Ω, \mathcal{B}, P) où \mathcal{B} est une *sous-tribu* de \mathcal{A} , c'est-à-dire une tribu telle que $\mathcal{B} \subset \mathcal{A}$; c'est en particulier le cas lorsque l'on étudie des phénomènes aléatoires qui dépendent du temps, l'information croissant avec le temps. Si Y est une variable aléatoire sur l'espace probabilisé (Ω, \mathcal{A}, P) , comment calculer sa moyenne sur des éléments de \mathcal{B} , uniquement à l'aide d'une variable aléatoire \mathcal{B} -mesurable? **L'espérance conditionnelle**, outil fondamental des probabilistes, permet de répondre à cette

question. On peut dire que son utilisation permet un calcul « progressif », comme on le verra constamment en étudiant les martingales (chap. 15) et les chaînes de Markov (chap. 16).

Dans ce paragraphe, on se donne un espace probabilisé (Ω, \mathcal{A}, P) et une sous-tribu \mathcal{B} de \mathcal{A} . On notera identiquement, sauf mention du contraire, une variable aléatoire X et sa classe \tilde{X} .

11.3.1. L'espérance conditionnelle comme projecteur orthogonal dans $L^2(\Omega, \mathcal{A}, P)$

Lemme 11.14. *Le sous-espace $L^2(\Omega, \mathcal{B}, P)$ est fermé dans l'espace hilbertien $L^2(\Omega, \mathcal{A}, P)$. Le projecteur orthogonal sur $L^2(\Omega, \mathcal{B}, P)$ est noté $E^{\mathcal{B}}$. La projection orthogonale $E^{\mathcal{B}}Y$ de $Y \in L^2(\Omega, \mathcal{A}, P)$ est caractérisée par la relation d'orthogonalité :*

$$\boxed{\forall Z \in L^2(\Omega, \mathcal{B}, P) \quad E(ZY) = E[Z(E^{\mathcal{B}}Y)]} \quad (11.12)$$

Démonstration. Le sous-espace $L^2(\Omega, \mathcal{B}, P)$ est complet, donc **fermé** dans $L^2(\Omega, \mathcal{A}, P)$. La relation (11.12) est l'écriture de l'orthogonalité de $Y - E^{\mathcal{B}}Y$ au sous-espace $L^2(\Omega, \mathcal{B}, P)$. \square

Remarque. L'unicité de la projection orthogonale sur un sous-espace fermé d'un espace hilbertien implique que $E^{\mathcal{B}}Y$ est l'**unique** classe U de variables aléatoires \mathcal{B} -mesurables qui satisfasse à la relation :

$$\forall Z \in L^2(\Omega, \mathcal{B}, P) \quad E(ZY) = E[ZU].$$

Cette unicité permet souvent d'identifier l'espérance conditionnelle.

Définition 11.15. *Si $Y \in L^2(\Omega, \mathcal{A}, P)$ (classe de variables aléatoires), la classe de variables aléatoires $E^{\mathcal{B}}Y$ est appelée **espérance conditionnelle de Y sachant \mathcal{B}** . Si $Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ (variable aléatoire), $E^{\mathcal{B}}\tilde{Y}$ est encore appelée **espérance conditionnelle de Y sachant \mathcal{B}** et notée $E^{\mathcal{B}}Y$.*

Remarques et notation. On parle donc indifféremment de l'espérance conditionnelle d'une variable aléatoire ou d'une **classe** de variables aléatoires, celle-ci étant toujours une **classe**. De plus, par abus de notation, s'il n'y a pas risque d'erreur, $E^{\mathcal{B}}Y$ pourra désigner un représentant quelconque de cette classe (souvent appelé **version de l'espérance conditionnelle**). Pour indiquer que U est une version de $E^{\mathcal{B}}Y$ on écrira

$$U = E^{\mathcal{B}}Y \quad P\text{-p.s.}$$

Proposition 11.16. *Soit $Y \in L^2(\Omega, \mathcal{A}, P)$; la relation (11.12) est équivalente à la relation*

$$\boxed{\forall B \in \mathcal{B} \quad E(\mathbf{1}_B Y) = E[\mathbf{1}_B (E^{\mathcal{B}}Y)]} \quad (11.13)$$

ce qui s'écrit encore :

$$\forall B \in \mathcal{B} \quad \int_B Y \, dP = \int_B E^{\mathcal{B}} Y \, dP. \quad (11.14)$$

Démonstration. Soit $Y \in L^2(\Omega, \mathcal{A}, P)$; l'implication (11.12) \Rightarrow (11.13) est évidente. Inversement, si (11.13) est vérifiée, par linéarité, (11.12) est vérifiée pour toute variable aléatoire Z étagée \mathcal{B} -mesurable; l'ensemble de ces variables aléatoires étant dense dans $L^2(\Omega, \mathcal{B}, P)$, on conclut par continuité, les applications qui à $Z \in L^2(\Omega, \mathcal{B}, P)$ donne l'un ou l'autre membre de l'égalité (11.12) étant, d'après l'inégalité de Schwarz, des formes linéaires continues. \square

Remarque. L'espérance conditionnelle est donc encore caractérisée comme l'**unique** classe $U \in L^2(\Omega, \mathcal{B}, P)$ de variables aléatoires \mathcal{B} -mesurables qui satisfasse à la relation :

$$\forall B \in \mathcal{B} \quad \int_B Y \, dP = \int_B U \, dP.$$

Note préliminaire sur l'ordre défini sur l'ensemble des classes de variables aléatoires. La relation d'équivalence « égalité P-p.s. » est compatible avec l'ordre partiel sur les variables aléatoires réelles (ou à valeurs dans $\overline{\mathbb{R}}$), elle induit un *ordre partiel* sur les classes encore noté \leq . En particulier, si $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ sont des suites de variables aléatoires telles que

$$\forall n \in \mathbb{N} \quad \text{P-p.s.} \quad X_n = Y_n,$$

on a, puisque une réunion d'ensembles de probabilité nulle est de probabilité nulle :

$$\text{P-p.s.} \quad \forall n \in \mathbb{N} \quad X_n = Y_n.$$

On a alors les égalités P-p.s. entre variables aléatoires à valeurs dans $\overline{\mathbb{R}}$:

$$\text{P-p.s.} \quad \sup_n X_n = \sup_n Y_n \quad \text{et} \quad \inf_n X_n = \inf_n Y_n.$$

Remarque. Il faut noter qu'il n'en est plus de même si on considère des familles *non dénombrables* de variables aléatoires, les sup et inf pouvant même ne plus être des variables aléatoires !

Proposition 11.17. *L'opérateur $E^{\mathcal{B}}$ est linéaire continu sur $L^2(\Omega, \mathcal{A}, P)$, de norme 1. Il est de plus **positif**, c'est-à-dire qu'il satisfait à l'implication :*

$$Y \geq 0 \implies E^{\mathcal{B}} Y \geq 0.$$

En particulier, si $Y_1, Y_2 \in L^2(\Omega, \mathcal{A}, P)$ sont telles que $Y_1 \leq Y_2$, on a :

$$E^{\mathcal{B}} Y_1 \leq E^{\mathcal{B}} Y_2.$$

Démonstration. C'est une propriété des projecteurs orthogonaux. La positivité vient de ce que si $Y \geq 0$, on a pour tout $B \in \mathcal{B}$, $\int_B E^{\mathcal{B}} Y dP \geq 0$, ce qui est équivalent à dire que $E^{\mathcal{B}} Y \geq 0$. \square

Proposition 11.18. Soit $Y \in L^2(\Omega, \mathcal{A}, P)$. On a les propriétés suivantes :

- (a) $E[E^{\mathcal{B}} Y] = EY$;
 (b) si Y est \mathcal{B} -mesurable, on a $E^{\mathcal{B}} Y = Y$;
 (c) si Z est \mathcal{B} -mesurable et bornée, on a

$$E^{\mathcal{B}}(ZY) = ZE^{\mathcal{B}} Y \quad P\text{-p.s.}$$

(d) « **Théorème des trois perpendiculaires** » : si \mathcal{B}_1 et \mathcal{B}_2 sont deux sous-tribus telles que $\mathcal{B}_1 \subset \mathcal{B}_2$:

$$E^{\mathcal{B}_1} Y = E^{\mathcal{B}_1} [E^{\mathcal{B}_2} Y] . \quad (11.15)$$

(e) $|E^{\mathcal{B}} Y| \leq E^{\mathcal{B}}(|Y|)$.

(f) L'opérateur $E^{\mathcal{B}}$ de $L^2(\Omega, \mathcal{A}, P)$ dans $L^2(\Omega, \mathcal{B}, P)$ est de norme 1 pour les normes L^1 , c'est-à-dire que l'on a, pour tout $Y \in L^2(\Omega, \mathcal{A}, P)$,

$$\|E^{\mathcal{B}} Y\|_1 \leq \|Y\|_1 . \quad (11.16)$$

Démonstration. (a) Il suffit de prendre $Z = 1$, qui est bien \mathcal{B} -mesurable, dans la relation (11.12).

(b) Y est dans le sous-espace $L^2(\Omega, \mathcal{B}, P)$.

(c) Si Z est bornée, $YZ \in L^2(\Omega, \mathcal{A}, P)$ et, pour tout $T \in L^2(\Omega, \mathcal{B}, P)$, on a par définition de $E^{\mathcal{B}}(ZY)$:

$$E[TE^{\mathcal{B}}(ZY)] = E[TZY] .$$

Mais ZT étant \mathcal{B} -mesurable, par définition de $E^{\mathcal{B}} Y$, on a

$$E[TE^{\mathcal{B}}(ZY)] = E[(TZ)E^{\mathcal{B}} Y] .$$

ce qui peut se lire :

$$E[TE^{\mathcal{B}}(ZY)] = E[T(ZE^{\mathcal{B}} Y)] .$$

Mais $ZE^{\mathcal{B}} Y \in L^2(\Omega, \mathcal{B}, P)$, la première remarque permet de conclure.

(d) C'est une propriété générale des espaces de Hilbert (connue en géométrie dans l'espace sous le nom de théorème des trois perpendiculaires). $L^2(\Omega, \mathcal{B}_1, P)$ étant un sous-espace fermé de $L^2(\Omega, \mathcal{B}_2, P)$. Redémontrons cette propriété dans ce contexte : pour tout $Z \in L^2(\Omega, \mathcal{B}_1, P)$, Z est \mathcal{B}_2 -mesurable et donc

$$E(ZY) = E[Z(E^{\mathcal{B}_2} Y)] .$$

Alors, par définition de $E^{\mathcal{B}_1} [E^{\mathcal{B}_2} Y]$, on a

$$E(ZY) = E \left[ZE^{\mathcal{B}_1} \left[E^{\mathcal{B}_2} Y \right] \right],$$

ce qui démontre le résultat.

(e) L'espérance conditionnelle étant linéaire, utilisons la **convexité** de la fonction valeur absolue en écrivant que c'est l'enveloppe supérieure de ses minorantes affines; plus précisément, en ne prenant que les extrémales, si $A = \{-1, 1\}$, on a :

$$\forall x \in \mathbb{R} \quad |x| = \sup_{a \in A} (ax)$$

(cet argument de convexité sera repris plus loin pour établir l'**inégalité de Jensen**). On a alors

$$\forall a \in A \quad \text{P-p.s.} \quad aE^{\mathcal{B}} Y = E^{\mathcal{B}}(aY) \leq E^{\mathcal{B}}|Y|,$$

et donc

$$\text{P-p.s.} \quad \forall a \in A \quad aE^{\mathcal{B}} Y \leq E^{\mathcal{B}}|Y|,$$

ce qui implique :

$$\text{P-p.s.} \quad |E^{\mathcal{B}} Y| = \sup_{a \in A} \left[aE^{\mathcal{B}} Y \right] \leq E^{\mathcal{B}}|Y|.$$

(f) Il suffit d'intégrer la relation précédente. □

11.3.2. Extension de la définition de l'espérance conditionnelle à $L^1(\Omega, \mathcal{A}, P)$.

Proposition 11.19. Soit $Y \in L^1(\Omega, \mathcal{A}, P)$ ou $\mathcal{L}^1(\Omega, \mathcal{A}, P)$. Il existe une **unique** classe de variables aléatoires \mathcal{B} -mesurables $U \in L^1(\Omega, \mathcal{B}, P)$ qui satisfasse à la relation :

$$\forall B \in \mathcal{B} \quad \int_B Y dP = \int_B U dP. \quad (11.17)$$

Elle est encore notée $E^{\mathcal{B}} Y$ et appelée **espérance conditionnelle de Y sachant \mathcal{B}** . Elle vérifie

$$\|E^{\mathcal{B}} Y\|_1 \leq \|Y\|_1. \quad (11.18)$$

Démonstration. On se ramène au cas $L^2(\Omega, \mathcal{A}, P)$ de la manière suivante : soit $(Y_n)_{n \in \mathbb{N}}$ la suite définie par

$$\forall n \in \mathbb{N} \quad Y_n = \mathbf{1}_{(|Y| \leq n)} Y.$$

Pour tout $n \in \mathbb{N}$, on a $Y_n \in L^2(\Omega, \mathcal{A}, P)$, et

$$|Y_n - Y| \leq |Y|.$$

La suite $(Y_n)_{n \in \mathbb{N}}$ convergeant P-p.s. vers Y , il résulte du théorème de convergence dominée qu'elle converge aussi dans $L^1(\Omega, \mathcal{A}, P)$ vers Y . Soit

alors, pour tout $n \in \mathbb{N}$, $Z_n = E^{\mathcal{B}} Y_n \in L^2(\Omega, \mathcal{B}, P)$; d'après l'inégalité (11.16), on a, pour tout $n, m \in \mathbb{N}$,

$$\|Z_n - Z_m\|_1 \leq \|Y_n - Y_m\|_1.$$

La suite $(Y_n)_{n \in \mathbb{N}}$ convergeant dans $L^1(\Omega, \mathcal{A}, P)$ est de Cauchy; il en est alors de même pour la suite $(Z_n)_{n \in \mathbb{N}}$; l'espace $L^1(\Omega, \mathcal{A}, P)$ étant complet, la suite $(Z_n)_{n \in \mathbb{N}}$ converge dans $L^1(\Omega, \mathcal{A}, P)$ vers Z ; plus précisément, puisque, pour tout $n \in \mathbb{N}$, Z_n est \mathcal{B} -mesurable, on a $Z \in L^1(\Omega, \mathcal{B}, P)$. Par ailleurs, pour tout $B \in \mathcal{B}$ et tout $n \in \mathbb{N}$, on a :

$$\int_B Y_n dP = \int_B Z_n dP.$$

La convergence des suites $(Y_n)_{n \in \mathbb{N}}$ et $(Z_n)_{n \in \mathbb{N}}$ dans $L^1(\Omega, \mathcal{A}, P)$ permet de passer à la limite, ce qui donne la relation :

$$\forall B \in \mathcal{B} \quad \int_B Y dP = \int_B Z dP.$$

Nous avons démontré l'existence; l'unicité est triviale. Reste à démontrer l'inégalité (11.18). Pour cela, on applique, pour tout n , l'inégalité (11.16) à Y_n et on « passe à la limite » : il en résulte que :

$$E|Z| = \lim_n E|Z_n| \leq \lim_n E|Y_n| = E \left[\lim_n |Y_n| \right] = E|Y|. \quad \square$$

Proposition 11.20. *L'opérateur $E^{\mathcal{B}}$ est linéaire continu sur $L^1(\Omega, \mathcal{A}, P)$ de norme 1. Il est positif.*

Soit $Y \in L^1(\Omega, \mathcal{A}, P)$. On a les propriétés suivantes :

- (a) $E[E^{\mathcal{B}} Y] = EY$;
- (b) si Y est \mathcal{B} -mesurable, on a $E^{\mathcal{B}} Y = Y$;
- (c) si Z est \mathcal{B} -mesurable et bornée, on a

$$E^{\mathcal{B}}(ZY) = ZE^{\mathcal{B}} Y \quad P\text{-p.s.};$$

- (d) si \mathcal{B}_1 et \mathcal{B}_2 sont deux sous-tribus telles que $\mathcal{B}_1 \subset \mathcal{B}_2$, on a

$$E^{\mathcal{B}_1} Y = E^{\mathcal{B}_1} [E^{\mathcal{B}_2} Y]; \quad (11.19)$$

- (e) $|E^{\mathcal{B}} Y| \leq E^{\mathcal{B}}(|Y|)$.

Démonstration. La linéarité résulte de la caractérisation (11.17) de $E^{\mathcal{B}} Y$. L'inégalité (11.16) montre que $E^{\mathcal{B}}$ est continu de norme inférieure ou égale à 1; cette norme est de fait égale à 1 puisque si $Y \in L^1(\Omega, \mathcal{B}, P)$, il résulte de la caractérisation (11.17) que $E^{\mathcal{B}} Y = Y$. Toutes les autres propriétés se déduisent alors par continuité des propriétés analogues dans $L^2(\Omega, \mathcal{A}, P)$ ou directement en utilisant la caractérisation (11.17) et les mêmes arguments qu'à la proposition 11.18. \square

Remarque. Puisque $E^{\mathcal{B}}$ est continu sur $L^1(\Omega, \mathcal{A}, P)$, si une suite $(X_n)_{n \in \mathbb{N}}$ converge vers X dans $L^1(\Omega, \mathcal{A}, P)$, la suite $(E^{\mathcal{B}}X_n)_{n \in \mathbb{N}}$ converge vers $E^{\mathcal{B}}X$ dans $L^1(\Omega, \mathcal{A}, P)$.

Proposition 11.21. *Si $Y \in L^1(\Omega, \mathcal{A}, P)$ et si Y et \mathcal{B} sont indépendantes (c'est-à-dire si les tribus $\sigma(Y)$ et \mathcal{B} le sont), on a :*

$$E^{\mathcal{B}}Y = EY \quad P\text{-p.s.}$$

Démonstration. Puisque pour tout $B \in \mathcal{B}$, les variables aléatoires $\mathbf{1}_B$ et Y sont indépendantes on a

$$E(\mathbf{1}_B Y) = E(\mathbf{1}_B)E(Y),$$

soit

$$E(\mathbf{1}_B Y) = E[\mathbf{1}_B E(Y)] :$$

reste à utiliser la caractérisation (11.17) de $E^{\mathcal{B}}Y$. □

Remarque. Bien noter qu'ici, on a une égalité entre classes et que EY représente la classe des variables aléatoires P -p.s. égales à EY .

La **généralisation** suivante de cette propriété est aussi **très souvent utilisée dans les calculs**.

Proposition 11.22. *Soient (X, Y) une variable aléatoire à valeurs dans l'espace probabilisable $(E \times F, \mathcal{E} \otimes \mathcal{F})$ et $f \in \mathcal{L}^1(F \times F, \mathcal{E} \otimes \mathcal{F}, P_{(X,Y)})$. On suppose que X est \mathcal{B} -mesurable et que Y et \mathcal{B} sont **indépendantes**. La fonction \hat{f} définie par*

$$\forall x \in E \quad \hat{f}(x) = E[f(x, Y)].$$

est \mathcal{E} -mesurable et on a :

$$\boxed{E^{\mathcal{B}}[f(X, Y)] = \hat{f} \circ X \quad P\text{-p.s.}} \quad (11.20)$$

Démonstration. Les variables aléatoires X et Y étant indépendantes, on a $P_{(X,Y)} = P_X \otimes P_Y$. Remarquante que

$$\forall x \in E \quad \hat{f}(x) = \int_F f(x, y) dP_Y(y),$$

la propriété de mesurabilité de \hat{f} résulte du théorème de Fubini. Par ailleurs, l'espace vectoriel engendré par les fonctions $(x, y) \mapsto g(x)h(y)$ où $g \in \mathcal{L}^1(E, \mathcal{E}, P_X)$ et $h \in \mathcal{L}^1(F, \mathcal{F}, P_Y)$ est dense dans $\mathcal{L}^1(E \times F, \mathcal{E} \otimes \mathcal{F}, P_{(X,Y)})$. De plus les applications

$$f \mapsto \hat{f} \circ X \quad \text{et} \quad f \mapsto E^{\mathcal{B}}[f(X, Y)]$$

sont continues de $\mathcal{L}^1(E \times F, \mathcal{E} \otimes \mathcal{F}, P_{(X,Y)})$ dans $L^1(\Omega, \mathcal{A}, P)$. En effet, il résulte des théorèmes de transfert puis de Fubini que l'on a

$$\begin{aligned} \|\widehat{f} \circ X\|_1 &= \int_{\Omega} \left| \int_F f(X, y) dP_Y(y) \right| dP \\ &\leq \int_E \left[\int_F |f(x, y)| dP_Y(y) \right] dP_X(x) = \int_{E \times F} |f| d(P_X \otimes P_Y); \end{aligned}$$

puisque $P_{(X,Y)} = P_X \otimes P_Y$, on a donc :

$$\|\widehat{f} \circ X\|_1 \leq \|f\|_1.$$

Quant à la continuité de l'application $f \mapsto E^{\mathcal{B}}[f(X, Y)]$, elle résulte, par application du théorème de transfert, des relations

$$\|E^{\mathcal{B}}[f(X, Y)]\|_1 \leq \|f(X, Y)\|_1 = \|f\|_1.$$

Reste donc à démontrer, en raison de la linéarité, la relation (11.20), pour f produit direct de $g \in \mathcal{L}^1(E, \mathcal{E}, P_X)$ et $h \in \mathcal{L}^1(F, \mathcal{F}, P_Y)$. Mais $g \circ X$ est \mathcal{B} -mesurable et $h \circ Y$ est indépendante de \mathcal{B} ; on a donc

$$E^{\mathcal{B}}[f(X, Y)] = E^{\mathcal{B}}[(g \circ X)(h \circ Y)] = (g \circ X)E^{\mathcal{B}}[h \circ Y] = (g \circ X)E[h \circ Y],$$

soit :

$$E^{\mathcal{B}}[f(X, Y)] = (g \circ X) \int_F [h(y)] dP_Y(y) = \int_F (g \circ X)[h(y)] dP_Y(y) = \widehat{f} \circ X. \quad \square$$

Exemple 11.10. Soient X et Y deux variables aléatoires réelles, Y étant de loi de Poisson de paramètre $\lambda > 0$; on suppose que X est \mathcal{B} -mesurable et que Y et \mathcal{B} sont indépendantes. Calculer $E^{\mathcal{B}}[\cos(XY)]$.

Solution. On a

$$\widehat{f}(x) = E \cos(xY) = \exp(-\lambda) \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \cos(kx);$$

or

$$\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \cos(kx) = \Re \left(\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \exp(ikx) \right) = \Re \exp[\lambda \exp(ix)].$$

ce qui donne :

$$E^{\mathcal{B}} \cos(XY) = \exp[-\lambda(1 - \cos X)] \times \cos(\lambda \sin X).$$

11.3.3. Extension de la définition de l'espérance conditionnelle à $\mathcal{M}^+(\mathcal{A})$

On note $\mathcal{M}^+(\mathcal{B})$ l'ensemble des variables aléatoires à valeurs dans $\overline{\mathbb{R}}^+$ et \mathcal{B} -mesurables.

Proposition 11.23. Soit $Y \in \mathcal{M}^+(\mathcal{A})$. Il existe une **unique** classe U d'éléments de $\mathcal{M}^+(\mathcal{B})$ qui satisfasse à la relation :

$$\forall B \in \mathcal{B} \quad \int_B Y dP = \int_B U dP. \quad (11.21)$$

Elle est encore notée $E^{\mathcal{B}}Y$ et appelée **espérance conditionnelle de Y sachant \mathcal{B}** .

Démonstration. Rien de changé pour l'unicité. Pour l'existence, soient, pour tout $n \in \mathbb{N}$, la variable aléatoire bornée $Y_n = \inf(Y, n)$ et U_n une version de $E^{\mathcal{B}}Y_n$. La suite $(Y_n)_{n \in \mathbb{N}}$ converge en croissant vers Y et P-p.s. la suite $(U_n)_{n \in \mathbb{N}}$ est croissante, donc convergente dans $\overline{\mathbb{R}}^+$ vers une limite \mathcal{B} -mesurable U . La propriété de Beppo Levi assure de plus que, pour tout $B \in \mathcal{B}$, on a

$$\int_B Y dP = \lim_n \int_B Y_n dP = \lim_n \int_B U_n dP = \int_B U dP. \quad \square$$

Proposition 11.24. Pour tous $Y, Z \in \mathcal{M}^+(\mathcal{A})$ telles que $Y \leq Z$, on a $E^{\mathcal{B}}Y \leq E^{\mathcal{B}}Z$.

De plus on a la propriété de **Beppo Levi conditionnelle** : si une suite $(Y_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{M}^+(\mathcal{A})$ converge **en croissant** vers Y , la suite $(E^{\mathcal{B}}Y_n)_{n \in \mathbb{N}}$ converge **en croissant** vers $E^{\mathcal{B}}Y$.

Démonstration. Pour la première propriété, on a, pour tout $n \in \mathbb{N}$, $\inf(Y, n) \leq \inf(Z, n)$ et donc

$$E^{\mathcal{B}}[\inf(Y, n)] \leq E^{\mathcal{B}}[\inf(Z, n)].$$

Il suffit de passer à la limite dans $\overline{\mathbb{R}}^+$ et de revenir à la définition de $E^{\mathcal{B}}Y$ et $E^{\mathcal{B}}Z$.

Pour la propriété de Beppo Levi conditionnelle, la croissance de la suite $(E^{\mathcal{B}}Y_n)_{n \in \mathbb{N}}$ résulte de la première propriété : cette suite converge alors dans $\overline{\mathbb{R}}^+$ et, d'après la propriété de Beppo Levi usuelle, pour tout $B \in \mathcal{B}$, on a

$$\int_B E^{\mathcal{B}}Y dP = \int_B Y dP = \lim_n \int_B Y_n dP = \lim_n \int_B E^{\mathcal{B}}Y_n dP = \int_B \lim_n E^{\mathcal{B}}Y_n dP.$$

Il en résulte que :

$$E^{\mathcal{B}}Y = \lim_n E^{\mathcal{B}}Y_n. \quad \square$$

Remarque. En corollaire, les quatre premières propriétés de $E^{\mathcal{B}}$ citées à la proposition 11.20 sont encore vraies sur $\mathcal{M}^+(\mathcal{A})$.

11.3.4. Théorèmes de convergence

Ayant obtenu une propriété de Beppo Levi conditionnelle, on obtient selon la même démarche qu'en théorie de l'intégration, un **lemme de Fatou** et un **théorème de convergence dominée** conditionnels.

Lemme 11.25 (Lemme de Fatou conditionnel). Soit $(X_n)_{n \in \mathbb{N}}$ une suite d'éléments de $\mathcal{M}^+(\mathcal{A})$. On a :

$$\mathbb{E}^{\mathcal{B}} \left[\liminf_n X_n \right] \leq \liminf_n \mathbb{E}^{\mathcal{B}}(X_n).$$

Démonstration. On a, pour tout $n \in \mathbb{N}$,

$$\forall p \geq n \quad \inf_{k \geq n} X_k \leq X_p$$

et donc, par croissance de l'espérance conditionnelle :

$$\forall p \geq n \quad \mathbb{E}^{\mathcal{B}} \left[\inf_{k \geq n} X_k \right] \leq \mathbb{E}^{\mathcal{B}} X_p.$$

Il en résulte que :

$$\mathbb{E}^{\mathcal{B}} \left[\inf_{k \geq n} X_k \right] \leq \inf_{p \geq n} \mathbb{E}^{\mathcal{B}} X_p.$$

Reste à appliquer la propriété de Beppo Levi conditionnelle. \square

Théorème 11.26 (Théorème de convergence dominée de Fatou-Lebesgue conditionnel). Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires finies P-p.s. et $Y \in \mathcal{L}^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ telles que

$$\forall n \in \mathbb{N} \quad |X_n| \leq Y \quad \text{P-p.s.}$$

(a) On a :

$$\mathbb{E}^{\mathcal{B}} \left[\liminf_n X_n \right] \leq \liminf_n \mathbb{E}^{\mathcal{B}}(X_n) \leq \limsup_n \mathbb{E}^{\mathcal{B}}(X_n) \leq \mathbb{E}^{\mathcal{B}} \left[\limsup_n X_n \right].$$

(b) De plus, si la suite $(X_n)_{n \in \mathbb{N}}$ est P-p.s. convergente, la suite $(\mathbb{E}^{\mathcal{B}} X_n)_{n \in \mathbb{N}}$ est P-p.s. convergente et on a

$$\mathbb{E}^{\mathcal{B}} \left[\lim_n X_n \right] = \lim_n \mathbb{E}^{\mathcal{B}}(X_n) \quad \text{P-p.s.}$$

Démonstration. Les variables aléatoires $Y + X_n$ et $Y - X_n$ sont définies et positives P-p.s., Y et X_n étant P-p.s. finies; notons de la même manière leur prolongement mesurable par 0. Le lemme de Fatou conditionnel donne alors

$$\mathbb{E}^{\mathcal{B}} \left[\liminf_n (Y + X_n) \right] \leq \liminf_n \mathbb{E}^{\mathcal{B}}(Y + X_n),$$

soit

$$E^{\mathcal{B}}Y + E^{\mathcal{B}}\left[\liminf_n X_n\right] \leq E^{\mathcal{B}}Y + \liminf_n E^{\mathcal{B}}X_n,$$

ce qui donne la première inégalité puisque $E^{\mathcal{B}}Y$ est intégrable, donc fini P-p.s. Pour la seconde inégalité, on procède de même avec $Y - X_n$.

Enfin, si la suite $(X_n)_{n \in \mathbb{N}}$ est P-p.s. convergente, ce qui est équivalent à $\liminf_n X_n = \limsup_n X_n = \lim_n X_n$, on a (théorème de convergence dominée) $\lim_n X_n \in \mathcal{L}^1_{\mathbb{R}}(\Omega, \mathcal{A}, P)$ et

$$E^{\mathcal{B}}\left[\lim_n X_n\right] \leq \liminf_n E^{\mathcal{B}}(X_n) \leq \limsup_n E^{\mathcal{B}}(X_n) \leq E^{\mathcal{B}}\left[\lim_n X_n\right].$$

ce qui donne le résultat annoncé. \square

Corollaire 11.27. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires finies P-p.s. telle que

$$\sum_{n=0}^{+\infty} E|X_n| < +\infty.$$

Alors, P-p.s. la série $\sum X_n$ est absolument convergente, sa somme appartient à $\mathcal{L}^1_{\mathbb{R}}(\Omega, \mathcal{A}, P)$ et

$$E^{\mathcal{B}}\left[\sum_{n=0}^{+\infty} X_n\right] = \sum_{n=0}^{+\infty} E^{\mathcal{B}}X_n.$$

Démonstration. Appliquer le théorème de Lebesgue conditionnel à la suite des sommes partielles. \square

Exemple 11.11. On reprend l'exemple 11.10. En supposant d'abord que X est bornée par M , on calcule $E^{\mathcal{B}}[\cos(XY)]$ en développant en série entière le cosinus ; le cas général est alors résolu par passage à la limite.

Remarque. Il s'avère que cette méthode est beaucoup plus longue et montre le chemin que l'on a parcouru en établissant la proposition 11.22.

Solution : On a

$$\cos(XY) = \sum_{r=0}^{+\infty} (-1)^r \frac{(XY)^{2r}}{(2r)!}.$$

On va vérifier que, sous l'hypothèse $|X| \leq M$, on a

$$\sum_{r=0}^{+\infty} E \frac{|XY|^{2r}}{(2r)!} < +\infty.$$

Puisque Y suit une loi de Poisson, on peut écrire dans $\overline{\mathbb{R}}^+$:

$$\begin{aligned}
\sum_{r=0}^{+\infty} E \frac{|XY|^{2r}}{(2r)!} &\leq \sum_{r=0}^{+\infty} \frac{M^{2r}}{(2r)!} \left(\sum_{k=0}^{+\infty} k^{2r} \exp(-\lambda) \frac{\lambda^k}{k!} \right) \\
&\leq \exp(-\lambda) \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \left(\sum_{r=0}^{+\infty} \frac{(Mk)^{2r}}{(2r)!} \right) \\
&= \exp(-\lambda) \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \operatorname{ch}(Mk) \\
&= \frac{\exp(-\lambda)}{2} [\exp(\lambda \exp(M)) + \exp(\lambda \exp(-M))] < +\infty.
\end{aligned}$$

On a donc :

$$E^{\mathcal{B}} \cos(XY) = \sum_{r=0}^{+\infty} (-1)^r E^{\mathcal{B}} \left[\frac{(XY)^{2r}}{(2r)!} \right].$$

Mais, puisque X^{2r} est \mathcal{B} -mesurable et que Y^{2r} et \mathcal{B} sont indépendantes, il vient :

$$E^{\mathcal{B}}(X^{2r} Y^{2r}) = X^{2r} E^{\mathcal{B}}(Y^{2r}) = X^{2r} E(Y^{2r}).$$

Un calcul identique au précédent, les interversions de signes « somme » étant justifiées par l'absolue convergence de la série double donne :

$$\begin{aligned}
E^{\mathcal{B}} \cos(XY) &= \sum_{r=0}^{+\infty} \frac{(-1)^r}{(2r)!} X^{2r} \left(\sum_{k=0}^{+\infty} k^{2r} \exp(-\lambda) \frac{\lambda^k}{k!} \right) \\
&= \exp(-\lambda) \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \left(\sum_{r=0}^{+\infty} \frac{(-1)^r (kX)^{2r}}{(2r)!} \right) \\
&= \exp(-\lambda) \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \cos(kX);
\end{aligned}$$

ou

$$\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \cos(kX) = \Re \left(\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \exp(ikX) \right) = \Re \exp[\lambda \exp(iX)],$$

ce qui donne :

$$\boxed{E^{\mathcal{B}} \cos(XY) = \exp[-\lambda(1 - \cos X)] \times \cos(\lambda \sin X).} \quad (11.22)$$

Si X est quelconque, on définit, pour tout $n \in \mathbb{N}$, $X_n = \mathbf{1}_{\{|X| \leq n\}} X$. Alors, la suite $(\cos(X_n Y))_{n \in \mathbb{N}}$ est P-p.s. convergente, et on a, pour tout $n \in \mathbb{N}$,

$$|\cos(X_n Y)| \leq 1.$$

Il résulte du théorème de convergence dominée conditionnel que la suite $(E^{\mathcal{B}} \cos(X_n Y))_{n \in \mathbb{N}}$ est P-p.s. convergente et que

$$E^{\mathcal{B}} \cos(XY) = \lim_n E^{\mathcal{B}}(\cos(X_n Y)) \quad \text{P-p.s.}$$

La formule (11.22) est donc vraie pour X quelconque.

11.3.5. Inégalité de Jensen

C'est une **inégalité de convexité** souvent utilisée. Nous en donnons d'abord une version élémentaire, puis une version améliorée.

Proposition 11.28 (Inégalité de Jensen). Soient g une fonction convexe² sur \mathbb{R} et $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ telles que $g \circ Y \in \mathcal{L}^1(\Omega, \mathcal{B}, P)$. On a

$$g[E^{\mathcal{B}}Y] \leq E^{\mathcal{B}}[g \circ Y]. \quad (11.23)$$

Démonstration. La fonction g étant convexe, il existe deux suites de réels telles que

$$\forall x \in \mathbb{R} \quad g(x) = \sup_{n \in \mathbb{N}} (a_n x + b_n).$$

On a alors

$$\forall n \in \mathbb{N} \quad \text{P-p.s.} \quad a_n E^{\mathcal{B}}Y + b_n = E^{\mathcal{B}}[a_n Y + b_n] \leq E^{\mathcal{B}}[g \circ Y],$$

et donc (une réunion dénombrable d'ensembles de probabilité nulle est de probabilité nulle)

$$\text{P-p.s.} \quad \forall n \in \mathbb{N} \quad a_n E^{\mathcal{B}}Y + b_n \leq E^{\mathcal{B}}[g \circ Y],$$

ce qui implique :

$$\text{P-p.s.} \quad \sup_{n \in \mathbb{N}} (a_n E^{\mathcal{B}}Y + b_n) \leq E^{\mathcal{B}}[g \circ Y].$$

L'inégalité est démontrée. \square

Corollaire 11.29. Soit $p \in \mathbb{N}^*$. Si $Y \in L^p(\Omega, \mathcal{A}, P)$, on a $E^{\mathcal{B}}Y \in L^p(\Omega, \mathcal{B}, P)$ et

$$\|E^{\mathcal{B}}Y\|_p \leq \|Y\|_p. \quad (11.24)$$

Autrement dit, $E^{\mathcal{B}}$ est une **contraction** de $L^p(\Omega, \mathcal{A}, P)$ sur $L^p(\Omega, \mathcal{B}, P)$.

Démonstration. On applique l'inégalité de Jensen à la fonction convexe $x \mapsto \|x\|^p$. \square

Proposition 11.30. (a) Soit $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ prenant ses valeurs dans un **convexe fermé** K de \mathbb{R} (c'est-à-dire un intervalle fermé). L'espérance conditionnelle $E^{\mathcal{B}}Y$ est P-p.s. à valeurs dans K .

Pour toute fonction g **convexe continue** sur K , à valeurs dans $\mathbb{R} \cup \{+\infty\}$, positive ou telle que $g \circ Y \in \mathcal{L}^1_{\mathbb{R}}(\Omega, \mathcal{A}, P)$, l'inégalité de Jensen (11.23) est satisfaite.

2. On rappelle que toute fonction réelle convexe définie sur un intervalle ouvert de \mathbb{R} est continue. Ceci est faux sur un intervalle non ouvert : prendre g définie sur $[0, +\infty[$ par $g(0) = 1$ et $g(x) = 0$ si $x > 0$.

(b) Soit Y une variable aléatoire à valeurs dans $\overline{\mathbb{R}}^+$. Pour toute fonction g convexe continue sur $\overline{\mathbb{R}}^+$, telle que $g(+\infty) = +\infty$, positive ou telle que $g \circ Y \in \mathcal{L}_{\mathbb{R}}^1(\Omega, \mathcal{A}, \mathbb{P})$, l'inégalité de Jensen (11.23) est satisfaite.

Démonstration. Elle est en tout point identique à celle de la proposition précédente. \square

Remarque. Il n'est pas difficile de généraliser la notion d'espérance conditionnelle au cas où la variable aléatoire Y est à valeurs dans un espace euclidien. La proposition précédente est encore vraie dans ce contexte, tout convexe fermé étant intersection dénombrable de demi-espaces fermés.

11.3.6. Calcul d'espérance conditionnelle

On a déjà donné un exemple de calcul. Un cas particulièrement fréquent est celui où la sous-tribu \mathcal{B} est engendrée par une variable aléatoire X et où il existe une loi conditionnelle de Y sachant X .

Proposition 11.31. Soient X une variable aléatoire à valeurs dans un espace probabilisable quelconque (E, \mathcal{E}) et $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. On suppose qu'existe une loi conditionnelle $\mathbb{P}_Y^{X=x}$ de Y sachant X . Alors $m_Y^{X=x} \circ X$ est une version de l'espérance conditionnelle $E^{\sigma(X)}Y$, $m_Y^{X=x}$ désignant la moyenne conditionnelle de Y sachant X , ce qui s'écrit³ :

$$E^{\sigma(X)}Y = m_Y^{X=x} \circ X \quad \text{P-p.s.}$$

Démonstration. Notons que sous ces hypothèses, la moyenne conditionnelle existe bien et rappelons que $\sigma(X) = \{X^{-1}(C) \mid C \in \mathcal{E}\}$ et que $\mathbf{1}_{X^{-1}(C)} = \mathbf{1}_C \circ X$. Pour tout $C \in \mathcal{E}$, il résulte successivement des théorèmes de transfert et de Fubini généralisé que

$$\begin{aligned} \int_{X^{-1}(C)} m_Y^{X=x} \circ X d\mathbb{P} &= \int_E \mathbf{1}_C(x) m_Y^{X=x} d\mathbb{P}_X(x) = \int_E \mathbf{1}_C(x) \left[\int_{\mathbb{R}} y d\mathbb{P}_Y^{X=x}(y) \right] d\mathbb{P}_X(x) \\ &= \int_{E \times \mathbb{R}} \mathbf{1}_C(x) y d\mathbb{P}_{(X,Y)}(x, y) = \int_{X^{-1}(C)} Y d\mathbb{P}, \end{aligned}$$

ce qui démontre le résultat. \square

Exemple 11.12. Reprenons l'exemple 11.5 où X et Y sont indépendantes de même loi $\exp(\lambda)$. On a vu que pour tout $s > 0$, $\mathbb{P}_X^{S=s}$ est la loi uniforme sur $[0, s]$ et donc que $m_X^{S=s} = s/2$: $S/2$ est alors une version de l'espérance conditionnelle $E^{\sigma(S)}X$.

3. Certains auteurs écrivent l'espérance conditionnelle $E^{\sigma(X)}Y$ sous la forme $E(Y \mid X)$. Nous emploierons quelquefois cette écriture, lorsque le contexte typographique nous y incitera.

Comme le montre l'exemple ci-dessous, ce résultat est indépendant de la nature de la loi de ces variables aléatoires.

Exemple 11.13. Soient X_1 et X_2 deux variables aléatoires réelles indépendantes et de même loi μ . Soit $S = X_1 + X_2$. Démontrer que $E^{\sigma(S)}X_1 = E^{\sigma(S)}X_2$ et en déduire $E^{\sigma(S)}X_1$.

Remarque. Cet exemple sera généralisé ci-dessous en exercice.

Solution : Tenant compte de l'indépendance de X_1 et X_2 , on a, pour tout borélien C de \mathbb{R} ,

$$\int_{S^{-1}(C)} X_1 dP = \int_{\mathbb{R}^2} \mathbf{1}_C(x_1 + x_2)x_1 d(P_{X_1} \otimes P_{X_2})(x_1, x_2)$$

et, puisque X_1 et X_2 ont même loi,

$$\int_{S^{-1}(C)} X_1 dP = \int_{\mathbb{R}^2} \mathbf{1}_C(x_1 + x_2)x_1 d(P_{X_2} \otimes P_{X_1})(x_1, x_2).$$

Il en résulte que :

$$\forall C \in \mathcal{B}_{\mathbb{R}} \quad \int_{S^{-1}(C)} X_1 dP = \int_{S^{-1}(C)} X_2 dP.$$

ce qui démontre l'égalité :

$$\boxed{E^{\sigma(S)}X_1 = E^{\sigma(S)}X_2.}$$

Alors, S étant $\sigma(S)$ -mesurable, on a

$$E^{\sigma(S)}(X_1 + X_2) = S = E^{\sigma(S)}X_1 + E^{\sigma(S)}X_2 \quad \text{P-p.s.}$$

et donc :

$$\boxed{E^{\sigma(S)}X_1 = \frac{S}{2} \quad \text{P-p.s.}}$$

résultat bien naturel.

Exercices

Sauf mention spéciale, toutes les variables aléatoires seront définies sur un même espace probabilisé (Ω, \mathcal{A}, P) .

Exercice 11.1. Lois de Poisson et multinomiale. Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes de loi de Poisson respective $\mathcal{P}(\lambda_i)$, $i = 1, 2, \dots, n$. On note $X = (X_1, X_2, \dots, X_n)$, variable aléatoire à valeurs dans \mathbb{N}^n , et $S_n = \sum_{i=1}^n X_i$. Déterminer une loi conditionnelle $P_X^{S_n = \cdot}$ de X sachant S_n .

Solution. On rappelle que S_n suit une loi de Poisson $\mathcal{P}(\sum_{i=1}^n \lambda_i)$. De plus, pour tout $(k_1, k_2, \dots, k_n, x) \in \mathbb{N}^{n+1}$, on a

$$P\left[\left(\bigcap_{i=1}^n (X_i = k_i)\right) \cap (S_n = x)\right] = \mathbf{1}_{(\sum_{i=1}^n l_i = x)}(k_1, \dots, k_n) P\left[\bigcap_{i=1}^n (X_i = k_i)\right]$$

soit, par indépendance des X_i ,

$$\begin{aligned} P\left[\left(\bigcap_{i=1}^n (X_i = k_i)\right) \cap (S_n = x)\right] \\ = \mathbf{1}_{(\sum_{i=1}^n l_i = x)}(k_1, \dots, k_n) \exp\left(-\sum_{i=1}^n \lambda_i\right) \prod_{i=1}^n \frac{\lambda_i^{k_i}}{k_i!}. \end{aligned}$$

On a donc, pour tout $(k_1, k_2, \dots, k_n, x) \in \mathbb{N}^{n+1}$,

$$\begin{aligned} P_X^{S_n=x}((k_1, k_2, \dots, k_n)) \\ = \mathbf{1}_{(\sum_{i=1}^n l_i = x)}(k_1, k_2, \dots, k_n) \frac{\exp\left(-\sum_{i=1}^n \lambda_i\right) \prod_{i=1}^n \frac{\lambda_i^{k_i}}{k_i!}}{\exp\left(-\sum_{i=1}^n \lambda_i\right) \frac{(\sum_{i=1}^n \lambda_i)^x}{x!}}. \end{aligned}$$

soit

$$\boxed{P_X^{S_n=x}((k_1, k_2, \dots, k_n)) = \mathbf{1}_{(\sum_{i=1}^n l_i = x)}(k_1, k_2, \dots, k_n) \frac{x!}{k_1! k_2! \dots k_n!} \prod_{i=1}^n \left[\left(\frac{\lambda_i}{\sum_{j=1}^n \lambda_j}\right)^{k_i}\right].}$$

c'est-à-dire que, pour tout $x \in \mathbb{N}^*$, $P_X^{S_n=x}$ est la **loi multinomiale**

$$M\left(x; \frac{\lambda_1}{\sum_{j=1}^n \lambda_j}, \frac{\lambda_2}{\sum_{j=1}^n \lambda_j}, \dots, \frac{\lambda_n}{\sum_{j=1}^n \lambda_j}\right).$$

Si $x = 0$, $P_X^{S_n=x}$ est la mesure de Dirac en 0.

Remarque. On retrouvera cette propriété des lois de Poisson ci-dessous, dans l'exercice sur le processus de Poisson : elle sera alors interprétée.

Exercice 11.2. Lois de Bernoulli et uniforme. Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes de même loi de Bernoulli $\mathcal{B}(1, p)$ où $0 < p < 1$. On note $X = (X_1, X_2, \dots, X_n)$, variable aléatoire à valeurs dans \mathbb{N}^n , et $S_n = \sum_{i=1}^n X_i$. Déterminer une loi conditionnelle $P_X^{S_n=x}$ de X sachant S_n .

Solution. On rappelle que S_n suit une loi binomiale $\mathcal{B}(n, p)$. De plus, pour tout $(k_1, k_2, \dots, k_n, x) \in \{0, 1\}^n \times \mathbb{N}$, on a

$$\begin{aligned} P\left[\left(\bigcap_{i=1}^n (X_i = k_i)\right) \cap (S_n = x)\right] \\ = \mathbf{1}_{(\sum_{i=1}^n l_i = x)}(k_1, k_2, \dots, k_n) P\left[\bigcap_{i=1}^n (X_i = k_i)\right]. \end{aligned}$$

soit, par indépendance des X_i ,

$$\begin{aligned} P\left[\left(\bigcap_{i=1}^n (X_i = k_i)\right) \cap (S_n = x)\right] \\ = \mathbf{1}_{(\sum_{i=1}^n l_i = x)}(k_1, k_2, \dots, k_n) \prod_{i=1}^n \left[p^{k_i} (1-p)^{1-k_i}\right] \\ = \mathbf{1}_{(\sum_{i=1}^n l_i = x)}(k_1, k_2, \dots, k_n) p^x (1-p)^{1-x}. \end{aligned}$$

On a donc, pour tout $(k_1, k_2, \dots, k_n, x) \in \{0, 1\}^n \times \mathbb{N}$,

$$P_X^{S_n=x}((k_1, k_2, \dots, k_n)) = \mathbf{1}_{(\sum_{i=1}^n l_i = x)}(k_1, k_2, \dots, k_n) \frac{p^x (1-p)^x}{\binom{n}{x} p^x (1-p)^x},$$

soit

$$P_X^{S_n=x}((k_1, k_2, \dots, k_n)) = \mathbf{1}_{(\sum_{i=1}^n l_i = x)}(k_1, k_2, \dots, k_n) \frac{1}{\binom{n}{x}}.$$

c'est-à-dire que, pour tout $x \in \mathbb{N}^*$, $P_X^{S_n=x}$ est la **loi uniforme** sur l'ensemble $\{(k_1, k_2, \dots, k_n) \in \{0, 1\}^n \mid \sum_{i=1}^n k_i = x\}$. Si $x = 0$, $P_X^{S_n=x}$ est la **mesure de Dirac** en 0.

Exercice 11.3. Processus de Poisson. Soit $(W_n)_{n \in \mathbb{N}^*}$ une suite croissante de variables aléatoires positives telle que $W_0 = 0$. Soit, pour $n \in \mathbb{N}^*$, la variable aléatoire $T_n = W_n - W_{n-1}$. On suppose que les variables aléatoires $T_n, n \in \mathbb{N}^*$, forment une famille de variables aléatoires *indépendantes*, de même **loi exponentielle** $\exp(-\lambda)$, où $\lambda > 0$. On pose $X_0 = 0$ et, pour tout $t > 0$,

$$X_t = \sum_{n \in \mathbb{N}^*} \mathbf{1}_{(W_n \leq t)}.$$

La famille de variables aléatoires $(X_t)_{t \in \mathbb{R}^+}$ est appelée **processus de Poisson d'intensité λ** .

1. Soient s, t tels que $0 \leq s < t$. Calculer par récurrence l'intégrale définie pour tout $n \in \mathbb{N}^*$ par

$$I_n(s, t) = \int_{\mathbb{R}^n} \mathbf{1}_{(s \leq x_1 \leq x_2 \leq \dots \leq x_n \leq t)} d\lambda_n(x_1, x_2, \dots, x_n).$$

2. Calculer, pour tout $n \in \mathbb{N}^*$ et toute famille $(f_j)_{1 \leq j \leq n}$ de fonctions mesurables positives bornées sur \mathbb{R} , la quantité

$$\mathbb{E}\left[\mathbf{1}_{(X_t=n)} \prod_{j=1}^n f_j(W_j)\right].$$

En déduire la loi de X_t et une loi conditionnelle de (W_1, W_2, \dots, W_n) sachant $(X_t = n)$.

3. Soient $t > 0$ puis un entier $k \geq 1$ quelconque et une suite finie quelconque de réels tels que $0 = t_0 \leq t_1 \leq \dots \leq t_k = t$. Déterminer la loi de la variable aléatoire $(X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}})$ et justifier l'indépendance des variables aléatoires $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}$.

On dit que le processus $(X_t)_{t \in \mathbb{R}^+}$ est à accroissements indépendants.

Quelle est, pour tout s, t tels que $0 \leq s < t$, la loi de la variable aléatoire $X_t - X_s$? En déduire sa moyenne $E(X_t - X_s)$.

4. Soit $k \in \mathbb{N}^*$ tel que $1 \leq k \leq n$. Déterminer une loi conditionnelle $P_{W_k}^{X_t = n}$ de W_k sachant $X_t = n$; l'identifier.

Remarque. Le processus de Poisson est un cas particulier des **processus de comptage** : une propriété apparaît aléatoirement au cours du temps : W_n est la date de la n^{e} réalisation, T_n est le temps écoulé entre les $(n-1)^{\text{e}}$ et n^{e} réalisations, X_t est le nombre de réalisations de la propriété dans l'intervalle de temps $[0, t]$; il apparaît en particulier dans les modèles de **file d'attente**.

Solution.

1. Par le théorème de Fubini :

$$I_2(s, t) = \int_{[s, t]} \left(\int_{\{x_1, j\}} d\lambda(x_2) \right) d\lambda(x_1) = \int_s^t (t - x_1) dx_1 = \frac{(t-s)^2}{2}.$$

Supposons que, pour tous s, t tels que $0 \leq s < t$, on ait

$$I_n(s, t) = \frac{(t-s)^n}{n!}. \quad (11.25)$$

Par le théorème de Fubini, on a

$$\begin{aligned} I_{n+1}(s, t) &= \int_{\mathbb{R}} \mathbf{1}_{(s \leq x_1 \leq t)} \left(\int_{\mathbb{R}^n} \mathbf{1}_{(x_1 \leq x_2 \leq \dots \leq x_{n+1} \leq t)} d\lambda_n(x_2, \dots, x_{n+1}) \right) d\lambda(x_1) \\ &= \int_{\mathbb{R}} \mathbf{1}_{(s \leq x_1 \leq t)} I_n(x_1, t) d\lambda(x_1) = \int_s^t \frac{(t-x_1)^n}{n!} dx_1 = \frac{(t-s)^{n+1}}{(n+1)!}; \end{aligned}$$

il en résulte que, pour tout $n \in \mathbb{N}^*$:

$$\boxed{I_n(s, t) = \frac{(t-s)^n}{n!}}.$$

2. Par définition de X_t , on a

$$E \left[\mathbf{1}_{(X_t = n)} \prod_{j=1}^n f_j(W_j) \right] = E \left[\mathbf{1}_{(W_n \leq t) \cap (W_{n+1} > t)} \prod_{j=1}^n f_j(W_j) \right]$$

soit, en utilisant les variables aléatoires T_n (lesquelles portent l'information probabiliste) :

$$E \left[\mathbf{1}_{(X_t = n)} \prod_{j=1}^n f_j(W_j) \right] = E \left[\mathbf{1}_{(\sum_{i=1}^n T_i \leq t) \cap (\sum_{i=1}^{n+1} T_i > t)} \prod_{j=1}^n f_j \left(\sum_{i=1}^j T_i \right) \right].$$

En utilisant le théorème de transfert et le fait que les variables aléatoires T_n sont indépendantes de loi exponentielle de paramètre λ , il vient :

$$\begin{aligned} \mathbb{E}\left[\mathbf{1}_{(X_t=n)} \prod_{j=1}^n f_j(W_j)\right] &= \int_{\mathbb{R}^{n+1}} \left[\mathbf{1}_{(\sum_{i=1}^n t_i \leq t) \cap (\sum_{i=1}^{n+1} t_i > t)} \prod_{j=1}^n f_j\left(\sum_{i=1}^j t_i\right)\right] \\ &\times \left[\prod_{j=1}^{n+1} \mathbf{1}_{\mathbb{R}^+}(t_j)\right] \lambda^{n+1} \exp\left(-\lambda \sum_{i=1}^{n+1} t_i\right) d\lambda_{n+1}(t_1, t_2, \dots, t_{n+1}). \end{aligned}$$

Par le changement de variables sur \mathbb{R}^{n+1} , de jacobien 1, défini par

$$\begin{cases} w_1 = t_1 \\ w_2 = t_1 + t_2 \\ \dots\dots\dots \\ w_{n+1} = t_1 + t_2 + \dots + t_{n+1} \end{cases} \iff \begin{cases} t_1 = w_1 \\ t_2 = w_2 - w_1 \\ \dots\dots\dots \\ t_{n+1} = w_{n+1} - w_n, \end{cases}$$

il vient

$$\begin{aligned} \mathbb{E}\left[\mathbf{1}_{(X_t=n)} \prod_{j=1}^n f_j(W_j)\right] &= \int_{\mathbb{R}^{n+1}} \left[\mathbf{1}_{(w_n \leq t) \cap (w_{n+1} > t)} \prod_{j=1}^n f_j(w_j)\right] \\ &\times \left[\mathbf{1}_{\mathbb{R}^+}(w_1) \prod_{j=2}^{n+1} \mathbf{1}_{\mathbb{R}^+}(w_j - w_{j-1})\right] \lambda^{n+1} \exp(-\lambda w_{n+1}) d\lambda_{n+1}(w_1, w_2, \dots, w_{n+1}), \end{aligned}$$

et, par le théorème de Fubini,

$$\mathbb{E}\left[\mathbf{1}_{(X_t=n)} \prod_{j=1}^n f_j(W_j)\right] = \int_{[t, +\infty[} \lambda \exp(-\lambda w_{n+1}) \times \Phi(t) d\lambda(w_{n+1}),$$

où

$$\Phi(t) = \int_{\mathbb{R}^n} \lambda^n \mathbf{1}_{(w_n \leq t)} \mathbf{1}_{(0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t)} \prod_{j=1}^n f_j(w_j) d\lambda_n(w_1, w_2, \dots, w_n).$$

On a donc :

$$\boxed{\begin{aligned} \mathbb{E}\left[\mathbf{1}_{(X_t=n)} \prod_{j=1}^n f_j(W_j)\right] &= \lambda^n \exp(-\lambda t) \\ &\times \left[\int_{\mathbb{R}^n} \prod_{j=1}^n f_j(w_j) \mathbf{1}_{(0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t)} d\lambda_n(w_1, w_2, \dots, w_n)\right]. \end{aligned}} \quad (11.26)$$

En particulier, si $f_j = 1$ pour tout j , il vient

$$\mathbb{E}\left[\mathbf{1}_{(X_t=n)}\right] = \lambda^n \exp(-\lambda t) \times J_n(0, t),$$

ce qui donne

$$\boxed{P(X_t = n) = \exp(-\lambda t) \frac{(\lambda t)^n}{n!}}$$

c'est-à-dire que X_t suit la loi de Poisson de paramètre λt . On a alors

$$\boxed{EX_t = \lambda t.}$$

De plus, en prenant par exemple $f_j = \mathbf{1}_{A_j}$ où $A_j \in \mathcal{B}_{\mathbb{R}}$, il résulte de l'égalité (11.26) que la variable aléatoire (W_1, W_2, \dots, W_n) admet une densité conditionnelle sachant $(X_t = n)$, $f_{(W_1, W_2, \dots, W_n)}^{X_t=n}$, donnée par en tout $(w_1, w_2, \dots, w_n) \in \mathbb{R}^n$ par

$$\boxed{f_{(W_1, W_2, \dots, W_n)}^{X_t=n}(w_1, w_2, \dots, w_n) = \frac{n!}{t^n} \mathbf{1}_{(0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t)},}$$

c'est-à-dire que la loi conditionnelle de (W_1, W_2, \dots, W_n) sachant $(X_t = n)$ est la loi de Dirichlet.

3. Soient des entiers positifs quelconques $\alpha_1, \alpha_2, \dots, \alpha_k$; notons n leur somme. Définissons, pour j tel que $1 \leq j \leq k$:

$$l_j = \alpha_1 + \alpha_2 + \dots + \alpha_j.$$

Remarquons que, puisque $l_k = n$, et que $\sum_{j=1}^k (t_j - t_{j-1}) = t$, on a

$$\bigcap_{j=1}^k (X_{t_j} - X_{t_{j-1}} = \alpha_j) \subset (X_t = n).$$

et donc:

$$\begin{aligned} \bigcap_{j=1}^k (X_{t_j} - X_{t_{j-1}} = \alpha_j) &= (X_t = n) \cap \left[\bigcap_{j=1}^{k-1} (X_{t_j} = l_j) \right] \\ &= (X_t = n) \cap \left[\bigcap_{j=1}^{k-1} (W_{l_j} \leq t_j) \cap (W_{l_{j+1}} > t_j) \right], \end{aligned}$$

Il résulte alors de l'égalité (11.26) que

$$P \left[\bigcap_{j=1}^k (X_{t_j} - X_{t_{j-1}} = \alpha_j) \right] = \lambda^n \exp(-\lambda t) \times \Psi(t_1, t_2, \dots, t_k),$$

où

$$\begin{aligned} \Psi(t_1, t_2, \dots, t_k) &= \int_{\mathbb{R}^n} \prod_{j=1}^{k-1} \mathbf{1}_{(w_{l_j} \leq t_j) \cap (w_{l_{j+1}} > t_j)} \\ &\quad \mathbf{1}_{(0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t)} d\lambda_n(w_1, w_2, \dots, w_n); \end{aligned}$$

or, en posant $l_0 = 0$, on a

$$\prod_{j=1}^{k-1} \mathbf{1}_{(w_{l_j} \leq t_j) \cap (w_{l_{j+1}} > t_j)} \mathbf{1}_{(0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t)} = \prod_{j=1}^k \mathbf{1}_{(t_{j-1} \leq w_{l_{j-1}+1} \leq \dots \leq w_{l_j} \leq t_j)},$$

compte tenu de ce que $t_j - t_{j-1} = \alpha_j$, le théorème de Fubini permet donc d'écrire

$$P\left[\bigcap_{j=1}^k (X_{t_j} - X_{t_{j-1}} = \alpha_j)\right] = \lambda^n \exp(-\lambda t) \times \prod_{j=1}^k \mathbf{1}_{\alpha_j}(t_{j-1}, t_j),$$

soit

$$P\left[\bigcap_{j=1}^k (X_{t_j} - X_{t_{j-1}} = \alpha_j)\right] = \prod_{j=1}^k \exp[-\lambda(t_j - t_{j-1})] \frac{[\lambda(t_j - t_{j-1})]^{\alpha_j}}{\alpha_j!}. \quad (11.27)$$

Ceci démontre que **les variables aléatoires $X_{t_j} - X_{t_{j-1}}$ sont indépendantes et de loi respective la loi de Poisson de paramètre $\lambda(t_j - t_{j-1})$.**

Les t_j étant quelconques, il en résulte que la loi de $X_t - X_s$ est la loi de Poisson de paramètre $\lambda(t - s)$ et que

$$\lambda = \frac{E(X_t - X_s)}{t - s},$$

d'où le nom, pour le paramètre λ , d'**intensité du processus**.

La variable aléatoire X_t suivant une loi de Poisson de paramètre λt , il résulte de l'égalité (11.27) que, après simplifications, on a

$$P^{X_t=n} \left[\bigcap_{j=1}^k (X_{t_j} - X_{t_{j-1}} = \alpha_j) \right] = \frac{n!}{\alpha_1! \alpha_2! \dots \alpha_k!} \prod_{j=1}^k \left(\frac{t_j - t_{j-1}}{t} \right)^{\alpha_j}.$$

c'est-à-dire que la loi conditionnelle de $(X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}})$ sachant $(X_t = n)$ est la **loi multinomiale** $M\left(n; \frac{t_1}{t}, \frac{t_2 - t_1}{t}, \dots, \frac{t_k - t_{k-1}}{t}\right)$.

Interprétation intuitive : soient n variables aléatoires *indépendantes* X_1, X_2, \dots, X_n de loi *uniforme* sur $]0, t]$ et

$$(Y_1, Y_2, \dots, Y_k) = \left(\sum_{j=1}^n \mathbf{1}_{]0, t_1]}(X_j), \sum_{j=1}^n \mathbf{1}_{]t_1, t_2]}(X_j), \dots, \sum_{j=1}^n \mathbf{1}_{]t_{k-1}, t_k]}(X_j) \right)$$

la variable aléatoire qui indique le **nombre de « points »** dans chaque intervalle $]t_{j-1}, t_j]$; sa loi est la loi multinomiale

$$M\left(n; \frac{t_1}{t}, \frac{t_2 - t_1}{t}, \dots, \frac{t_k - t_{k-1}}{t}\right).$$

On vient de montrer pour le processus de Poisson que, *sachant que la propriété s'est réalisée exactement n fois dans l'intervalle de temps $]0, t]$* , la variable aléatoire donnant le nombre de réalisations de la propriété dans chacun des intervalles de temps $]t_{j-1}, t_j]$ (qui forment une partition de $]0, t]$) a la même loi que celle de (Y_1, Y_2, \dots, Y_k) , et ceci quel que soit la partition choisie! Cela traduit une **uniformité** dans le temps pour la réalisation de cette propriété.

4. Soit $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R})$; en prenant dans l'égalité (11.26) toutes les fonctions f_j égales à 1 sauf f_k que l'on prend égale à f , il vient

$$\begin{aligned} & \mathbb{E}[\mathbf{1}_{\{X_t = n\}} f(W_k)] \\ &= \lambda^n \exp(-\lambda t) \times \left\{ \int_{\mathbb{R}^n} f(w_k) \mathbf{1}_{\{0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t\}} d\lambda_n(w_1, w_2, \dots, w_n) \right\}; \end{aligned}$$

après avoir remarqué que

$$\begin{aligned} & f(w_k) \mathbf{1}_{\{0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t\}} \\ &= \mathbf{1}_{\{0 \leq w_1 \leq w_2 \leq \dots \leq w_{k-1} \leq w_k\}} f(w_k) \mathbf{1}_{\{w_k \leq w_{k+1} \leq \dots \leq w_n \leq t\}}. \end{aligned}$$

le théorème de Fubini permet d'écrire, en intégrant d'abord par rapport aux $k-1$ premières variables, et en utilisant l'égalité (11.25),

$$\mathbb{E}[\mathbf{1}_{\{X_t = n\}} f(W_k)] = \lambda^n \exp(-\lambda t) \times R(t), \quad (11.28)$$

où

$$R(t) = \int_{\mathbb{R}^{n-k+1}} \mathbf{1}_{\{0 \leq w_k \leq t\}} \frac{(w_k)^{k-1}}{(k-1)!} f(w_k) \mathbf{1}_{\{w_k \leq \dots \leq w_n \leq t\}} d\lambda_{n-k+1}(w_k, w_{k+1}, \dots, w_n).$$

Mais, toujours par le théorème de Fubini, on a

$$\begin{aligned} R(t) &= \int_{\mathbb{R}} f(w_k) \mathbf{1}_{\{0 \leq w_k \leq t\}} \frac{(w_k)^{k-1}}{(k-1)!} \\ &\quad \times \left[\int_{\mathbb{R}^{n-k}} \mathbf{1}_{\{w_k \leq w_{k+1} \leq \dots \leq w_n \leq t\}} d\lambda_{n-k}(w_{k+1}, \dots, w_n) \right] d\lambda(w_k), \end{aligned}$$

soit, par définition des intégrales $I_n(s, t)$,

$$R(t) = \int_{\mathbb{R}} f(w_k) \mathbf{1}_{\{0 \leq w_k \leq t\}} \frac{(w_k)^{k-1}}{(k-1)!} \times I_{n-k}(w_k, t) d\lambda(w_k).$$

ou encore, d'après l'égalité (11.25),

$$R(t) = \int_{\mathbb{R}} f(w_k) \mathbf{1}_{\{0 \leq w_k \leq t\}} \frac{(w_k)^{k-1}}{(k-1)!} \frac{(t-w_k)^{n-k}}{(n-k)!} d\lambda(w_k).$$

Ainsi, en reportant dans l'égalité (11.28), on a

$$\begin{aligned} \mathbb{E}[\mathbf{1}_{\{X_t = n\}} f(W_k)] &= \lambda^n \exp(-\lambda t) \\ &\quad \times \int_{\mathbb{R}} f(w_k) \mathbf{1}_{\{0 \leq w_k \leq t\}} \frac{(w_k)^{k-1}}{(k-1)!} \frac{(t-w_k)^{n-k}}{(n-k)!} d\lambda(w_k), \end{aligned}$$

ce qui s'écrit encore

$$\begin{aligned} \mathbb{E}[\mathbf{1}_{\{X_t = n\}} f(W_k)] &= \frac{n!}{t^n} \times \mathbb{P}(X_t = n) \\ &\quad \times \left\{ \int_{\mathbb{R}} f(w_k) \mathbf{1}_{\{0 \leq w_k \leq t\}} \frac{(w_k)^{k-1}}{(k-1)!} \frac{(t-w_k)^{n-k}}{(n-k)!} d\lambda(w_k) \right\}. \end{aligned}$$

Il existe donc une **densité conditionnelle** $f_{W_k}^{X_t=n}$ de W_k sachant $(X_t = n)$ donnée par

$$\forall w \in \mathbb{R} \quad f_{W_k}^{X_t=n}(w) = \mathbf{1}_{[0,t]}(w) \frac{n!}{(k-1)!(n-k)!} \frac{1}{t} \left(\frac{w}{t}\right)^{k-1} \left(1 - \frac{w}{t}\right)^{n-k}.$$

La loi conditionnelle de W_k sachant $(X_t = n)$ est donc une loi bêta de première espèce $B(k, n - k + 1)$ sur $[0, t]$.

Interprétation intuitive. Soit toujours n variables aléatoires indépendantes X_1, X_2, \dots, X_n de loi uniforme sur $]0, t]$ et $X_{(k)}$ la k^{e} « statistique d'ordre » (cf. l'exercice 8 « Loi de Dirichlet et statistique d'ordre » du chapitre 9). On vient de montrer pour le processus de Poisson que, sachant que la propriété s'est réalisée exactement n fois dans l'intervalle de temps $]0, t]$, la date de la k^{e} réalisation de la propriété est une variable aléatoire qui a même loi que celle de $X_{(k)}$. Cela traduit encore une **uniformité** dans le temps pour la réalisation de cette propriété.

Exercice 11.4. Tirage uniforme et intervalle de longueur aléatoire. Soit $\{L, (X_n)_{n \in \mathbb{N}^*}\}$ une famille de variables aléatoires réelles indépendantes de même loi uniforme sur $[0, 1]$. On définit, pour tout $n \in \mathbb{N}^*$, l'application S_n par

$$\forall \omega \in \Omega \quad S_n(\omega) = \sum_{j=1}^n \mathbf{1}_{[0, L(\omega)]}(X_j(\omega)).$$

1. Vérifier que S_n est une variable aléatoire et déterminer une loi conditionnelle $P_{S_n}^{L=\cdot}$ de S_n sachant L .
2. En déduire la loi de S_n .
3. Déterminer une loi conditionnelle $P_L^{S_n=\cdot}$ de L sachant S_n . Calculer la moyenne conditionnelle $m_L^{S_n=\cdot}$ de L sachant S_n et retrouver la moyenne EL de L .

Solution.

1. Les variables aléatoires $Y = (X_1, X_2, \dots, X_n)$ et L sont indépendantes. Si f est l'application définie sur $\mathbb{R}^n \times [0, 1]$ par

$$\forall (y, l) \in \mathbb{R}^n \times [0, 1] \quad f(y, l) = \sum_{j=1}^n \mathbf{1}_{[0, l]}(y_j),$$

on a P.-p.s. $S_n = f(Y, L)$, et donc, pour P_L -presque tout l :

$$P_{S_n}^{L=l} = P_{f(Y, L)}^{L=l} = P_{f(Y, L)},$$

la dernière égalité résultant de l'indépendance de Y et L . Or pour tout $l \in]0, 1]$, $P_{f(Y, L)}$ est la loi binomiale $\mathcal{B}(n, l)$, c'est-à-dire que :

$$\text{pour } P_L\text{-presque tout } l \quad P_{S_n}^{L=l} = \mathcal{B}(n, l).$$

2. On a alors, pour tout $A \in \mathcal{B}_{\mathbb{R}}$:

$$\begin{aligned} P_{S_n}(A) &= \int_{\mathbb{E}} P_{S_n}^{L=l}(A) dP_L(l) \\ &= \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(l) \left[\sum_{k=0}^n \binom{n}{k} l^k (1-l)^{n-k} \delta_k(A) \right] d\lambda(l) \\ &= \sum_{k=0}^n \binom{n}{k} \delta_k(A) \left[\int_{[0,1]} l^k (1-l)^{n-k} d\lambda(l) \right] \\ &= \sum_{k=0}^n \binom{n}{k} \delta_k(A) B(k+1, n-k+1); \end{aligned}$$

puisque

$$\binom{n}{k} B(k+1, n-k+1) = \frac{n!}{k!(n-k)!} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{1}{n+1}$$

on a :

$$P_{S_n}(A) = \sum_{k=0}^n \frac{1}{n+1} \delta_k(A),$$

c'est-à-dire que **la loi de S_n est la loi uniforme sur $\{0, 1, 2, \dots, n\}$.**

3. La loi du couple (S_n, L) est alors déterminée par la donnée, pour tous $A, B \in \mathcal{B}_{\mathbb{R}}$, de $P_{(S_n, L)}(A \times B)$, soit :

$$\begin{aligned} P_{(S_n, L)}(A \times B) &= \int_{\mathbb{B}} P_{S_n}^{L=l}(A) dP_L(l) \\ &= \int_{\mathbb{B}} \left[\sum_{k=0}^n \binom{n}{k} l^k (1-l)^{n-k} \delta_k(A) \right] \mathbf{1}_{[0,1]}(l) d\lambda(l) \\ &= \sum_{k=0}^n \frac{1}{n+1} \delta_k(A) \left[\int_{\mathbb{B}} \mathbf{1}_{[0,1]}(l) \binom{n}{k} (n+1) l^k (1-l)^{n-k} d\lambda(l) \right] \\ &= \int_A \beta_1(k+1, n-k+1)(B) dP_{S_n}(k), \end{aligned}$$

où on note $\beta_1(k+1, n-k+1)(\cdot)$ la loi bêta de première espèce sur $[0, 1]$ de paramètres $k+1$ et $n-k+1$.

Pour tout $k \in \{0, 1, 2, \dots, n\}$, la loi conditionnelle $P_L^{S_n=k}$ de L sachant $S_n = k$ est la loi $\beta_1(k+1, n-k+1)$.

La moyenne conditionnelle de L sachant S_n est alors donnée (cf les tables de loi) par

$$\forall k \in \{0, 1, 2, \dots, n\} \quad m_L^{S_n=k} = \frac{k+1}{n+2}.$$

On retrouve la moyenne de L , puisque

$$EL = \int_{\mathbb{R}} m_L^{S_n=k} dP_{S_n}(k) = \frac{1}{n+1} \sum_{k=0}^n \frac{k+1}{n+2},$$

soit, en tenant compte de ce que

$$\sum_{k=0}^n (k+1) = \frac{(n+1)(n+2)}{2},$$

$$\boxed{EL = \frac{1}{2}.}$$

Exercice 11.5. Parties entières et décimales. Soit X une variable aléatoire positive de densité f_X . Soit $Y = X - [X]$ où $[\cdot]$ désigne la partie entière.

1. Déterminer la loi de la variable aléatoire $([X], Y)$ en fonction de f_X . En déduire les lois de $[X]$ et Y . Retrouver directement la loi de $[X]$.

2. Déterminer les lois conditionnelles de Y sachant $[X]$ et $[X]$ sachant Y , $P_{Y}^{[X]=\cdot}$ et $P_{[X]}^{Y=\cdot}$.

3. On suppose que X suit la loi gamma $\gamma(a, p)$ où $a > 0$ et $p > 0$; pour quelles valeurs du couple (a, p) les variables aléatoires $[X]$ et Y sont-elles indépendantes? Déterminer les lois de $[X]$ et Y dans le cas où X suit la loi exponentielle $\exp(p)$, $p > 0$.

4. On suppose que la densité f_X de X est donnée par

$$\forall x \in \mathbb{R} \quad f_X(x) = \sum_{n=0}^{+\infty} \mathbf{1}_{[n, n+1[}(x) \exp(-\lambda) \frac{\lambda^n}{n!}.$$

Démontrer que les variables aléatoires $[X]$ et Y sont indépendantes et déterminer leur loi.

5. On suppose à nouveau que X suit la loi gamma $\gamma(a, p)$. Calculer les moyennes conditionnelles $m_Y^{[X]=\cdot}$ et $m_{[X]}^{Y=\cdot}$ (on n'explicitera ni les intégrales, ni les sommes de séries intervenant dans les résultats).

Solution.

1. On a, par le théorème de transfert, pour tout $A, B \in \mathcal{B}_{\mathbb{R}}$:

$$\begin{aligned} P_{([X], Y)}(A \times B) &= \int_{\mathbb{R}} \mathbf{1}_A([x]) \mathbf{1}_B(x - [x]) f_X(x) d\lambda(x) \\ &= \sum_{n=0}^{+\infty} \int_{[n, n+1[} \mathbf{1}_A([x]) \mathbf{1}_B(x - [x]) f_X(x) d\lambda(x) \\ &= \sum_{n=0}^{+\infty} \delta_n(A) \int_{[n, n+1[} \mathbf{1}_B(x - n) f_X(x) d\lambda(x). \end{aligned}$$

On a donc, pour tous $A, B \in \mathcal{B}_{\mathbb{R}}$,

$$\boxed{P_{([X], Y)}(A \times B) = \sum_{n=0}^{+\infty} \delta_n(A) \int_{[0, 1[} \mathbf{1}_B(x) f_X(x + n) d\lambda(x).} \quad (11.29)$$

On obtient la loi de $[X]$ en prenant $B = \mathbb{R}$: on a, pour tout $A \in \mathcal{B}_{\mathbb{R}}$,

$$P_{[X]}(A) = P_{([X], Y)}(A \times \mathbb{R}) = \sum_{n=0}^{+\infty} \delta_n(A) \int_{[0,1[} f_X(x+n) d\lambda(x),$$

c'est-à-dire que $[X]$ est une variable aléatoire discrète à valeurs dans \mathbb{N} telle que :

$$\forall n \in \mathbb{N} \quad P([X] = n) = \int_{[0,1[} f_X(x+n) d\lambda(x),$$

résultat que l'on obtient directement en écrivant que

$$([X] = n) = (n \leq X < n+1),$$

ce qui donne

$$P([X] = n) = P_X([n, n+1[) = \int_{[n, n+1[} f_X(x) d\lambda(x),$$

et, par changement de variables,

$$P([X] = n) = \int_{[0,1[} f_X(x+n) d\lambda(x).$$

On obtient la loi de Y en prenant $A = \mathbb{R}$: on a, pour tout $B \in \mathcal{B}_{\mathbb{R}}$,

$$P_Y(B) = P_{([X], Y)}(\mathbb{R} \times B) = \int_B \mathbf{1}_{[0,1[}(x) \left[\sum_{n=0}^{+\infty} f_X(x+n) \right] d\lambda(x).$$

La marginale Y admet donc une densité f_Y donnée par

$$\forall y \in \mathbb{R} \quad f_Y(y) = \mathbf{1}_{[0,1[}(y) \sum_{n=0}^{+\infty} f_X(y+n).$$

2. Pour tous $A, B \in \mathcal{B}_{\mathbb{R}}$, l'égalité (11.29) s'écrit encore :

$$P_{([X], Y)}(A \times B) = \sum_{n \in \text{val}([X])} \delta_n(A) P_{[X]}(\{n\}) \int_B \frac{\mathbf{1}_{[0,1[}(x) f_X(x+n)}{P_{[X]}(\{n\})} d\lambda(x).$$

Pour tout $n \in \text{val}([X])$, Y admet donc une **densité conditionnelle** sachant $[X] = n$ donnée par

$$\forall y \in \mathbb{R} \quad f_Y^{[X]=n}(y) = \mathbf{1}_{[0,1[}(y) \frac{f_X(y+n)}{\int_{[0,1[} f_X(x+n) d\lambda(x)}$$

De même, pour tous $A, B \in \mathcal{B}_{\mathbb{R}}$, l'égalité (11.29) s'écrit encore, avec la convention d'écriture $\frac{0}{0} = 0$,

$$P_{([X], Y)}(A \times B) = \int_B \mathbf{1}_{[0,1]}(y) \frac{\sum_{n=0}^{+\infty} \delta_n(A) f_X(y+n)}{\sum_{n=0}^{+\infty} f_X(y+n)} f_Y(y) d\lambda(y).$$

Pour tout y tel que $f_Y(y) \neq 0$, $[X]$ admet donc une **loi conditionnelle** sachant $Y = y$ donnée par

$$\forall A \in \mathcal{B}_{\mathbb{R}} \quad P_{[X]}^{Y=y}(A) = \sum_{n=0}^{+\infty} \delta_n(A) \frac{f_X(y+n)}{\sum_{n=0}^{+\infty} f_X(y+n)}.$$

3. Si $P_X = \gamma(a, p)$, on a, pour tout $n \in \mathbb{N}$,

$$\int_{[0,1]} f_X(x+n) d\lambda(x) > 0,$$

ce qui implique que $\text{val}([X]) = \mathbb{N}$; on a donc, pour tout $n \in \mathbb{N}$ et tout $y \in \mathbb{R}$,

$$f_X^{[X]=n}(y) = \mathbf{1}_{[0,1]}(y) \frac{\exp(-py)(y+n)^{a-1}}{\int_{[0,1]} \exp(-px)(x+n)^{a-1} d\lambda(x)}.$$

Cette expression n'est indépendante de n que si $a = 1$; **dans ce cas uniquement, c'est-à-dire si $P_X = \exp(p)$, les variables aléatoires Y et $[X]$ sont indépendantes.**

Si $P_X = \exp(p)$, on a

$$P([X] = n) = \exp(-pn) \int_{[0,1]} p \exp(-px) d\lambda(x),$$

soit

$$P([X] = n) = \exp(-pn)(1 - \exp(-p)),$$

c'est-à-dire que $[X]$ suit la loi géométrique sur \mathbb{N} de paramètre $1 - \exp(-p)$, et

$$\forall y \in \mathbb{R} \quad f_Y(y) = \mathbf{1}_{[0,1]}(y) \sum_{n=0}^{+\infty} p \exp(-p(y+n))$$

soit

$$\forall y \in \mathbb{R} \quad f_Y(y) = \mathbf{1}_{[0,1]}(y) \frac{p}{1 - \exp(-p)} \exp(-py).$$

4. Dans ce cas, on a, d'après l'égalité (11.29), pour tous $A, B \in \mathcal{B}_{\mathbb{R}}$:

$$P_{([X], Y)}(A \times B) = \sum_{n=0}^{+\infty} \delta_n(A) \int_B \mathbf{1}_{[0,1]}(\lambda) \cdot \left[\sum_{k=0}^{+\infty} \mathbf{1}_{\{k, k+1\}}(tx + n) \exp(-\lambda) \frac{\lambda^k}{k!} \right] d\lambda(x),$$

soit

$$P_{([X], Y)}(A \times B) = \left[\sum_{n=0}^{+\infty} \delta_n(A) \exp(-\lambda) \frac{\lambda^n}{n!} \right] \left[\int_B \mathbf{1}_{[0,1]}(x) d\lambda(x) \right],$$

ce qui démontre qu'alors Y et $[X]$ sont indépendantes et de loi respective la loi uniforme sur $[0, 1]$ et la loi de Poisson $\mathcal{P}(\lambda)$.

5. On a

$$\forall n \in \text{val}([X]) \quad m_Y^{[X]=n} = \int_{\mathbb{R}} y f_Y^{[X]=n}(y) d\lambda(y),$$

soit

$$\forall n \in \mathbb{N} \quad m_Y^{[X]=n} = \frac{\int_{[0,1]} y \exp(-py)(y+n)^{n-1} d\lambda(y)}{\int_{[0,1]} \exp(-px)(x+n)^{a-1} d\lambda(x)};$$

de plus, pour tout y tel que $f_Y(y) \neq 0$, on a

$$m_{[X]}^{Y=y} = \int_{\mathbb{R}} x dP_{[X]}^{Y=y}(x),$$

soit :

$$\forall y \in [0, 1[\quad m_{[X]}^{Y=y} = \frac{\sum_{n=0}^{+\infty} n \exp(-pn)(y+n)^{a-1}}{\sum_{n=0}^{+\infty} \exp(-pn)(y+n)^{a-1}}.$$

Remarque. Dans le cas où $a = 2$, $m_{[X]}^{Y=y}$ est la restriction à l'intervalle $[0, 1[$ d'une fonction homographe.

Exercice 11.6. Espérance conditionnelle et variable aléatoire gaussienne; différentes méthodes de calcul. Soient \mathcal{B} une sous-tribu de \mathcal{A} , X et Y deux variables aléatoires réelles telles que X soit \mathcal{B} -mesurable et que les tribus \mathcal{B} et $\sigma(Y)$ soient indépendantes. On suppose de plus que Y a la loi normale $\mathcal{N}_{\mathbb{R}}(0, 1)$.

1. Démontrer l'équivalence des trois propriétés suivantes :

- (i) $\exp(\frac{X^2}{2})$ est P -intégrable
- (ii) $\exp(XY)$ est P -intégrable
- (iii) $\exp(|XY|)$ est P -intégrable

2. On suppose que $\exp(\frac{X^2}{2})$ est P -intégrable.

(a) Sans calculer l'espérance conditionnelle, démontrer que

$$E^{\mathcal{B}} \exp(XY) \geq 1 \quad \text{P-p.s.}$$

(b) Dans le cas où $\mathcal{B} = \sigma(X)$, calculer $E^{\mathcal{B}} \exp(XY)$.

(c) Dans le cas général (a priori $\sigma(X) \subsetneq \mathcal{B}$), calculer par deux méthodes l'espérance conditionnelle $E^{\mathcal{B}} \exp(XY)$, d'abord en faisant un développement en série de l'exponentielle, puis en utilisant la proposition 11.22.

3. On ne suppose plus que $\exp(\frac{X^2}{2})$ soit P-intégrable. Calculer $E^{\mathcal{B}} \exp(XY)$.

Solution.

1. Les variables aléatoires X et Y sont indépendantes. Il résulte alors des théorèmes de transfert et de Fubini que l'on a, dans \mathbb{R}^+ ,

$$\begin{aligned} \int_{\Omega} \exp(XY) dP &= \int_{\mathbb{R}^2} \exp(xy) dP_X \otimes P_Y(x, y) \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \exp(xy) dP_Y(y) \right] dP_X(x). \end{aligned}$$

Mais, Y étant gaussienne, en tenant compte de l'égalité

$$\exp(xy) \exp\left(-\frac{y^2}{2}\right) = \exp\left(\frac{x^2}{2}\right) \exp\left(-\frac{(y-x)^2}{2}\right),$$

on a, pour tout $x \in \mathbb{R}$:

$$\boxed{\int_{\mathbb{R}} \exp(xy) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) d\lambda(y) = \exp\left(\frac{x^2}{2}\right).} \quad (11.30)$$

Il en résulte que

$$\int_{\Omega} \exp(XY) dP = \int_{\Omega} \exp\left(\frac{X^2}{2}\right) dP. \quad (11.31)$$

L'équivalence de (i) et (ii) est alors claire.

Par ailleurs, en tenant compte de l'égalité, pour tout $x \in \mathbb{R}$,

$$\begin{aligned} \exp(x|y|) \exp\left(-\frac{y^2}{2}\right) &= \mathbf{1}_{\mathbb{R}^+}(y) \exp\left(\frac{x^2}{2}\right) \exp\left(-\frac{(y-x)^2}{2}\right) \\ &+ \mathbf{1}_{\mathbb{R}^-}(y) \exp\left(\frac{x^2}{2}\right) \exp\left(-\frac{(y+x)^2}{2}\right). \end{aligned}$$

on a l'inégalité (intégrer sur \mathbb{P}^+ et \mathbb{R}^{-*}) :

$$\int_{\mathbb{R}} \exp(|xy|) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) d\lambda(y) \leq 2 \exp\left(\frac{x^2}{2}\right).$$

Il en résulte que

$$\int_{\Omega} \exp(|XY|) dP \leq 2 \int_{\Omega} \exp\left(\frac{X^2}{2}\right) dP,$$

ce qui démontre que (i) implique (iii). Enfin (iii) implique (ii) puisque $\exp(XY) \leq \exp(|XY|)$.

(a) Il suffit d'appliquer l'inégalité de Jensen pour obtenir :

$$\exp \left[E^{\mathcal{B}}(XY) \right] \leq E^{\mathcal{B}} \exp(XY).$$

Mais, X étant \mathcal{B} -mesurable, on a

$$E^{\mathcal{B}}(XY) = X E^{\mathcal{B}}(Y).$$

Les tribus \mathcal{B} et $\sigma(Y)$ étant indépendantes et Y étant centrée, il vient

$$E^{\mathcal{B}}(XY) = X E(Y) = 0,$$

et donc :

$$1 \leq E^{\mathcal{B}} \exp(XY) \quad \text{P-p.s.}$$

(b) Si $\mathcal{B} = \sigma(X)$, une version de $E^{\mathcal{B}} \exp(XY)$ est obtenue en composant avec X la moyenne conditionnelle de $\exp(XY)$ sachant X . Mais, par le théorème de transfert conditionnel, on a, pour P_X -presque tout x ,

$$m_{\exp(XY)}^{X=x} = m_{\exp(xY)}^{X=x} = E \exp(xY).$$

(la dernière égalité résultant de l'indépendance de X et Y), ce qui donne, d'après l'égalité (11.30) :

$$m_{\exp(XY)}^{X=x} = \exp\left(\frac{x^2}{2}\right).$$

Il en résulte que :

$$E^{\mathcal{B}} \exp(XY) = \exp\left(\frac{X^2}{2}\right) \quad \text{P-p.s.}$$

(c) **Cas général.**

• **Première méthode.** Puisque, pour tout $p \in \mathbb{N}$, on a

$$\left| \sum_{n=0}^p \frac{(XY)^n}{n!} \right| \leq \sum_{n=0}^{+\infty} \frac{|XY|^n}{n!} = \exp(|XY|),$$

et que $\exp(|XY|)$ est P -intégrable, il résulte du théorème de convergence dominée pour l'espérance conditionnelle que

$$E^{\mathcal{B}} \exp(XY) = \sum_{n=0}^{+\infty} E^{\mathcal{B}} \frac{(XY)^n}{n!}.$$

Mais, X^n étant \mathcal{B} -mesurable, on a

$$E^{\mathcal{B}}[(XY)^n] = X^n E^{\mathcal{B}}(Y^n).$$

Les tribus \mathcal{B} et $\sigma(Y)$ étant indépendantes, il vient

$$E^{\mathcal{B}}[(XY)^n] = X^n E(Y^n).$$

La variable aléatoire Y étant **gaussienne, centrée réduite**, un calcul classique sur les moments (intégration par parties) conduit aux relations, pour tout $p \in \mathbb{N}^*$:

$$E(Y^{2p+1}) = 0 \quad E(Y^{2p}) = 1 \cdot 3 \cdot 5 \cdots (2p-1).$$

Par conséquent, on a

$$E^{\mathcal{B}} \exp(XY) = \sum_{n=0}^{+\infty} \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{(2n)!} X^{2n} = \sum_{n=0}^{+\infty} \frac{1}{n!} \left(\frac{X^2}{2}\right)^n,$$

soit encore :

$$E^{\mathcal{B}} \exp(XY) = \exp\left(\frac{X^2}{2}\right) \quad \text{P-p.s.}$$

• **Deuxième méthode.** Si f est la fonction définie par

$$\forall x \in \mathbb{R} \quad f(x) = E \exp(xY),$$

l'égalité (11.30) s'écrit

$$\forall x \in \mathbb{R} \quad f(x) = \exp\left(\frac{x^2}{2}\right),$$

et la proposition 11.22 affirme que $f \circ X$ est une version de $E^{\mathcal{B}} \exp(XY)$.

Remarque. On retrouve alors le résultat précédent; en effet, en raison de l'inclusion des tribus $\mathcal{B} \supset \sigma(X)$ et de la $\sigma(X)$ -mesurabilité de la variable aléatoire $\exp\left(\frac{X^2}{2}\right)$, on a

$$E^{\sigma(X)} \exp(XY) = E^{\sigma(X)} \left[E^{\mathcal{B}} \exp(XY) \right] = \exp\left(\frac{X^2}{2}\right) \quad \text{P-p.s.}$$

2. Soit, pour tout $n \in \mathbb{N}$, la variable aléatoire $X_n = \mathbf{1}_{\{|X| \leq n\}} X$: elle est bornée par n , et $\exp\left(\frac{X_n^2}{2}\right)$ est alors P-intégrable. On a donc, pour tout $n \in \mathbb{N}$,

$$E^{\mathcal{B}} \exp(X_n Y) = \exp\left(\frac{X_n^2}{2}\right) \quad \text{P-p.s.}$$

La suite de terme général positif $\exp(X_n^2/2)$ converge P-p.s. en croissant vers $\exp(X^2/2)$. De plus, le théorème de Fatou-Lebesgue conditionnel 11.26 donne

$$\begin{aligned} E^{\mathcal{B}} \exp(XY) &= E^{\mathcal{B}} [\liminf_n \exp(X_n Y)] \leq \liminf_n E^{\mathcal{B}} \exp(X_n Y) \\ &\leq \limsup_n E^{\mathcal{B}} \exp(X_n Y) \leq E^{\mathcal{B}} [\limsup_n \exp(X_n Y)] = E^{\mathcal{B}} \exp(XY). \end{aligned}$$

Il en résulte que la suite de terme général $E^{\mathcal{B}} \exp(X_n Y)$ converge P-p.s. et que

$$E^{\mathcal{B}} \exp(XY) = \lim_n E^{\mathcal{B}} \exp(X_n Y).$$

Il vient alors

$$E^{\mathcal{B}} \exp(XY) = \exp\left(\frac{X^2}{2}\right) \quad \text{P-p.s.}$$

Exercice 11.7. Espérance conditionnelle et indépendance. Soient $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ et deux sous-tribus \mathcal{A}_1 et \mathcal{A}_2 telles que les tribus $\mathcal{A}_1 \vee \sigma(X)$ et \mathcal{A}_2 soient indépendantes ($\mathcal{A}_1 \vee \sigma(X)$ désigne la tribu engendrée par \mathcal{A}_1 et $\sigma(X)$, c'est-à-dire la plus petite tribu contenant \mathcal{A}_1 et $\sigma(X)$). Démontrer l'égalité

$$E^{\mathcal{A}_1 \vee \mathcal{A}_2} X = E^{\mathcal{A}_1} X.$$

Solution. Pour tous $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$, puisque $\mathbf{1}_{A_1} X$ et $\mathbf{1}_{A_2}$ sont indépendantes, on a

$$E[\mathbf{1}_{A_1 \cap A_2} X] = E[\mathbf{1}_{A_1} \mathbf{1}_{A_2} X] = E[\mathbf{1}_{A_1} X] E[\mathbf{1}_{A_2}],$$

et donc, par définition de $E^{\mathcal{A}_1 \vee \mathcal{A}_2} X$,

$$E[\mathbf{1}_{A_1 \cap A_2} X] = E[\mathbf{1}_{A_1} E^{\mathcal{A}_1 \vee \mathcal{A}_2} X] E[\mathbf{1}_{A_2}].$$

Mais $\mathbf{1}_{A_1} E^{\mathcal{A}_1 \vee \mathcal{A}_2} X$ et $\mathbf{1}_{A_2}$ sont indépendantes : il en résulte que

$$E[\mathbf{1}_{A_1 \cap A_2} X] = E[(\mathbf{1}_{A_1} E^{\mathcal{A}_1 \vee \mathcal{A}_2} X) \mathbf{1}_{A_2}] = E[\mathbf{1}_{A_1 \cap A_2} E^{\mathcal{A}_1 \vee \mathcal{A}_2} X].$$

Puisque $\{A_1 \cap A_2 \mid A_1 \in \mathcal{A}_1 \text{ et } A_2 \in \mathcal{A}_2\}$ est un π -système qui engendre la tribu $\mathcal{A}_1 \vee \mathcal{A}_2$, il résulte alors du théorème de prolongement par mesurabilité que l'on a :

$$\forall A \in \mathcal{A}_1 \vee \mathcal{A}_2 \quad E[\mathbf{1}_A X] = E[\mathbf{1}_A E^{\mathcal{A}_1 \vee \mathcal{A}_2} X].$$

Pour conclure, il reste à remarquer que $E^{\mathcal{A}_1 \vee \mathcal{A}_2} X$ est $\mathcal{A}_1 \vee \mathcal{A}_2$ -mesurable.

Exercice 11.8. Sur la voie d'une loi forte des grands nombres. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes et de même loi μ . Soit $S_n = \sum_{j=1}^n X_j$. Démontrer que pour tout i tel que $1 \leq i \leq n$, on a

$$E^{\sigma(S_n)} X_i = E^{\sigma(S_n)} X_1$$

et en déduire $E^{\sigma(S_n)} X_1$, puis $E^{\mathcal{A}_n} X_1$, où $\mathcal{A}_n = \sigma(S_{n+j} \mid j \in \mathbb{N})$ (utiliser le résultat de l'exercice précédent).

Remarque. Cet exercice généralise l'exemple 11.13.

Solution. Tenant compte de l'indépendance des X_i , on a pour tout borélien C de \mathbb{R} :

$$\int_{S_n^{-1}(C)} X_i dP = \int_{\mathbb{R}^n} \mathbf{1}_C(x_1 + x_2 + \dots + x_n) \nu_i dP_{X_1} \otimes P_{X_2} \otimes \dots \otimes P_{X_n}(x_1, x_2, \dots, x_n)$$

et, puisque les X_i ont même loi μ ,

$$\begin{aligned} \int_{S_n^{-1}(C)} X_i dP &= \int_{\mathbb{R}^n} \mathbf{1}_C(x_1 + x_2 + \dots + x_n) x_i d\mu^{\otimes n}(x_1, x_2, \dots, x_n) \\ &= \int_{\mathbb{R}^n} \mathbf{1}_C(x_1 + x_2 + \dots + x_n) x_1 d\mu^{\otimes n}(x_1, x_2, \dots, x_n). \end{aligned}$$

Il en résulte que :

$$\forall C \in \mathcal{B}_{\mathbb{R}} \quad \int_{S_n^{-1}(C)} X_i dP = \int_{S_n^{-1}(C)} X_1 dP,$$

ce qui démontre l'égalité demandée.

On a alors S_n étant $\sigma(S_n)$ -mesurable,

$$E^{\sigma(S_n)}(S_n) = S_n = nE^{\sigma(S_n)}X_1,$$

et donc :

$$E^{\sigma(S_n)}X_1 = \frac{S_n}{n}.$$

Puisque pour tout $k \in \mathbb{N}^*$

$$S_{n+k} = S_n + \sum_{j=n+1}^{n+k} X_j.$$

on a l'égalité de tribus :

$$\mathcal{A}_n = \sigma(S_n) \vee \sigma\left(\sum_{j=n+1}^{n+k} X_j \mid k \in \mathbb{N}^*\right).$$

Il résulte de l'exercice précédent que

$$E^{\mathcal{A}_n}X_1 = E^{\sigma(S_n)}X_1,$$

ce qui donne

$$E^{\mathcal{A}_n}X_1 = \frac{S_n}{n}.$$

Exercice 11.9. Méthode de simulation par rejet et introduction aux méthodes de chaînes de Markov. Soient f et g deux densités de probabilité sur \mathbb{R} (par rapport à la mesure de Lebesgue). L'objectif est de simuler une variable aléatoire Y de densité f de forme analytique a priori « compliquée » en utilisant g choisie « voisine » de f et de forme analytique « plus simple ». On pose (avec la convention $\frac{0}{0} = 0$) et suppose :

$$t(x) = \frac{f(x)}{g(x)} \quad \text{et} \quad 1 < \bar{t} = \sup_{x \in \mathbb{R}} t(x) < +\infty.$$

On considère une famille de variables aléatoires indépendantes $\{X_n, Y_p \mid n \in \mathbb{N}^*, p \in \mathbb{N}^*\}$ telle que, pour tout $n \in \mathbb{N}^*$, X_n soit de densité g et Y_n de loi uniforme

sur l'intervalle $[0, \bar{t}]$. On introduit les variables aléatoires à valeurs respectivement dans \mathbb{R}^2 et $\mathbb{R} \cup \{+\infty\}$:

$$M_n = (X_n, Y_n) \quad \text{et} \quad X_\infty = \limsup_n X_n.$$

On considère l'ensemble $G = \{(x, y) \in \mathbb{R}^2 \mid t(x) \geq y\}$ et les applications T et X_T à valeurs respectivement dans $\bar{\mathbb{N}}$ et $\mathbb{R} \cup \{+\infty\}$ définies en tout $\omega \in \Omega$ par

$$T(\omega) = \inf\{n \in \mathbb{N}^* \mid M_n(\omega) \in G\} \quad \text{et} \quad X_T(\omega) = X_{T(\omega)}(\omega),$$

où on fait la convention $\inf \emptyset = +\infty$.

1. Démontrer que pour toute fonction φ mesurable bornée sur \mathbb{R} , la quantité

$$I(\varphi) \equiv E\left[\mathbf{1}_{(M_n \in G)} \varphi(X_n)\right]$$

est égale, à une constante près que l'on déterminera, à l'intégrale $\int_{\mathbb{R}} \varphi f \, d\lambda$. En déduire la probabilité $P(M_n \in G)$.

2. Vérifier que T et X_T sont des variables aléatoires; déterminer la loi de T et en déduire que X_T est P-p.s. finie. Calculer la moyenne de T .

3. Déterminer la loi de X_T .

4. Toujours avec la convention $\inf \emptyset = +\infty$, on définit la suite d'applications à valeurs dans $\bar{\mathbb{N}}$ définies par $T_1 = T$ et, pour tout $k \in \mathbb{N}^*$,

$$\forall \omega \in \Omega \quad T_{k+1}(\omega) = \inf\{n > T_k(\omega) \mid M_n \in G\}.$$

On note \mathcal{A}_n la tribu $\sigma(M_j \mid 1 \leq j \leq n)$.

(a) Démontrer, par récurrence sur k , que pour tout $k \in \mathbb{N}^*$, pour tout $n \in \mathbb{N}^*$, $(T_k = n) \in \mathcal{A}_n$ et que T_k est P-p.s. fini. On définit pour tout $k \in \mathbb{N}^*$ la famille d'événements

$$\mathcal{A}_{T_k} = \{A \in \mathcal{A} \mid A \cap (T_k = n) \in \mathcal{A}_n \quad \forall n \in \mathbb{N}^*\}.$$

Vérifier que, pour tout $k \in \mathbb{N}^*$, on a l'inclusion :

$$\mathcal{A}_{T_k} \subset \mathcal{A}_{T_{k+1}}. \quad (11.32)$$

(b) Soit, pour tout $k \in \mathbb{N}^*$, f_k une fonction mesurable positive bornée quelconque; calculer pour tout $n \in \mathbb{N}^*$, l'espérance conditionnelle

$$E^{\mathcal{A}_n} \left[\mathbf{1}_{(T_k = n)} f_{k+1}(X_{T_{k+1}}) \right],$$

En déduire l'espérance conditionnelle

$$E^{\mathcal{A}_{T_k}} \left[f_{k+1}(X_{T_{k+1}}) \right],$$

puis la loi de $X_{T_{k+1}}$.

- (c) Démontrer que pour tout $k \in \mathbb{N}^*$, la variable aléatoire $f_k(X_{T_k})$ est \mathcal{A}_{T_k} -mesurable et en déduire que $(X_{T_k})_{k \in \mathbb{N}^*}$ est une suite de variables aléatoires indépendantes.
- (d) **Application numérique.** Prendre, pour $a > 2$ et $b > 0$ tels que $ab > 1$, f et g définies par

$$f(x) = \mathbf{1}_{\mathbb{R}^+}(x) \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) \quad \text{et} \quad g(x) = \mathbf{1}_{\mathbb{R}^+}(x) \frac{1}{a} \exp\left(-\frac{x}{a}\right).$$

Vérifier que t est bornée, déterminer \hat{x} tel que $t(\hat{x}) = \bar{t}$. Dans les deux cas suivants, $b = 1$ et $a = \frac{5}{2}$, puis $b = 1$ et $a = \frac{9}{2}$, vérifier que $\bar{t} > 1$ et déterminer une valeur numérique à 10^{-2} près de $P(M_n \in G)$ et ET. (On rappelle que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$).

Solution.

1. Le théorème de transfert, s'applique, φ étant bornée; de plus, X_n et Y_n étant indépendantes, de densité respectives g et $\mathbf{1}_{[0, \bar{t}]}$, on a

$$I(\varphi) \equiv E[\mathbf{1}_{(M_n \in G)} \varphi(X_n)] = \int_{\mathbb{R}^2} \mathbf{1}_G(x, y) \varphi(x) \frac{1}{\bar{t}} \mathbf{1}_{[0, \bar{t}]}(y) g(x) d\lambda_2(x, y),$$

et, par le théorème de Fubini (φ est bornée) :

$$I(\varphi) = \int_{\mathbb{R}} \varphi(x) g(x) \left[\int_{[0, t(x)]} \frac{1}{\bar{t}} d\lambda(y) \right] d\lambda(x) = \frac{1}{\bar{t}} \int_{\{g \neq 0\}} \varphi(x) f(x) d\lambda(x).$$

L'inclusion $\{g = 0\} \subset \{f = 0\}$ résulte des inégalités : pour tout $x \in \mathbb{R}$, $0 \leq f(x) \leq \bar{t}g(x)$, et $\bar{t} > 0$; il s'en suit que l'on a

$$\int_{\{g=0\}} |\varphi(x)| f(x) d\lambda(x) \leq \int_{\{f=0\}} |\varphi(x)| f(x) d\lambda(x) = 0,$$

et donc :

$$I(\varphi) = \frac{1}{\bar{t}} \int_{\mathbb{R}} \varphi(x) f(x) d\lambda(x).$$

En particulier, si $\varphi = 1$, on obtient :

$$P(M_n \in G) = \frac{1}{\bar{t}}.$$

2. T est bien une variable aléatoire à valeurs dans $\overline{\mathbb{N}}$, puisque l'on a $(T = 1) = (M_1 \in G) \in \mathcal{A}$ et que, pour tout $n \geq 2$, on a

$$(T = n) = \left[\bigcap_{j=1}^{n-1} (M_j \notin G) \cap (M_n \in G) \right] \in \mathcal{A};$$

il en résulte bien sûr que $(T = +\infty) = (T \in \mathbb{N}^*)^c \in \mathcal{A}$. De plus, les variables aléatoires M_n étant indépendantes de même loi, on a

$$P(T = 1) = P(M_1 \in G) = \frac{1}{\bar{t}},$$

et, pour tout $n \geq 2$,

$$P(T = n) = \prod_{j=1}^{n-1} P(M_j \notin G) \times P(M_n \in G) = \frac{1}{\bar{t}} \left(1 - \frac{1}{\bar{t}}\right)^{n-1}.$$

c'est-à-dire que T a la loi géométrique sur \mathbb{N}^* , $\mathcal{G}_{\mathbb{N}^*}\left(\frac{1}{\bar{t}}\right)$. Il en résulte que T est P-p.s. finie, et, puisque

$$(X_T = +\infty) \subset (T = +\infty),$$

que X_T est P-p.s. finie. De plus, on a

$$\boxed{ET = \bar{t}.}$$

3. Pour tout $\varphi \in \mathcal{C}_{\mathcal{X}}^+(\mathbb{R})$, on a alors :

$$\begin{aligned} E\varphi(X_T) &= \sum_{n \in \mathbb{N}^*} \int_{(T=n)} \varphi(X_n) dP \\ &= \int_{(M_1 \in G)} \varphi(X_1) dP + \sum_{n \geq 2} \int_{\Omega} \left[\prod_{j=1}^{n-1} \mathbf{1}_{(M_j \notin G)} \right] \mathbf{1}_{(M_n \in G)} \varphi(X_n) dP. \end{aligned}$$

Les variables aléatoires $\prod_{j=1}^{n-1} \mathbf{1}_{(M_j \notin G)}$ et $\mathbf{1}_{(M_n \in G)} \varphi(X_n)$ étant indépendantes, il résulte de la première question que l'on a

$$\begin{aligned} E\varphi(X_T) &= I(\varphi) \left[1 + \sum_{n \geq 2} \prod_{j=1}^{n-1} P(M_j \notin G) \right] \\ &= I(\varphi) \left[1 + \sum_{n \geq 2} \left(1 - \frac{1}{\bar{t}}\right)^{n-1} \right] \\ &= I(\varphi) \frac{1}{1 - \left(1 - \frac{1}{\bar{t}}\right)} \\ &= \bar{t} I(\varphi). \end{aligned}$$

On a ainsi démontré que :

$$\boxed{\forall \varphi \in \mathcal{C}_{\mathcal{X}}^+(\mathbb{R}) \quad E\varphi(X_T) = \int_{\mathbb{R}} \varphi(x) f(x) d\lambda(x);}$$

autrement dit, X_T admet la densité f .

(a) La propriété est déjà démontrée pour T_1 . Supposons la vérifiée jusqu'à l'entier k . Si $n < k$, on a $(T_{k+1} = n) = \emptyset \in \mathcal{A}_n$. Si $n \geq k+1$, on a

$$(T_{k+1} = n) = \bigcup_{j=1}^{n-1} \left[(T_k = j) \cap \bigcap_{i=j+1}^{n-1} (M_i \notin G) \cap (M_n \in G) \right], \quad (11.33)$$

où on pose

$$\bigcap_{i=n}^{n-1} (M_i \notin G) = \Omega :$$

en vertu de l'hypothèse de récurrence et du fait que, si $1 \leq j \leq n-1$, on a $(T_k = j) \in \mathcal{A}_j \subset \mathcal{A}_n$, il en résulte que $(T_{k+1} = n) \in \mathcal{A}_n$.

On sait déjà que T_k est P.-p.s. fini. **Supposons que T_k soit P.-p.s. fini.** Il résulte des égalités ensemblistes ci-dessus que l'on a, pour tout $n \in \mathbb{N}^*$,

$$P(T_{k+1} = n) = \frac{1}{l} P(T_k = n-1) + \sum_{j=1}^{n-2} \left[P(T_k = j) \frac{1}{l} \left(1 - \frac{1}{l}\right)^{n-j-1} \right];$$

en sommant sur \mathbb{N}^* , il vient

$$\begin{aligned} \sum_{n \in \mathbb{N}^*} P(T_{k+1} = n) &= \frac{1}{l} \sum_{n \in \mathbb{N}^*} P(T_k = n-1) + \sum_{n \in \mathbb{N}^*} \sum_{j=1}^{n-2} \left[P(T_k = j) \frac{1}{l} \left(1 - \frac{1}{l}\right)^{n-j-1} \right] \\ &= \frac{1}{l} P(T_k < +\infty) + \frac{1}{l} \sum_{j \in \mathbb{N}^*} \left[P(T_k = j) \sum_{n \geq j+2} \left(1 - \frac{1}{l}\right)^{n-j-1} \right] \\ &= \frac{1}{l} P(T_k < +\infty) + \frac{1}{l} \sum_{j \in \mathbb{N}^*} \left[P(T_k = j) \left(1 - \frac{1}{l}\right)^j \right] \\ &= P(T_k < +\infty), \end{aligned}$$

et donc $P(T_{k+1} < +\infty) = 1$. Il en est donc ainsi pour tout k .

Soit $A \in \mathcal{A}_{T_k}$: l'égalité (11.33) montre que l'on a, pour tout $n \in \mathbb{N}^*$,

$$A \cap (T_{k+1} = n) \in \mathcal{A}_n,$$

c'est-à-dire que $A \in \mathcal{A}_{T_{k+1}}$; par conséquent on a : $\mathcal{A}_{T_k} \subset \mathcal{A}_{T_{k+1}}$.

(b) On a alors

$$\begin{aligned} \mathbf{1}_{(T_k = n)} f_{k+1}(X_{T_{k+1}}) &= \sum_{l \in \mathbb{N}^*} \mathbf{1}_{(T_k = n)} f_{k+1}(X_{n+l}) \mathbf{1}_{(T_{k+1} = n+l)} \\ &= \sum_{l \in \mathbb{N}^*} \left[\mathbf{1}_{(T_k = n)} f_{k+1}(X_{n+l}) \prod_{j=1}^{l-1} \mathbf{1}_{(M_{n+j} \notin G)} \mathbf{1}_{(M_{n+l} \in G)} \right], \end{aligned}$$

avec la convention $\prod_{j=1}^0 \mathbf{1}_{(M_{n+j} \notin G)} = 1$.

En prenant l'espérance conditionnelle par rapport à \mathcal{A}_n , et en tenant compte de ce que $(T_k = n) \in \mathcal{A}_n$, on a

$$\begin{aligned} E^{\mathcal{A}_n} [\mathbf{1}_{(T_k = n)} f_{k+1}(X_{T_{k+1}})] \\ &= \sum_{l \in \mathbb{N}^*} \mathbf{1}_{(T_k = n)} E^{\mathcal{A}_n} \left[f_{k+1}(X_{n+l}) \prod_{j=1}^{l-1} \mathbf{1}_{(M_{n+j} \notin G)} \mathbf{1}_{(M_{n+j} \in G)} \right] \\ &= \sum_{l \in \mathbb{N}^*} \mathbf{1}_{(T_k = n)} E \left[f_{k+1}(X_{n+l}) \prod_{j=1}^{l-1} \mathbf{1}_{(M_{n+j} \notin G)} \mathbf{1}_{(M_{n+j} \in G)} \right], \end{aligned}$$

la dernière égalité résultant de l'indépendance des tribus $\sigma(M_{n+j} \mid j \in \mathbb{N}^*)$ et \mathcal{A}_n . Il vient alors

$$\mathbf{1}_{(T_k = n)} E^{\mathcal{A}_n} f_{k+1}(X_{T_{k+1}}) = \mathbf{1}_{(T_k = n)} \sum_{l \in \mathbb{N}^*} \left(1 - \frac{1}{l}\right)^{l-1} l(f_{k+1}),$$

soit

$$\boxed{\mathbf{1}_{(T_k = n)} E^{\mathcal{A}_n} f_{k+1}(X_{T_{k+1}}) = \mathbf{1}_{(T_k = n)} \bar{l} I(f_{k+1})}.$$

Soit $A \in \mathcal{A}_{T_k}$ quelconque. Puisque $A \cap (T_k = n) \in \mathcal{A}_n$ et que T_k est fini P-p.s. on a

$$\begin{aligned} E[\mathbf{1}_A f_{k+1}(X_{T_{k+1}})] &= \sum_{n \in \mathbb{N}^*} E[\mathbf{1}_{A \cap (T_k = n)} f_{k+1}(X_{T_{k+1}})] \\ &= \sum_{n \in \mathbb{N}^*} E[\mathbf{1}_A \mathbf{1}_{(T_k = n)} \bar{l} I(f_{k+1})]. \end{aligned}$$

Ainsi, pour tout $A \in \mathcal{A}_{T_k}$, on a

$$E[\mathbf{1}_A f_{k+1}(X_{T_{k+1}})] = E[\mathbf{1}_A \bar{l} I(f_{k+1})],$$

ce qui prouve, $\bar{l} I(f_{k+1})$ étant constant et donc \mathcal{A}_{T_k} -mesurable, que

$$\boxed{E^{\mathcal{A}_{T_k}} [f_{k+1}(X_{T_{k+1}})] = \bar{l} I(f_{k+1})}.$$

En prenant l'espérance, on a alors

$$E[f_{k+1}(X_{T_{k+1}})] = E[E^{\mathcal{A}_{T_k}} [f_{k+1}(X_{T_{k+1}})]] = \bar{l} I(f_{k+1}).$$

c'est-à-dire

$$\boxed{E[f_{k+1}(X_{T_{k+1}})] = \int_{\mathbb{R}} f_{k+1}(x) f(x) d\lambda(x)}.$$

Autrement dit, $X_{T_{k+1}}$ admet encore la densité f .

- (c) Remarquons que pour tout $k \in \mathbb{N}^*$, $f_k(X_{T_k})$ est \mathcal{A}_{T_k} -mesurable. En effet, pour tout borélien B de \mathbb{R} , on a, pour tout $n \in \mathbb{N}^*$,

$$(\Gamma_k = n) \cap [f_k(X_{T_k})]^{-1}(B) = (\Gamma_k = n) \cap [f_k(X_n)]^{-1}(B) \in \mathcal{A}_n,$$

puisque $(\Gamma_k = n) \in \mathcal{A}_n$ et que $[f_k(X_n)]^{-1}(B) \in \mathcal{A}_n$. Compte tenu de l'inclusion (11.32), le produit $\prod_{j=1}^{k-1} f_j(X_{T_j})$ est alors $\mathcal{A}_{T_{k-1}}$ -mesurable. On peut alors écrire

$$\begin{aligned} \mathbb{E}\left[\prod_{j=1}^k f_j(X_{T_j})\right] &= \mathbb{E}\left[\mathbb{E}^{\mathcal{A}_{T_{k-1}}}\left[\prod_{j=1}^k f_j(X_{T_j})\right]\right] \\ &= \mathbb{E}\left[\prod_{j=1}^{k-1} f_j(X_{T_j}) \mathbb{E}^{\mathcal{A}_{T_{k-1}}}[f_k(X_{T_k})]\right] \\ &= \mathbb{E}\left[\prod_{j=1}^{k-1} f_j(X_{T_j})\right] \bar{1}(f_k), \end{aligned}$$

soit

$$\mathbb{E}\left[\prod_{j=1}^k f_j(X_{T_j})\right] = \mathbb{E}\left[\prod_{j=1}^{k-1} f_j(X_{T_j})\right] \mathbb{E}[f_k(X_{T_k})],$$

et par itération rétrograde :

$$\mathbb{E}\left[\prod_{j=1}^k f_j(X_{T_j})\right] = \prod_{j=1}^k \mathbb{E}[f_j(X_{T_j})].$$

Ceci étant vrai pour toutes fonctions f_k mesurables positives bornées, **la suite des X_{T_j} est une suite de variables aléatoires indépendantes (de densité f)**.

- (d) On a

$$\forall x \in \mathbb{R}^+ \quad t(x) = \frac{ab^a}{\Gamma(a)} x^{a-1} \exp\left[-x\left(b - \frac{1}{a}\right)\right].$$

Soit $h = \ln \circ t$. On a :

$$\forall x \in \mathbb{R}^{+*} \quad h'(x) = \frac{a-1}{x} - \left(b - \frac{1}{a}\right) \quad \text{et} \quad h''(x) = \frac{1-a}{x^2} < 0;$$

on a de plus :

$$h'(0^+) = +\infty \quad \text{et} \quad \lim_{x \rightarrow +\infty} h'(x) = -\frac{ab-1}{a} < 0.$$

Il en résulte que h , et donc t , admet un maximum unique sur \mathbb{R}^+ en \hat{x} solution de l'équation $h'(x) = 0$, soit :

$$\hat{x} = \frac{a(a-1)}{ab-1};$$

on a alors

$$\bar{t} \approx t(\hat{x}) = \frac{(ab)^a}{\Gamma(a)} \left(\frac{a-1}{ab-1} \right)^{a-1} \exp(-a+1).$$

- Si $b = 1$ et $a = \frac{5}{2}$, on a

$$\bar{t} = \frac{\left(\frac{5}{2}\right)^{\frac{5}{2}}}{\frac{3}{2} \frac{1}{2} \sqrt{\pi}} \exp\left(-\frac{3}{2}\right) \simeq 1,66.$$

ce qui donne :

$$\mathbf{P}(M_n \in G) \simeq 0,6 \quad \text{et} \quad \mathbf{ET} = 1,66.$$

- Si $b = 1$ et $a = \frac{9}{2}$, un calcul analogue conduit à :

$$\mathbf{P}(M_n \in G) \simeq 0,44 \quad \text{et} \quad \mathbf{ET} = 2,26.$$

Chapitre 12

Transformées de Fourier et fonctions caractéristiques

La transformation de Fourier associe une fonction à toute mesure bornée définie sur \mathbb{R}^d . Opérant sur l'ensemble des mesures bornées définies sur \mathbb{R}^d , cette transformation est injective; elle permet donc, sans perte d'information, de substituer à l'étude d'une famille de mesures celle de la famille des fonctions associées. Plus précisément, la puissance de la transformation de Fourier vient de ce qu'elle transforme le produit de convolution des mesures en produit de fonctions, et que les propriétés de convergence des mesures se traduisent en termes de convergence de leurs transformées de Fourier.

12.1. Définition et propriétés immédiates

Sauf mention du contraire, dans cette section μ est une mesure bornée sur \mathbb{R}^d muni de sa tribu borélienne et X une variable aléatoire définie sur l'espace probabilisé (Ω, \mathcal{A}, P) à valeurs dans \mathbb{R}^d .

On note $\langle \cdot, \cdot \rangle$ le produit scalaire euclidien usuel sur \mathbb{R}^d .

Puisque, pour tout $t \in \mathbb{R}^d$, $|\exp(i\langle x, t \rangle)| = 1$ et que μ est une mesure bornée, la fonction $x \mapsto \exp(i\langle x, t \rangle)$ est μ -intégrable.

Définition 12.1. On appelle **transformée de Fourier** de la mesure bornée μ l'application $\hat{\mu}$ de \mathbb{R}^d dans \mathcal{C} définie par

$$\forall t \in \mathbb{R}^d \quad \hat{\mu}(t) = \int_{\mathbb{R}^d} \exp(i\langle x, t \rangle) d\mu(x). \quad (12.1)$$

On appelle **fonction caractéristique** de la variable aléatoire X la transformée de Fourier de sa loi P_X . Elle est notée φ_X .

Remarque. Il faut bien noter que la notion de fonction caractéristique est relative à la loi de la variable aléatoire X et non à l'application X elle-même.

La formule fondamentale suivante résulte du théorème de transfert :

$$\forall t \in \mathbb{R}^d \quad \varphi_X(t) = E \exp(i\langle X, t \rangle). \quad (12.2)$$

Les notions de transformée de Fourier et de fonction caractéristique se généralisent immédiatement, et sans changement dans les formules, au

cas où \mathbb{R}^d est remplacé par un espace vectoriel de dimension finie E : le crochet est alors la forme bilinéaire de dualité¹ entre E et son dual E^* et $\widehat{\mu}$ est défini comme une fonction sur E^* . Si Φ est un isomorphisme de E sur \mathbb{R}^d , et si $\widehat{\Phi[\mu]}$ désigne la mesure image de μ par Φ , un calcul immédiat montre que $\widehat{\Phi[\mu]} = \widehat{\mu} \circ \Phi^*$, où Φ^* désigne l'adjoint de Φ , défini sur \mathbb{R}^d et à valeurs dans E^* . Si X est une variable aléatoire à valeurs dans E , sa fonction caractéristique est alors définie, par la même formule (12.2), comme fonction sur le dual E^* . Les propriétés démontrées dans le cas de \mathbb{R}^d se transportent alors au cas « général ». On pourra traiter cette extension en exercice.

Nous donnons en parallèle les propriétés immédiates de $\widehat{\mu}$ et φ_X .

Proposition 12.2. *Avec les notations précédentes, on a :*

1. $\widehat{\mu}(0) = \mu(\mathbb{R}^d)$ et $\varphi_X(0) = 1$,
2. $\forall t \in \mathbb{R}^d \quad |\widehat{\mu}(t)| \leq \mu(\mathbb{R}^d)$ et $|\varphi_X(t)| \leq 1$,
3. $\forall t \in \mathbb{R}^d \quad \widehat{\mu}(-t) = \overline{\widehat{\mu}(t)}$ et $\varphi_X(-t) = \overline{\varphi_X(t)}$.
4. Soient $A \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^k)$ et $b \in \mathbb{R}^k$; on a :

$$\forall t \in \mathbb{R}^k \quad \varphi_{AX+b}(t) = \varphi_X(A^*t) \exp(i\langle b, t \rangle), \quad (12.3)$$

où A^* désigne l'adjoint de A .

5. Les fonctions $\widehat{\mu}$ et φ_X sont uniformément continues sur \mathbb{R}^d .

Démonstration. Les trois premières propriétés sont immédiates et la démonstration en est laissée au lecteur.

Pour la quatrième, on a, pour tout $t \in \mathbb{R}^k$,

$$\varphi_{AX+b}(t) = E \exp(i\langle AX + b, t \rangle).$$

Par définition de l'adjoint de A , on a

$$\langle AX + b, t \rangle = \langle X, A^*t \rangle + \langle b, t \rangle$$

et donc :

$$\varphi_{AX+b}(t) = \exp(i\langle b, t \rangle) E \exp(i\langle X, A^*t \rangle) = \varphi_X(A^*t) \exp(i\langle b, t \rangle).$$

Démontrons que $\widehat{\mu}$ est uniformément continue. Soit $\varepsilon > 0$ fixé ; puisque μ est une mesure bornée, on peut choisir un entier n tel que

$$\mu(B(0, n)^c) \leq \frac{\varepsilon}{4}.$$

1. Rappelons que si E est muni d'une structure euclidienne, E^* s'identifie naturellement à E .

Pour tous $u, t \in \mathbb{R}^d$, on a

$$|\hat{\mu}(u) - \hat{\mu}(t)| \leq \int_{\mathbb{B}(0, n)} |\exp(i \langle x, u \rangle) - \exp(i \langle x, t \rangle)| d\mu(x) + 2\mu(\mathbb{B}(0, n)^c).$$

Mais l'inégalité des accroissements finis donne

$$|\exp(i \langle x, u \rangle) - \exp(i \langle x, t \rangle)| \leq \|u - t\| \|x\|$$

et donc :

$$|\hat{\mu}(u) - \hat{\mu}(t)| \leq n\mu(\mathbb{R}^d)\|u - t\| + 2\mu(\mathbb{B}(0, n)^c).$$

Si $\eta = \frac{\varepsilon}{2n\mu(\mathbb{R}^d)}$, on a alors :

$$\forall u, t \in \mathbb{R}^d \quad \text{tel que } \|u - t\| \leq \eta \quad |\hat{\mu}(u) - \hat{\mu}(t)| \leq \varepsilon.$$

ce qui démontre le résultat, ε étant quelconque. \square

12.2. Le théorème d'injectivité

Notation. Dans ce chapitre, on note $\int_{\mathbb{R}^d} g(x) dx$ l'intégrale de Lebesgue de g sur \mathbb{R}^d .

Définition 12.3. Si μ est une mesure bornée sur \mathbb{R}^d , et f une fonction borélienne telle que, pour tout x , la fonction $y \mapsto f(x - y)$ soit μ -intégrable, la convolution de f et μ est la fonction $f * \mu$ définie par

$$\forall x \in \mathbb{R}^d \quad (f * \mu)(x) = \int_{\mathbb{R}^d} f(x - y) d\mu(y).$$

Si g est Lebesgue-intégrable, on note \hat{g} sa transformée de Fourier, c'est-à-dire la fonction définie sur \mathbb{R}^d par

$$\forall t \in \mathbb{R}^d \quad \hat{g}(t) = \int_{\mathbb{R}^d} g(x) \exp(i \langle x, t \rangle) dx.$$

Pour nous, la propriété fondamentale de la transformation de Fourier est son injectivité : autrement dit, une mesure bornée μ sur \mathbb{R}^d est déterminée par sa transformée de Fourier $\hat{\mu}$. Cette propriété sera démontrée par une succession de lemmes.

Lemme 12.4 (Propriétés du noyau gaussien). Soit, pour tout $\sigma > 0$, la fonction g_σ , appelée **noyau gaussien**, définie sur \mathbb{R}^d par

$$\forall x \in \mathbb{R}^d \quad g_\sigma(x) = \frac{1}{(\sigma\sqrt{2\pi})^d} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right),$$

où $\|\cdot\|$ est la norme euclidienne usuelle de \mathbb{R}^d .

- (a) La fonction g_σ est une densité de probabilité sur \mathbb{R}^d .
 (b) Pour tout $\varepsilon > 0$, on a

$$\lim_{\sigma \rightarrow 0} \int_{\{\|x\| > \varepsilon\}} g_\sigma(x) dx = 0.$$

- (c) Pour tout $f \in \mathcal{C}_b(\mathbb{R}^d)$ et tout $x \in \mathbb{R}^d$

$$(f * g_\sigma)(x) \xrightarrow{\sigma \rightarrow 0} f(x). \quad (12.4)$$

- (d) La transformée de Fourier de g_1 est donnée par

$$\forall t \in \mathbb{R}^d \quad \widehat{g}_1(t) = \exp\left(-\frac{\|t\|^2}{2}\right) \equiv (\sqrt{2\pi})^d g_1(t). \quad (12.5)$$

Démonstration. (a) Commençons par le cas $d = 1$. Le changement de variables $x = \sigma y$ permet d'écrire

$$\int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = 1$$

(pour la dernière égalité, voir tome 1, ch. 6, p. 213). Dans le cas général, le théorème de Fubini donne alors

$$\int_{\mathbb{R}^d} g_\sigma(x) dx = \prod_{j=1}^d \left(\int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x_j^2}{2\sigma^2}\right) dx_j \right) = 1.$$

(b) On note provisoirement $\|\cdot\|_2$ la norme euclidienne de \mathbb{R}^d et $\|\cdot\|_\infty$ la norme max; il existe une constante $c > 0$ telle que, pour tout $x \in \mathbb{R}^d$, on ait $c\|x\|_2 \leq \|x\|_\infty$ et donc, pour tout $\varepsilon > 0$

$$\{\|x\|_\infty \leq c\varepsilon\} \subset \{\|x\|_2 \leq \varepsilon\}.$$

Alors

$$\int_{\{\|x\|_\infty \leq c\varepsilon\}} g_\sigma(x) dx \leq \int_{\{\|x\|_2 \leq \varepsilon\}} g_\sigma(x) dx \leq 1.$$

Mais, par le théorème de Fubini

$$\begin{aligned} \int_{\{\|x\|_\infty \leq c\varepsilon\}} g_\sigma(x) dx &= \prod_{j=1}^d \left(\int_{\{|x_j| \leq c\varepsilon\}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x_j^2}{2\sigma^2}\right) dx_j \right) \\ &= \left(\int_{\{|y| \leq \frac{c\varepsilon}{\sigma}\}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \right)^d, \end{aligned}$$

et donc

$$\int_{\{\|x\|_\infty \leq c\varepsilon\}} g_\sigma(x) dx \xrightarrow{\sigma \rightarrow 0} 1,$$

ce qui assure le résultat.

Remarque. On peut aussi démontrer ce résultat en passant en coordonnées sphériques; en effet le calcul montre qu'il existe une constante $c_d > 0$ telle que

$$\int_{\{\|x\|_2 > \varepsilon\}} g_\sigma(x) dx = c_d \int_{\frac{\varepsilon}{\sigma}}^{+\infty} \exp\left(-\frac{r^2}{2}\right) r^{d-1} dr.$$

(c) Soit $x \in \mathbb{R}^d$. Le changement de variables $z = \frac{x-y}{\sigma}$, de jacobien σ^d , donne

$$\begin{aligned} (f * g_\sigma)(x) &= \int_{\mathbb{R}^d} f(y) \frac{1}{(\sigma\sqrt{2\pi})^d} \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) dy \\ &= \int_{\mathbb{R}^d} f(x-\sigma z) g_1(z) dz. \end{aligned} \quad (12.6)$$

Mais, f étant continue bornée, on a $\lim_{\sigma \rightarrow 0} f(x-\sigma z) = f(x)$ et $|f(x-\sigma z)| \leq \|f\|_\infty$, constante intégrable par rapport à la probabilité de densité g_1 ; le théorème de convergence dominée (appliqué à une suite positive quelconque qui converge vers 0) donne le résultat.

(d) Soit $t \in \mathbb{R}^d$; il résulte du théorème de Fubini (applicable en raison de l'intégrabilité de la fonction $x \mapsto \exp(i\langle x, t \rangle)$ par rapport à la probabilité de densité g_1) que

$$\widehat{g}_1(t) = \prod_{j=1}^d \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[ix_j t_j - \frac{x_j^2}{2}\right] dx_j \right).$$

On en déduit

$$\widehat{g}_1(t) = \exp\left(-\frac{\|t\|^2}{2}\right),$$

pourvu que l'on montre que

$$\forall u \in \mathbb{R} \quad \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[ixu - \frac{x^2}{2}\right] dx = \exp\left(-\frac{u^2}{2}\right).$$

Démontrons ce dernier résultat, utile par ailleurs: on a

$$\forall z \in \mathbb{R} \quad \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-z)^2}{2}\right] dx = 1,$$

ce qui donne, en développant le carré,

$$\forall z \in \mathbb{R} \quad \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[zx - \frac{x^2}{2}\right] dx = \exp\left(\frac{z^2}{2}\right). \quad (12.7)$$

De plus, pour tout $z \in \mathbb{C}$, on a

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[|z|x - \frac{x^2}{2}\right] dx < +\infty.$$

ce qui démontre d'une part l'intégrabilité de l'application $x \mapsto \exp\left[zx - \frac{x^2}{2}\right]$, puisque

$$\left| \exp\left[zx - \frac{x^2}{2}\right] \right| \leq \exp\left[|zx| - \frac{x^2}{2}\right],$$

et d'autre part l'inégalité suivante, par convergence monotone :

$$\sum_{n=0}^{+\infty} \int_{\mathbb{P}} \frac{1}{\sqrt{2\pi}} \frac{|zx|^n}{n!} \exp\left[-\frac{x^2}{2}\right] dx < +\infty.$$

Le corollaire du théorème de convergence dominée relatif aux séries (cf. annexe, corollaire A.33) assure alors que, pour tout $z \in \mathbb{C}$,

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[zx - \frac{x^2}{2}\right] dx = \sum_{n=0}^{+\infty} \frac{z^n}{n!} \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x^n \exp\left[-\frac{x^2}{2}\right] dx \right).$$

La fonction $z \mapsto \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[zx - \frac{x^2}{2}\right] dx$ est donc une fonction entière qui coïncide sur \mathbb{P} avec la fonction entière $z \mapsto \exp\left(\frac{z^2}{2}\right)$; en vertu du principe du prolongement analytique, ces fonctions coïncident sur \mathbb{C} . Il reste à faire $z = iu$ dans l'égalité (12.7). \square

Le lemme suivant est la clef de la démonstration du théorème d'injectivité; il affirme que la donnée de $\hat{\mu}$ détermine les produits de convolution $g_\sigma * \mu$ ($\sigma > 0$). La démonstration du théorème 12.6 consistera ensuite à montrer comment la donnée des $g_\sigma * \mu$ détermine à son tour la mesure μ .

Lemme 12.5. Soit μ une mesure bornée sur \mathbb{R}^d . Pour tout $\sigma > 0$ et tout $y \in \mathbb{R}^d$, la fonction $g_\sigma(\cdot - y)$ est μ -intégrable et

$$(g_\sigma * \mu)(y) = (\sqrt{2\pi})^{-d} \int_{\mathbb{R}^d} \hat{\mu}(v) g_1(\sigma v) \exp(-i \langle y, v \rangle) dv. \quad (12.8)$$

Démonstration. Soit $y \in \mathbb{R}^d$. La fonction $g_\sigma(\cdot - y)$ est μ -intégrable puisque bornée (ainsi que μ). La relation (12.5) entre g_1 et sa transformée de Fourier \hat{g}_1 permet d'écrire

$$g_\sigma(y - x) = g_\sigma(x - y) = \frac{1}{(\sigma \sqrt{2\pi})^d} \hat{g}_1\left(\frac{x - y}{\sigma}\right),$$

soit, par le changement de variable $v = \frac{z}{\sigma}$ de jacobien $\frac{1}{\sigma^d}$,

$$g_\sigma(y - x) = \frac{1}{(\sqrt{2\pi})^d} \int_{\mathbb{R}^d} g_1(\sigma v) \exp(i \langle x - y, v \rangle) dv.$$

On a donc

$$(g_\sigma * \mu)(y) = \int_{\mathbb{R}^d} \left[\frac{1}{(\sqrt{2\pi})^d} \int_{\mathbb{R}^d} g_1(\sigma v) \exp(i \langle x - y, v \rangle) dv \right] d\mu(x).$$

Comme on a

$$|g_1(\sigma v) \exp(i \langle x - y, v \rangle)| \leq \exp\left(-\sigma^2 \frac{\|v\|^2}{2}\right),$$

et comme la fonction déterminée par le membre de droite est $\mu \otimes \lambda_d$ -intégrable (là encore l'hypothèse μ bornée est importante!), il résulte du théorème de Fubini que

$$(g_\sigma * \mu)(y) = (\sqrt{2\pi})^{-d} \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \exp(i \langle x, v \rangle) d\mu(x) \right] g_1(\sigma v) \exp(-i \langle y, v \rangle) dv.$$

ce qui est le résultat annoncé. \square

Théorème 12.6 (Théorème d'injectivité de la transformation de Fourier).
Deux mesures bornées sur \mathbb{R}^d qui ont même transformée de Fourier sont égales.

Démonstration. On rappelle qu'une mesure μ bornée sur \mathbb{R}^d est déterminée par la donnée, pour tout $f \in \mathcal{C}_b(\mathbb{R}^d)$, des intégrales $\int_{\mathbb{R}^d} f d\mu$ (cf. chap. 8, corollaire 8.5). Soit donc $f \in \mathcal{C}_b(\mathbb{R}^d)$ quelconque: on montre que son intégrale par rapport à μ est fonction de $\widehat{\mu}$, ce qui assure l'injectivité de la transformation de Fourier.

Soit une suite positive $(\sigma_n)_{n \in \mathbb{N}}$ quelconque qui converge vers 0. Il résulte de la relation (12.4) que, tout $x \in \mathbb{R}^d$,

$$\lim_n (f * g_{\sigma_n})(x) = f(x). \quad (12.9)$$

Notons que, d'après (12.6), on a, pour tout $n \in \mathbb{N}$,

$$|(f * g_{\sigma_n})(x)| \leq \|f\|_\infty : \quad (12.10)$$

puisque toute fonction constante est μ -intégrable, le théorème de convergence dominée assure alors que

$$\int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} \lim_n (f * g_{\sigma_n})(x) d\mu(x) = \lim_n \int_{\mathbb{R}^d} (f * g_{\sigma_n})(x) d\mu(x).$$

soit encore :

$$\int_{\mathbb{R}^d} f d\mu = \lim_n \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} f(y) g_{\sigma_n}(x-y) dy \right] d\mu(x). \quad (12.11)$$

On peut, dans cette dernière expression, changer l'ordre des intégrations, après avoir observé que la fonction $(x, y) \mapsto f(y) g_{\sigma_n}(x-y)$ est $\lambda_d \otimes \mu$ -intégrable. On a en effet, par application du théorème de Fubini pour les fonctions positives,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |f(y) g_{\sigma_n}(x-y)| d\lambda_d \otimes \mu(x, y) = \int_{\mathbb{R}^d} |f(y)| \left[\int_{\mathbb{R}^d} g_{\sigma_n}(x-y) d\lambda_d(x) \right] d\mu(y).$$

ce qui implique que

$$\begin{aligned} \int_{\mathbb{R}^d \times \mathbb{R}^d} |f(y)g_{\sigma_n}(x-y)| d\lambda_d \otimes \mu(x, y) \\ \leq \int_{\mathbb{R}^d} \|f\|_\infty \left[\int_{\mathbb{R}^d} g_{\sigma_n}(x-y) d\lambda_d(x) \right] d\mu(y). \end{aligned}$$

Puisque, pour tout $y \in \mathbb{R}^d$, on a

$$\int_{\mathbb{R}^d} g_{\sigma_n}(x-y) d\lambda_d(x) = \int_{\mathbb{R}^d} g_{\sigma_n}(x) d\lambda_d(x) = 1,$$

on obtient alors les inégalités

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |f(y)g_{\sigma_n}(x-y)| d\lambda_d \otimes \mu(x, y) \leq \|f\|_\infty \mu(\mathbb{R}^d) < +\infty,$$

ce qui permet d'appliquer le théorème de Fubini aux intégrales apparaissant dans l'égalité (12.11). En tenant compte de la parité de g_{σ_n} , on obtient alors

$$\int_{\mathbb{R}^d} f d\mu = \lim_n \int_{\mathbb{R}^d} f(y) \left[\int_{\mathbb{R}^d} g_{\sigma_n}(x-y) d\mu(x) \right] dy = \lim_n \int_{\mathbb{R}^d} f(y) (g_{\sigma_n} * \mu)(y) dy,$$

soit, d'après l'égalité (12.8),

$$\int_{\mathbb{R}^d} f d\mu = (\sqrt{2\pi})^{-d} \lim_n \int_{\mathbb{R}^d} f(y) \left[\int_{\mathbb{R}^d} \hat{\mu}(v) g_1(\sigma_n v) \exp(-i\langle y, v \rangle) dv \right] dy.$$

Ceci montre, comme annoncé, que $\int_{\mathbb{R}^d} f d\mu$ est fonction de $\hat{\mu}$ et le théorème est démontré. \square

Remarque importante. Il résulte de la propriété d'injectivité que **la fonction caractéristique d'une variable aléatoire à valeurs dans \mathbb{R}^d caractérise entièrement la loi de cette variable aléatoire** (d'où son nom): en particulier, le tableau du chapitre 8 donnant les lois de probabilité et leur transformée de Fourier se lit dans les deux sens! Cette propriété a été démontrée pour la première fois par **Paul Lévy**², en 1922, pour les variables aléatoires réelles et en termes de leur fonction de répartition.

Pour préciser le théorème d'injectivité, il existe une formule donnant explicitement la **fonction cumulative**³ d'une mesure bornée sur \mathbb{R}^d en

2. Paul Lévy (1886-1971), né à Paris, enseigne à l'École des mines de Saint-Étienne, puis à celle de Paris, en enseignant parallèlement à l'École polytechnique. Ses premiers travaux portent sur l'analyse fonctionnelle, mais il se tourne vite vers le calcul des probabilités. Sa contribution à la théorie des probabilités est très importante, en particulier, dans le domaine des processus aléatoires et du mouvement brownien. Ses œuvres sont rassemblées dans une série de trois livres édités par Gauthier-Villars en 1976 et 1980.

3. La **fonction cumulative** d'une mesure bornée μ sur \mathbb{R}^d est la fonction définie sur \mathbb{R}^d : $x \mapsto \mu(\{y \mid y \leq x\})$, le signe \leq désignant l'ordre partiel usuel de \mathbb{R}^d . Si μ est la loi d'une variable aléatoire X à valeurs dans \mathbb{R}^d , la fonction cumulative de μ n'est autre que la **fonction de répartition** de X .

fonction de sa transformée de Fourier (voir l'exercice 10 de ce chapitre). Nous nous contentons d'étudier ici le cas où la mesure est **absolument continue**.

Proposition 12.7. *Soit μ une mesure bornée sur \mathbb{R}^d telle que sa transformée de Fourier $\widehat{\mu}$ soit Lebesgue-intégrable. Alors μ est absolument continue par rapport à la mesure de Lebesgue et sa densité est donnée λ_d -p.p. par la fonction continue h définie par*

$$\forall x \in \mathbb{R}^d \quad h(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{\mu}(t) \exp(-i \langle x, t \rangle) dt.$$

Démonstration. Pour identifier la mesure μ il suffit de calculer les intégrales $\int_{\mathbb{R}^d} f d\mu$ pour tout $f \in \mathcal{C}_X(\mathbb{R}^d)$. Reprenons la démonstration du théorème d'injectivité à son terme; on a, pour tout $f \in \mathcal{C}_X(\mathbb{R}^d)$ et toute suite positive $(\sigma_n)_{n \in \mathbb{N}}$ qui converge vers 0 :

$$\int_{\mathbb{R}^d} f d\mu = \lim_n \int_{\mathbb{R}^d} f(y) (g_{\sigma_n} * \mu)(y) dy.$$

Mais, d'après (12.8), on a

$$|g_{\sigma_n} * \mu(y)| \leq \int_{\mathbb{R}^d} |\widehat{\mu}(v)| g_1(\sigma_n v) dv \leq \int_{\mathbb{R}^d} |\widehat{\mu}(v)| dv < +\infty,$$

et donc

$$|f(y) g_{\sigma_n} * \mu(y)| \leq |f(y)| \|\widehat{\mu}\|_{1,1}.$$

Le membre de droite est alors Lebesgue-intégrable, puisque f est continue à support compact et on peut appliquer le théorème de convergence dominée :

$$\int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} f(y) \left[\lim_n (g_{\sigma_n} * \mu)(y) \right] dy,$$

ce qui démontre que μ est absolument continue par rapport à la mesure de Lebesgue et que sa densité est donnée λ_d -p.p. par la fonction continue h définie par $\lim_n (g_{\sigma_n} * \mu)(y)$. On obtient le résultat en notant que

$$\lim_n \widehat{\mu}(v) g_1(\sigma_n v) \exp(-i \langle y, v \rangle) = (\sqrt{2\pi})^{-d} \widehat{\mu}(v) \exp(-i \langle y, v \rangle)$$

et en appliquant une nouvelle fois le théorème de convergence dominée, ce qui est licite puisque $\widehat{\mu}$ est intégrable et que, d'après la relation (12.8), on a

$$|\widehat{\mu}(v) g_1(\sigma_n v) \exp(-i \langle y, v \rangle)| \leq |\widehat{\mu}(v)|.$$

La continuité de h est obtenue par le théorème de continuité des intégrales à paramètre, corollaire du théorème de convergence dominée. \square

12.3. Propriétés relatives à l'indépendance

Sur $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ le produit scalaire euclidien usuel vérifie, pour tous (x_1, x_2) et (t_1, t_2) de $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$:

$$\langle (x_1, x_2), (t_1, t_2) \rangle = \langle x_1, t_1 \rangle + \langle x_2, t_2 \rangle .$$

Il en résulte immédiatement la proposition suivante :

Proposition 12.8. *Si μ_1 et μ_2 sont des mesures bornées respectivement sur \mathbb{R}^{d_1} et \mathbb{R}^{d_2} , la transformée de Fourier de la mesure produit est le produit direct des transformées de Fourier de μ_1 et μ_2 , ce qui signifie que :*

$$\forall (t_1, t_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \quad \widehat{\mu_1 \otimes \mu_2}(t_1, t_2) = \widehat{\mu_1}(t_1) \widehat{\mu_2}(t_2) .$$

Démonstration. La fonction $(x_1, x_2) \mapsto \exp[i \langle (x_1, x_2), (t_1, t_2) \rangle]$ est bornée donc $\mu_1 \otimes \mu_2$ -intégrable ; le théorème de Fubini donne le résultat en tenant compte de la relation

$$\exp[i \langle (x_1, x_2), (t_1, t_2) \rangle] = \exp[i \langle x_1, t_1 \rangle] \exp[i \langle x_2, t_2 \rangle] . \quad \square$$

On obtient alors un **critère d'indépendance** de variables aléatoires en termes de fonctions caractéristiques.

Corollaire 12.9 (Critère d'indépendance). *Soit $X = (X_1, X_2)$ une variable aléatoire à valeurs dans $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Pour que X_1 et X_2 soient indépendantes, il faut et il suffit que*

$$\forall (t_1, t_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \quad \varphi_{(X_1, X_2)}(t_1, t_2) = \varphi_{X_1}(t_1) \varphi_{X_2}(t_2) . \quad (12.12)$$

Démonstration. Pour que X_1 et X_2 soient indépendantes, il faut et il suffit que $P_{(X_1, X_2)} = P_{X_1} \otimes P_{X_2}$, soit, par le théorème d'injectivité, que $\widehat{P}_{(X_1, X_2)} = \widehat{P}_{X_1} \otimes \widehat{P}_{X_2}$. La proposition précédente donne le résultat. \square

Remarque. La **fonction caractéristique d'une marginale** s'obtient très facilement ; avec les notations du corollaire 12.9, on a

$$\begin{aligned} \forall t_1 \in \mathbb{R}^{d_1} \quad \varphi_{X_1}(t_1) &= \varphi_{(X_1, X_2)}(t_1, 0) , \\ \forall t_2 \in \mathbb{R}^{d_2} \quad \varphi_{X_2}(t_2) &= \varphi_{(X_1, X_2)}(0, t_2) . \end{aligned} \quad (12.13)$$

Le corollaire 12.9 peut donc encore s'énoncer sous la forme :

Corollaire 12.10 (Critère d'indépendance bis). *Soit $X = (X_1, X_2)$ une variable aléatoire à valeurs dans $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Pour que X_1 et X_2 soient indépendantes, il faut et il suffit que*

$$\forall (t_1, t_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \quad \varphi_{(X_1, X_2)}(t_1, t_2) = \varphi_{(X_1, X_2)}(t_1, 0) \varphi_{(X_1, X_2)}(0, t_2) . \quad (12.14)$$

Exemple 12.1. Soient X_1 et X_2 deux variables aléatoires réelles indépendantes de même loi de **Laplace**, de fonction caractéristique φ donnée par

$$\forall t \in \mathbb{R} \quad \varphi(t) = \frac{1}{1+t^2}.$$

On définit les variables aléatoires Y_1 et Y_2 par

$$Y_1 = X_1 - X_2 \quad Y_2 = X_1 + X_2,$$

c'est-à-dire

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

D'après l'égalité (12.12), la fonction caractéristique de (X_1, X_2) est définie par

$$\forall (t_1, t_2) \in \mathbb{R}^2 \quad \varphi_{(X_1, X_2)}(t_1, t_2) = \frac{1}{1+t_1^2} \frac{1}{1+t_2^2}$$

et, d'après l'égalité (12.3), la fonction caractéristique de (Y_1, Y_2) est définie par

$$\begin{aligned} \forall (t_1, t_2) \in \mathbb{R}^2 \quad \varphi_{(Y_1, Y_2)}(t_1, t_2) &= \varphi_{(X_1, X_2)}(t_1 + t_2, -t_1 + t_2) \\ &= \frac{1}{1+(t_1+t_2)^2} \frac{1}{1+(-t_1+t_2)^2}. \end{aligned}$$

On obtient les fonctions caractéristiques des marginales Y_1 et Y_2 par les relations (12.13) :

$$\forall t_1 \in \mathbb{R} \quad \varphi_{Y_1}(t_1) = \frac{1}{(1+t_1^2)^2} \quad \forall t_2 \in \mathbb{R} \quad \varphi_{Y_2}(t_2) = \frac{1}{(1+t_2^2)^2}.$$

Les variables aléatoires Y_1 et Y_2 ont même fonction caractéristique, donc même loi; elles ne sont pas indépendantes puisque $\varphi_{(Y_1, Y_2)}(1, 1) = \frac{1}{5} \neq \frac{1}{16} = \varphi_{Y_1}(1)\varphi_{Y_2}(1)$.

On peut toutefois remarquer que ces variables aléatoires sont *non corrélées* puisque, X_1 et X_2 ayant même loi, donc mêmes moments, on a

$$E[Y_1 Y_2] = E(X_1^2) - E(X_2^2) = 0 \quad \text{et} \quad E[Y_1] = E(X_1) - E(X_2) = 0$$

et donc $\text{cov}(Y_1, Y_2) = 0$.

L'un des intérêts essentiels de cette théorie est que transformée de Fourier de **convolution** de mesures bornées et fonction caractéristique de **somme de variables aléatoires indépendantes** se calculent facilement, comme nous allons le voir maintenant.

Proposition 12.11. Si μ_1 et μ_2 sont des mesures bornées sur \mathbb{R}^d , la transformée de Fourier du produit de convolution de μ_1 et μ_2 est le produit de leur transformée de Fourier

$$\widehat{\mu_1 * \mu_2} = \widehat{\mu_1} \widehat{\mu_2}.$$

Démonstration. La convolution $\mu_1 * \mu_2$ étant la mesure image de $\mu_1 \otimes \mu_2$ par l'application somme et l'exponentielle complexe étant bornée, le théorème de transfert donne :

$$\forall t \in \mathbb{R}^d \quad \widehat{\mu_1 * \mu_2}(t) = \int_{\mathbb{R}^d} \exp(i \langle x_1 + x_2, t \rangle) d\mu_1 \otimes \mu_2(x_1, x_2).$$

Le résultat s'obtient en factorisant l'exponentielle et en appliquant le théorème de Fubini. \square

Corollaire 12.12. Soient X_1 et X_2 deux variables aléatoires indépendantes à valeurs dans \mathbb{R}^d . La fonction caractéristique de leur somme est donnée par

$$\forall t \in \mathbb{R}^d \quad \varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \varphi_{X_2}(t).$$

Démonstration. Il suffit de se rappeler que, du fait de l'indépendance, la loi de $X_1 + X_2$ est la convolution des lois de X_1 et X_2 . \square

On obtient alors, compte tenu de l'injectivité de la transformation de Fourier, un **moyen de calcul de la loi d'une somme finie de variables aléatoires indépendantes**.

Exemple 12.2. On rappelle le cadre de modélisation de la **loi multinomiale**. Soit $k \in \mathbb{N}^*$ fixé. Pour tout $n \in \mathbb{N}^*$, on considère une partition $(A_j^n)_{1 \leq j \leq k}$ de Ω , où $A_j^n \in \mathcal{A}$. On suppose que les familles d'événements, indexées sur n , constituées par les éléments de ces partitions sont **indépendantes**. On suppose de plus que :

$$\forall n \in \mathbb{N}^*, \forall j = 1, 2, \dots, k \quad P(A_j^n) = p_j,$$

où $p_j > 0$ et $\sum_{j=1}^k p_j = 1$. On définit les variables aléatoires X^n et Y^n à valeurs dans \mathbb{R}^k par

$$X^n = \begin{pmatrix} \mathbf{1}_{A_1^n} \\ \vdots \\ \mathbf{1}_{A_k^n} \end{pmatrix} \quad \text{et} \quad Y^n = \sum_{j=1}^k X^j.$$

Les variables aléatoires X^j étant indépendantes de même loi, la fonction caractéristique de Y^n est donnée par

$$\forall t \in \mathbb{R}^k \quad \varphi_{Y^n}(t) = [\varphi_{X^1}(t)]^n.$$

De plus, puisque $(A_j^1)_{1 \leq j \leq k}$ forme une partition de Ω , on a

$$\forall t \in \mathbb{R}^k \quad \varphi_{X^1}(t) = \sum_{j=1}^k \int_{A_j^1} \exp(i \langle \mathbf{X}^1, t \rangle) dP = \sum_{j=1}^k p_j \exp(it_j),$$

et donc

$$\forall t \in \mathbb{R}^k \quad \varphi_{Y^n}(t) = \left[\sum_{j=1}^k p_j \exp(it_j) \right]^n. \quad (12.15)$$

Remarque. Il en résulte que si Z_1 et Z_2 sont deux variables aléatoires, à valeurs dans \mathbb{R}^k , indépendantes, de lois multinomiales respectives $M(n_1; p_1, p_2, \dots, p_{k-1})$ et $M(n_2; p_1, p_2, \dots, p_{k-1})$, la variable aléatoire $Z_1 + Z_2$ est de loi multinomiale $M(n_1 + n_2; p_1, p_2, \dots, p_{k-1})$. Autrement dit, **la famille des lois multinomiales $M(n; p_1, p_2, \dots, p_{k-1})$, $n \in \mathbb{N}^*$, est stable par convolution.**

12.4. Fonction caractéristique et moments

On étudie les relations entre les propriétés de dérivabilité de la fonction caractéristique et l'existence des moments d'une variable aléatoire. On rappelle, sans démonstration et sous une forme vectorielle, le théorème de dérivation d'une intégrale dépendant d'un paramètre.

Théorème 12.13. *Soit μ une mesure σ -finie sur l'espace probabilisable (Ω, \mathcal{A}) . Soient E et F deux espaces vectoriels normés de dimension finie et O un ouvert de E . Soit f une application de $O \times \Omega$ dans F vérifiant :*

1. *Pour tout $\omega \in \Omega$, l'application partielle $f(\cdot, \omega)$ est de classe C^1 dans O et il existe $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ telle que*

$$\forall x \in E \quad \left\| \frac{\partial}{\partial x} f(x, \cdot) \right\|_{\mathcal{L}(E, F)} \leq g.$$

2. *Pour tout $x \in O$, l'application partielle $f(x, \cdot)$ est μ -intégrable.*

Alors l'application de O dans F , $x \mapsto \int_{\Omega} f(x, \omega) d\mu(\omega)$ est différentiable et on a pour tout $x \in O$:

$$\frac{\partial}{\partial x} \int_{\Omega} f(x, \omega) d\mu(\omega) = \int_{\Omega} \frac{\partial}{\partial x} f(x, \omega) d\mu(\omega).$$

Démonstration. La démonstration résulte d'une simple application de l'inégalité des accroissements finis et du théorème de convergence dominée. □

Dans un premier temps, pour éviter les « difficultés » inhérentes au calcul différentiel, on s'intéresse aux variables aléatoires réelles.

Proposition 12.14. *Soit X une variable aléatoire réelle et φ_X sa fonction caractéristique.*

(a) Si X admet un moment d'ordre $n \in \mathbb{N}^*$, φ_X est de classe C^n et, pour tout entier k tel que $1 \leq k \leq n$, on a

$$\forall t \in \mathbb{R} \quad \varphi_X^{(k)}(t) = i^k \int_{\Omega} X^k \exp(itX) dP, \quad (12.16)$$

et, en particulier,

$$\boxed{\varphi_X^{(k)}(0) = i^k EX^k.} \quad (12.17)$$

(b) Inversement, si φ_X est k fois dérivable en 0 ($k \geq 2$), X admet des moments jusqu'à l'ordre $2\left[\frac{k}{2}\right]$; ils sont donnés par la formule (12.17).

Démonstration. (a) Puisque

$$\frac{d^k}{dt^k} \exp(itX) = (iX)^k \exp(itX),$$

on a

$$\left| \frac{d^k}{dt^k} \exp(itX) \right| \leq |X|^k,$$

et on peut appliquer n fois le théorème 12.13.

(b) Démontrons le résultat pour $k = 2$. Dans ce cas, φ_X admet un développement limité de Taylor-Young à l'ordre deux et donc :

$$\lim_{t \rightarrow 0} \frac{\varphi_X(t) + \varphi_X(-t) - 2}{t^2} = \varphi_X''(0).$$

Alors, puisque

$$\varphi_X(t) + \varphi_X(-t) = 2\Re(\varphi_X(t)) = 2E \cos(tX),$$

on a

$$\lim_{t \rightarrow 0} E \left[\frac{1 - \cos(tX)}{t^2} \right] = -\frac{1}{2} \varphi_X''(0).$$

L'intégrande étant positive, il résulte alors du lemme de Fatou que, si $(t_n)_{n \in \mathbb{N}}$ est une suite convergeant vers 0, on a :

$$\begin{aligned} \int_{\Omega} X^2 dP &= E \left[2 \liminf_n \frac{1 - \cos(t_n X)}{t_n^2} \right] \\ &\leq 2 \liminf_n E \left[\frac{1 - \cos(t_n X)}{t_n^2} \right] < +\infty. \end{aligned}$$

Supposons avoir démontré l'existence de tous les moments jusqu'à l'ordre $2(n-1) = 2\left[\frac{k}{2}\right] - 2$; il nous faut démontrer que le moment d'ordre $2n = 2\left[\frac{k}{2}\right]$ existe. Il résulte de la proposition directe que

$$\varphi_X^{2(n-1)}(t) + \varphi_X^{2(n-1)}(-t) = (-1)^{n-1} 2E[X^{2(n-1)} \cos(tX)]$$

et

$$\varphi_X^{2(n-1)}(0) = (-1)^{n-1} E[X^{2(n-1)}].$$

Par ailleurs, $\varphi_X^{2(n-1)}$ étant par hypothèse deux fois dérivable en 0, admet un développement limité de Taylor-Young à l'ordre deux, et donc :

$$\lim_{t \rightarrow 0} \frac{\varphi_X^{2(n-1)}(t) + \varphi_X^{2(n-1)}(-t) - 2\varphi_X^{2(n-1)}(0)}{t^2} = \varphi_X^{(2n)}(0).$$

De ces trois dernières relations, on déduit que

$$\lim_{t \rightarrow 0} E \left[X^{2(n-1)} \frac{1 - \cos(tX)}{t^2} \right] = \frac{(-1)^n}{2} \varphi_X^{(2n)}(0).$$

On conclut avec le lemme de Fatou, de la même manière que ci-dessus. \square

Remarque. Comme le montre l'exemple ci-dessous (cf. Stoyanov, p. 64, ou Jeulin, chap. 2, p. 20), **la fonction caractéristique peut être dérivable à l'origine (et même en tout point) sans que la variable aléatoire admette une moyenne.**

Exemple 12.3. Soit X une variable aléatoire réelle de loi $P_X = \sum_{k \in \mathbb{Z}} a_k \delta_k$ symétrique, c'est-à-dire telle que $a_k = a_{-k}$, et telle que $\sum_{k \in \mathbb{N}} ka_k = +\infty$. On a

$$\int_{\Omega} |X| dP = 2 \sum_{k \in \mathbb{N}^*} ka_k = +\infty \quad \text{et} \quad \varphi_X(t) = a_0 + 2 \sum_{k=1}^{+\infty} a_k \cos(kt).$$

On choisit la suite $(a_k)_{k \in \mathbb{Z}}$ telle que la suite $(ka_k)_{k \in \mathbb{N}^*}$ tende **en décroissant** vers 0. On rappelle la majoration :

$$\left| \sum_{k=0}^n \sin(kx) \right| \leq \left| \sum_{k=0}^n \exp(ikx) \right| = \left| \frac{1 - \exp(i(n+1)x)}{1 - \exp(ix)} \right| \leq \frac{2}{\left| \sin\left(\frac{x}{2}\right) \right|}.$$

On a donc, si $\alpha \in]0, 2\pi[$:

$$\forall n \in \mathbb{N}^*, \forall x \in [\alpha, 2\pi - \alpha] \quad \left| \sum_{k=0}^n \sin(kx) \right| \leq \frac{2}{\left| \sin\left(\frac{\alpha}{2}\right) \right|}.$$

Le **critère d'Abel** assure la convergence uniforme sur l'intervalle $[\alpha, 2\pi - \alpha]$ de la série de fonctions de terme général $ka_k \sin kt$; la fonction φ_X est donc dérivable sur cet intervalle et donc aussi sur $\mathbb{R} \setminus 2\pi\mathbb{Z}$, puisqu'elle est 2π -périodique. Il reste à choisir convenablement la suite $(a_k)_{k \in \mathbb{Z}}$ pour obtenir la dérivabilité en 0. Prenons la suite définie par

$$a_0 = a_1 = a_{-1} = 0 \quad \text{et} \quad \forall k \geq 2 \quad a_k = a_{-k} = \frac{c}{k^2 \ln k},$$

où $c = \frac{1}{2} \left(\sum_{k=2}^{+\infty} \frac{1}{k^2 \ln k} \right)^{-1}$ (la série de terme général a_k est une **série de Bertrand** convergente). Toutes les conditions requises précédemment sont

satisfaites. De plus, pour tout $t \neq 0$, on a, par le théorème de transfert :

$$0 \leq \frac{1 - \varphi_X(t)}{t} = \frac{1}{t} \mathbb{E}[1 - \cos(tX)] = \frac{2c}{t} \sum_{k=2}^{+\infty} \frac{1}{k^2 \ln k} (1 - \cos(tk)).$$

Pour tout t tel que $0 < t < 1/2$, éclatons cette somme en deux, suivant que k est plus petit ou plus grand que t^{-1} . Les fonctions $x \mapsto (\ln x)^{-1}$ et $x \mapsto x^{-2}$ étant décroissantes (comparaison séries et intégrales), on a

$$\begin{aligned} \frac{1}{t} \sum_{k \geq \frac{1}{t}} \frac{1}{k^2 \ln k} (1 - \cos(tk)) &\leq -\frac{2}{t \ln t} \sum_{k \geq \left[\frac{1}{t}\right]} \frac{1}{k^2} \leq -\frac{2}{t \ln t} \int_{\left[\frac{1}{t}\right]-1}^{+\infty} \frac{1}{x^2} dx \\ &= -\frac{2}{t \left(\left[\frac{1}{t}\right] - 1\right) \ln t} \leq -2 \frac{\left[\frac{1}{t}\right] + 1}{\left[\frac{1}{t}\right] - 1} \cdot \frac{1}{\ln t} \xrightarrow{t \rightarrow 0} 0. \end{aligned}$$

Par ailleurs, en utilisant l'inégalité

$$\forall x \in \mathbb{R} \quad 1 - \cos x \leq \frac{x^2}{2},$$

il vient de même :

$$\begin{aligned} \frac{1}{t} \sum_{2 \leq k < \frac{1}{t}} \frac{1}{k^2 \ln k} (1 - \cos(tk)) &\leq t \sum_{2 \leq k < \frac{1}{t}} \frac{1}{\ln k} \leq \frac{t}{\ln 2} + t \sum_{3 \leq k \leq \left[\frac{1}{t}\right]} \frac{1}{\ln k} \\ &\leq \frac{t}{\ln 2} + t \sum_{3 \leq k \leq \left[\frac{1}{t}\right]} \int_{k-1}^k \frac{1}{\ln x} dx \leq \frac{t}{\ln 2} + t \int_2^{\frac{1}{t}} \frac{1}{\ln x} dx. \end{aligned}$$

De plus, classiquement, par une intégration par parties, on a :

$$\int_2^y \frac{1}{\ln x} dx = \left[\frac{x}{\ln x} \right]_2^y + \int_2^y \frac{1}{(\ln x)^2} dx.$$

Mais puisque, pour x tendant vers l'infini, $\frac{1}{(\ln x)^2} = o\left(\frac{1}{\ln x}\right)$, on a aussi, pour y tendant vers l'infini $\int_2^y \frac{1}{(\ln x)^2} dx = o\left(\int_2^y \frac{1}{\ln x} dx\right)$: il en résulte que

$$\lim_{t \rightarrow 0} t \int_2^{\frac{1}{t}} \frac{1}{\ln x} dx = 0,$$

ce qui achève de démontrer que

$$\lim_{t \rightarrow 0} \frac{1 - \varphi_X(t)}{t} = 0.$$

c'est-à-dire que φ_X est dérivable en 0 de dérivée nulle. **En résumé, pour un tel choix de loi, X n'admet pas de moyenne, cependant que φ_X est dérivable partout.**

La proposition 12.14 se généralise au cas de variables aléatoires à valeurs dans \mathbb{R}^d . Par souci de simplification, nous n'étudions dans ce cas que les moments d'ordre inférieur ou égal à deux.

Proposition 12.15. *Soit X une variable aléatoire à valeurs dans \mathbb{R}^d et soit φ_X sa fonction caractéristique.*

(a) *Si X admet une **moyenne** (c'est-à-dire est de norme intégrable), φ_X est différentiable; sa différentielle en t , application linéaire de \mathbb{R}^d dans \mathbb{C} est donnée par*

$$\forall t \in \mathbb{R}^d, \forall x \in \mathbb{R}^d \quad \varphi'_X(t)(x) = i \int_{\Omega} \langle X, x \rangle \exp(i \langle X, t \rangle) dP;$$

en particulier, on a :

$$\boxed{\forall x \in \mathbb{R}^d \quad \varphi'_X(0)(x) = i \langle EX, x \rangle .}$$

(b) *Si la norme de X est de carré intégrable, φ_X est deux fois différentiable et sa différentielle seconde en t , application bilinéaire de \mathbb{R}^d dans \mathbb{C} est donnée par*

$$\forall t \in \mathbb{R}^d, \forall x, y \in \mathbb{R}^d \quad \varphi''_X(t)(x, y) = - \int_{\Omega} \langle X, x \rangle \langle X, y \rangle \exp(i \langle X, t \rangle) dP;$$

en particulier, on a :

$$\boxed{\forall x \in \mathbb{R}^d \quad \varphi''_X(0)(x, x) = -E \langle X, x \rangle^2 .}$$

La **variance** de X est alors donnée par la relation

$$\boxed{\forall x \in \mathbb{R}^d \quad \sigma_X^2(x) = -\varphi''_X(0)(x, x) + [\varphi'_X(0)(x)]^2 ,}$$

et la **matrice de covariance** C_X de X est donnée par

$$\boxed{C_X = \left(-\frac{\partial^2}{\partial t_i \partial t_j} \varphi_X(0) \right) + \left(\frac{\partial}{\partial t_i} \varphi_X(0) \frac{\partial}{\partial t_j} \varphi_X(0) \right) .}$$

Démonstration. (a) Puisque

$$\frac{\partial}{\partial t} \exp(i \langle X, t \rangle) = i \langle X, \cdot \rangle \exp(i \langle X, t \rangle),$$

on a

$$\left\| \frac{\partial}{\partial t} \exp(i \langle X, t \rangle) \right\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{C})} \leq \|X\| ,$$

et on peut appliquer le théorème 12.13.

(b) De même, pour $x, y \in \mathbb{R}^d$,

$$\left[\frac{\partial^2}{\partial t^2} \exp(i \langle X, t \rangle) \right] (x, y) = - \langle X, x \rangle \langle X, y \rangle \exp(i \langle X, t \rangle),$$

et donc

$$\left\| \frac{\partial^2}{\partial t^2} \exp(i \langle X, t \rangle) \right\|_{\mathcal{L}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{C})} \leq \|X\|^2,$$

et on peut encore appliquer le théorème 12.13.

Enfin, pour calculer la variance, il suffit de se rappeler que

$$\forall x \in \mathbb{R}^d \quad \sigma_X^2(x) = E \langle X, x \rangle^2 - [E \langle X, x \rangle]^2.$$

L'opérateur d'auto-covariance de X , Λ_X , étant obtenu par bilinéarisation de la variance, vérifie

$$\forall x, y \in \mathbb{R}^d \quad \langle \Lambda_X x, y \rangle = -\varphi_X''(0)(x, y) + [\varphi_X'(0)(x)] [\varphi_X'(0)(y)],$$

ce qui donne la matrice de covariance C_X , représentation matricielle de Λ_X dans la base canonique (les dérivées partielles en 0 sont les valeurs de la différentielle en 0 prises aux différents vecteurs de cette base). \square

Exemple 12.4. (Suite de l'exemple 12.2.) Calculons la moyenne et la matrice de covariance de Y_n ($n \geq 2$). On a, d'après l'égalité (12.15),

$$\frac{\partial}{\partial t_j} \varphi_{Y_n}(t) = i n p_j \exp(it_j) \left[\sum_{m=1}^k p_m \exp(it_m) \right]^{n-1},$$

d'où

$$\frac{\partial}{\partial t_j} \varphi_{Y_n}(0) = i n p_j,$$

et donc

$$E(Y_n) = n p_j.$$

Si $j \neq l$, on a

$$\frac{\partial^2}{\partial t_l \partial t_j} \varphi_{Y_n}(t) = -n(n-1) p_l p_j \exp(it_l) \exp(it_j) \left[\sum_{m=1}^k p_m \exp(it_m) \right]^{n-2},$$

et donc

$$\frac{\partial^2}{\partial t_l \partial t_j} \varphi_{Y_n}(0) = -n(n-1) p_l p_j,$$

ce qui donne, après calcul :

$$\text{si } j \neq l \quad (C_{Y_n})_{jl} = -n p_l p_j.$$

Enfin, on a

$$\frac{\partial^2}{\partial t_j^2} \varphi_{X_n}(t) = in p_j \exp(it_j) \left[i \left[\sum_{m=1}^k p_m \exp(it_m) \right]^{n-1} + i(n-1) p_j \exp(it_j) \left[\sum_{m=1}^k p_m \exp(it_m) \right]^{n-2} \right],$$

et donc

$$\frac{\partial^2}{\partial t_j^2} \varphi_{X_n}(0) = -n p_j [1 + (n-1) p_j],$$

ce qui donne, après calcul :

$$(C_{X_n})_{jj} = n p_j (1 - p_j).$$

L'existence de moments permet d'obtenir un **développement limité** de la fonction caractéristique en zéro, ce qui est particulièrement utile pour l'étude des **convergences en loi** (voir au ch. 14 la démonstration du théorème limite central), mais aussi pour calculer les moments, par identification des coefficients du développement limité). On donne aussi une condition suffisante pour que la fonction caractéristique soit **développable en série entière**.

Proposition 12.16. *Soit X une variable aléatoire réelle et soit φ_X sa fonction caractéristique.*

(a) *Si X admet un moment d'ordre $n \in \mathbb{N}^*$, φ_X admet un **développement de Taylor en 0 avec reste intégral** qui s'écrit sous la forme, pour tout $t \in \mathbb{R}$,*

$$\varphi_X(t) = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \mathbb{E} X^k + \frac{(it)^n}{(n-1)!} \mathbb{E} \left[X^n \int_0^1 (1-u)^{n-1} \exp(ituX) du \right]. \quad (12.18)$$

Il en résulte que

$$\forall t \in \mathbb{R} \quad \varphi_X(t) = \sum_{k=0}^n \frac{(it)^k}{k!} \mathbb{E} X^k + \frac{(it)^n}{n!} \varepsilon_n(t), \quad (12.19)$$

où

$$|\varepsilon_n(t)| \leq 2\mathbb{E}|X^n| \quad \text{et} \quad \lim_{t \rightarrow 0} \varepsilon_n(t) = 0.$$

On obtient ainsi en particulier un **développement limité d'ordre n** de φ_X au voisinage de 0.

(b) *Si X admet des moments de tout ordre et si*

$$\limsup_n \frac{\|X\|_n}{n} = \frac{1}{R} < +\infty. \quad (12.20)$$

où $\|X\|_n$ est la norme⁴ n de X (cette condition est en particulier satisfaite si la variable aléatoire X est bornée), alors φ_X est **développable en série entière** au voisinage de tout réel, le rayon de convergence étant $\geq R/e$ (la fonction φ_X est donc analytique). En particulier, φ_X admet dans l'intervalle $]-R/e, R/e[$ le développement

$$\forall t \in \left] -\frac{R}{e}, \frac{R}{e} \right[\quad \varphi_X(t) = \sum_{k=0}^{+\infty} \frac{(it)^k}{k!} \mathbf{E}X^k. \quad (12.21)$$

Démonstration. (a) La formule de Taylor avec reste intégral appliquée à l'exponentielle complexe donne, pour tout $y \in \mathbb{R}$,

$$\exp(iy) = \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + \frac{(iy)^n}{(n-1)!} \int_0^1 (1-u)^{n-1} \exp(iuy) du,$$

et donc, pour tout $t \in \mathbb{R}$,

$$\exp(itX) = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} X^k + \frac{(it)^n}{(n-1)!} X^n \int_0^1 (1-u)^{n-1} \exp(iutX) du. \quad (12.22)$$

On obtient la formule (12.18) en intégrant par rapport à P . Par ailleurs, en remarquant que

$$\int_0^1 (1-u)^{n-1} du = \frac{1}{n},$$

on a, pour tout $y \in \mathbb{R}$,

$$\exp(iy) = \sum_{k=0}^n \frac{(iy)^k}{k!} + \frac{(iy)^n}{(n-1)!} \int_0^1 [(1-u)^{n-1} (\exp(iuy) - 1)] du,$$

ce qui donne, pour tout $t \in \mathbb{R}$,

$$\exp(itX) = \sum_{k=0}^n \frac{(it)^k}{k!} X^k + \frac{(it)^n}{(n-1)!} X^n \int_0^1 [(1-u)^{n-1} (\exp(iutX) - 1)] du.$$

En intégrant par rapport à P , il vient, pour tout $t \in \mathbb{R}$,

$$\varphi_X(t) = \sum_{k=0}^n \frac{(it)^k}{k!} \mathbf{E}X^k + \frac{(it)^n}{n!} \varepsilon_n(t),$$

où

$$\varepsilon_n(t) = n \int_{\Omega} \left[X^n \int_0^1 (1-u)^{n-1} [\exp(iutX) - 1] du \right] dP.$$

4. On rappelle que la norme n d'une variable aléatoire X est définie par $\|X\|_n = (\int_{\Omega} |X|^n dP)^{1/n}$, dans la mesure où cette quantité est finie.

En utilisant le théorème de Fubini, on obtient la majoration :

$$|\varepsilon_n(t)| \leq 2nE|X^n| \int_0^1 (1-u)^{n-1} du = 2E|X^n|.$$

De plus, on a

$$|X^n(1-u)^{n-1} [\exp(iutX) - 1]| \leq 2|X|^n(1-u)^{n-1},$$

majoration indépendante de t par une application $\lambda_{[0,1]} \otimes P$ -intégrable. Après application du théorème de Fubini, il résulte du théorème de convergence dominée (prendre une suite quelconque qui converge vers 0) que

$$\lim_{t \rightarrow 0} \varepsilon_n(t) = 0.$$

(b) Soit un réel quelconque t_0 . La variable aléatoire X admettant des moments de tout ordre, sa fonction caractéristique φ_X (à variable réelle...) est \mathcal{C}^∞ . Celle-ci admet un développement de Taylor de tout ordre n , donné, pour tout réel t , par

$$\varphi_X(t) = \varphi_X(t_0) + \sum_{k=1}^n \frac{(t-t_0)^k}{k!} \varphi_X^{(k)}(t_0) + R_n(t_0, t),$$

où le reste est défini par

$$R_n(t_0, t) = \int_{t_0}^t \frac{(t-u)^n}{n!} \varphi_X^{(n+1)}(u) du.$$

Il s'agit de démontrer que ce reste tend vers 0. Remarquons qu'il résulte de (12.16) que l'on a

$$|R_n(t_0, t)| \leq \frac{[|t-t_0| \|X\|_{n+1}]^{n+1}}{(n+1)!}.$$

Soit $\varepsilon > 0$ quelconque ; la condition (12.20) implique qu'il existe N tel que, pour tout $n \geq N$, on ait

$$\frac{\|X\|_n}{n} \leq \frac{1}{R} + \varepsilon,$$

soit, en utilisant la formule de Stirling :

$$\left[\frac{|t-t_0|^n \|X\|_n^n}{n!} \right]^{\frac{1}{n}} \leq |t-t_0| \left(\frac{1}{R} + \varepsilon \right) n \left[\left(\frac{n}{e} \right)^n \sqrt{2\pi n} \left(1 + \frac{1}{12n} + o\left(\frac{1}{n}\right) \right) \right]^{-\frac{1}{n}}.$$

Le membre de droite convergeant vers $|t-t_0| \left(\frac{1}{R} + \varepsilon \right) e$, il résulte de l'arbitraire de ε que

$$\limsup_n \left[\frac{|t-t_0|^n \|X\|_n^n}{n!} \right]^{\frac{1}{n}} \leq |t-t_0| \frac{e}{R}.$$

Alors, pour tout t tel que $|t - t_0| < R/e$, le test de Cauchy montre que la série de terme général $|t - t_0|^n \|X\|_n^n/n!$ est convergente. Il en résulte que le reste de Taylor $R_n(t_0, t)$ tend vers 0 quand n tend vers l'infini, ce qui prouve l'existence du développement de Taylor en t_0 pour φ_X et aussi l'analyticité de φ_X . On obtient alors le développement (12.21) en prenant $t_0 = 0$ et en tenant compte de la valeur des dérivées de φ_X en 0 donnée par (12.17). \square

Exercices

Toutes les variables aléatoires introduites sont définies sur un même espace probabilisé (Ω, \mathcal{A}, P) .

Exercice 12.1. Fonction caractéristique et injectivité, loi triangulaire ; des fonctions caractéristiques peuvent coïncider sur un intervalle sans être égales (Stoyanov). Soit Φ la fonction définie sur \mathbb{R} par, pour tout réel t :

$$\Phi(t) = \begin{cases} 1 - |t| & \text{si } 0 \leq |t| \leq 1 \\ 0 & \text{sinon.} \end{cases}$$

Soit X une variable aléatoire de loi triangulaire sur l'intervalle $[-1, 1]$, c'est-à-dire que X est de densité $f_X = \Phi$.

1. Calculer la fonction caractéristique φ_X de X . On note ρ la loi uniforme sur $[-\frac{1}{2}, \frac{1}{2}]$. Justifier le fait que la loi P_X de X est le produit de convolution $\rho * \rho$.
2. Démontrer que Φ est la transformée de Fourier d'une probabilité $\mu = f \cdot \lambda$ sur \mathbb{R} où f est une densité de probabilité à déterminer.
3. Soient Y et Z deux variables aléatoires telles que Y soit de densité f et Z soit à valeurs dans l'ensemble des entiers relatifs \mathbb{Z} , de loi donnée par

$$P_Z = \frac{1}{2} \delta_0 + \sum_{k \in \mathbb{Z}} \frac{2}{(2k-1)^2 \pi^2} \delta_{(2k-1)\pi}.$$

Il s'agit de démontrer que les fonctions caractéristiques de Y et Z coïncident sur l'intervalle $[-1, 1]$ et ne sont pas égales. Pour cela, développer en série de Fourier la fonction Ψ , périodique de période 2 et égale à $1 - \Phi$ sur l'intervalle $[-1, 1]$ et conclure.

Solution.

1. La densité de X étant paire, on a

$$\varphi_X(t) = \int_{-1}^{+1} (1 - |t|) \exp(itx) dx = 2 \int_0^{+1} (1 - t) \cos(tx) dx,$$

ce qui donne, après une intégration par parties, valable pour tout $t \neq 0$,

$$\forall t \neq 0 \quad \varphi_X(t) = 2 \frac{1 - \cos t}{t^2} \quad \text{et} \quad \varphi_X(0) = 1.$$

Ceci s'écrit encore

$$\forall t \neq 0 \quad \varphi_X(t) = 4 \frac{\sin^2 \frac{t}{2}}{t^2} = [\widehat{\rho}(t)]^2 \quad \text{et} \quad \varphi_X(0) = 1,$$

puisque la transformée de Fourier de ρ vaut, en tout $t \neq 0$,

$$\widehat{\rho}(t) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} \exp(itx) dx = 2 \frac{\sin \frac{t}{2}}{t}.$$

Il résulte de la propriété d'injectivité de la transformée de Fourier que la loi de X est la convolution de la loi uniforme sur $[-\frac{1}{2}, \frac{1}{2}]$ par elle-même.

2. Puisque Φ est intégrable, il résulte de la proposition 12.7 que si μ est une probabilité telle que $\widehat{\mu} = \Phi$, elle est de densité f donnée par

$$\forall x \in \mathbb{R} \quad f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \widehat{\mu}(t) \exp(-ixt) dt \equiv \frac{1}{\sqrt{2\pi}} \widehat{\Phi}(-x).$$

Puisque la variable aléatoire X est de densité $f_X = \Phi$, sa fonction caractéristique φ_X vaut $\widehat{\Phi}$; on doit donc avoir

$$f(x) = \frac{1}{\sqrt{2\pi}} \varphi_X(-x),$$

et donc, puisque φ_X est paire,

$$f(x) = \frac{1}{\sqrt{2\pi}} \varphi_X(x).$$

Il reste à vérifier que la probabilité $\mu = f \cdot \lambda$ ainsi définie vérifie bien $\widehat{\mu} = \Phi$. On a, pour tout réel t :

$$\widehat{\mu}(t) = \int_{\mathbb{R}} f(x) \exp(ixt) dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi_X(x) \exp(ixt) dx;$$

on a donc, toujours par la proposition 12.7 (φ_X est intégrable), que $\widehat{\mu}(t) = f_X(-t)$, et puisque f_X est paire, que $\widehat{\mu} = f_X = \Phi$.

3. La fonction caractéristique de Z vaut en tout réel t :

$$\varphi_Z(t) = \frac{1}{2} + \frac{2}{\pi^2} \sum_{k \in \mathbb{Z}} \frac{2}{(2k-1)^2} \exp[it(2k-1)\pi].$$

soit

$$\varphi_Z(t) = \frac{1}{2} + \frac{4}{\pi^2} \sum_{k=1}^{+\infty} \frac{1}{(2k-1)^2} \cos[(2k-1)\pi t].$$

Par ailleurs, la fonction Ψ , est paire, périodique de période 2, continue et \mathcal{C}^1 par morceaux. Le théorème de Dirichlet assure que, pour tout réel t ,

$$\Psi(t) = \frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n \cos\left(2\pi n \frac{t}{2}\right).$$

où

$$a_0 = \int_{-1}^1 |t| dt = 1 \quad \text{et}$$

$$a_n = \int_{-1}^1 |t| \cos(2\pi n \frac{t}{2}) dt = 2 \int_0^1 t \cos(\pi n t) dt \quad \text{si } n \in \mathbb{N}^*.$$

Si $n \in \mathbb{N}^*$, on a, après intégration par parties,

$$a_n = -\frac{2}{\pi^2 n^2} (1 - \cos(\pi n)) = \begin{cases} -\frac{4}{\pi^2 n^2} & \text{si } n \text{ est pair} \\ 0 & \text{si } n \text{ est impair;} \end{cases}$$

on a donc, pour tout t réel :

$$\Psi(t) = \frac{1}{2} - \frac{4}{\pi^2} \sum_{k=1}^{+\infty} \frac{1}{(2k-1)^2} \cos[(2k-1)\pi t].$$

Il en résulte que, sur l'intervalle $[-1, 1]$:

$$\varphi_Z(t) = 1 - \Psi(t) = \Phi(t) = \varphi_Y(t).$$

Il est à remarquer que la variable aléatoire Z est discrète tandis que la variable aléatoire Y admet une densité.

Exercice 12.2. Fonction caractéristique d'un produit de variables aléatoires indépendantes. Soient X et Y deux variables aléatoires réelles indépendantes.

1. Démontrer que la fonction caractéristique du produit XY est donnée par la relation : pour tout $t \in \mathbb{R}$,

$$\varphi_{XY}(t) = \int_{\mathbb{R}} \varphi_X(ty) dP_Y(y).$$

Si de plus X et Y ont même loi normale $\mathcal{N}_{\mathbb{R}}(0, 1)$, déterminer la fonction caractéristique de XY .

2. Soient X_1, X_2, X_3, X_4 , quatre variables aléatoires réelles indépendantes, de loi normale $\mathcal{N}_{\mathbb{R}}(0, 1)$. Déterminer la fonction caractéristique et la loi de la variable aléatoire $X_1 X_2 + X_3 X_4$.

3. Quelle est la loi de la variable aléatoire $|X_1 X_2 + X_3 X_4|$?

Solution.

1. L'application $(x, y) \mapsto \exp(itxy)$ étant bornée, le théorème de transfert assure que

$$\varphi_{XY}(t) = \int_{\mathbb{R}^2} \exp(itxy) dP_{(X,Y)}(x, y).$$

Les variables aléatoires X et Y étant indépendantes, la loi du couple (X, Y) est le produit des lois de X et Y . L'application $(x, y) \mapsto \exp(itxy)$ étant bornée, donc $P_X \otimes P_Y$ -intégrable, le théorème de Fubini permet alors d'écrire

$$\varphi_{XY}(t) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \exp(itxy) dP_X(x) \right] dP_Y(y).$$

ce qui démontre la formule demandée. Dans le cas de variables aléatoires de loi normale $\mathcal{N}_{\mathbb{R}}(0, 1)$, la fonction caractéristique de XY est alors donnée par, pour tout $t \in \mathbb{R}$:

$$\begin{aligned}\varphi_{XY}(t) &= \int_{\mathbb{R}} \exp\left(-\frac{t^2 y^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(1+t^2)y^2}{2}\right) dy.\end{aligned}$$

soit :

$$\boxed{\varphi_{XY}(t) = \frac{1}{\sqrt{1+t^2}}}.$$

2. Les variables aléatoires $X_1 X_2$ et $X_3 X_4$ sont indépendantes et de même loi que celle de XY ; on a donc, pour tout $t \in \mathbb{R}$,

$$\varphi_{X_1 X_2 + X_3 X_4}(t) = \varphi_{X_1 X_2}(t) \varphi_{X_3 X_4}(t) = \frac{1}{1+t^2}.$$

Il résulte du théorème d'injectivité de la transformation de Fourier que $X_1 X_2 + X_3 X_4$ suit la **loi de Laplace** de densité la fonction $x \mapsto \frac{1}{2} \exp(-|x|)$ (voir le tableau des lois).

3. Pour toute $f \in \mathcal{C}_{\mathcal{K}}^+(\mathbb{R})$, on a alors

$$\int_{\mathbb{R}} f(|u|) dP_{|X_1 X_2 + X_3 X_4|}(u) = \int_{\mathbb{R}} f(|u|) \frac{1}{2} \exp(-|u|) du.$$

soit, en utilisant la parité de l'intégrande :

$$\int_{\mathbb{R}} f(|u|) dP_{|X_1 X_2 + X_3 X_4|}(u) = \int_{\mathbb{R}} f(u) \mathbf{1}_{\mathbb{R}^+}(u) \exp(-u) du.$$

La variable aléatoire $|X_1 X_2 + X_3 X_4|$ suit donc la **loi exponentielle** de paramètre 1, notée $\exp(1)$.

Exercice 12.3. Fonction caractéristique, convolution et moments. Soit U une variable aléatoire réelle de densité f_U donnée par, pour tout $u \in \mathbb{R}$:

$$f_U(u) = \sum_{n=0}^{+\infty} \mathbf{1}_{[n, n+1[}(u) \exp(-\lambda) \frac{\lambda^n}{n!}.$$

1. Calculer la fonction caractéristique φ_U de U et en déduire que la loi de U est la convolution de deux lois à préciser.

2. Déterminer alors sans calcul la moyenne et la variance de U .

3. Soit T une variable aléatoire indépendante de U et de loi uniforme sur l'intervalle $[0, 1]$. Déterminer la fonction caractéristique de la variable aléatoire $W = T + U$ (penser à utiliser la première question). Justifier sa dérivabilité et donner, sans calcul, sa dérivée en 0.

Solution

1. Soit t un réel quelconque. On a

$$\varphi_U(t) = \int_{\mathbb{R}} \exp(itu) f_U(u) du.$$

Puisque

$$\sum_{n=0}^{+\infty} \int_{\mathbb{R}} |\exp(itu)| \mathbf{1}_{[n, n+1]}(u) \exp(-\lambda) \frac{\lambda^n}{n!} du = \sum_{n=0}^{+\infty} \exp(-\lambda) \frac{\lambda^n}{n!} = 1,$$

il résulte du corollaire du théorème de convergence dominée sur les séries de fonctions (cf. annexe, corollaire A.33) que, pour tout $t \neq 0$,

$$\begin{aligned} \varphi_U(t) &= \sum_{n=0}^{+\infty} \left[\exp(-\lambda) \frac{\lambda^n}{n!} \int_{\mathbb{R}} \exp(itu) \mathbf{1}_{[n, n+1]}(u) du \right] \\ &= \sum_{n=0}^{+\infty} \exp(-\lambda) \frac{\lambda^n}{n!} \frac{\exp(it(n+1)) - \exp(itn)}{it} \\ &= \frac{\exp(-\lambda) [\exp(it) - 1]}{it} \sum_{n=0}^{+\infty} \frac{[\lambda \exp(it)]^n}{n!}, \end{aligned}$$

soit

$$\varphi_U(t) = \frac{\exp(it) - 1}{it} \exp[\lambda(\exp(it) - 1)].$$

Par la propriété d'injectivité de la transformation de Fourier, il en résulte que la loi de U est **convolution** de la loi uniforme sur l'intervalle $[0, 1]$ et de la loi de Poisson $\mathcal{P}(\lambda)$.

2. Soient deux variables aléatoires indépendantes X et N , la première de loi uniforme sur l'intervalle $[0, 1]$ et la deuxième de loi de Poisson $\mathcal{P}(\lambda)$. La loi de $X + N$ est alors la loi de U . On a donc

$$EU = EX + EN = \frac{1}{2} + \lambda.$$

et, puisque X et N sont indépendantes,

$$\sigma_U^2 = \sigma_X^2 + \sigma_N^2 = \frac{1}{12} + \lambda.$$

3. Supposons de plus que les variables aléatoires X et N introduites ci-dessus soient aussi indépendantes de T . On a

$$P_W = P_T * P_U = P_T * (P_X * P_N).$$

En prenant les transformées de Fourier, il en résulte que :

$$\varphi_W = \varphi_T \varphi_X \varphi_N$$

(remarque qu'ainsi, en utilisant à nouveau l'injectivité de la transformation de Fourier, on démontre que le **produit de convolution est associatif**). On a donc, pour tout $t \neq 0$,

$$\varphi_W(t) = -\frac{[\exp(it) - 1]^2}{t^2} \exp[\lambda(\exp(it) - 1)].$$

Les variables aléatoires T , X et N admettent une moyenne; il en est de même pour W . La fonction caractéristique φ_W de W est donc dérivable et on a

$$\varphi'_W(0) = iEW = iE[T + X + N] = i(1 + \lambda).$$

Exercice 12.4. Fonction caractéristique; développement limité et développement en série entière. Soient X , Y et Z trois variables aléatoires indépendantes de même loi normale $\mathcal{N}_{\mathbb{R}}(0, 1)$.

- Déterminer les moments de tout ordre de X à partir de sa fonction caractéristique.
- En déduire les moments de tout ordre de la variable aléatoire $U = XY$; trouver ainsi la fonction caractéristique de U (voir le premier exercice pour une autre méthode qui est d'ailleurs plus rapide).
- On note V la variable aléatoire YZ . Déterminer la fonction caractéristique du couple (U, V) . En déduire la fonction caractéristique de la variable aléatoire $\frac{U+V}{\sqrt{2}}$. Comparer la loi de cette dernière variable aléatoire à celle de U .
- Les variables aléatoires U et V sont-elles indépendantes? Sont-elles corrélées?

Solution.

1. La fonction caractéristique φ_X de X admet un développement limité de tout ordre, et on a pour tout n :

$$\forall t \in \mathbb{R} \quad \varphi_X(t) = \exp\left(-\frac{t^2}{2}\right) = \sum_{k=0}^n \frac{(-1)^k t^{2k}}{2^k k!} + o(t^{2n}).$$

La variable aléatoire X admet donc des moments de tout ordre (ce que l'on peut voir directement); ils sont donnés, pour tout $k \in \mathbb{N}$, par

$$E(X^{2k+1}) = 0 \quad \text{et} \quad E(X^{2k}) = \frac{(2k)!}{2^k k!} = 1 \cdot 3 \cdots (2k-3)(2k-1).$$

On remarque de plus que φ_X est analytique sur \mathbb{R} .

2. La fonction caractéristique de U admet un développement de Taylor de tout ordre n , donné, pour tout réel t , par

$$\varphi_U(t) = 1 + \sum_{k=1}^n \frac{(it)^k}{k!} E(U^k) + R_n(t).$$

où le reste est défini par

$$R_n(t) = \int_0^t \frac{(t-u)^n}{n!} \varphi_U^{(n+1)}(u) du.$$

Les variables aléatoires X et Y étant indépendantes, U admet des moments de tout ordre donnés par

$$E(U^n) = E(X^n) E(Y^n),$$

soit, pour tout $k \in \mathbb{N}$,

$$E(U^{2k+1}) = 0 \quad \text{et} \\ E(U^{2k}) = \left[\frac{(2k)!}{2^k k!} \right]^2 = [1 \cdot 3 \cdots (2k-3)(2k-1)]^2.$$

Il s'agit alors de démontrer que le reste de Taylor $R_n(t)$ tend vers 0. Remarquons qu'il résulte de (12.16) que l'on a

$$|R_n(t)| \leq \frac{|t|^{n+1} E|U|^{n+1}}{(n+1)!}.$$

Par l'inégalité de Schwarz, on a

$$E|U|^{2k+1} \leq (E|U|^{2k})^{\frac{1}{2}} (E|U|^{2k+2})^{\frac{1}{2}},$$

soit, en tenant compte des valeurs de ces moments :

$$E|U|^{2k+1} \leq (2k+1) E(U^{2k}),$$

ce qui donne la majoration

$$|R_{2k}(t)| \leq \frac{|t|^{2k+1} E(U^{2k})}{(2k)!} = |t|^{2k+1} \frac{1 \cdot 3 \cdots (2k-3)(2k-1)}{2^k k!};$$

on a de plus la majoration

$$|R_{2k-1}(t)| \leq |t|^{2k} \frac{1 \cdot 3 \cdots (2k-3)(2k-1)}{2^k k!}.$$

Le reste de Taylor $R_n(t)$ tend donc vers 0 dès que $|t| < 1$ et on a alors

$$\varphi_U(t) = \sum_{k=0}^{+\infty} (-1)^k \frac{1 \cdot 3 \cdots (2k-3)(2k-1)}{2^k k!} t^{2k},$$

soit

$$\varphi_U(t) = \frac{1}{\sqrt{1+t^2}}.$$

Par le principe du prolongement analytique, cette formule est alors vraie pour tout réel.

3. Soit $(a, b) \in \mathbb{R}^2$ quelconque. Les variables aléatoires X, Y, Z étant indépendantes, les théorèmes de transfert, puis de Fubini (l'intégrande est bornée) conduisent à l'expression de la fonction caractéristique de (U, V) suivante :

$$\begin{aligned}\varphi_{(U,V)}(a, b) &= \int_{\mathbb{R}^2} \left[\int_{\mathbb{R}} \exp(i(ax + bz)y) dP_Y(y) \right] dP_X \otimes dP_Z(x, z) \\ &= \int_{\mathbb{R}^2} \varphi_Y(ax + bz) dP_X \otimes dP_Z(x, z).\end{aligned}$$

c'est-à-dire

$$\begin{aligned}\varphi_{(U,V)}(a, b) &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \exp\left[-\frac{(ax + bz)^2}{2}\right] \exp\left[-\frac{x^2 + z^2}{2}\right] dx dz \\ &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \exp\left[-\frac{(1+a^2)x^2 + 2abxz + (1+b^2)z^2}{2}\right] dx dz.\end{aligned}$$

Le théorème de Fubini donne

$$\begin{aligned}\varphi_{(U,V)}(a, b) &= \\ \frac{1}{2\pi} \int_{\mathbb{R}} \exp\left[-\frac{(1+a^2)x^2}{2}\right] &\left[\int_{\mathbb{R}} \exp\left[-\frac{(1+b^2)z^2 + 2abxz}{2}\right] dz \right] dx.\end{aligned}$$

Un simple calcul conduit à démontrer que, pour tout $(a, b) \in \mathbb{R}^2$,

$$\int_{\mathbb{R}} \exp\left[-\frac{au^2 + 2bu}{2}\right] du = \exp\left(\frac{b^2}{2a}\right) \int_{\mathbb{R}} \exp\left[-\frac{v^2}{2}\right] \frac{1}{\sqrt{a}} dv = \sqrt{\frac{2\pi}{a}} \exp\left(\frac{b^2}{2a}\right).$$

Il vient alors

$$\varphi_{(U,V)}(a, b) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp\left[-\frac{(1+a^2)x^2}{2}\right] \sqrt{\frac{2\pi}{1+b^2}} \exp\left(\frac{a^2 b^2 x^2}{2(1+b^2)}\right) dx,$$

ce qui donne après réduction :

$$\varphi_{(U,V)}(a, b) = \frac{1}{\sqrt{1+a^2+b^2}}.$$

La fonction caractéristique de $\frac{U+V}{\sqrt{2}}$ est donnée par, pour tout $t \in \mathbb{R}$,

$$\varphi_{\frac{U+V}{\sqrt{2}}}(t) = \varphi_{(U,V)}\left(\frac{t}{\sqrt{2}}, \frac{t}{\sqrt{2}}\right) = \frac{1}{\sqrt{1+t^2}}.$$

Le théorème d'injectivité de la transformation de Fourier assure que **les variables aléatoires $\frac{U+V}{\sqrt{2}}$ et U ont même loi.**

4. On a

$$\varphi_{(U,V)}(a, b) \neq \varphi_{(U,V)}(a, 0)\varphi_{(U,V)}(0, b) = \varphi_U(a)\varphi_V(b),$$

ce qui démontre que **les variables aléatoires U et V ne sont pas indépendantes.** Par contre, les variables aléatoires X, Y, Z étant indépendantes, on a

$$E(UV) = E(XY^2Z) = E(X)E(Y^2)E(Z) = 0.$$

Puisque U et V sont centrées, il en résulte que **les variables aléatoires U et V sont non corrélées.**

Exercice 12.5. Critère d'indépendance de variables aléatoires bornées (M. Kac).

Soient X et Y deux variables aléatoires réelles bornées. Démontrer que pour que X et Y soient indépendantes, il faut et il suffit que

$$\forall (k, l) \in \mathbb{N}^2 \quad \mathbb{E}[X^k Y^l] = \mathbb{E}(X^k) \mathbb{E}(Y^l). \quad (12.23)$$

Solution. La condition est nécessaire, puisque si X et Y sont indépendantes, les variables aléatoires X^k et Y^l le sont aussi, ce qui donne la relation (12.23). Inversement, supposons la vraie. La fonction caractéristique de (X, Y) est donnée en tout $(u, v) \in \mathbb{R}^2$ par

$$\varphi_{(X, Y)}(u, v) = \mathbb{E}[\exp(iuX) \exp(ivY)] = \mathbb{E}\left[\left(\sum_{k=0}^{+\infty} \frac{(iu)^k X^k}{k!}\right) \left(\sum_{l=0}^{+\infty} \frac{(iv)^l Y^l}{l!}\right)\right].$$

Soit C un majorant de $|X|$ et $|Y|$. On a

$$\sum_{(k, l) \in \mathbb{N}^2} \frac{|u|^k |v|^l C^{k+l}}{k! l!} = \exp(|u|C) \exp(|v|C) < +\infty.$$

Il en résulte que la famille $\left\{\frac{(iu)^k X^k}{k!} \frac{(iv)^l Y^l}{l!} \mid (k, l) \in \mathbb{N}^2\right\}$ est sommable et, puisqu'elle est dénombrable, que l'on a, par application du théorème de convergence dominée :

$$\varphi_{(X, Y)}(u, v) = \sum_{(k, l) \in \mathbb{N}^2} i^{k+l} \frac{u^k v^l}{k! l!} \mathbb{E}[X^k Y^l].$$

En tenant compte de l'hypothèse, il vient

$$\varphi_{(X, Y)}(u, v) = \sum_{(k, l) \in \mathbb{N}^2} i^{k+l} \frac{u^k v^l}{k! l!} \mathbb{E}(X^k) \mathbb{E}(Y^l).$$

Cette dernière famille est encore sommable et il vient, par application de la propriété de Fubini, puis à nouveau du théorème de convergence dominée,

$$\begin{aligned} \varphi_{(X, Y)}(u, v) &= \left[\sum_{k \in \mathbb{N}} i^k \frac{u^k}{k!} \mathbb{E}(X^k) \right] \left[\sum_{l \in \mathbb{N}} i^l \frac{v^l}{l!} \mathbb{E}(Y^l) \right] \\ &= \mathbb{E}\left(\sum_{k \in \mathbb{N}} i^k \frac{u^k}{k!} X^k\right) \mathbb{E}\left(\sum_{l \in \mathbb{N}} i^l \frac{v^l}{l!} Y^l\right), \end{aligned}$$

c'est-à-dire

$$\varphi_{(X, Y)}(u, v) = \varphi_X(u) \varphi_Y(v).$$

Ceci démontre l'indépendance de X et Y .

Exercice 12.6. On peut avoir $\varphi_{X+Y} = \varphi_X \varphi_Y$ sans que les variables aléatoires X et Y soient indépendantes. Soit f la fonction définie sur \mathbb{R}^2 par

$$\forall (x, y) \in \mathbb{R}^2 \quad f(x, y) = \frac{1}{4} \mathbf{1}_C(x, y) [1 + xy(x^2 - y^2)],$$

où $C = [-1, 1]^2$.

1. Vérifier que f est une densité de probabilité sur \mathbb{R}^2 relativement à la mesure de Lebesgue.
2. Soit (X, Y) une variable aléatoire à valeurs dans \mathbb{R}^2 de densité f . On définit la variable aléatoire $Z = X + Y$. Calculer les densités des variables aléatoires X, Y, Z . Préciser les lois de X et Y .
3. Démontrer que si U est une variable aléatoire réelle admettant une densité paire, sa fonction caractéristique vérifie :

$$\forall t \in \mathbb{R} \quad \varphi_U(t) = 2 \int_0^{+\infty} \cos(tu) f_U(u) du.$$

Calculer alors les fonctions caractéristiques de X et Y . Exprimer la fonction caractéristique de Z en fonction de celles de X et Y .

4. Calculer le coefficient de corrélation de X et Y .
5. Remarques sur cet exercice.

Solution.

1. La fonction f est positive ; en effet, pour tout $(x, y) \in C$, on a

$$-1 \leq -y^2 \leq x^2 - y^2 \leq x^2 \leq 1 \quad \text{et donc} \quad |xy(x^2 - y^2)| \leq 1;$$

il en résulte que

$$1 + xy(x^2 - y^2) \geq 0.$$

La fonction f est mesurable ; c'est une densité, puisqu'en utilisant la linéarité puis les symétries, on trouve que :

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \frac{1}{4} \int_C [1 + xy(x^2 - y^2)] dx dy = 1.$$

2. La variable aléatoire marginale X admet la densité f_X donnée par, pour tout réel x ,

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy = \frac{1}{4} \mathbf{1}_{[-1, 1]}(x) \int_{-1}^{+1} [1 + xy(x^2 - y^2)] dy.$$

Il en résulte que :

$$f_X = \frac{1}{2} \mathbf{1}_{[-1, 1]}.$$

La variable aléatoire X est de loi uniforme sur l'intervalle $[-1, 1]$. Par symétrie, il en est de même pour la variable aléatoire Y . Remarquons que f n'est pas le produit direct des densités de X et de Y ; **les variables aléatoires X et Y ne sont donc pas indépendantes.**

Étudions maintenant la loi de Z . La variable aléatoire $(X + Y, X)$ est transformée par un difféomorphisme linéaire de la variable aléatoire (X, Y) ; le jacobien étant de valeur absolue 1, la variable aléatoire (Z, X) admet la densité $f_{(Z, X)}$ donnée en tout $(z, t) \in \mathbb{R}^2$ par

$$f_{(Z, X)}(z, t) = f_{(X, Y)}(t, z - t).$$

La variable aléatoire Z admet la densité f_Z donnée en tout $z \in \mathbb{R}$ par

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(t, z-t) dt \\ &= \frac{1}{4} \int_{-\infty}^{+\infty} [1 + t(z-t)(t^2 - (z-t)^2)] \mathbf{1}_{\mathbb{C}}(t, z-t) dt, \end{aligned}$$

soit

$$f_Z(z) = \frac{1}{4} \mathbf{1}_{[-2,2]}(z) \int_{\max(z-1, -1)}^{\min(z+1, 1)} [1 + t(z-t)(t^2 - (z-t)^2)] dt.$$

Si $0 < z < 2$, on a

$$f_Z(z) = \frac{1}{4} \int_{z-1}^1 [1 + zt(z-t)(2t-z)] dt.$$

En faisant le changement de variables $u = 2t - z$, on obtient

$$\begin{aligned} f_Z(z) &= \frac{1}{8} \int_{-(2-z)}^{2-z} \left[1 + zu \frac{u+z}{2} \frac{z-u}{2} \right] du \\ &= \frac{1}{8} \int_{-(2-z)}^{2-z} \left[\frac{z^2 - u^2}{4} \right] zu du + \frac{1}{8} \int_{-(2-z)}^{2-z} du, \end{aligned}$$

soit, la première intégrale étant nulle,

$$f_Z(z) = \frac{2-z}{4}.$$

Les variables aléatoires (X, Y) et $(-X, -Y)$ ayant même loi, il en est de même des variables aléatoires Z et $-Z$; il en résulte que f_Z est paire. On a donc

$$\forall z \in \mathbb{R} \quad f_Z(z) = \mathbf{1}_{[-2,2]}(z) \frac{2-|z|}{4}.$$

3. Si U a une densité paire, sa fonction caractéristique φ_U vérifie, pour tout $t \in \mathbb{R}$,

$$\varphi_U(t) = \int_{\mathbb{R}} [\cos(tu) + i \sin(tu)] f_U(u) du,$$

soit

$$\varphi_U(t) = 2 \int_0^{+\infty} \cos(tu) f_U(u) du.$$

La fonction caractéristique de X est alors donnée par

$$\forall t \in \mathbb{R} \quad \varphi_X(t) = 2 \int_0^1 \cos(tu) \frac{1}{2} du.$$

soit

$$\forall t \neq 0, \quad \varphi_X(t) = \varphi_Y(t) = \frac{\sin t}{t} \quad \text{et} \quad \varphi_X(0) = \varphi_Y(0) = 1.$$

De même, la fonction caractéristique de Z est donnée par

$$\forall t \neq 0 \quad \varphi_Z(t) = 2 \int_0^2 \cos(tz) \frac{2-z}{4} dz = \frac{\sin 2t}{t} - \frac{1}{2} \int_0^2 z \cos(tz) dz,$$

soit, après une intégration par parties,

$$\forall t \neq 0 \quad \varphi_Z(t) = \frac{1 - \cos 2t}{2t^2} = \frac{\sin^2 t}{t^2}.$$

On a donc

$$\boxed{\varphi_Z = \varphi_X \varphi_Y.}$$

autrement dit, par la propriété d'injectivité de la transformation de Fourier, on a $P_{X+Y} = P_X * P_Y$.

4. On a $EX = EY = 0$. De plus

$$E(XY) = \frac{1}{4} \int_C xy [1 + xy(x^2 - y^2)] dx dy,$$

soit

$$E(XY) = \frac{1}{4} \left[\int_C x^4 y^2 dx dy - \int_C x^2 y^4 dx dy \right] = 0.$$

Le coefficient de corrélation de X et Y est donc nul.

5. En résumé, on a un exemple de variable aléatoire (X, Y) de loi non uniforme sur C dont les deux marginales sont de loi uniforme et sont non indépendantes tout en étant non corrélées. Toutefois, ces marginales vérifient $\varphi_{X+Y} = \varphi_X \varphi_Y$ (et donc $P_{X+Y} = P_X * P_Y$).

Exercice 12.7. Encore un exemple où l'on a $\varphi_{X+Y} = \varphi_X \varphi_Y$ sans que les variables aléatoires soient indépendantes; lois de Cauchy. La loi de Cauchy μ_a de paramètre $a > 0$ est la probabilité sur \mathbb{R} de densité f_a définie par, pour tout réel x ,

$$f_a(x) = \frac{a}{\pi(a^2 + x^2)}.$$

1. Démontrer la relation sur les transformées de Fourier :

$$\forall t \in \mathbb{R} \quad \widehat{\mu}_a(t) = \widehat{\mu}_1(at);$$

en déduire que si une variable aléatoire Z suit une loi de Cauchy de paramètre 1, la variable aléatoire aZ suit une loi de Cauchy de paramètre a .

2. Soient U et V deux variables aléatoires de loi de Cauchy de paramètre 1. On rappelle que $\widehat{\mu}_1(t) = \exp(-|t|)$, comme un simple calcul par résidus peut le montrer. Soient quatre réels strictement positifs a, b, c, d et X et Y les variables aléatoires définies par

$$X = aU + bV \quad Y = cU + dV.$$

Calculer la fonction caractéristique de la variable aléatoire (X, Y) et en déduire que X et Y ne sont pas indépendantes.

3. Calculer la fonction caractéristique de $X + Y$ et en déduire l'égalité des lois

$$P_{X+Y} = P_X * P_Y.$$

Solution.

1. Pour tout réel t , on a, par changement de variables,

$$\widehat{\mu}_a(t) = \int_{\mathbb{R}} \exp(itx) \frac{a}{\pi(a^2 + x^2)} dx = \int_{\mathbb{R}} \exp(iat \frac{x}{a}) \frac{1}{\pi(1 + \frac{x^2}{a^2})} \frac{dx}{a}.$$

ce qui montre que

$$\widehat{\mu}_a(t) = \widehat{\mu}_1(at).$$

La fonction caractéristique φ_{aZ} de aZ est alors donnée par, pour tout réel t ,

$$\varphi_{aZ}(t) = \varphi_Z(at) = \widehat{\mu}_1(at) = \widehat{\mu}_a(t).$$

La propriété d'injectivité de la transformation de Fourier assure alors que la loi de aZ est la loi μ_a .

2. On a

$$\begin{pmatrix} X \\ Y \end{pmatrix} = A \begin{pmatrix} U \\ V \end{pmatrix},$$

où A est la matrice

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Pour tout $(\alpha, \beta) \in \mathbb{R}^2$, on a

$$A^* \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} a\alpha + c\beta \\ b\alpha + d\beta \end{pmatrix};$$

il en résulte que, pour tout $(\alpha, \beta) \in \mathbb{R}^2$:

$$\varphi_{(X,Y)}(\alpha, \beta) = \varphi_{(U,V)}(a\alpha + c\beta, b\alpha + d\beta).$$

soit, puisque les variables aléatoires U et V sont indépendantes,

$$\varphi_{(X,Y)}(\alpha, \beta) = \varphi_U(a\alpha + c\beta) \varphi_V(b\alpha + d\beta) = \exp(-[|a\alpha + c\beta| + |b\alpha + d\beta|]).$$

On a

$$\varphi_X(\alpha) = \varphi_{(X,Y)}(\alpha, 0) = \exp(-[a + b]|\alpha|).$$

et donc,

$$\varphi_{(X,Y)}(\alpha, \beta) \neq \varphi_X(\alpha) \varphi_Y(\beta),$$

ce qui démontre que **les variables aléatoires X et Y ne sont pas indépendantes.**

3. Toutefois, on a, pour tout réel α ,

$$\begin{aligned} \varphi_{X+Y}(\alpha) &= \varphi_{(X,Y)}(\alpha, \alpha) = \exp(-[(a+c+b+d)|\alpha|]) \\ &= \varphi_X(\alpha) \varphi_Y(\alpha), \end{aligned}$$

ce qui démontre, par injectivité de la transformation de Fourier, que

$$P_{X+Y} = P_X * P_Y.$$

Exercice 12.8. Fonction caractéristique et support de loi. Soit X une variable aléatoire réelle de fonction caractéristique φ_X .

1. Démontrer que s'il existe un réel $t_0 \neq 0$ tel que $|\varphi_X(t_0)| = 1$, alors il existe un réel a tel que

$$P_X\left(a + \mathbb{Z} \frac{2\pi}{t_0}\right) = 1.$$

2. S'il existe deux réels t_1 et t_2 non nuls tels que t_1/t_2 soit irrationnel et tels que $|\varphi_X(t_1)| = |\varphi_X(t_2)| = 1$, alors la variable aléatoire X est **dégénérée** (c'est-à-dire qu'elle est P-p.s. égale à une constante).

3. Démontrer que pour que la variable aléatoire X soit dégénérée, il faut et il suffit que $|\varphi_{X_0}| = 1$.

Solution.

1. Soit un réel $t_0 \neq 0$ tel que $\varphi_X(t_0) = \exp(it_0 a)$, c'est-à-dire tel que $E[\exp(it_0(X - a))] = 1$. On a alors

$$E[1 - \exp(it_0(X - a))] = 0,$$

et, en prenant la partie réelle,

$$E[1 - \cos(t_0(X - a))] = 0;$$

l'intégrande étant positive, il en résulte que P-p.s. $\cos(t_0(X - a)) = 1$, et donc que $P_X\left(a + \mathbb{Z} \frac{2\pi}{t_0}\right) = 1$.

2. Supposons qu'existent deux réels t_1 et t_2 non nuls tels que $\frac{t_1}{t_2}$ soit irrationnel et tels que $|\varphi_X(t_1)| = |\varphi_X(t_2)| = 1$. D'après la question précédente, il existe deux réels a et b tels que

$$P_X\left(a + \mathbb{Z} \frac{2\pi}{t_1}\right) = P_X\left(b + \mathbb{Z} \frac{2\pi}{t_2}\right) = 1.$$

Si la variable aléatoire X était non dégénérée, les ensembles $a + \mathbb{Z} \frac{2\pi}{t_1}$ et $b + \mathbb{Z} \frac{2\pi}{t_2}$ auraient au moins deux points communs distincts, c'est-à-dire qu'il existerait des entiers $k \neq k'$ et $l \neq l'$ tels que $a + k \frac{2\pi}{t_1} = b + l \frac{2\pi}{t_2}$ et $a + k' \frac{2\pi}{t_1} = b + l' \frac{2\pi}{t_2}$. On aurait alors

$$a - b = l \frac{2\pi}{t_2} - k \frac{2\pi}{t_1} = l' \frac{2\pi}{t_2} - k' \frac{2\pi}{t_1},$$

et donc

$$(l - l') \frac{2\pi}{t_2} = (k - k') \frac{2\pi}{t_1},$$

ce qui est impossible si t_1/t_2 est irrationnel. **La variable aléatoire X est donc dégénérée.**

3. Si la variable aléatoire X est dégénérée et vaut P-p.s. a , on a, pour tout réel t , $\varphi_X(t) = \exp(it a)$ et donc $|\varphi_X| = 1$. La réciproque résulte de la question précédente.

Exercice 12.9. Fonction caractéristique et espérance conditionnelle. Jeu de pile ou face (variante). Soit $(U_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes de même loi $(\delta_{-1} + \delta_1)/2$. On pose $U_{-1} = 0$. On définit la suite de variables

aléatoires $(Y_n)_{n \in \mathbb{N}}$ par

$$Y_n = \sum_{j=0}^n U_{j-1} U_j.$$

On note, pour tout $n \in \mathbb{N}$, \mathcal{F}_n la tribu engendrée par les variables aléatoires U_j , $0 \leq j \leq n$. La variable aléatoire Y_n peut représenter le gain algébrique, après le n -ième jet, d'un joueur jouant à pile ou face avec une pièce équilibrée, avec la **règle de gain suivante** : il gagne une unité après le n -ième lancer si le résultat est le même qu'au lancer précédent; dans le cas contraire, il perd une unité (c'est aussi, avec la règle habituelle, le gain d'un joueur qui mise toujours sur la face qui vient de sortir).

1. Calculer, pour tout réel t , l'espérance conditionnelle

$$E^{\mathcal{F}_{n-1}}[\exp(it U_{n-1} U_n)].$$

2. Calculer, pour tout réel t et tout entier l tel que $1 \leq l \leq n$, l'espérance conditionnelle $E^{\mathcal{F}_{n-l}}[\exp(it Y_n)]$ (on pourra procéder par récurrence sur l).

3. En déduire la fonction caractéristique φ_{Y_n} de Y_n .

4. Déterminer la loi de Y_n .

5. Étudier, pour tout réel t , la suite de terme général $\varphi_{\frac{Y_n}{n}}(t)$.

Solution.

1. Les variables aléatoires $(U_0, U_1, \dots, U_{n-1})$ et U_n étant indépendantes, on a

$$E^{\mathcal{F}_{n-1}}[\exp(it U_{n-1} U_n)] = f(U_0, U_1, \dots, U_{n-1}),$$

où l'application f est définie sur \mathbb{R}^n par

$$\forall (u_0, u_1, \dots, u_{n-1}) \in \mathbb{R}^n \quad f(u_0, u_1, \dots, u_{n-1}) = E[\exp(it u_{n-1} U_n)],$$

soit

$$f(u_0, u_1, \dots, u_{n-1}) = \frac{1}{2} [\exp(it u_{n-1}) + \exp(-it u_{n-1})] = \cos(t u_{n-1}).$$

Il en résulte que :

$$E^{\mathcal{F}_{n-1}}[\exp(it U_{n-1} U_n)] = \cos(t U_{n-1}).$$

2. Puisque Y_{n-1} est \mathcal{F}_{n-1} -mesurable, on a

$$\begin{aligned} E^{\mathcal{F}_{n-1}}[\exp(it Y_n)] &= \exp(it Y_{n-1}) E^{\mathcal{F}_{n-1}}[\exp(it U_{n-1} U_n)] \\ &= \exp(it Y_{n-1}) \cos(t U_{n-1}). \end{aligned}$$

Un calcul à l'ordre deux et trois permet de poser l'hypothèse de récurrence à l'ordre l :

$$(HR_l) \quad E^{\mathcal{F}_{n-l}}[\exp(it Y_n)] = \exp(it Y_{n-l}) \cos(t U_{n-l}) \cos^{l-1}(t).$$

Vérifions que cette formule est vraie à l'ordre $l+1$. En vertu de l'inclusion $\mathcal{F}_{n-l-1} \subset \mathcal{F}_{n-l}$, on a :

$$E^{\mathcal{F}_{n-(l+1)}}[\exp(itY_n)] = E^{\mathcal{F}_{n-l-1}}[E^{\mathcal{F}_{n-l}}[\exp(itY_n)]],$$

soit, d'après l'hypothèse de récurrence,

$$E^{\mathcal{F}_{n-(l+1)}}[\exp(itY_n)] = E^{\mathcal{F}_{n-l-1}}[\exp(itY_{n-l}) \cos(tU_{n-l}) \cos^{l-1}(t)].$$

soit encore, puisque $Y_{n-(l+1)}$ est \mathcal{F}_{n-l-1} -mesurable,

$$E^{\mathcal{F}_{n-(l+1)}}[\exp(itY_n)] = \exp(itY_{n-(l+1)}) E^{\mathcal{F}_{n-l-1}}[\exp(itU_{n-(l+1)}U_{n-l}) \cos(tU_{n-l}) \cos^{l-1}(t)].$$

Un calcul analogue à celui de la question précédente conduit à l'égalité

$$E^{\mathcal{F}_{n-(l+1)}}[\exp(itY_n)] = \exp(itY_{n-(l+1)}) \times \left[\frac{1}{2} \exp(itU_{n-(l+1)}) \cos(t) + \frac{1}{2} \exp(-itU_{n-(l+1)}) \cos(-t) \right] \cos^{l-1}(t),$$

soit

$$E^{\mathcal{F}_{n-(l+1)}}[\exp(itY_n)] = \exp(itY_{n-(l+1)}) \cos(tU_{n-(l+1)}) \cos^l(t);$$

ceci démontre la formule pour tout l tel que $1 \leq l \leq n$.

3. La fonction caractéristique de Y_n est alors donnée en tout réel t par

$$\varphi_{Y_n}(t) = E[E^{\mathcal{F}_0}[\exp(itY_n)]] = E[\exp(itY_0) \cos(tU_0)] \cos^{n-1}(t),$$

soit, puisque $Y_0 = 0$:

$$\varphi_{Y_n}(t) = \frac{1}{2} [\cos(t) + \cos(-t)] \cos^{n-1}(t);$$

on a donc :

$$\boxed{\varphi_{Y_n}(t) = \cos^n(t).}$$

4. On remarque que la transformée de Fourier en t de la probabilité $(\delta_{-1} + \delta_1)/2$ est $\cos t$. Il résulte de la propriété d'injectivité de la transformation de Fourier que Y_n a même loi que la somme de n variables aléatoires indépendantes de loi $(\delta_{-1} + \delta_1)/2$, c'est-à-dire même loi que la variable aléatoire $\sum_{j=1}^n U_j$. Ainsi, pour tout entier k tel que $-n \leq k \leq n$, on a, en tenant compte de l'indépendance des U_j :

$$P(Y_n = k) = P \left[\bigoplus_{\substack{J \subset \{1, \dots, n\} \\ |J| = \frac{k+n}{2}}} \left\{ \bigcap_{j \in J} (U_j = 1) \cap \bigcap_{j \in J^c} (U_j = -1) \right\} \right],$$

soit

$$P(Y_n = k) = \sum_{\substack{J \subset \{1, \dots, n\} \\ |J| = \frac{k+n}{2}}} \frac{1}{2^n}.$$

On a donc, pour tout k tel que $-n \leq k \leq n$:

$$\mathbb{P}(Y_n = k) = \begin{cases} \binom{n}{\frac{k+n}{2}} & \text{si } k + n \text{ pair,} \\ 0 & \text{si } k + n \text{ impair.} \end{cases}$$

5. On a

$$\varphi_{\frac{Y_n}{n}}(t) = \varphi_{Y_n}\left(\frac{t}{n}\right) = \cos^n\left(\frac{t}{n}\right)$$

et donc

$$\ln \varphi_{\frac{Y_n}{n}}(t) = n \ln \left[1 - \frac{t^2}{2n^2} + o\left(\frac{1}{n^2}\right) \right].$$

Il en résulte que

$$\varphi_{\frac{Y_n}{n}}(t) = \exp \left[-\frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right],$$

et donc que

$$\lim_n \varphi_{\frac{Y_n}{n}}(t) = 1.$$

Remarque. On vient de démontrer que la suite de terme général Y_n/n converge en loi vers 0 (on étudiera la notion de convergence en loi au chapitre 14).

Exercice 12.10. Formule d'inversion de la transformée de Fourier d'une probabilité.

Soit μ une probabilité sur $(\mathbb{P}, \mathcal{B}_{\mathbb{R}})$ de transformée de Fourier φ . Il s'agit, pour commencer, d'établir la formule d'inversion suivante : pour tous réels a et b tels que $a < b$, on a

$$\lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^{+T} \frac{\exp(-it a) - \exp(-it b)}{it} \varphi(t) dt = \frac{1}{2} \mu(\{a, b\}) + \mu([a, b]). \quad (12.24)$$

Les résultats des questions 2 et 3 sont aussi intéressants et utiles.

1. Pour tout réel positif T , on définit la fonction complexe de la variable réelle x par

$$\forall x \in \mathbb{P} \quad \mathbf{I}_T(x) = \frac{1}{2\pi} \int_{-T}^{+T} \frac{\exp(it(x-a)) - \exp(it(x-b))}{it} dt.$$

Démontrer que, pour tout réel x ,

$$\lim_{T \rightarrow +\infty} \mathbf{I}_T(x) = \frac{1}{2} \mathbf{1}_{(a,b)}(x) + \mathbf{1}_{[a,b]}(x).$$

2. En déduire la formule d'inversion (12.24).

3. Suivre la même démarche pour démontrer que, pour tout réel b , on a

$$\mu(\{b\}) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} \exp(-itb) \varphi(t) dt.$$

4. Démontrer l'égalité dans $\overline{\mathbb{R}}^+$

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} |\varphi(t)|^2 dt = \sum_{x \in \mathbb{P}} \mu(\{x\})^2$$

Pour cela on introduira deux variables aléatoires indépendantes X et Y de loi μ et on appliquera les résultats précédents à la fonction caractéristique de la variable aléatoire $X - Y$.

Solution.

1. Pour tout $x \notin \{a, b\}$ on a, par changement de variables,

$$I_T(x) = \frac{1}{2\pi} \int_{-T(x-a)}^{+T(x-a)} \frac{\sin(u)}{u} du - \frac{1}{2\pi} \int_{-T(x-b)}^{+T(x-b)} \frac{\sin(u)}{u} du.$$

Il résulte de l'égalité

$$\lim_{\alpha \rightarrow +\infty} \int_0^\alpha \frac{\sin(u)}{u} du = \frac{\pi}{2}$$

que

- si $x < a$ ou $x > b$, on a $\lim_{T \rightarrow +\infty} I_T(x) = 0$.
- si $a < x < b$, on a $\lim_{T \rightarrow +\infty} I_T(x) = \frac{1}{2\pi}(\pi + \pi) = 1$.
- si $x = a$, on a

$$\begin{aligned} I_T(a) &= \frac{1}{2\pi} \int_{-T}^{+T} \frac{1 - \exp(it(a-b))}{it} dt \\ &= \frac{1}{2\pi} \left[\int_{-T}^{+T} \frac{1 - \cos(t(b-a))}{it} dt + \int_{-T}^{+T} \frac{\sin(t(b-a))}{t} dt \right], \end{aligned}$$

soit, par changement de variables et propriété de parité,

$$I_T(a) = \frac{1}{2\pi} \int_{-T(b-a)}^{+T(b-a)} \frac{\sin(u)}{u} du.$$

Il en résulte que : $\lim_{T \rightarrow +\infty} I_T(a) = \frac{1}{2}$. On a de même $\lim_{T \rightarrow +\infty} I_T(b) = \frac{1}{2}$.

2. En vertu de l'inégalité des accroissements finis, la fonction mesurable $(t, x) \mapsto \frac{\exp(-ita) - \exp(-itb)}{it} \exp(itx)$ est bornée sur $[-T, T] \times \mathbb{R}$. Il résulte alors du théorème de Fubini que

$$\frac{1}{2\pi} \int_{-T}^{+T} \frac{\exp(-ita) - \exp(-itb)}{it} \varphi(t) dt = \frac{1}{2\pi} \int_{\mathbb{R}} I_T(x) d\mu(x).$$

Par ailleurs, la fonction $\alpha \mapsto \int_0^\alpha \frac{\sin(u)}{u} du$ est uniformément continue sur \mathbb{P} et tend vers $\pi/2$ quand α tend vers $+\infty$; il en résulte qu'il existe un réel M tel que, pour tout (x, T) , on ait $|I_T(x)| \leq M$. Le théorème de convergence dominée assure que

$$\lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^{+T} \frac{\exp(-ita) - \exp(-itb)}{it} \varphi(t) dt = \frac{1}{2\pi} \int_{\mathbb{P}} \lim_{T \rightarrow +\infty} I_T(x) d\mu(x),$$

ce qui, compte tenu de la question précédente, démontre la formule d'inversion (12.24).

3. Il résulte du théorème de Fubini que

$$\int_{-T}^{+T} \exp(-itb) \varphi(t) dt = \int_{\mathbb{R}} \left[\int_{-T}^{+T} \exp(it(x-b)) dt \right] d\mu(x).$$

Mais, si $x \neq b$, on a

$$\int_{-T}^{+T} \exp(it(x-b)) dt = 2 \frac{\sin(T(x-b))}{x-b},$$

ce qui donne

$$\frac{1}{2T} \int_{-T}^{+T} \exp(-itb) \varphi(t) dt = \mu(\{b\}) + \int_{\mathbb{R} \setminus \{b\}} \left[\frac{\sin(T(x-b))}{T(x-b)} \right] d\mu(x).$$

Puisque la fonction $u \mapsto \frac{\sin u}{u}$ (prolongée par 1 en 0) est bornée et tend vers 0 quand u tend vers $\pm\infty$, le théorème de convergence dominée assure que l'intégrale du membre de droite tend vers 0 quand T tend vers $+\infty$, ce qui démontre que

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} \exp(-itb) \varphi(t) dt = \mu(\{b\}).$$

4. Soient deux variables aléatoires indépendantes X et Y de loi μ . La fonction caractéristique φ_{X-Y} de la variable aléatoire $X - Y$ est donnée par

$$\forall t \in \mathbb{R} \quad \varphi_{X-Y}(t) = \varphi_X(t) \varphi_Y(-t) = |\varphi(t)|^2.$$

Le résultat de la question précédente assure alors que

$$P(X - Y = 0) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} |\varphi(t)|^2 dt.$$

Mais, les variables aléatoires X et Y étant indépendantes de loi μ , il résulte du théorème de Fubini que

$$P(X - Y = 0) = \mu \otimes \mu(\{x = y\}) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbf{1}_{\{x=y\}} d\mu(y) \right] d\mu(x) = \int_{\mathbb{R}} \mu(\{x\}) d\mu(x).$$

L'ensemble $S = \{x \mid \mu(\{x\}) \neq 0\}$ étant dénombrable, on a

$$\int_{\mathbb{R}} \mu(\{x\}) d\mu(x) = \int_S \mu(\{x\}) d\mu(x) = \sum_{x \in S} \mu(\{x\})^2 = \sum_{x \in \mathbb{R}} \mu(\{x\})^2.$$

Il en résulte que

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} |\varphi(t)|^2 dt = \sum_{x \in \mathbb{R}} \mu(\{x\})^2.$$

Exercice 12.11. Fonctions caractéristiques de variables aléatoires vectorielles, calcul de moyennes conditionnelles et injectivité de la transformation de Fourier. Soient deux variables aléatoires réelles X et Y telles que Y soit de loi de Bernoulli $\mathcal{B}(1, 1 - \rho)$ et qu'une loi conditionnelle de X sachant Y soit donnée par

$$P_X^{Y=0} = \delta_0 \quad \text{et} \quad P_X^{Y=1} = \exp(\lambda),$$

où δ_0 est la mesure de Dirac en 0 et $\exp(\lambda)$ la loi exponentielle de paramètre $\lambda > 0$.

1. Calculer la fonction caractéristique φ_X de X . On note μ la loi de X (on ne cherche pas à la calculer).

2. En déduire les moyenne et variance de X .

On considère une famille de variables aléatoires indépendantes $X_0, \varepsilon_n, n \in \mathbb{N}^*$. On suppose que les ε_n sont de même loi μ et que X_0 est de loi $\exp(\lambda)$. On définit par récurrence les variables aléatoires X_n par

$$\forall n \in \mathbb{N}^* \quad X_n = \rho X_{n-1} + \varepsilon_n.$$

3. Justifier l'indépendance, pour tout $n \in \mathbb{N}^*$, des variables aléatoires ε_n et $(X_0, X_1, \dots, X_{n-1})$ et démontrer par récurrence que les variables aléatoires X_n ont même fonction caractéristique; identifier leur loi.

Solution.

1. Le théorème de transfert donne

$$\forall t \in \mathbb{R} \quad \varphi_X(t) = \int_{\mathbb{R}^2} \exp(itx) dP_{(X,Y)}(x, y).$$

La fonction $(x, y) \mapsto \exp(itx)$ étant bornée, on peut appliquer le théorème de Fubini généralisé (chap. 11, théorème 11.3) et obtenir successivement, puisque la loi de Y est $P_Y = \rho\delta_0 + (1 - \rho)\delta_1$:

$$\begin{aligned} \varphi_X(t) &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \exp(itx) dP_X^{Y=y}(x) \right] dP_Y(y) \\ &= \rho \int_{\mathbb{R}} \exp(itx) dP_X^{Y=0}(\lambda) + (1 - \rho) \int_{\mathbb{R}} \exp(itx) dP_X^{Y=1}(x) \\ &= \rho \exp(it \cdot 0) + (1 - \rho) \int_{\mathbb{R}} \exp(itx) \mathbf{1}_{\mathbb{R}^+}(x) \lambda \exp(-\lambda x) dx. \end{aligned}$$

On obtient :

$$\forall t \in \mathbb{R} \quad \varphi_X(t) = \rho + (1 - \rho) \frac{\lambda}{\lambda - it}.$$

2. La fonction caractéristique de X est deux fois dérivable; la variable aléatoire X admet donc un moment d'ordre deux et on a

$$E(X) = -i \varphi_X'(0) \quad \text{et} \quad E(X^2) = -\varphi_X''(0),$$

ce qui donne, puisque

$$\varphi_X'(t) = (1 - \rho) \frac{\lambda i}{(\lambda - it)^2} \quad \text{et} \quad \varphi_X''(t) = -2(1 - \rho) \frac{\lambda i}{(\lambda - it)^3},$$

$$E(X) = \frac{1-\rho}{\lambda} \quad E(X^2) = \frac{2(1-\rho)}{\lambda^2} \quad \sigma_X^2 = \frac{1-\rho^2}{\lambda^2}.$$

3. La variable aléatoire $(X_0, X_1, \dots, X_{n-1})$ est une fonction linéaire de $(X_0, \varepsilon_1, \dots, \varepsilon_{n-1})$; elle est donc indépendante de ε_n , d'après les hypothèses d'indépendance faites.

La fonction caractéristique de X_0 vaut, en tout réel t , $\varphi_{X_0}(t) = \frac{\lambda}{\lambda - it}$. Supposons que $\varphi_{X_{n-1}} = \varphi_{X_0}$. Les variables aléatoires X_{n-1} et ε_n étant indépendantes, on a, en appliquant l'hypothèse de récurrence, pour tout réel t ,

$$\varphi_{X_n}(t) = \varphi_{\rho X_{n-1}}(t) \varphi_{\varepsilon_n}(t) = \varphi_{X_{n-1}}(\rho t) \varphi_{\varepsilon_n}(t),$$

et donc

$$\varphi_{X_n}(t) = \frac{\lambda}{\lambda - i\rho t} \left[\rho + (1-\rho) \frac{\lambda}{\lambda - it} \right],$$

soit encore

$$\varphi_{X_n}(t) = \frac{\lambda}{\lambda - it}.$$

Il en résulte que, pour tout n , $\varphi_{X_n} = \varphi_{X_0}$, et donc que **les variables aléatoires X_n ont même loi exponentielle $\exp(\lambda)$** .

4. Puisque X_{n-1} et ε_n sont indépendantes, la moyenne conditionnelle $m_{X_n}^{X_{n-1}}$ de X_n sachant X_{n-1} est donnée par, pour tout réel x_{n-1} ,

$$m_{X_n}^{X_{n-1}=x_{n-1}} = \rho x_{n-1} + E(\varepsilon_n),$$

soit

$$m_{X_n}^{X_{n-1}=x_{n-1}} = \rho x_{n-1} + \frac{1-\rho}{\lambda}.$$

5. La fonction caractéristique $\varphi_{(X_{n-1}, X_n)}$ de la variable aléatoire (X_{n-1}, X_n) vaut, en tout $(u, v) \in \mathbb{R}^2$,

$$\varphi_{(X_{n-1}, X_n)}(u, v) = E[\exp(i(uX_{n-1} + vX_n))] = E[\exp(i(u + \rho v)X_{n-1}) \exp(iv\varepsilon_n)],$$

soit, par indépendance des variables aléatoires X_{n-1} et ε_n :

$$\varphi_{(X_{n-1}, X_n)}(u, v) = \varphi_{X_{n-1}}(u + \rho v) \varphi_{\varepsilon_n}(v).$$

Il vient

$$\forall (u, v) \in \mathbb{R}^2 \quad \varphi_{(X_{n-1}, X_n)}(u, v) = \frac{\lambda}{\lambda - i(u + \rho v)} \left[\rho + (1-\rho) \frac{\lambda}{\lambda - iv} \right].$$

6. Le couple (X_{n-1}, X_n) admettant une moyenne, sa fonction caractéristique est différentiable et on a

$$\frac{\partial}{\partial u} \varphi_{(X_{n-1}, X_n)}(u, v) = i E[X_{n-1} \exp(i(uX_{n-1} + vX_n))].$$

Il en résulte que

$$\frac{\partial}{\partial u} \varphi_{(X_{n-1}, X_n)}(0, v) = i E [X_{n-1} \exp(ivX_n)] ,$$

soit, d'après le théorème de Fubini généralisé,

$$\forall v \in \mathbb{R} \quad \frac{\partial}{\partial u} \varphi_{(X_{n-1}, X_n)}(0, v) = i \int_{\mathbb{R}} \exp(ivx) m_{X_{n-1}}^{X_n=x} dP_{X_n}(x) ,$$

ou encore, puisque X_n est de loi exponentielle $\exp(\lambda)$,

$$\forall v \in \mathbb{R} \quad \frac{\partial}{\partial u} \varphi_{(X_{n-1}, X_n)}(0, v) = i \int_{\mathbb{R}} \exp(ivx) f(x) dx = i \widehat{f}(v) ,$$

où la fonction f , intégrable positive, (de transformée de Fourier \widehat{f}) est définie par, pour tout réel x ,

$$f(x) = \mathbf{1}_{\mathbb{R}^+}(x) \lambda \exp(-\lambda x) m_{X_{n-1}}^{X_n=x} .$$

Par ailleurs, le calcul direct de la dérivée partielle $\varphi_{(X_{n-1}, X_n)}$ donne

$$\forall v \in \mathbb{R} \quad \frac{\partial}{\partial u} \varphi_{(X_{n-1}, X_n)}(0, v) = \frac{\lambda i}{(\lambda - i\rho v)(\lambda - iv)} .$$

Il en résulte que la transformée de Fourier \widehat{f} de f vaut en tout réel v :

$$\widehat{f}(v) = \frac{\lambda}{(\lambda - i\rho v)(\lambda - iv)} ,$$

fonction Lebesgue-intégrable. La formule d'inversion assure alors que

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\lambda}{(\lambda - i\rho v)(\lambda - iv)} \exp(-ixv) dv .$$

Reste à calculer cette intégrale ; on a

$$\frac{\lambda}{(\lambda - i\rho v)(\lambda - iv)} = -\frac{1}{1-\rho} \frac{1}{\frac{\lambda}{\rho} - iv} + \frac{1}{1-\rho} \frac{1}{\lambda - iv}$$

et, d'après le théorème d'injectivité de la transformation de Fourier (appliqué à la transformée de Fourier de la loi $\exp(\lambda)$),

$$\frac{1}{2\pi} \int_{\mathbb{R}} \frac{\lambda}{\lambda - iv} \exp(-ixv) dv = \mathbf{1}_{\mathbb{R}^+}(x) \exp(-\lambda x) .$$

Il en résulte que :

$$f(x) = -\frac{1}{1-\rho} \mathbf{1}_{\mathbb{R}^+}\left(\frac{x}{\rho}\right) \exp\left(-\frac{\lambda}{\rho}x\right) + \frac{1}{1-\rho} \mathbf{1}_{\mathbb{R}^+}(x) \exp(-\lambda x) .$$

La définition de f et un calcul facile montrent que

$$\forall x \in \mathbb{R} \quad m_{X_{n-1}}^{X_n=x} = \mathbf{1}_{\mathbb{R}^+}(x) \frac{1}{\lambda(1-\rho)} \left[1 - \exp\left(-\frac{\lambda(1-\rho)}{\rho}x\right) \right] .$$

Chapitre 13

Variables aléatoires gaussiennes

Ce chapitre est consacré à l'étude des variables aléatoires gaussiennes à valeurs dans un espace vectoriel de dimension finie E ; cette étude est faite de manière intrinsèque, c'est-à-dire indépendamment du choix d'une base de E . Bien que cela ne joue aucun rôle dans la définition des variables aléatoires gaussiennes à valeurs dans E , il est utile de supposer E muni d'un produit scalaire : cela évite l'intervention explicite du dual, et cela permet notamment de considérer la variance comme une forme quadratique sur E (cf. chap. 8). La structure euclidienne peut d'ailleurs apparaître naturellement : par exemple, dans l'étude de problèmes d'estimation, ou de tests statistiques, qui conduisent à l'étude de variables aléatoires à valeurs matricielles.

Le lecteur pourra, s'il le désire, supposer $E = \mathbb{R}^d$. On se ramène aisément à ce cas en choisissant une base orthonormée de E .

En fin de chapitre, on étudie le problème de régression dans le contexte gaussien. En particulier, on résout, dans le cadre du modèle linéaire gaussien, le problème d'estimation des paramètres et les problèmes de test et de détermination d'intervalles de confiance relatifs à ces paramètres.

Dans ce chapitre, sauf mention du contraire, toutes les variables aléatoires seront définies sur un même espace probabilisé (Ω, \mathcal{A}, P) . On désignera par E un espace euclidien réel fixé, de dimension d , où le produit scalaire est noté $\langle \cdot, \cdot \rangle$. L'espace E est identifié à son dual; ainsi une forme linéaire u sur E sera notée aussi $\langle \cdot, u \rangle$. On note \mathcal{E} la tribu borélienne de E , c'est à dire la tribu engendrée par la famille des ouverts de E . L'espace E sera toujours supposé muni de la tribu \mathcal{E} .

Rappels (cf. chap. 8). La loi gaussienne sur \mathbb{R} , appelée encore **loi de Laplace-Gauss** ou **loi normale, de paramètres** $m \in \mathbb{R}$ et $\sigma^2 > 0$, est la probabilité de densité f par rapport à la mesure de Lebesgue, où f est définie par

$$\forall x \in \mathbb{R} \quad f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right);$$

cette loi est notée indifféremment $\mathcal{N}(m, \sigma^2)$ ou $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$. Sa transformée de Fourier est donnée par la relation

$$\forall t \in \mathbb{R} \quad \widehat{\mathcal{N}(m, \sigma^2)}(t) = \exp(itm) \exp\left(-\frac{t^2 \sigma^2}{2}\right). \quad (13.1)$$

Une **variable aléatoire réelle** X est dite **gaussienne** (ou **normale**) si sa loi est gaussienne. Si la variable aléatoire X est de loi $\mathcal{N}(m, \sigma^2)$, sa fonction caractéristique est définie par

$$\forall t \in \mathbb{R} \quad \varphi_X(t) = \exp(itm) \exp\left(-\frac{t^2 \sigma^2}{2}\right); \quad (13.2)$$

sa moyenne est m , sa variance σ^2 . Elle admet des moments de tout ordre, que l'on peut par exemple obtenir par des développements limités de φ_X . En particulier (cf. chap. 12. ex. 4), si X est de loi $\mathcal{N}(0, 1)$, on a, pour tout $n \in \mathbb{N}$,

$$EX^{2n+1} = 0 \text{ et si } n \geq 1, EX^{2n} = \frac{(2n)!}{2^n n!} = 1 \cdot 3 \cdots (2n-3)(2n-1).$$

La généralisation à l'espace euclidien de la notion de **loi (ou mesure) gaussienne** conduit à considérer une **mesure de Dirac** comme une **loi gaussienne dégénérée** : des variables aléatoires gaussiennes portées par un sous-espace affine s'introduisent en effet naturellement dans cette étude ; une variable aléatoire P-p.s. **constante** (donc de variance nulle), est alors encore **gaussienne**.

13.1. Définition et propriétés

Définition 13.1. On appelle **loi gaussienne** (ou **normale**) sur E une probabilité μ sur (E, \mathcal{E}) telle que la mesure image de μ par toute forme linéaire sur E soit une loi gaussienne sur \mathbb{R} .

Une variable aléatoire X à valeurs dans E est dite **gaussienne** (ou **normale**) si sa loi P_X est gaussienne sur E .

Remarque. Une variable aléatoire X à valeurs dans E est gaussienne si et seulement si pour tout $u \in E$, la variable aléatoire réelle $\langle X, u \rangle$ est gaussienne (cela résulte de l'égalité des probabilités $\langle \cdot, u \rangle(P_X)$ et $P_{\langle X, u \rangle}$).

Nous allons étudier quelques conséquences immédiates et importantes de ces définitions. Elles nous permettront de démontrer l'**existence** de lois et variables aléatoires gaussiennes de moyenne et de variance données.

Notation. On note $\mathcal{L}^+(E)$ l'ensemble des opérateurs (ou endomorphismes) **auto-adjoints et positifs** : un endomorphisme Λ de E appartient à $\mathcal{L}^+(E)$ si et seulement s'il vérifie $\Lambda = \Lambda^*$ et si $\langle \Lambda x, x \rangle \geq 0$, pour tout $x \in E$.

Proposition 13.2. Soit X une variable aléatoire à valeurs dans E .

(a) Si X est gaussienne, la variable aléatoire $\|X\|$ est dans \mathcal{L}^2 , et X admet donc une moyenne m (élément de E) et une variances σ_X^2 , forme quadratique sur E . On note Λ_X l'opérateur d'auto-covariance de X , unique opérateur auto-adjoint positif tel que

$$\forall x \in E \quad \langle \Lambda_X x, x \rangle = \sigma_X^2(x).$$

On rappelle que l'on a :

$$\boxed{\forall x \in E \quad E\langle X, x \rangle = \langle m, x \rangle \quad \text{et} \quad \sigma_X^2(x) = \sigma_{\langle X, x \rangle}^2 = \langle \Lambda_X x, x \rangle.} \quad (13.3)$$

(b) La variable aléatoire X est gaussienne si et seulement si sa fonction caractéristique φ_X est donnée par

$$\boxed{\forall t \in E \quad \varphi_X(t) = \exp(i \langle m, t \rangle) \exp\left[-\frac{1}{2} \langle Ct, t \rangle\right],} \quad (13.4)$$

où $m \in E$ et $C \in \mathcal{L}^+(E)$. Dans ce cas, on a $m = EX$ et $C = \Lambda_X$.

En conséquence, EX et Λ_X caractérisent entièrement la loi de la variable aléatoire gaussienne X .

(c) Une mesure μ sur (E, \mathcal{E}) est gaussienne si et seulement si sa transformée de Fourier $\hat{\mu}$ est donnée par la relation

$$\boxed{\forall t \in E \quad \hat{\mu}(t) = \exp(i \langle m, t \rangle) \exp\left[-\frac{1}{2} \langle Ct, t \rangle\right].} \quad (13.5)$$

où $m \in E$ et $C \in \mathcal{L}^+(E)$; m et C sont alors **uniquement** déterminés; c'est en fait une probabilité sur E . Cette loi de probabilité est notée $\mathcal{N}_E(m, C)$ et appelée¹ **loi gaussienne** (ou **mesure gaussienne**) de paramètres m et C .

(d) Soit F un autre espace euclidien; si X est gaussienne de loi $\mathcal{N}_E(m, \Lambda_X)$, pour tout $A \in \mathcal{L}(E, F)$ et tout $b \in F$, la variable aléatoire $AX + b$, à valeurs dans F , est gaussienne de loi $\mathcal{N}_F(Am + b, A\Lambda_X A^*)$.

Démonstration. (a) Pour tout $x \in E$, la variable aléatoire réelle $\langle X, x \rangle$ est gaussienne, donc dans \mathcal{L}^2 , ce qui est équivalent à dire que $\|X\|$ est dans \mathcal{L}^2 (cf. chap. 8, prop. 8.31); le reste découle immédiatement des définitions.

(b) Pour tout $t \in E$ et tout $\alpha \in \mathbb{R}$, on a

$$\varphi_X(\alpha t) = E[\exp(i \langle X, \alpha t \rangle)] = \varphi_{\langle X, t \rangle}(\alpha); \quad (13.6)$$

si X est gaussienne, d'après (13.3), $\langle X, t \rangle$ est de loi gaussienne

$$\mathcal{N}(\langle m, t \rangle, \langle \Lambda_X t, t \rangle),$$

1. Pour l'instant, rien ne dit que cette mesure existe; son existence sera démontrée au théorème 13.4.

ce qui, en prenant $\alpha = 1$ et en utilisant (13.2) donne le résultat.

Inversement, si φ_X est donnée par (13.4), il résulte de (13.6) que, pour tout $t \in E$ et tout $\alpha \in \mathbb{R}$, on a

$$\varphi_{\langle X, t \rangle}(\alpha) = \exp(i\alpha \langle m, t \rangle) \exp\left[-\frac{\alpha^2}{2} \langle Ct, t \rangle\right],$$

ce qui démontre que $\langle X, t \rangle$ est gaussienne, et donc aussi X . Plus précisément, la loi de $\langle X, t \rangle$ est la loi $\mathcal{N}(\langle m, t \rangle, \langle Ct, t \rangle)$; on a donc

$$\forall t \in E \quad E \langle X, t \rangle = \langle m, t \rangle \quad \text{et} \quad \sigma_{\langle X, t \rangle}^2 = \langle Ct, t \rangle,$$

ce qui démontre, compte tenu de (13.3), que :

$$EX = m \quad \text{et} \quad \forall t \in E \quad \sigma_{\langle X, t \rangle}^2 = \langle Ct, t \rangle = \langle \Lambda_X t, t \rangle.$$

Autrement dit, on a $EX = m$ et $C = \Lambda_X$.

(c) L'application identique I sur E , vue comme une variable aléatoire définie sur l'espace probabilisé (E, \mathcal{E}, μ) à valeurs dans (E, \mathcal{E}) est de loi μ et de fonction caractéristique $\hat{\mu}$. Ainsi I est une variable aléatoire gaussienne si et seulement si la mesure μ est gaussienne; il suffit alors d'appliquer la caractérisation précédente.

(d) En utilisant la définition du transposé, on a, pour tout $t \in E$,

$$\varphi_{AX+b}(t) = \exp(i \langle b, t \rangle) \varphi_X(A^*t) = \exp(i \langle Am + b, t \rangle) \exp\left[-\frac{1}{2} \langle A \Lambda_X A^* t, t \rangle\right],$$

ce qui, ainsi qu'on vient de le voir, donne le résultat. \square

Remarque. On rappelle (cf. chap. 8) que la matrice C_X , qui représente l'opérateur Λ_X dans une base orthonormée $(e_j)_{1 \leq j \leq d}$, est la matrice de covariance de X dans cette base et que l'on a

$$(C_X)_{ij} = \text{cov}(\langle X, e_i \rangle, \langle X, e_j \rangle).$$

En particulier, si $E = \mathbb{R}^d$, la base usuellement choisie est la base canonique; dans ce cas, l'opérateur Λ_X est représenté par la matrice C_X des covariances $\text{cov}(X_i, X_j)$ des marginales X_j . La loi gaussienne de X est alors encore notée $\mathcal{N}_{\mathbb{R}^d}(m, C_X)$.

13.2. Existence des mesures gaussiennes. Condition d'absolue continuité

Lemme 13.3. *La mesure produit $[\mathcal{N}_{\mathbb{R}}(0, 1)]^{\otimes d}$ sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ est la mesure gaussienne $\mathcal{N}_{\mathbb{R}^d}(0, I_{\mathbb{R}^d})$, où $I_{\mathbb{R}^d}$ est la matrice identité de \mathbb{R}^d .*

Démonstration. La transformée de Fourier d'une mesure produit étant le produit direct des transformées de Fourier des mesures facteurs, on a, pour tout $t \in \mathbb{R}^d$,

$$\begin{aligned} \widehat{[\mathcal{N}_{\mathbb{R}}(0, 1)]^{\otimes d}}(t) &= \prod_{j=1}^d [\mathcal{N}_{\mathbb{R}}(0, 1)](t_j) \\ &= \prod_{j=1}^d \exp\left(-\frac{t_j^2}{2}\right) = \exp\left(-\frac{\|t\|^2}{2}\right) = \widehat{\mathcal{N}_{\mathbb{R}^d}(0, I_{\mathbb{R}^d})}(t), \end{aligned}$$

ce qui démontre le résultat par injectivité de la transformation de Fourier. \square

Remarque. Ceci assure l'existence de la mesure gaussienne $\mathcal{N}_{\mathbb{R}^d}(0, I_{\mathbb{R}^d})$. Par ailleurs, la mesure produit $[\mathcal{N}_{\mathbb{R}}(0, 1)]^{\otimes d}$ admet une densité f , produit direct des densités des mesures marginales; elle est définie par

$$\forall x \in \mathbb{R}^d \quad f(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|x\|^2}{2}\right).$$

Ainsi, la mesure gaussienne $\mathcal{N}_{\mathbb{R}^d}(0, I_{\mathbb{R}^d})$ admet la densité f .

De ce dernier lemme, on déduit le théorème 13.4 qui assure l'existence d'une mesure gaussienne de moyenne m et d'opérateur d'autocovariance Λ donnés. Ce théorème a un contenu purement algébrique: il s'agit de montrer l'existence d'un opérateur B tel que $BB^* = \Lambda$. Nous donnons deux démonstrations de ce fait, la première s'appuyant sur le théorème spectral relatif aux opérateurs auto-adjoints, la deuxième n'utilisant que la décomposition en carrés des formes quadratiques.

Théorème 13.4 (Théorème d'existence). *Pour tout vecteur $m \in E$ et tout opérateur auto-adjoint et positif² Λ , il existe une unique mesure gaussienne $\mathcal{N}_E(m, \Lambda)$.*

Démonstration. Il suffit d'exhiber une variable aléatoire X à valeurs dans E , gaussienne, de moyenne m et d'opérateur de covariance Λ .

En vertu de la proposition 13.2, toute variable aléatoire de la forme

$$X = m + BX_0,$$

où X_0 est une variable aléatoire à valeurs dans \mathbb{R}^k , de loi $\mathcal{N}_{\mathbb{R}^k}(0, I_{\mathbb{R}^k})$, et où $B \in \mathcal{L}(\mathbb{R}^k, E)$ est tel que

$$BB^* = \Lambda \tag{13.7}$$

répond à la question.

2. Mais pas nécessairement défini positif!

Il est toujours possible de prendre pour X_0 l'application identique de \mathbb{R}^k sur lui-même, considérée comme une variable aléatoire définie sur l'espace probabilisable $(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k})$ muni de la probabilité $\mathcal{N}_{\mathbb{R}^k}(0, \mathbf{I}_{\mathbb{R}^k})$.

Pour montrer l'existence d'un opérateur B satisfaisant l'égalité (13.7), il y a deux méthodes, la première ((a) ci-dessous) basée sur la propriété des opérateurs auto-adjoints de pouvoir être diagonalisés dans une base orthonormée, la seconde ((b) ci-dessous) basée sur la décomposition en carrés des formes quadratiques (décomposition de Gauss).

(a) On définit d'abord un opérateur auto-adjoint positif Δ tel que

$$\Delta^2 = \Lambda .$$

(On peut démontrer l'unicité d'un tel Δ . On dit que Δ est la **racine carrée positive** de Λ .)

Pour cela, puisque Λ est auto-adjoint, il existe une base orthonormée $(e_i)_{1 \leq i \leq d}$ de E formée de vecteurs propres de Λ . L'opérateur Λ s'écrit donc

$$\Lambda = \sum_{i=1}^d \lambda_i \langle \cdot, e_i \rangle e_i ,$$

où les λ_i sont les valeurs propres de Λ , répétées avec leur ordre de multiplicité (elles sont positives). L'opérateur

$$\Delta = \sum_{i=1}^d \sqrt{\lambda_i} \langle \cdot, e_i \rangle e_i$$

est lui aussi auto-adjoint, positif, et vérifie $\Delta^2 = \Lambda$. Soit Φ l'isomorphisme de \mathbb{R}^d sur E associé à la base $(e_i)_{1 \leq i \leq d}$. Il est défini par

$$\forall (a_1, \dots, a_d) \in \mathbb{R}^d \quad \Phi(a_1, \dots, a_d) = \sum_{i=1}^d a_i e_i .$$

On peut prendre $B = \Delta\Phi$. En effet, l'adjoint de Φ , isomorphisme de E sur \mathbb{R}^d , est défini par

$$\forall y \in E \quad \Phi^*(y) = (\langle y, e_1 \rangle, \dots, \langle y, e_d \rangle),$$

(cela résulte immédiatement de la définition) et on a

$$\Phi\Phi^* = \mathbf{1}_E$$

(identité sur E). On a donc

$$BB^* = \Delta\Phi\Phi^*\Delta^* = \Delta^2 = \Lambda .$$

(b) Le théorème de décomposition en carrés des formes quadratiques (**décomposition de Gauss**), appliqué à la forme quadratique $x \mapsto \langle \Lambda x, x \rangle$,

affirme qu'on peut écrire, pour tout $x \in E$,

$$\langle \Lambda x, x \rangle = \sum_{i=1}^r \langle u_i, x \rangle^2,$$

où r est le rang de Λ et où les u_i sont des formes linéaires (identifiées à des éléments de E) indépendantes sur E .

Définissons $A \in \mathcal{L}(E, \mathbb{R}^r)$ en posant, pour tout $x \in E$,

$$Ax = (\langle u_1, x \rangle, \dots, \langle u_r, x \rangle).$$

On a immédiatement, pour tout $x \in E$,

$$\langle \Lambda x, x \rangle = \langle Ax, Ax \rangle = \langle A^* Ax, x \rangle.$$

En observant que A admet pour adjointe l'application linéaire $B \in \mathcal{L}(\mathbb{R}^r, E)$ définie, pour tout $\alpha \in \mathbb{R}^r$, par

$$B\alpha = \sum_{i=1}^r \alpha_i u_i,$$

on obtient, pour tout $x \in E$,

$$\langle \Lambda x, x \rangle = \langle BB^* x, x \rangle,$$

et donc $\Lambda = BB^*$. □

Remarques importantes. 1. Dans les deux cas ((a) avec $k = d$, ((b) avec $k = r$), on observe que X est portée par le sous-espace affine $m + \text{Im } B$.

- Avec (a), il est immédiat que $\text{Im } B = \text{Im } \Lambda$.
- Avec (b), on écrit que $\text{Im } B = (\text{Ker } B^*)^\perp = (\text{Ker } A)^\perp$. Or

$$\text{Ker } A = \{x \mid \langle u_1, x \rangle = \dots = \langle u_r, x \rangle = 0\} = \{x \mid \langle \Lambda x, x \rangle = 0\}.$$

d'où³

$$\text{Ker } A = \text{Ker } \Lambda \quad \text{et} \quad \text{Im } B = \text{Im } \Lambda.$$

2. Avec (b), on a un peu plus, sans effort.

• La mesure gaussienne $\mathcal{N}_E(m, \Lambda)$ est l'image de la mesure standard $\mathcal{N}_{\mathbb{R}^r}(0, \mathbf{1}_{\mathbb{R}^r})$ par l'application

$$\alpha \mapsto m + \sum_{i=1}^r \alpha_i u_i. \tag{13.8}$$

3. Rappelons qu'il ne faut pas confondre le **cône isotrope** d'une forme quadratique q (ensemble des vecteurs x tels que $q(x) = 0$) et le **noyau** de q (ensemble des vecteurs x tels que $\varphi(x, y) = 0$ pour tout y , où φ est la forme bilinéaire associée à q). Toutefois ces deux ensembles coïncident dans le cas d'une forme positive, en raison de l'inégalité de Schwarz $|\varphi(x, y)| \leq q(x)q(y)$.

Si $r = d$, c'est-à-dire si l'opérateur Λ est *défini* positif, cette application est un difféomorphisme, et la mesure $\mathcal{N}_E(m, \Lambda)$ possède une densité par rapport à la mesure de Lebesgue⁴ sur E (voir prop. 13.6 pour le calcul de cette densité).

• Si $r < d$, l'application (13.8) est un difféomorphisme de \mathbb{R}^r sur le sous-espace affine $m + \text{Im } \Lambda$. Dans ce cas, la mesure $\mathcal{N}_E(m, \Lambda)$ possède une densité par rapport à la mesure de Lebesgue r -dimensionnelle sur $m + \text{Im } \Lambda$. (cette mesure de Lebesgue est bien définie grâce à la structure euclidienne...)

3. Les deux méthodes proposées pour la démonstration de ce théorème sont constructives et permettent, de manière évidente, d'écrire deux **algorithmes de simulation** d'une variable aléatoire gaussienne de loi $\mathcal{N}_{\mathbb{R}^d}(m, \Lambda)$, le deuxième (issu de (b)) consommant a priori moins d'appels au générateur aléatoire que le premier (issu de (a)) dès que $r < d$ (cela ne veut pas dire qu'il soit plus rapide).

Définition 13.5. Avec les notations employées ci-dessus, une variable aléatoire gaussienne de loi $\mathcal{N}_E(m, \Lambda)$ est dite **dégénérée** si le sous-espace affine $m + \text{Im } \Lambda$ (encore égal à $m + (\text{Ker } \Lambda)^\perp$) est un sous-espace strict de E .

Nous avons vu (cf. chap. 8, exercice 7) qu'une variable aléatoire X de carré de norme intégrable prend P-p.s. ses valeurs dans le sous-espace affine $EX + (\text{Ker } \Lambda_X)^\perp$ (sous-espace d'ailleurs identique à $EX + \text{Im } \Lambda_X$). Ainsi pour qu'une variable aléatoire gaussienne à valeurs dans \mathbb{R}^d admette une **densité**, il est nécessaire que ce noyau soit réduit à $\{0\}$. Nous allons voir que cette condition est aussi suffisante.

Proposition 13.6. Soient $m \in \mathbb{R}^d$ et C une matrice $d \times d$ symétrique positive et soit X une variable aléatoire à valeurs dans \mathbb{R}^d de loi gaussienne $\mathcal{N}_{\mathbb{R}^d}(m, C)$.

(a) Si C est **défini positif**, alors X admet une **densité** f_X donnée par, pour tout $x \in \mathbb{R}^d$,

$$f_X(x) = \frac{1}{(\sqrt{2\pi})^d} (\det C)^{-1/2} \exp\left(-\frac{1}{2} \langle C^{-1}(x - m), (x - m) \rangle\right). \quad (13.9)$$

(b) Si C n'est pas **défini positif**, alors X prend P-p.s. ses valeurs dans le sous-espace affine $m + \text{Im } C$ et, en conséquence, n'admet pas de densité (autrement dit, sa loi n'est pas absolument continue; elle est même **étrangère** à la mesure de Lebesgue sur \mathbb{R}^d).

Démonstration. (a) Supposons C **défini positif**. Soit Y une variable aléatoire à valeurs dans \mathbb{R}^d de loi gaussienne $\mathcal{N}_{\mathbb{R}^d}(0, \mathbf{1}_{\mathbb{R}^d})$. Il résulte du lemme

4. Pour une définition de la **mesure de Lebesgue** sur l'espace euclidien E , voir le complément en fin de cette section.

13.3 que les marginales de Y sont **indépendantes** de même loi gaussienne $\mathcal{N}_k(0, 1)$ et que la variable aléatoire Y admet une densité f_Y (produit direct des densités de ses marginales), donnée, pour tout $y \in \mathbb{R}^d$, par

$$f_Y(y) = \frac{1}{(\sqrt{2\pi})^d} \exp\left(-\frac{\|y\|^2}{2}\right),$$

où $\|\cdot\|$ désigne la norme euclidienne usuelle de \mathbb{R}^d . Soit B la **racine carrée** positive de C . La variable aléatoire $Z = m + BY$ est de même loi gaussienne $\mathcal{N}_{\mathbb{R}^d}(m, C)$ que X . L'application $y \mapsto m + By$ est un difféomorphisme, puisque les matrices C , et donc B , sont inversibles. La variable aléatoire Z admet donc une densité f_Z donnée, pour tout $z \in \mathbb{R}^d$, par

$$\begin{aligned} f_Z(z) &= f_Y(B^{-1}(z - m)) |\det(B^{-1})| \\ &= \frac{1}{(\sqrt{2\pi})^d} \exp\left(-\frac{\|B^{-1}(z - m)\|^2}{2}\right) |\det(B^{-1})|, \end{aligned}$$

ce qui démontre le résultat après avoir remarqué que

$$\|B^{-1}(z - m)\|^2 = \langle C^{-1}(z - m), (z - m) \rangle \quad \text{et} \quad \det(B^{-1}) = (\det C)^{-\frac{1}{2}}$$

(on utilise le fait que B est auto-adjoint et que $B^2 = C$).

(b) Si C n'est pas **définie positive**, on a $P_X(m + \text{Im } C) = 1$ (puisque X prend P.-p.s. ses valeurs dans le sous-espace affine $m + \text{Im } C$), tandis que l'on a $\lambda_d(m + \text{Im } C) = 0$, puisque le sous-espace affine $m + \text{Im } C$ est stricte. Ainsi la loi de X est étrangère à la mesure de Lebesgue λ_d . \square

Complément. On définit la **mesure de Lebesgue sur l'espace euclidien** E de la manière suivante : on identifie \mathbb{R}^d à E au moyen de l'isomorphisme Φ introduit dans la démonstration du théorème 13.4, après avoir choisi une base *orthonormée* de E . La **mesure de Lebesgue sur** E est la mesure image μ de la mesure de Lebesgue λ_d sur \mathbb{R}^d par Φ . Elle est en fait indépendante du choix de la base *orthonormée*.

En effet, notons Ψ un autre isomorphisme correspondant à un autre choix de base *orthonormée* de E , et ν la mesure image de la mesure de Lebesgue λ_d sur \mathbb{R}^d par Ψ . On a, pour tout $B \in \mathcal{E}$,

$$\nu(B) = \lambda_d(\Psi^{-1}(B)) = \int_{\mathbb{R}^d} \mathbf{1}_B \circ \Psi \, d\lambda_d.$$

Puisque $\Phi \Phi^* = \mathbf{1}_E$, on a

$$\nu(B) = \int_{\mathbb{R}^d} (\mathbf{1}_B \circ \Phi) \circ (\Phi^* \circ \Psi) \, d\lambda_d,$$

soit, en faisant le changement de variables défini par le difféomorphisme $\Phi^* \circ \Psi$ de jacobien ± 1 ($\Phi^* \circ \Psi$ est une isométrie),

$$\nu(B) = \int_{\mathbb{R}^d} \mathbf{1}_B \circ \Phi \, d\lambda_d = \mu(B).$$

13.3. Marginales

On s'intéresse aux propriétés d'**indépendance** des marginales. Étudions d'abord le cas simple où $E = \mathbb{R}^d$ et où les marginales considérées sont toutes uni-dimensionnelles.

Proposition 13.7. *Soit $X = (X_1, \dots, X_d)$ une variable aléatoire à valeurs dans \mathbb{R}^d de loi gaussienne $\mathcal{N}_{\mathbb{R}^d}(m, C)$, où $m \in \mathbb{R}^d$ et où C est une matrice $d \times d$ symétrique positive. Alors, les variables aléatoires X_j , $1 \leq j \leq d$, sont gaussiennes.*

De plus, pour que les variables aléatoires X_j , $1 \leq j \leq d$, soient indépendantes il faut et il suffit qu'elles soient non corrélées deux à deux (ce qui est équivalent à dire que la matrice des covariances C de X est diagonale).

Démonstration. Les variables aléatoires X_j , $1 \leq j \leq d$, sont gaussiennes comme transformées linéaires de la variable aléatoire gaussienne X .

Si les variables aléatoires X_j , $1 \leq j \leq d$, sont indépendantes, leur covariance deux à deux est nulle, et la matrice des covariances C de X est diagonale.

Inversement, supposons que la matrice des covariances C de X soit diagonale. La fonction caractéristique de X vérifie alors, pour tout u de \mathbb{R}^d ,

$$\begin{aligned} \varphi_X(u) &= \exp(i \langle m, u \rangle) \exp\left(-\frac{\langle Cu, u \rangle}{2}\right) \\ &= \prod_{j=1}^d \exp\left(i \sum_{j=1}^d m_j u_j\right) \exp\left(-\frac{1}{2} \sum_{j=1}^d C_{jj} u_j^2\right), \end{aligned}$$

ce qui démontre que, pour tout u de \mathbb{R}^d ,

$$\varphi_X(u) = \prod_{j=1}^d \varphi_{X_j}(u_j),$$

la fonction caractéristique de la marginale X_j s'obtenant en effet par le calcul suivant :

$$\varphi_{X_j}(u_j) = \varphi_X(0, \dots, 0, u_j, 0, \dots, 0) = \exp\left(i \sum_{j=1}^d m_j u_j\right) \exp\left(-\frac{1}{2} \sum_{j=1}^d C_{jj} u_j^2\right).$$

Ainsi les variables aléatoires X_j sont indépendantes. □

On s'intéresse maintenant aux propriétés d'**indépendance** des marginales d'une variable aléatoire à valeurs dans l'espace euclidien E .

Pour cela, on rappelle la définition de l'opérateur d'intercovariance de deux variables aléatoires à valeurs dans des espaces euclidiens (cf. chap. 8, exercice 8).

Définition 13.8. Soient F et G deux espaces euclidiens et soient deux variables aléatoires $X \in \mathcal{L}_F^2(\Omega, \mathcal{A}, P)$ et $Y \in \mathcal{L}_G^2(\Omega, \mathcal{A}, P)$. L'opérateur d'intercovariance de X et Y est l'unique opérateur $\Lambda_{X,Y} \in \mathcal{L}(F, G)$ vérifiant

$$\forall (x, y) \in F \times G \quad \langle \Lambda_{X,Y} x, y \rangle = E[\langle \overset{\circ}{X}, x \rangle \langle \overset{\circ}{Y}, y \rangle] = \text{cov}(\langle X, x \rangle, \langle Y, y \rangle).$$

Les variables aléatoires X et Y sont dites **non corrélées** si $\Lambda_{X,Y} = 0$.

Remarque. L'opérateur $\Lambda_{X,X}$ n'est autre que l'opérateur d'auto-covariance de X . Cette notion de variables aléatoires non corrélées coïncide, dans le cas où $E = F = \mathbb{R}$, avec la notion de variables aléatoires réelles non corrélées, définie au chapitre 8.

Proposition 13.9. Soit X une variable aléatoire à valeurs dans l'espace euclidien E de loi gaussienne $\mathcal{N}_E(m, \Lambda)$, où $m \in E$ et $\Lambda \in \mathcal{L}^+(\mathbb{R})$.

Soit $E = \bigoplus_{j=1}^n E_j$, $n \leq d$, une décomposition en somme directe de E (tout $x \in E$ s'écrit donc de manière unique $x = \sum_{j=1}^n x_j$, où $x_j \in E_j$, pour tout $j = 1, \dots, n$). La variable aléatoire s'écrit ainsi $X = \sum_{j=1}^n X_j$, où, pour tout $j = 1, \dots, n$, X_j est une variable aléatoire à valeurs dans le sous-espace E_j .

Alors, les variables aléatoires X_j sont gaussiennes.

De plus, pour que les variables aléatoires X_j , $1 \leq j \leq n$, soient indépendantes il faut et il suffit qu'elles soient non corrélées deux à deux.

Démonstration. Les variables aléatoires X_j sont gaussiennes, comme transformées linéaires (par les projections sur les E_j) de la variable aléatoire gaussienne X .

Soient j et k deux entiers distincts, $1 \leq j \leq n$, $1 \leq k \leq n$. Si les variables aléatoires X_j , $1 \leq j \leq n$, sont indépendantes, il en est de même, pour tout $x_j \in E_j$ et tout $y_k \in E_k$, des variables aléatoires $\langle X_j, x_j \rangle$ et $\langle X_k, y_k \rangle$ et par conséquent, on a $\Lambda_{X_j, X_k} = 0$.

Inversement, on suppose que les variables aléatoires X_j , $1 \leq j \leq n$, sont non corrélées deux à deux. Pour tout choix de u_j dans E_j , $1 \leq j \leq n$, la variable aléatoire $(\langle X_1, u_1 \rangle, \langle X_2, u_2 \rangle, \dots, \langle X_n, u_n \rangle)$ est gaussienne à valeurs dans \mathbb{R}^n . Par hypothèse, sa matrice des covariances est diagonale; il résulte alors de la proposition 13.7 que les variables aléatoires $\langle X_j, u_j \rangle$, $1 \leq j \leq n$, sont indépendantes. On a alors

$$\begin{aligned}\varphi_{(X_1, X_2, \dots, X_n)}(u_1, u_2, \dots, u_n) &= \mathbb{E} \left[\exp \left(i \sum_{j=1}^n \langle X_j, u_j \rangle \right) \right] \\ &= \mathbb{E} \left[\prod_{j=1}^n \exp(i \langle X_j, u_j \rangle) \right],\end{aligned}$$

soit, par indépendance,

$$\varphi_{(X_1, X_2, \dots, X_n)}(u_1, u_2, \dots, u_n) = \prod_{j=1}^n \mathbb{E} \left[\exp(i \langle X_j, u_j \rangle) \right] :$$

puisque la fonction caractéristique de X_j vérifie, pour tout $u_j \in E_j$,

$$\varphi_{X_j}(u_j) = \varphi_{(X_1, X_2, \dots, X_n)}(0, \dots, 0, u_j, 0, \dots, 0) = \mathbb{E} \left[\exp(i \langle X_j, u_j \rangle) \right],$$

on a montré que, pour tout $(u_1, u_2, \dots, u_n) \in \prod_{j=1}^n E_j$,

$$\varphi_{(X_1, X_2, \dots, X_n)}(u_1, u_2, \dots, u_n) = \prod_{j=1}^n \varphi_{X_j}(u_j).$$

ce qui est équivalent à l'indépendance des X_j . □

Remarque. On obtient une proposition analogue à celle de la proposition 13.9, en remplaçant dans celle-ci la décomposition en somme directe $\bigoplus_{j=1}^n E_j$ par le produit cartésien $\prod_{j=1}^n E_j$, puisqu'en fait ces ensembles sont isomorphes pour la structure euclidienne.

En particulier, on obtient le corollaire suivant relatif aux marginales d'une variable aléatoire gaussienne à valeurs dans un espace \mathbb{R}^d . Nous l'énonçons (évidemment sans nouvelle démonstration), étant donné son importance « pratique ».

Corollaire 13.10. *Soient des variables aléatoires X_j , $1 \leq j \leq n$, définies sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans \mathbb{R}^{d_j} . Si la variable aléatoire $X = (X_1, X_2, \dots, X_n)$, à valeurs dans $\mathbb{R}^{d_1+d_2+\dots+d_n}$, est gaussienne, et si les X_j sont non corrélées, alors les variables aléatoires X_j sont gaussiennes et indépendantes.*

Remarque. Comme le montre le contre-exemple suivant, dans la proposition 13.9 (et donc aussi dans son corollaire 13.10), il ne faut surtout pas oublier l'hypothèse que la variable aléatoire **globale** X est gaussienne. Par ailleurs, il est équivalent de dire que les X_j sont non corrélées ou de dire que la matrice de covariance de X est diagonale par blocs.

Contre-exemple. Soit X une variable aléatoire réelle de loi symétrique à densité f_X et admettant un moment d'ordre deux. Pour tout réel **positif** a ,

on définit la variable aléatoire Y_a par

$$Y_a = -X \mathbf{1}_{(|X| \leq a)} + X \mathbf{1}_{(|X| > a)} \equiv X(2\mathbf{1}_{(|X| > a)} - 1).$$

Les variables aléatoires Y_a et X ont même loi. En effet, pour tout $f \in \mathcal{C}_c^1(\mathbb{R})$, on a, d'après le théorème de transfert,

$$Ef(Y_a) = \int_{(|x| \leq a)} f(-x) f_X(x) dx + \int_{(|x| > a)} f(x) f_X(x) dx,$$

soit, en faisant le changement de variables $x \mapsto -x$ dans la première intégrale, et en utilisant la parité de f_X ,

$$\begin{aligned} Ef(Y_a) &= \int_{(|y| \leq a)} f(y) f_X(y) dy + \int_{(|x| > a)} f(x) f_X(x) dx \\ &= \int_{\mathbb{R}} f(x) f_X(x) dx = Ef(X), \end{aligned}$$

ce qui démontre que Y_a et X ont même loi. Puisque la loi de X est symétrique, Y_a et X sont centrées et leur covariance est donnée par

$$\text{cov}(X, Y_a) = E(XY_a) = E[X^2(2\mathbf{1}_{(|X| > a)} - 1)] = 4 \int_{(x > a)} x^2 f_X(x) dx - EX^2.$$

En particulier, si X est de loi $\mathcal{N}_{\mathbb{R}}(0, 1)$, il en est de même de Y_a . Par contre la variable aléatoire $X + Y_a = 2X\mathbf{1}_{(|X| > a)}$ n'est pas gaussienne, puisque l'on a $P(X + Y_a = 0) = P(|X| \leq a) > 0$; par conséquent **la variable aléatoire (X, Y_a) n'est pas gaussienne.** Enfin, on peut choisir a positif tel que $\text{cov}(X, Y_a) = 0$; en effet, sous les hypothèses précédentes, on a

$$\text{cov}(X, Y_a) = 0 \iff 4 \int_{(x > a)} x^2 f_X(x) dx = 1.$$

Ainsi, puisque $\int_{\mathbb{R}^+} x^2 f_X(x) dx = \frac{1}{2}$, et que la fonction $a \mapsto \int_{(x > a)} x^2 f_X(x) dx$ est strictement décroissante sur \mathbb{R}^+ et tend vers 0 quand a tend vers $+\infty$, **il existe un unique a positif tel que $\text{cov}(X, Y_a) = 0$.** Pour cet a , Y_a et X ont même loi gaussienne, sont de covariance nulle, et cependant, le couple (X, Y_a) n'est pas gaussien.

On a toutefois la proposition importante suivante :

Proposition 13.11. *Soit une variable aléatoire $Z = (X, Y)$ à valeurs dans $E \times F$, où E et F sont des espaces euclidiens. On suppose que X et Y sont indépendantes. Pour que Z soit gaussienne, il faut et il suffit que X et Y le soient.*

Démonstration. Si Z est gaussienne, X et Y le sont comme transformées linéaires de Z . Inversement, si X et Y sont gaussiennes, leurs fonctions

caractéristiques sont données par

$$\forall u \in E \quad \varphi_X(u) = \exp(i \langle EX, u \rangle_E) \exp\left[-\frac{1}{2} \langle \Lambda_X u, u \rangle_E\right]$$

et

$$\forall v \in F \quad \varphi_Y(v) = \exp(i \langle EY, v \rangle_F) \exp\left[-\frac{1}{2} \langle \Lambda_Y v, v \rangle_F\right].$$

L'indépendance de X et Y implique que la fonction caractéristique φ_Z de Z est le produit direct de φ_X et φ_Y , ce qui donne, pour tout $(u, v) \in E \times F$,

$$\varphi_Z(u, v) = \exp[i(\langle EX, u \rangle_E + \langle EY, v \rangle_F)] \exp\left[-\frac{1}{2}(\langle \Lambda_X u, u \rangle_E + \langle \Lambda_Y v, v \rangle_F)\right].$$

Le produit scalaire sur $E \times F$ étant défini par la relation

$$\begin{aligned} &\text{pour tout } (u, v) \text{ et tout } (u', v') \text{ de } E \times F, \\ &\langle (u, v), (u', v') \rangle_{E \times F} = \langle u, u' \rangle_E + \langle v, v' \rangle_F, \end{aligned}$$

si on définit l'opérateur $\Lambda \in \mathcal{L}^+(E \times F)$ par

$$\langle \Lambda(u, v), (u, v) \rangle_{E \times F} = \langle \Lambda_X u, u \rangle_E + \langle \Lambda_Y v, v \rangle_F,$$

on a bien

$$\varphi_Z(u, v) = \exp[i \langle (EX, EY), (u, v) \rangle_{E \times F}] \exp\left[-\frac{1}{2} \langle \Lambda(u, v), (u, v) \rangle_{E \times F}\right],$$

ce qui démontre que Z est gaussienne. \square

Voici deux autres corollaires de la proposition 13.9.

Corollaire 13.12. Soit X une variable aléatoire à valeurs dans l'espace euclidien E de loi gaussienne $\mathcal{N}_E(m, \Lambda)$, où $m \in E$ et $\Lambda \in \mathcal{L}^+(E)$ et soit (e_1, \dots, e_d) une base orthogonale de E .

Pour que les variables aléatoires $\langle X, e_i \rangle$, $1 \leq i \leq d$, soient indépendantes il faut et il suffit que, pour tout $i = 1, \dots, d$, e_i soit vecteur propre de Λ (ce qui est équivalent à dire que la matrice de covariance C_X , représentation matricielle de Λ dans la base (e_1, \dots, e_d) est diagonale).

Démonstration. Notons que

$$\langle \Lambda e_i, e_j \rangle = \text{cov}(\langle X, e_i \rangle, \langle X, e_j \rangle). \quad (13.10)$$

Si les variables aléatoires $\langle X, e_i \rangle$, $1 \leq i \leq d$, sont indépendantes, elles sont non corellées deux à deux et on a, si $i \neq j$,

$$\langle \Lambda e_i, e_j \rangle = 0,$$

ce qui implique que, pour tout $i = 1, \dots, d$,

$$\Lambda e_i = \sum_{j=1}^d \langle \Lambda e_i, e_j \rangle e_j = \langle \Lambda e_i, e_i \rangle e_i.$$

Ainsi, e_i est vecteur propre de Λ associé à la valeur propre $\langle \Lambda e_i, e_i \rangle = \sigma_{\langle X, e_i \rangle}^2$.

Inversement, supposons que, pour tout $i = 1, \dots, d$, e_i soit vecteur propre de Λ associé à la valeur propre positive λ_i , c'est à dire que $\Lambda e_i = \lambda_i e_i$. La base (e_1, \dots, e_d) étant orthogonale, on a, si $i \neq j$,

$$\langle \Lambda e_i, e_j \rangle = 0,$$

et donc, en vertu de (13.10), les variables aléatoires $\langle X, e_i \rangle$, $1 \leq i \leq d$, sont non corellées deux à deux. Leur indépendance résulte alors de la proposition 13.9. \square

Corollaire 13.13. *Soit X une variable aléatoire à valeurs dans l'espace euclidien E de loi gaussienne $\mathcal{N}_E(m, \Lambda)$, où $m \in E$ et $\Lambda \in \mathcal{L}^+(E)$. Soient E_j , $j = 1, \dots, n$, les sous-espaces propres de Λ (ils sont orthogonaux et stables par Λ et forment une décomposition en somme directe de E); on note Π_j le projecteur orthogonal sur E_j . Les marginales $X_j = \Pi_j \circ X$ de X sur les sous-espaces E_j , $j = 1, \dots, k$, sont indépendantes de lois respectives $\mathcal{N}_{E_j}(\Pi_j m, \Lambda_j)$, où $\Lambda_j \in \mathcal{L}^+(E_j)$ est la restriction de Λ à E_j .*

Démonstration. L'indépendance résulte de ce que les E_j , $j = 1, \dots, n$, forment une décomposition en somme directe de E (cf. proposition 13.9). Les X_j sont gaussiennes, de moyenne $\Pi_j m$ et d'opérateur d'autocovariance $\Pi_j \Lambda \Pi_j^* = \Lambda_j$, puisque les E_j sont stables par Λ . \square

On a vu dans le contre-exemple ci-dessus (p. 246) qu'une variable aléatoire pouvait avoir des marginales gaussiennes sans être elle-même gaussienne. La proposition suivante donne une **caractérisation de variables aléatoires gaussiennes**.

Proposition 13.14. *Soit une variable aléatoire $Z = (X, Y)$ à valeurs dans $E \times F$, où E et F sont des espaces euclidiens. On suppose que Z est de carré de norme intégrable. On note Λ_X (resp. Λ_Y) l'opérateur de covariance de X (resp. Y) et $\Lambda_{X,Y} \in \mathcal{L}(E, F)$ l'opérateur d'inter-covariance de X et Y . On suppose que Λ_X est inversible.*

Alors, la variable aléatoire Z est gaussienne si et seulement si les deux conditions suivantes sont satisfaites :

- (i) *la marginale X est gaussienne;*
- (ii) *il existe $A \in \mathcal{L}(E, F)$, $b \in F$ et $\Lambda \in \mathcal{L}^+(F)$ tels que, pour P_X -presque tout $x \in E$, la loi conditionnelle $P_Y^{X=x}$ de Y sachant $X = x$ est la loi gaussienne $\mathcal{N}_F(Ax + b, \Lambda)$.*

Dans ce cas, on a

$$m_Y^{X=x} = EY + \Lambda_{X,Y} \Lambda_X^{-1} (x - EX) \quad \text{et} \quad \Lambda = \Lambda_Y - \Lambda_{X,Y} \Lambda_X^{-1} \Lambda_{X,Y}^*.$$

où $m_Y^{X=x}$ est la moyenne conditionnelle de Y sachant $X = x$.

Démonstration. **Supposons Z gaussienne.** La marginale X est alors gaussienne. Pour $C \in \mathcal{L}(E, F)$, on définit $Y' = Y - CX$; la variable aléatoire (X, Y') , transformée linéaire de Z est alors gaussienne. Un calcul simple montre que

$$\Lambda_{X, Y'} = \Lambda_{X, Y} - C\Lambda_X;$$

par conséquent, $\Lambda_{X, Y'} = 0$ si et seulement si $C = \Lambda_{X, Y}\Lambda_X^{-1}$. Choisissons pour C cette valeur; il résulte de la remarque 1 suivant la proposition 13.10 que les variables aléatoires X et Y' sont indépendantes. On a donc les égalités de lois conditionnelles suivantes: pour P_X -presque tout $x \in E$,

$$P_Y^{X=x} = P_{Y'+CX}^{X=x} = P_{Y'+Cx}^{X=x},$$

soit, puisque X et Y' sont indépendantes,

$$P_Y^{X=x} = P_{Y'+Cx} = \mathcal{N}_F(EY' + Cx, \Lambda_{Y'}),$$

ce qui achève de démontrer la condition nécessaire. En fait, on a alors, pour P_X -presque tout $x \in E$,

$$\boxed{P_Y^{X=x} = \mathcal{N}_F(EY + \Lambda_{X, Y}\Lambda_X^{-1}(x - EX), \Lambda_Y - \Lambda_{X, Y}\Lambda_X^{-1}\Lambda_{X, Y}^*):}$$

en effet, on a $EY' = EY - CEX$ et, par indépendance de Y' et CX ,

$$\Lambda_Y = \Lambda_{Y'} + \Lambda_{CX} = \Lambda_{Y'} + C\Lambda_X C^* = \Lambda_{Y'} + (\Lambda_{X, Y}\Lambda_X^{-1})\Lambda_X(\Lambda_{X, Y}\Lambda_X^{-1})^*.$$

ce qui donne, en tenant compte du fait que Λ_X est auto-transposé,

$$\Lambda_{Y'} = \Lambda_Y - \Lambda_{X, Y}\Lambda_X^{-1}\Lambda_{X, Y}^*.$$

Inversement, supposons que X est gaussienne et qu'existent $A \in \mathcal{L}(E, F)$, $b \in F$ et $\Lambda \in \mathcal{L}^+(F)$ tels que, pour P_X -presque tout $x \in E$, la loi conditionnelle $P_Y^{X=x}$ de Y sachant $X = x$ est la loi gaussienne $\mathcal{N}_F(Ax + b, \Lambda)$. Soit $Y'' = Y - AX - b$. Il résulte du théorème de transfert conditionnel que, pour P_X -presque tout $x \in E$,

$$P_{Y''}^{X=x} = P_{Y-AX-b}^{X=x} = \mathcal{N}_F(0, \Lambda),$$

loi indépendante de x ; les variables aléatoires Y'' et X sont donc indépendantes et la loi de Y'' est la loi gaussienne $\mathcal{N}_F(0, \Lambda)$. Il résulte alors de la proposition 13.11 que la variable aléatoire (X, Y'') est gaussienne, et donc aussi la variable aléatoire (X, Y) , transformée linéaire de (X, Y'') . \square

13.4. Régression; le modèle linéaire

Nous avons étudié au chapitre 8 le **problème de régression linéaire** entre variables aléatoires réelles et sa généralisation au cas de variables aléatoires à valeurs dans un espace euclidien (chap. 8, ex. 8). Rappelons la formulation et la solution de ces problèmes.

Cas de variables aléatoires réelles

Les variables aléatoires réelles X et $Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ étant données, la « meilleure » approximation de Y comme fonction affine de X au sens des moindres carrés, identifiée à la solution en le couple $(a, b) \in \mathbb{R}^2$ du problème de minimisation

$$\inf(\Phi(a, b) \mid (a, b) \in \mathbb{R}^2),$$

où $\Phi(a, b) = E[Y - (aX + b)]^2$, est

$$EY + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (X - EX).$$

Le couple optimal (\hat{a}, \hat{b}) est donné par

$$\begin{cases} \hat{a} = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \\ \hat{b} = EY - EX \cdot \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}. \end{cases} \quad (13.11)$$

La droite de régression linéaire de Y en X est la droite d'équation

$$(y - EY) - \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (x - EX) = 0,$$

et l'erreur de prédiction est

$$\Phi(\hat{a}, \hat{b}) = E[\overset{\circ}{Y} - \hat{a}\overset{\circ}{X}]^2 = \sigma_Y^2 - 2\hat{a} \operatorname{cov}(X, Y) + \hat{a}^2 \sigma_X^2 = \sigma_Y^2 (1 - \rho_{X,Y}^2).$$

En particulier, si la variable aléatoire est de loi uniforme sur l'ensemble des n points du plan $\{(x_i, y_i)\}_{1 \leq i \leq n}$, on a $\Phi(a, b) = \frac{1}{n} \sum_{i=1}^n [y_i - (ax_i + b)]^2$ et on retrouve la droite d'approximation des moindres carrés des physiciens.

Cas de variables aléatoires à valeurs dans un espace euclidien

Soient $X \in \mathcal{L}_F^2(\Omega, \mathcal{A}, P)$ et $Y \in \mathcal{L}_G^2(\Omega, \mathcal{A}, P)$ deux variables aléatoires prenant respectivement leurs valeurs dans des espaces euclidiens F et G ; on suppose que l'opérateur d'auto-covariance Λ_X est inversible. La « meilleure » approximation de Y comme fonction affine de X au sens des moindres carrés, identifiée à la solution en le couple $(A, b) \in \mathcal{L}(F, G) \times G$ du problème de minimisation

$$\inf(\Phi(A, b) \mid (A, b) \in \mathcal{L}(F, G) \times G),$$

où $\Phi(A, b) = E \|Y - (AX + b)\|^2$, est

$$EY + \Lambda_{X,Y} \Lambda_X^{-1} (X - EX);$$

dans cette formule, $\Lambda_{X,Y}$ désigne l'opérateur d'intercovariance de X et Y . Le couple optimal (\hat{A}, \hat{b}) est donné par

$$(\hat{A}, \hat{b}) = (\Lambda_{X,Y} \Lambda_X^{-1}, EY - \Lambda_{X,Y} \Lambda_X^{-1}(EX)).$$

La surface de régression linéaire de Y en X est la surface (sous-espace affine) d'équation

$$(y - EY) - \Lambda_{X,Y} \Lambda_X^{-1}(x - EX) = 0,$$

et l'erreur de prédiction est

$$\Phi(\hat{A}, \hat{b}) = E \|\hat{Y} - \hat{A}\hat{X}\|^2 = \text{tr}[\Lambda_Y + \hat{A}\Lambda_X\hat{A}^* - \Lambda_{X,Y}\hat{A}^* - \hat{A}\Lambda_{Y,X}].$$

soit, en tenant compte de ce que $\hat{A} = \Lambda_{X,Y}\Lambda_X^{-1}$,

$$\Phi(\hat{A}, \hat{b}) = \text{tr}[\Lambda_Y - \Lambda_{X,Y}\Lambda_X^{-1}\Lambda_{X,Y}^*].$$

Ces résultats sont théoriques et nécessitent, pour être appliqués, la connaissance des « moments » d'ordre deux du couple (X, Y) . Se posent alors différents problèmes statistiques; au vu de résultats expérimentaux $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, valeurs du couple (X, Y) obtenues lors d'expériences indépendantes, on veut avoir des renseignements sur **la droite de régression de Y en X** . Plusieurs attitudes peuvent être adoptées, donner une **estimation** de la droite de régression, c'est-à-dire en fait donner une estimation des paramètres a et b , **tester** les « bonnes » valeurs de a et b , ou donner un **intervalle de confiance** pour ces paramètres. On peut aussi se poser la question de **prédiction** de valeurs de Y connaissant une réalisation de X .

Nous abordons ci-dessous ces différents problèmes. Les notions introduites pour les formuler précisément et les traiter ont une portée générale en statistique, mais nous n'en donnons de définition que dans le cadre limité du problème de régression. Le problème d'estimation sera introduit de manière plus systématique au chapitre 14, section 14.4. La notion de test⁵ est abordée à différents endroits de ce livre (voir dans l'index à : test du chi-deux, de Student, de Kolmogorov). Dans la suite, par souci de simplification, nous ne considérons que des variables aléatoires réelles.

13.4.1. Estimation des paramètres de régression

Problème 1. Suite à la modélisation probabiliste d'un phénomène aléatoire, on s'intéresse au couple de variables aléatoires réelles $Z = (X, Y)$ censé représenter deux « grandeurs réelles » liées à ce phénomène. La loi de Z est **inconnue** de l'expérimentateur; toutefois, celui-ci, à l'issue de calculs

5. Concernant la mise en œuvre pratique de ces problèmes statistiques, on pourra consulter le livre de Gilbert Saporta (1990), *Probabilités, analyse des données et statistique*, Technip, Paris.

et raisonnements, est conduit à formuler des **hypothèses** sur cette loi (en particulier, que les variables aléatoires X et Y sont d'ordre deux). Il s'agit d'**estimer la droite de régression de Y en X** au vu d'un **échantillon** de taille n de Z , à savoir le vecteur $\underline{z}_n = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ de \mathbb{R}^{2n} , obtenu en observant n réalisations « indépendantes » de ce phénomène. Cet échantillon est censé être la réalisation (c'est-à-dire la valeur pour une réalisation ω) de n variables aléatoires Z_1, Z_2, \dots, Z_n , **indépendantes, de même loi**⁶ que Z . La méthode est celle des **moindres carrés** ; elle consiste à choisir, pour cet ω , la droite d'équation $y = \widehat{a}_n(\omega)x + \widehat{b}_n(\omega)$ où le couple $(\widehat{a}_n(\omega), \widehat{b}_n(\omega))$ est solution du problème de minimisation

$$\inf(\Theta(a, b) \mid (a, b) \in \mathbb{R}^2),$$

où

$$\Theta(a, b) = \sum_{j=1}^n [Y_j(\omega) - (aX_j(\omega) + b)]^2.$$

La droite d'équation $y = \widehat{a}_n(\omega)x + \widehat{b}_n(\omega)$ est appelée **estimée** (au sens des moindres carrés) de **la droite de régression de Y en X** . Une autre justification de l'emploi de cette estimée sera apportée par le théorème de **Gauss-Markov** ci-dessous. Cette estimée est déterminée (pour tout ω) par la proposition suivante.

Proposition 13.15. *Soient deux variables aléatoires réelles X et $Y \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ et soit $\underline{Z}_n = [(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)]$ un échantillon empirique de taille n de la variable aléatoire $Z = (X, Y)$. Les coefficients de l'estimée (au sens des moindres carrés) de la droite de régression de Y en X sont donnés par*

$$\begin{cases} \widehat{a}_n &= r_n \frac{s_{Y,n}}{s_{X,n}} \\ \widehat{b}_n &= \bar{Y}_n - \bar{X}_n \cdot r_n \frac{s_{Y,n}}{s_{X,n}}, \end{cases} \quad (13.12)$$

où on note les moments empiriques associés à cet échantillon de la manière suivante : \bar{X}_n et \bar{Y}_n sont les **moyennes empiriques** de X et Y , $s_{X,n}$, $s_{Y,n}$ sont les **variances empiriques** de X et Y , et r_n est le **coefficient de corrélation empirique** de X et Y . Ces quantités sont ainsi définies :

6. La variable aléatoire $\underline{Z}_n = (Z_1, Z_2, \dots, Z_n)$ est appelée « **échantillon empirique** » de taille n de la variable aléatoire Z .

$$\begin{aligned}\bar{X}_n &= \frac{1}{n} \sum_{j=1}^n X_j, & \bar{Y}_n &= \frac{1}{n} \sum_{j=1}^n Y_j, \\ s_{X,n}^2 &= \frac{1}{n} \sum_{j=1}^n X_j^2 - (\bar{X}_n)^2, & s_{Y,n}^2 &= \frac{1}{n} \sum_{j=1}^n Y_j^2 - (\bar{Y}_n)^2, \\ r_n &= \frac{\frac{1}{n} \sum_{j=1}^n X_j Y_j - \bar{X}_n \bar{Y}_n}{s_{X,n} s_{Y,n}}.\end{aligned}$$

Démonstration. Il suffit d'appliquer, de la manière suivante, les résultats sur la régression linéaire rappelés ci-dessus. Pour tout ω fixé, considérons l'espace probabilisé $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2}, \mu_\omega)$, où $\mu_\omega = \frac{1}{n} \sum_{j=1}^n \delta_{(X_j(\omega), Y_j(\omega))}$ est la **mesure empirique** associée à l'échantillon, et considérons sur ce nouvel espace probabilisé les variables aléatoires U et V , projections canoniques de \mathbb{R}^2 sur \mathbb{R} . Puisqu'on a

$$\Theta(a, b) = \sum_{j=1}^n [Y_j(\omega) - (aX_j(\omega) + b)]^2 = n \int_{\mathbb{R}^2} [V - (aU + b)]^2 d\mu_\omega,$$

les formules (13.11) donnent le résultat. \square

Remarque. Les coefficients \hat{a}_n et \hat{b}_n introduits dans cette proposition sont en fait des variables aléatoires dont les valeurs en chaque ω déterminent une estimée de la droite de régression ; ce sont des estimateurs (c'est-à-dire des fonctions mesurables de l'échantillon \underline{Z}_n) des vrais coefficients de régression linéaire a et b .

Problème 2. Dans bien des situations, la variable aléatoire est déterministe ; par exemple, lors d'une réaction chimique, X est la dose de catalyseur et Y est la quantité d'un certain produit formée par cette réaction. Le **modèle linéaire** consiste alors à considérer que Y s'écrit sous la forme

$$Y = ax + b + \varepsilon, \quad (13.13)$$

où ε est une variable aléatoire **centrée** d'ordre deux, représentant une **erreur** d'approximation ou de mesure. Le problème est alors, au vu de résultats y_1, y_2, \dots, y_n d'expériences indépendantes faites respectivement aux « **niveaux** » x_1, x_2, \dots, x_n de valeurs de x , d'avoir une estimation des coefficients a et b .

Le modèle statistique associé est le suivant. On définit n **observations indépendantes** Y_1, Y_2, \dots, Y_n de Y faites aux **niveaux** x_1, x_2, \dots, x_n de valeurs de x ; autrement dit, les Y_i sont des variables aléatoires qui s'écrivent

$$Y_i = ax_i + b + \varepsilon_i, \quad (13.14)$$

où les variables aléatoires ε_i , $1 \leq i \leq n$, sont indépendantes **centrées** d'ordre deux, de même variance inconnue σ^2 . On cherche des estimateurs

de a et b en termes des Y_i . On peut utiliser une technique de moindres carrés en ramenant ce problème à un problème de type 1 : on considère que la variable aléatoire X_i est constante et égale à x_i et on minimise la somme des carrés des erreurs. Le théorème de Gauss-Markov donne une justification de l'emploi des estimateurs de moindres carrés trouvés par cette démarche. Dans la suite, sauf mention du contraire, $\langle \cdot, \cdot \rangle$ désigne le produit scalaire usuel de \mathbb{R}^n et $\| \cdot \|$ la norme associée.

Définition 13.16. Un estimateur linéaire du paramètre inconnu $(a, b) \in \mathbb{R}^2$ est une transformée linéaire du vecteur $\underline{Y} = (Y_1, Y_2, \dots, Y_n)$ du type $T_{u,v} = (\langle \underline{Y}, u \rangle, \langle \underline{Y}, v \rangle)$, où $u, v \in \mathbb{R}^n$. Un estimateur du paramètre inconnu $(a, b) \in \mathbb{R}^2$ est **sans biais** si sa moyenne est égale à (a, b) , pour tout $(a, b) \in \mathbb{R}^2$. Un estimateur linéaire $T_{u,v}$ du paramètre inconnu (a, b) est de **variance minimum** parmi tous les estimateurs linéaires sans biais de (a, b) s'il est solution du problème de minimisation

$$\min \left\{ \sigma_{\langle \underline{Y}, u \rangle}^2 + \sigma_{\langle \underline{Y}, v \rangle}^2 \mid u, v \in \mathbb{R}^n \right\}. \quad (13.15)$$

Théorème 13.17 (Théorème de Gauss-Markov). Soient n observations indépendantes Y_1, Y_2, \dots, Y_n , de Y faites aux niveaux x_1, x_2, \dots, x_n de valeurs de x ; autrement dit, supposons que les variables aléatoires Y_i s'écrivent

$$Y_i = ax_i + b + \varepsilon_i, \quad (13.16)$$

où les variables aléatoires ε_i , $1 \leq i \leq n$, sont indépendantes centrées d'ordre deux, et de même variance inconnue σ^2 .

L'estimateur linéaire de **variance minimum** parmi tous les estimateurs linéaires sans biais de (a, b) est l'estimateur (\hat{a}_n, \hat{b}_n) , où \hat{a}_n et \hat{b}_n sont les estimateurs de moindres carrés de a et b donnés par la proposition 13.15 ; ils s'écrivent

$$\begin{cases} \hat{a}_n = r_n \frac{S_{Y,n}}{S_{X,n}} \\ \hat{b}_n = \bar{Y}_n - \bar{x}_n \cdot r_n \frac{S_{Y,n}}{S_{X,n}}. \end{cases} \quad (13.17)$$

où on note

$$\begin{aligned} \bar{x}_n &= \frac{1}{n} \sum_{j=1}^n x_j, & \bar{Y}_n &= \frac{1}{n} \sum_{j=1}^n Y_j, \\ S_{X,n}^2 &= \frac{1}{n} \sum_{j=1}^n x_j^2 - (\bar{x}_n)^2, & S_{Y,n}^2 &= \frac{1}{n} \sum_{j=1}^n Y_j^2 - (\bar{Y}_n)^2, \\ r_n &= \frac{\frac{1}{n} \sum_{j=1}^n x_j Y_j - \bar{x}_n \bar{Y}_n}{S_{X,n} S_{Y,n}}. \end{aligned}$$

Démonstration. Notons $e = (1, 1, \dots, 1)$, $x = (x_1, \dots, x_n)$, et $\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$. On a alors

$$\underline{Y} = ax + be + \underline{\varepsilon}, \quad (13.18)$$

et donc, puisque la variable aléatoire $\underline{\varepsilon}$ est centrée,

$$E(\underline{Y}) = ax + be. \quad (13.19)$$

Un estimateur **linéaire** $T_{u,v} = (\langle \underline{Y}, u \rangle, \langle \underline{Y}, v \rangle)$ de (a, b) est alors **sans biais** si et seulement si, on a, pour tout $(a, b) \in \mathbb{R}^2$,

$$E(\langle \underline{Y}, u \rangle) = a \quad \text{et} \quad E(\langle \underline{Y}, v \rangle) = b,$$

soit encore, d'après (13.19), si et seulement si $u \in A$ et $v \in B$, où on note

$$\begin{cases} A = \{(x, u) = 1 \quad \text{et} \quad (e, u) = 0\} \\ B = \{(x, v) = 0 \quad \text{et} \quad (e, v) = 1\}. \end{cases}$$

Remarquons que

$$\sigma_{\langle \underline{Y}, u \rangle}^2 = \langle \Lambda_{\underline{Y}} u, u \rangle = \sigma^2 \|u\|^2,$$

l'estimateur **linéaire sans biais** $T_{\hat{u}, \hat{v}}$ est alors de **variance minimum** (quelque soit σ) parmi tous les estimateurs linéaires **sans biais** de (a, b) lorsque \hat{u} et \hat{v} sont solutions des deux problèmes de minimisation sous contrainte

$$\min \{ \|u\|^2 \mid u \in A \} \quad (13.20)$$

et

$$\min \{ \|v\|^2 \mid v \in B \}. \quad (13.21)$$

Étudions le premier problème d'extremum lié (13.20); aux multiplicateurs de Lagrange λ et μ , associons la fonction $\Phi_{\lambda, \mu}$ définie en tout $u \in \mathbb{R}^n$ par

$$\Phi_{\lambda, \mu}(u) = \|u\|^2 - \lambda(\langle x, u \rangle - 1) - \mu \langle e, u \rangle. \quad (13.22)$$

Un point \hat{u} est solution du problème d'extremum relatif lié associé à (13.20) s'il existe λ et μ tels que

$$\Phi'_{\lambda, \mu}(\hat{u}) = 0, \quad \langle x, \hat{u} \rangle = 1 \quad \text{et} \quad \langle e, \hat{u} \rangle = 0; \quad (13.23)$$

puisque

$$\Phi'_{\lambda, \mu}(u) = 2 \langle u, \cdot \rangle - \lambda \langle x, \cdot \rangle - \mu \langle e, \cdot \rangle = 2 \langle u, \cdot \rangle - \langle \lambda x + \mu e, \cdot \rangle,$$

on a $\Phi'_{\lambda, \mu}(\hat{u}) = 0$ si et seulement si $\hat{u} = \frac{1}{2}(\lambda x + \mu e)$, et \hat{u} est donc solution du problème d'extremum s'il existe λ et μ solutions du système

$$\begin{cases} \left\langle x, \frac{1}{2}(\lambda x + \mu e) \right\rangle = 1 \\ \left\langle e, \frac{1}{2}(\lambda x + \mu e) \right\rangle = 0, \end{cases}$$

encore équivalent au système (puisque $\langle e, e \rangle = n$)

$$\begin{cases} \lambda \|x\|^2 + \mu \langle x, e \rangle = 2 \\ \lambda \langle x, e \rangle + n\mu = 0. \end{cases}$$

Ce système a pour solutions

$$\lambda = \frac{2n}{(n\|x\|^2 - \langle x, e \rangle^2)} \quad \mu = \frac{-2\langle x, e \rangle}{(n\|x\|^2 - \langle x, e \rangle^2)}.$$

L'**unique** solution \hat{u} trouvée est donc

$$\hat{u} = \frac{1}{2}(\lambda x + \mu e) = \frac{1}{2} \left(\frac{2n}{(n\|x\|^2 - \langle x, e \rangle^2)} x - \frac{2\langle x, e \rangle}{(n\|x\|^2 - \langle x, e \rangle^2)} e \right),$$

soit

$$\hat{u} = \frac{1}{(n\|x\|^2 - \langle x, e \rangle^2)} (nx - \langle x, e \rangle e),$$

et \hat{u} correspond donc à un extremum **global**. Il faut encore montrer que ce point correspond à un **minimum**. Pour cela, décomposons tout $u \in A$ sous la forme $u = \hat{u} + \delta$. Puisque $\hat{u} \in A$, on a que

$$\langle x, \delta \rangle = 0 \quad \text{et} \quad \langle e, \delta \rangle = 0,$$

ce qui implique que

$$\langle \hat{u}, \delta \rangle = \left\langle \frac{1}{2}(\lambda x + \mu e), \delta \right\rangle = \frac{1}{2} [\lambda \langle x, \delta \rangle + \mu \langle e, \delta \rangle] = 0.$$

Ainsi, on a, pour tout $u \in A$,

$$\|u\|^2 = \|\hat{u}\|^2 + \|\delta\|^2 \geq \|\hat{u}\|^2,$$

ce qui démontre que \hat{u} est l'unique solution du problème d'extremum lié (13.20). On a alors

$$\langle \underline{Y}, \hat{u} \rangle = \frac{n\langle \underline{Y}, x \rangle - \langle x, e \rangle \langle \underline{Y}, e \rangle}{(n\|x\|^2 - \langle x, e \rangle^2)}. \quad (13.24)$$

ce qui n'est autre que l'estimateur \hat{a}_n , comme un calcul simple le montre.

Étudions le deuxième problème d'extremum lié (13.21); aux multiplicateurs de Lagrange λ et μ , associons la fonction $\Psi_{\lambda, \mu}$ définie en tout $v \in \mathbb{R}^n$ par

$$\Psi_{\lambda, \mu}(v) = \|v\|^2 - \lambda \langle x, v \rangle - \mu (\langle e, v \rangle - 1). \quad (13.25)$$

Un point \hat{v} est solution du problème d'extremum relatif lié associé à (13.21) s'il existe λ et μ tels que

$$\Psi'_{\lambda, \mu}(\hat{v}) = 0, \quad \langle x, \hat{v} \rangle = 0 \quad \text{et} \quad \langle e, \hat{v} \rangle = 1: \quad (13.26)$$

puisque

$$\Psi'_{\lambda, \mu}(v) = 2 \langle v, \cdot \rangle - \lambda \langle v, \cdot \rangle - \mu \langle e, \cdot \rangle = 2 \langle v, \cdot \rangle - \langle \lambda x + \mu e, \cdot \rangle.$$

on a $\Psi'_{\lambda, \mu}(\hat{v}) = 0$ si et seulement si $\hat{v} = \frac{1}{2}(\lambda x + \mu e)$, et \hat{v} est donc solution du problème d'extremum s'il existe λ et μ solutions du système

$$\begin{cases} \left\langle x, \frac{1}{2}(\lambda x + \mu e) \right\rangle = 0 \\ \left\langle e, \frac{1}{2}(\lambda x + \mu e) \right\rangle = 1, \end{cases}$$

encore équivalent au système (puisque $\langle e, e \rangle = n$)

$$\begin{cases} \lambda \|x\|^2 + \mu \langle x, e \rangle = 0 \\ \lambda \langle x, e \rangle + n\mu = 2. \end{cases}$$

Ce système a pour solutions

$$\lambda = \frac{-2 \langle x, e \rangle}{(n \|x\|^2 - \langle x, e \rangle^2)} \quad \mu = \frac{2 \|x\|^2}{(n \|x\|^2 - \langle x, e \rangle^2)}.$$

L'unique solution \hat{v} trouvée est donc

$$\hat{v} = \frac{1}{2}(\lambda x + \mu e) = \frac{1}{2} \left(\frac{-2 \langle x, e \rangle}{(n \|x\|^2 - \langle x, e \rangle^2)} x + \frac{2 \|x\|^2}{(n \|x\|^2 - \langle x, e \rangle^2)} e \right),$$

soit

$$\hat{v} = \frac{1}{(n \|x\|^2 - \langle x, e \rangle^2)} (\|x\|^2 e - \langle x, e \rangle x),$$

et \hat{v} correspond donc à un extremum **global**. Il faut encore montrer que ce point correspond à un **minimum**. Pour cela, décomposons tout $v \in B$ sous la forme $v = \hat{v} + \delta$. Puisque $\hat{v} \in B$, on a que

$$\langle x, \delta \rangle = 0 \quad \text{et} \quad \langle e, \delta \rangle = 0,$$

ce qui implique que

$$\langle \hat{v}, \delta \rangle = \left\langle \frac{1}{2}(\lambda x + \mu e), \delta \right\rangle = \frac{1}{2} [\lambda \langle x, \delta \rangle + \mu \langle e, \delta \rangle] = 0.$$

Ainsi, on a, pour tout $v \in B$,

$$\|v\|^2 = \|\hat{v}\|^2 + \|\delta\|^2 \geq \|\hat{v}\|^2,$$

ce qui démontre que \hat{v} est l'unique solution du problème d'extremum lié (13.21). On a alors

$$\langle \underline{\underline{Y}}, \hat{v} \rangle = \frac{\|x\|^2 \langle \underline{\underline{Y}}, e \rangle - \langle x, e \rangle \langle \underline{\underline{Y}}, x \rangle}{(n \|x\|^2 - \langle x, e \rangle^2)}, \quad (13.27)$$

ce qui n'est autre que l'estimateur \widehat{b}_n , comme le montre un calcul un peu long, mais simple. \square

13.4.2. Le modèle linéaire gaussien

Nous étudions plus avant le modèle linéaire (13.13) et sa formulation statistique (13.16). Pour obtenir des renseignements quantitatifs sur les estimateurs trouvés, il nous faut renforcer les hypothèses, à savoir, faire une hypothèse sur la loi des erreurs **indépendantes** ε_i : nous supposons que les ε_i sont toutes de même loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$, de variance σ^2 inconnue. On parle dans ce cas du **modèle linéaire gaussien**. La loi de la variable aléatoire \underline{Y} définie en (13.18) est alors, avec les notations de la section précédente, la loi $\mathcal{N}_{\mathbb{R}^n}(ax + be, \sigma^2 \mathbf{1}_n)$, où $\mathbf{1}_n$ désigne la matrice identité de \mathbb{R}^n . Sa densité (appelée **vraisemblance** par les statisticiens) est alors donnée, pour tout $y \in \mathbb{R}^n$, par

$$f_{\underline{Y}}(y) = \frac{1}{(2\pi c)^{n/2}} \exp\left(-\frac{\|y - m(a, b)\|^2}{2c}\right),$$

où on note $m(a, b) = ax + be$ et $c = \sigma^2$. Dans un premier temps, nous définissons et donnons des **estimateurs⁸ du maximum de vraisemblance** de a , b et c . Ces estimateurs sont obtenus de la manière suivante : pour tout y , on détermine les paramètres qui maximisent la vraisemblance en y ; dans ce problème, ils existent et sont uniques, on les note respectivement $\widehat{a}(y)$, $\widehat{b}(y)$ et $\widehat{c}(y)$. Les estimateurs du maximum de vraisemblance de a , b et c sont alors les variables aléatoires $\widehat{a}(\underline{Y})$, $\widehat{b}(\underline{Y})$ et $\widehat{c}(\underline{Y})$.

Ici, il est plus facile de maximiser en a , b et c ce que l'on appelle la **log-vraisemblance** en y , à savoir la quantité

$$\ln [f_{\underline{Y}}(y)] = -\frac{n}{2} \ln(2\pi c) - \frac{\|y - m(a, b)\|^2}{2c}.$$

Cherchons les points stationnaires. On a $\frac{\partial}{\partial a} \ln [f_{\underline{Y}}(y)] = 0$ si et seulement si $\frac{\partial}{\partial a} \|y - m(a, b)\|^2 = 0$ et on a $\frac{\partial}{\partial b} \ln [f_{\underline{Y}}(y)] = 0$ si et seulement si $\frac{\partial}{\partial b} \|y - m(a, b)\|^2 = 0$. Puisque

$$\|y - m(a, b)\|^2 = \|y\|^2 - 2\langle y, m(a, b) \rangle + \|m(a, b)\|^2,$$

on a

$$\frac{\partial}{\partial a} \|y - m(a, b)\|^2 = -2\langle y, \cdot \rangle x + 2\langle m(a, b), \cdot \rangle x = 2[(m(a, b) - y, \cdot)] x,$$

7. Le paramètre à estimer est la variance, et non l'écart-type.

8. La notion d'**estimateur du maximum de vraisemblance** est introduite de manière plus systématique au chapitre 14.

et

$$\frac{\partial}{\partial b} \|y - m(a, b)\|^2 = -2 \langle y, \cdot \rangle e + 2 \langle m(a, b), \cdot \rangle e = 2 [\langle m(a, b) - y, \cdot \rangle] e;$$

par ailleurs, on a

$$\frac{\partial}{\partial c} \ln [f_{\underline{Y}}(y)] = -\frac{n}{2c} + \frac{\|y - m(a, b)\|^2}{2c^2}.$$

Un point stationnaire $(\hat{a}(y), \hat{b}(y), \hat{c}(y))$ doit donc vérifier

$$\hat{a}(y)x + \hat{b}(y)e = y, \quad (13.28)$$

et

$$\hat{c}(y) = \frac{\|y - m(\hat{a}(y), \hat{b}(y))\|^2}{n}. \quad (13.29)$$

En faisant successivement le produit scalaire des deux membres de (13.28) par x et e , on trouve que $\hat{a}(y)$ et $\hat{b}(y)$ doivent être solution du système

$$\begin{cases} \hat{a}(y) \|x\|^2 + \hat{b}(y) \langle x, e \rangle = \langle y, x \rangle \\ \hat{a}(y) \langle x, e \rangle + \hat{b}(y) \|e\|^2 = \langle y, e \rangle, \end{cases} \quad (13.30)$$

système qui a pour unique solution (tenant compte de l'égalité $\|e\|^2 = n$) :

$$\hat{a}(y) = \frac{n \langle y, x \rangle - \langle x, e \rangle \langle y, e \rangle}{n \|x\|^2 - \langle x, e \rangle^2} \quad \hat{b}(y) = \frac{\|x\|^2 \langle y, e \rangle - \langle x, e \rangle \langle y, x \rangle}{n \|x\|^2 - \langle x, e \rangle^2}. \quad (13.31)$$

Il en résulte, en comparant aux égalités (13.24) et (13.27) que $\hat{a}(\underline{Y}) = \hat{a}_n$ et $\hat{b}(\underline{Y}) = \hat{b}_n$.

Autrement dit, **pour le modèle linéaire gaussien, les estimateurs du maximum de vraisemblance de a et b sont aussi les estimateurs de moindres carrés et ceux de variance minimum parmi tous les estimateurs linéaires sans biais. L'estimateur du maximum de vraisemblance \hat{c}_n de la variance est alors**

$$\hat{c}_n = \frac{\|\underline{Y} - (\hat{a}_n x + \hat{b}_n e)\|^2}{n}. \quad (13.32)$$

Théorème 13.18. *La variable aléatoire $(\hat{a}_n, \hat{b}_n, \underline{Y} - (\hat{a}_n x + \hat{b}_n e))$ à valeurs dans \mathbb{R}^{n+2} est gaussienne. L'estimateur (\hat{a}_n, \hat{b}_n) est une variable aléatoire gaussienne indépendante de \hat{c}_n . Les moyennes et variances de \hat{a}_n et \hat{b}_n sont données par*

$$\begin{cases} E(\hat{a}_n) = a, & \sigma_{\hat{a}_n}^2 = \frac{n}{n \|x\|^2 - \langle x, e \rangle^2} \sigma^2 = \frac{\sigma^2}{n s_{x,n}^2} \\ E(\hat{b}_n) = b, & \sigma_{\hat{b}_n}^2 = \frac{\|x\|^2}{n \|x\|^2 - \langle x, e \rangle^2} \sigma^2 = \frac{\|x\|^2}{n^2 s_{x,n}^2} \sigma^2. \end{cases} \quad (13.33)$$

La variable aléatoire $\frac{\widehat{c}_n}{\sigma^2}$ suit la loi du chi-deux χ_{n-2}^2 . En conséquence, on a

$$E(\widehat{c}_n) = \frac{n-2}{n} \sigma^2 \quad \sigma_{\widehat{c}_n}^2 = \frac{2(n-2)}{n^2} \sigma^4. \quad (13.34)$$

Démonstration. La variable aléatoire $(\widehat{a}_n, \widehat{b}_n, \underline{Y} - (\widehat{a}_n x + \widehat{b}_n e))$ est une transformée linéaire de la variable aléatoire gaussienne \underline{Y} , comme le montre un examen des égalités (13.24) et (13.27); elle est donc elle-même gaussienne, et il en est bien sûr de même de la variable aléatoire $(\widehat{a}_n, \widehat{b}_n)$. Les moyennes et variances de \widehat{a}_n et \widehat{b}_n se calculent facilement à partir des égalités (13.24) et (13.27). Notons V le sous-espace vectoriel de \mathbb{R}^n engendré par les vecteurs x et e et déterminons la projection orthogonale $y_V = \widehat{\alpha}x + \widehat{\beta}e$ d'un vecteur y quelconque de \mathbb{R}^n . Elle est caractérisée par la relation d'orthogonalité

$$\forall (u, v) \in \mathbb{R}^2 \quad \langle y - (\widehat{\alpha}x + \widehat{\beta}e), ux + ve \rangle = 0,$$

relation équivalente à

$$\forall (u, v) \in \mathbb{R}^2 \quad u[\langle y, x \rangle - \widehat{\alpha}\|x\|^2 - \widehat{\beta}\langle x, e \rangle] + v[\langle y, e \rangle - \widehat{\alpha}\langle x, e \rangle - \widehat{\beta}\|e\|^2] = 0,$$

relation encore équivalente au système

$$\begin{cases} \widehat{\alpha}\|x\|^2 + \widehat{\beta}\langle x, e \rangle = \langle y, x \rangle \\ \widehat{\alpha}\langle x, e \rangle + \widehat{\beta}\|e\|^2 = \langle y, e \rangle. \end{cases}$$

Ainsi, $\widehat{\alpha}$ et $\widehat{\beta}$ sont solutions du système (13.30), ce qui prouve que la variable aléatoire $\underline{Y}_V = \Pi_V \underline{Y}$ (où Π_V est le projecteur orthogonal sur V) vérifie

$$\underline{Y}_V = \widehat{a}_n x + \widehat{b}_n e.$$

Il en résulte que l'on a (ponctuellement) l'égalité : $\underline{Y} - (\widehat{a}_n x + \widehat{b}_n e) = \underline{Y}_{V^\perp}$, projection orthogonale de \underline{Y} sur V^\perp . Puisque \underline{Y} est de loi gaussienne $\mathcal{N}_{\mathbb{R}^n}(ax + be, \sigma^2 \mathbf{1}_n)$, les variables aléatoires \underline{Y}_V et \underline{Y}_{V^\perp} sont indépendantes (d'après la proposition 13.9) et donc aussi les variables aléatoires \underline{Y}_{V^\perp} et $(\widehat{a}_n, \widehat{b}_n)$, puisque cette dernière est fonction mesurable de \underline{Y}_V . Ceci prouve l'indépendance de $(\widehat{a}_n, \widehat{b}_n)$ et \widehat{c}_n .

Par ailleurs, la dimension de V^\perp est $n-2$ et on a

$$\Pi_{V^\perp}(ax + be) = 0 \quad \text{et} \quad \Lambda_{\underline{Y}_{V^\perp}} = \Pi_{V^\perp} \Lambda_{\underline{Y}} \Pi_{V^\perp} = \sigma^2 \Pi_{V^\perp};$$

la loi de \underline{Y}_{V^\perp} est donc la loi $\mathcal{N}_{\mathbb{R}^n}(0, \sigma^2 \Pi_{V^\perp})$, ce qui entraîne que la loi de $\|\underline{Y}_{V^\perp}/\sigma\|^2$ est la loi du chi-deux χ_{n-2}^2 ; c'est aussi la loi de $n \frac{\widehat{c}_n}{\sigma^2}$, puisque ces deux variables aléatoires sont égales. On a alors

$$E\left(n \frac{\widehat{c}_n}{\sigma^2}\right) = n-2 \quad \sigma^2 \left(n \frac{\widehat{c}_n}{\sigma^2}\right) = 2(n-2).$$

ce qui donne immédiatement les égalités (13.34). \square

Le théorème 13.18 permet alors d'obtenir des **tests** et des **intervalles de confiance** pour les différents paramètres de régression, puis avec le modèle estimé, de faire de la prédiction.

Tests d'hypothèse

Par exemple, si on veut tester l'hypothèse H_0 que la vraie valeur de la variance de l'erreur ε est σ^2 , on se fixe un **seuil** α et on détermine, à l'aide d'une table ou d'un logiciel statistique, la valeur c_α telle que $\chi_{n-2}^2([c_\alpha, +\infty[) = \alpha$. On **rejette l'hypothèse** H_0 si $n \frac{\widehat{c}_n}{\sigma^2} > c_\alpha$, soit encore si $\widehat{c}_n > \frac{\sigma^2}{n} c_\alpha$. Compte tenu des égalités, (13.24), (13.27) et (13.32), la **zone de rejet** de l'hypothèse H_0 est donc la partie de \mathbb{R}^n :

$$\left\{ \left\| y - \left(\frac{n \langle y, x \rangle - \langle x, e \rangle \langle y, e \rangle}{n \|x\|^2 - \langle x, e \rangle^2} x + \frac{\|x\|^2 \langle y, e \rangle - \langle x, e \rangle \langle y, x \rangle}{n \|x\|^2 - \langle x, e \rangle^2} e \right) \right\|^2 > \sigma^2 c_\alpha \right\}.$$

De même, on peut tester une hypothèse sur le paramètre b . Pour cela, on introduit la variable aléatoire $\widehat{B}_n(b)$, qui est la variable aléatoire centrée réduite associée à l'estimateur \widehat{b}_n , mais dans laquelle σ^2 est remplacée par son estimation **sans biais** $\frac{n}{n-2} \widehat{c}_n$; tenant compte des égalités (13.33), elle est définie par

$$\widehat{B}_n(b) = \left[\frac{\|x\|^2}{n \|x\|^2 - \langle x, e \rangle^2} \frac{n}{n-2} \widehat{c}_n \right]^{-\frac{1}{2}} (\widehat{b}_n - b),$$

ou encore

$$\widehat{B}_n(b) = \left[\frac{\|x\|^2}{n^2 s_{x,n}^2} \frac{n}{n-2} \widehat{c}_n \right]^{-\frac{1}{2}} (\widehat{b}_n - b). \quad (13.35)$$

On rappelle alors que, si X et Y sont deux variables aléatoires indépendantes de lois respectives la loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$ et la loi du chi-deux χ_n^2 , la loi de la variable aléatoire $\frac{\sqrt{n}X}{\sqrt{Y}}$ est la loi de **Student** à n degrés de liberté (cf. ex. 4. chap. 9). En conséquence, il résulte facilement du théorème 13.18 que la variable aléatoire $\widehat{B}_n(b)$ suit la loi de Student t_{n-2} (nous laissons ce calcul à titre d'exercice).

Ainsi, pour tester l'hypothèse H_0 que la vraie valeur du paramètre b est b_0 contre l'hypothèse H_1 que $b > b_0$, on se fixe un **seuil** α et on détermine, à l'aide d'une table ou d'un logiciel statistique, la valeur $b_{1-\alpha}$ telle que $t_{n-2}(-\infty, b_{1-\alpha}] = 1 - \alpha$. On **rejette l'hypothèse** H_0 si $\widehat{B}_n(b_0) > b_{1-\alpha}$. Compte tenu des égalités, (13.24), (13.27) et (13.32), la **zone de rejet** de l'hypothèse H_0 contre l'hypothèse H_1 est donc la partie de \mathbb{R}^n :

$$\left\{ \left[\frac{\|x\|^2}{n\|x\|^2 - \langle x, e \rangle^2} \frac{\|y - (\widehat{a}(y)x + \widehat{b}(y)e)\|^2}{n-2} \right]^{-\frac{1}{2}} (\widehat{b}(y) - b_0) > b_{1-\alpha} \right\}.$$

Si on veut tester la même hypothèse H_0 contre l'hypothèse H_2 que $b \neq b_0$, on choisit $p, 0 < p < 1$, et on détermine, à l'aide d'une table ou d'un logiciel statistique, les valeurs $b_{1-\alpha p}$ telle que $t_{n-2}([-\infty, b_{1-\alpha p}]) = 1 - \alpha p$ et $b_{1-\alpha(1-p)}$ telle que $t_{n-2}([-\infty, b_{1-\alpha(1-p)}]) = 1 - \alpha(1-p)$. Puisque la loi t_{n-2} est symétrique, on a alors

$$\begin{aligned} t_{n-2}([-\infty, -b_{1-\alpha(1-p)}]) &= t_{n-2}([b_{1-\alpha(1-p)}, +\infty[) \\ &= 1 - t_{n-2}([-\infty, b_{1-\alpha(1-p)}]) = \alpha(1-p). \end{aligned}$$

On a alors

$$t_{n-2}([-\infty, -b_{1-\alpha(1-p)}]) \cup [b_{1-\alpha p}, +\infty[) = \alpha(1-p) + \alpha p = \alpha.$$

On rejette l'hypothèse H_0 si $\widehat{B}_n(b_0) > b_{1-\alpha p}$ ou si $\widehat{B}_n(b_0) < -b_{1-\alpha(1-p)}$. On écrirait de même que ci-dessus la **zone de rejet** pour ce test d'hypothèse de H_0 contre H_2 , mais cela n'apporte rien de plus.

On peut évidemment tester, de manière analogue, une hypothèse sur la vraie valeur du paramètre a .

Intervalle de confiance

Donnons juste en détail l'exemple de construction d'un intervalle de confiance pour b au niveau β . On détermine, à l'aide d'une table ou d'un logiciel statistique, la valeur $b_{1-\frac{\beta}{2}}$ telle que $t_{n-2}([-\infty, b_{1-\frac{\beta}{2}}]) = 1 - \frac{\beta}{2}$. On a alors

$$t_{n-2}([-\infty, -b_{1-\frac{\beta}{2}}]) \cup [b_{1-\frac{\beta}{2}}, +\infty[) = \beta.$$

Il résulte encore du théorème 13.18 que la variable aléatoire $\widehat{B}_n(b)$ suit la loi de Student t_{n-2} ; on a alors

$$P\left[|\widehat{B}_n(b)| \leq b_{1-\frac{\beta}{2}}\right] = 1 - \beta, \quad (13.36)$$

égalité qui donne, au niveau β , l'intervalle de confiance $[I, S]$, où

$$I = \widehat{b}_n - b_{1-\frac{\beta}{2}} \left[\frac{\|x\|^2}{n\|x\|^2 - \langle x, e \rangle^2} \frac{n}{n-2} \widehat{c}_n \right]^{\frac{1}{2}},$$

et

$$S = \widehat{b}_n + b_{1-\frac{\beta}{2}} \left[\frac{\|x\|^2}{n\|x\|^2 - \langle x, e \rangle^2} \frac{n}{n-2} \widehat{c}_n \right]^{\frac{1}{2}}.$$

En s'appuyant toujours sur le théorème 13.18, on peut aussi construire des intervalles de confiance pour les paramètres a et σ^2 . Par exemple, pour a ,

on introduit la variable aléatoire

$$\widehat{A}_n(a) = \left[\frac{1}{ns_{x,n}^2} \frac{n}{n-2} \widehat{c}_n \right]^{-\frac{1}{2}} (\widehat{a}_n - a), \quad (13.37)$$

et on détermine la valeur $a_{1-\frac{\alpha}{2}}$ telle que $t_{n-2}([-\infty, a_{1-\frac{\alpha}{2}}]) = 1 - \frac{\alpha}{2}$. On a alors de même

$$P \left[|\widehat{A}_n(a)| \leq a_{1-\frac{\alpha}{2}} \right] = 1 - \alpha, \quad (13.38)$$

et on termine de manière analogue.

Prédiction

Le modèle théorique étant toujours décrit par l'égalité (13.13), il s'agit maintenant de faire une prédiction sur le résultat d'une expérience qui serait faite au niveau \widetilde{x} de valeur de x . On travaille pour cela avec le **modèle estimé** à partir d'un échantillon de longueur n , et défini par l'égalité

$$Y_{n+1} = \widehat{a}_n \widetilde{x} + \widehat{b}_n + \varepsilon_{n+1}, \quad (13.39)$$

où les variables aléatoires $\widehat{a}_n, \widehat{b}_n, \varepsilon_n$ sont celles introduites précédemment. Il s'agit d'obtenir un **intervalle de confiance** pour Y_{n+1} .

On remarque d'abord que les variables aléatoires $\widehat{a}_n \widetilde{x} + \widehat{b}_n$ et ε_{n+1} sont gaussiennes et indépendantes; la variable aléatoire Y_{n+1} est donc aussi gaussienne. On évalue maintenant sa moyenne et sa variance. Puisque les estimateurs \widehat{a}_n et \widehat{b}_n sont sans biais et que ε_n est centrée, on a

$$E(Y_{n+1}) = a\widetilde{x} + b. \quad (13.40)$$

Par ailleurs, il résulte des égalités (13.17) que

$$\widehat{b}_n = \overline{Y}_n - \overline{x}_n \widehat{a}_n;$$

on a donc

$$\widehat{a}_n \widetilde{x} + \widehat{b}_n = \widehat{a}_n (\widetilde{x} - \overline{x}_n) + \overline{Y}_n.$$

Les variables aléatoires \widehat{a}_n et \overline{Y}_n ne sont pas indépendantes, mais on a la majoration suivante de la variance de $\widehat{a}_n \widetilde{x} + \widehat{b}_n$:

$$\sigma_{\widehat{a}_n \widetilde{x} + \widehat{b}_n}^2 \leq 2(\sigma_{\widehat{a}_n (\widetilde{x} - \overline{x}_n)}^2 + \sigma_{\overline{Y}_n}^2) = 2((\widetilde{x} - \overline{x}_n)^2 \sigma_{\widehat{a}_n}^2 + \sigma_{\overline{Y}_n}^2);$$

tenant compte de l'expression de la variance de \widehat{a}_n (cf. 13.33), et de l'égalité $\sigma_{\overline{Y}_n}^2 = \frac{\sigma^2}{n}$, on a alors la majoration

$$\sigma_{\widehat{a}_n \widetilde{x} + \widehat{b}_n}^2 \leq 2\sigma^2 \left(\frac{(\widetilde{x} - \overline{x}_n)^2}{ns_{x,n}^2} + \frac{1}{n} \right).$$

Les variables aléatoires $\hat{a}_n \tilde{x} + \hat{b}_n$ et ε_{n+1} étant indépendantes, on en déduit l'inégalité

$$\sigma_{\hat{Y}_{n+1}}^2 \leq \sigma^2 \left(1 + \frac{2(\tilde{x} - \bar{x}_n)^2}{n s_{x,n}^2} + \frac{2}{n} \right). \quad (13.41)$$

Notons $\overset{\circ}{Y}_{n+1}$ la variable aléatoire gaussienne centrée $Y_{n+1} - (a\tilde{x} + b)$. Il résulte du théorème 13.18 que les variables aléatoires $(\hat{a}_n \tilde{x} + \hat{b}_n, \varepsilon_{n+1})$ et $n\hat{c}_n/\sigma^2$, et donc aussi les variables aléatoires $\overset{\circ}{Y}_{n+1}$ et $n\hat{c}_n/\sigma^2$, sont indépendantes. Puisque $n\hat{c}_n/\sigma^2$ suit la loi du chi-deux χ_{n-2}^2 , la variable aléatoire

$$Z_n = \sqrt{n-2} \frac{\overset{\circ}{Y}_{n+1}}{\sigma_{Y_{n+1}}} \left[\frac{n\hat{c}_n}{\sigma^2} \right]^{-\frac{1}{2}}$$

suit la loi de Student t_{n-2} .

On peut alors construire de la manière suivante un intervalle de confiance pour Y_{n+1} à un niveau inférieur ou égal à $\alpha + \beta + \gamma$ (avec $0 < \alpha + \beta + \gamma < 1$). Comme précédemment, on détermine, à l'aide d'une table ou d'un logiciel statistique, la valeur $z_{1-\frac{\gamma}{2}}$ telle que

$$t_{n-2}([\!-\infty, -z_{1-\frac{\gamma}{2}}] \cup [z_{1-\frac{\gamma}{2}}, +\infty[) = \gamma.$$

On a alors

$$\mathbb{P}\left[|Z_n| \leq z_{1-\frac{\gamma}{2}}\right] = 1 - \gamma.$$

Par définition de Z_n , on a l'équivalence

$$|Z_n| \leq z_{1-\frac{\gamma}{2}} \iff |\overset{\circ}{Y}_{n+1}| \leq \frac{\sigma_{Y_{n+1}}}{\sigma} \left[\frac{n}{n-2} \hat{c}_n \right]^{\frac{1}{2}} z_{1-\frac{\gamma}{2}};$$

en tenant compte de la majoration (13.41), on a alors l'implication

$$|Z_n| \leq z_{1-\frac{\gamma}{2}} \Rightarrow |\overset{\circ}{Y}_{n+1}| \leq z_{1-\frac{\gamma}{2}} \left[\left(1 + \frac{2(\tilde{x} - \bar{x}_n)^2}{n s_{x,n}^2} + \frac{2}{n} \right) \left(\frac{n}{n-2} \hat{c}_n \right) \right]^{\frac{1}{2}}.$$

On a donc

$$\mathbb{P} \left[|\overset{\circ}{Y}_{n+1}| \leq z_{1-\frac{\gamma}{2}} \left[\left(1 + \frac{2(\tilde{x} - \bar{x}_n)^2}{n s_{x,n}^2} + \frac{2}{n} \right) \left(\frac{n}{n-2} \hat{c}_n \right) \right]^{\frac{1}{2}} \right] \geq 1 - \gamma. \quad (13.42)$$

Ainsi, en posant

$$I_n = a\tilde{x} + b - z_{1-\frac{\gamma}{2}} \left[\left(1 + \frac{2(\tilde{x} - \bar{x}_n)^2}{n s_{x,n}^2} + \frac{2}{n} \right) \left(\frac{n}{n-2} \hat{c}_n \right) \right]^{\frac{1}{2}},$$

et

$$S_n = a\tilde{x} + b + z_{1-\frac{\gamma}{2}} \left[\left(1 + \frac{2(\tilde{x} - \bar{x}_n)^2}{ns_{x,n}^2} + \frac{2}{n} \right) \left(\frac{n}{n-2} \hat{c}_n \right) \right]^{\frac{1}{2}}.$$

on a

$$P(Y_{n+1} \in [I_n, S_n]) \geq 1 - \gamma,$$

mais ceci ne donne pas un intervalle de confiance pour Y_{n+1} , puisque l'intervalle $[I_n, S_n]$ dépend des paramètres inconnus a et b . Pour déterminer un tel intervalle, il faut encore faire intervenir les valeurs estimées de ces paramètres. Pour alléger ce texte, nous ne donnons que le principe de la méthode de construction.

Après avoir déterminé, par la méthode ci-dessus, des intervalles de confiance pour a et b aux niveaux respectifs α et β (à l'aide des égalités (13.38) et (13.36)), et compte tenu de (13.42), on est dans la situation suivante : on a déterminé des variables aléatoires $u_n(\alpha)$, $v_n(\beta)$ et $w_n(\gamma)$ telles que l'on ait simultanément les inégalités

$$P(|a - \hat{a}_n| \leq u_n(\alpha)) \geq 1 - \alpha, \quad P(|b - \hat{b}_n| \leq v_n(\beta)) \geq 1 - \beta, \\ P(|Y_{n+1} - (a\tilde{x} + b)| \leq w_n(\gamma)) \geq 1 - \gamma.$$

où on a posé

$$\begin{cases} u_n(\alpha) = a_{1-\frac{\alpha}{2}} \left[\frac{1}{ns_{x,n}^2} \frac{n}{n-2} \hat{c}_n \right]^{\frac{1}{2}} \\ v_n(\beta) = b_{1-\frac{\beta}{2}} \left[\frac{\|x\|^2}{n^2 s_{x,n}^2} \frac{n}{n-2} \hat{c}_n \right]^{\frac{1}{2}} \\ w_n(\gamma) = z_{1-\frac{\gamma}{2}} \left[\left(1 + \frac{2(\tilde{x} - \bar{x}_n)^2}{ns_{x,n}^2} + \frac{2}{n} \right) \left(\frac{n}{n-2} \hat{c}_n \right) \right]^{\frac{1}{2}}. \end{cases}$$

Or, si A, B, C sont des événements vérifiant les inégalités

$$P(A) \geq 1 - \alpha, \quad P(B) \geq 1 - \beta, \quad P(C) \geq 1 - \gamma,$$

on a

$$P(A^c \cup B^c \cup C^c) \leq P(A^c) + P(B^c) + P(C^c) \leq \alpha + \beta + \gamma,$$

et donc

$$P(A \cap B \cap C) \geq 1 - (\alpha + \beta + \gamma).$$

Ainsi, par l'inégalité triangulaire, on a, avec une probabilité supérieure ou égale à $1 - (\alpha + \beta + \gamma)$,

$$|Y_{n+1} - (\hat{a}_n \tilde{x} + \hat{b}_n)| \leq |Y_{n+1} - (a\tilde{x} + b)| + |a - \hat{a}_n| |\tilde{x}| + |b - \hat{b}_n| \\ \leq w_n(\gamma) + u_n(\alpha) |\tilde{x}| + v_n(\beta),$$

ce qui permet de dire que l'intervalle $[\hat{I}_n, \hat{S}_n]$ est un intervalle de confiance pour Y_{n+1} à un niveau inférieur ou égal à $\alpha + \beta + \gamma$, où on a posé

$$\hat{I}_n = \hat{a}_n \tilde{x} + \hat{b}_n - (u_n(\alpha)|\tilde{x}| + v_n(\beta) + w_n(\gamma)),$$

et

$$\hat{S}_n = \hat{a}_n \tilde{x} + \hat{b}_n + (u_n(\alpha)|\tilde{x}| + v_n(\beta) + w_n(\gamma)).$$

Donnons quelques valeurs de la fonction inverse de la fonction de répartition d'une variable aléatoire X de loi de **Student** t_n pour différentes valeurs de n : pour n et y fixés, la table donne la valeur x telle que $P(X \leq x) = y$.

n	y	0.75	0.90	0.95	0.990	0.995
5		0,727	1,476	2,015	3,365	4,032
10		0,700	1,372	1,812	2,764	3,169
15		0,691	1,341	1,753	2,602	2,947
20		0,687	1,325	1,725	2,528	2,845

Pour conclure, signalons que, par souci de simplification, nous n'avons traité que des modèles linéaires à un seul facteur x . Un modèle linéaire à k facteurs x_j , est un modèle théorique de la forme

$$Y = \sum_{j=1}^k a_j x_j + b + \varepsilon, \quad (13.43)$$

où ε est une variable aléatoire **centrée** d'ordre deux, représentant un **erreur** d'approximation ou de mesure. On peut trouver l'étude statistique générale des modèles linéaires (et une bibliographie sur ce sujet) dans, par exemple, le premier chapitre du livre de A. Antoniadis⁹; en fait, ce livre traite essentiellement des modèles non linéaires.

Exercices

Sauf mention du contraire, toutes les variables aléatoires sont définies sur un même espace probabilisé (Ω, \mathcal{A}, P) .

Exercice 13.1. Une mesure non gaussienne dont les marginales le sont. Soit X une variable aléatoire réelle de loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$. On considère les variables aléatoires à valeurs dans \mathbb{R}^2 , $Y = (X, -X)$ et $Z = (X, X)$. On étudie la probabilité μ sur \mathbb{R}^2 définie par $\mu = (P_Y + P_Z)/2$. On note Π_1 et Π_2 les applications coordonnées définies par $\Pi_1(x, y) = x$ et $\Pi_2(x, y) = y$, pour tout $(x, y) \in \mathbb{R}^2$; enfin, on note $\mu_1 = \Pi_1(\mu)$ et $\mu_2 = \Pi_2(\mu)$ les **marginales** de μ , c'est à dire les mesures images de μ par Π_1 et Π_2 .

9. Antoniadis A., Berruyer J., Carmona R. (1992), *Régression non linéaire et applications*, Economica, collection Économie et statistiques avancées. Paris.

Démontrer que μ_1 et μ_2 sont égales à la mesure gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$. Calculer la transformée de Fourier de μ et en déduire que μ n'est pas gaussienne.

Solution. Pour tout $f \in \mathcal{C}_{\mathcal{X}}^+(\mathbb{R})$, il résulte du théorème de transfert, de la définition de μ et du fait que $\Pi_1(Y) = \Pi_1(Z) = X$ que l'on a

$$\int_{\mathbb{R}} f d\mu_1 = \int_{\mathbb{R}^2} f \circ \Pi_1 d\mu = \frac{1}{2} \int_{\Omega} [f \circ \Pi_1(Y) + f \circ \Pi_1(Z)] dP = \int_{\Omega} f(X) dP,$$

et donc

$$\int_{\mathbb{R}} f d\mu_1 = \int_{\mathbb{P}} f dP_X,$$

ce qui prouve que $\mu_1 = P_X = \mathcal{N}_{\mathbb{R}}(0, 1)$. On a de même

$$\int_{\mathbb{R}} f d\mu_2 = \int_{\mathbb{R}^2} f \circ \Pi_2 d\mu = \frac{1}{2} \int_{\Omega} [f \circ \Pi_2(Y) + f \circ \Pi_2(Z)] dP,$$

soit

$$\int_{\mathbb{R}} f d\mu_2 = \frac{1}{2} \int_{\Omega} [f(-X) + f(X)] dP,$$

et donc, puisque P_X est symétrique,

$$\int_{\mathbb{R}} f d\mu_2 = \int_{\mathbb{R}} f dP_X,$$

ce qui prouve que $\mu_2 = P_X = \mathcal{N}_{\mathbb{R}}(0, 1)$.

Par ailleurs, la transformée de Fourier $\widehat{\mu}$ de μ est donnée par

$$\widehat{\mu} = \frac{1}{2}(\widehat{P_Y} + \widehat{P_Z}) = \frac{1}{2}(\varphi_Y + \varphi_Z),$$

ce qui donne, pour tout $u \in \mathbb{R}^2$,

$$\widehat{\mu}(u) = \frac{1}{2} \left[\exp\left(-\frac{(u_1 - u_2)^2}{2}\right) + \exp\left(-\frac{(u_1 + u_2)^2}{2}\right) \right],$$

soit, après réduction,

$$\widehat{\mu}(u) = \frac{1}{2} \exp\left(-\frac{(u_1^2 + u_2^2)}{2}\right) [\exp(u_1 u_2) + \exp(-u_1 u_2)],$$

ou encore

$$\widehat{\mu}(u) = \exp\left(-\frac{(u_1^2 + u_2^2)}{2}\right) \operatorname{ch}(u_1 u_2).$$

Ainsi, la mesure μ n'est pas gaussienne.

Remarque. Par injectivité de la transformée de Fourier, on retrouve que μ_1 et μ_2 sont gaussiennes, puisque que l'on a

$$\widehat{\mu}_1(u_1) = \widehat{\mu}(u_1, 0) = \exp\left(-\frac{u_1^2}{2}\right) \quad \text{et} \quad \widehat{\mu}_2(u_2) = \widehat{\mu}(0, u_2) = \exp\left(-\frac{u_2^2}{2}\right).$$

Exercice 13.2. Une transformée non linéaire de variable aléatoire gaussienne peut être gaussienne. Soient X , Y et Z trois variables aléatoires réelles indépendantes, gaussiennes de loi $\mathcal{N}_{\mathbb{R}}(0, 1)$. On définit la variable aléatoire U par

$$U = \frac{X + YZ}{\sqrt{1 + Z^2}}.$$

Déterminer une loi conditionnelle $P_U^{Z=z}$ de U sachant Z ; en déduire que U et Z sont indépendantes et déterminer la loi de U . Conclure.

Solution. Une loi conditionnelle $P_U^{Z=z}$ de U sachant Z est donnée par le noyau défini pour P_Z -presque tout réel z par

$$P_U^{Z=z} = P_{\frac{X+YZ}{\sqrt{1+z^2}}}^{Z=z},$$

soit, par indépendance des variables aléatoires $X + Yz$ et Z ,

$$P_U^{Z=z} = P_{\frac{X+YZ}{\sqrt{1+z^2}}} = \mathcal{N}_{\mathbb{R}}\left(0, \frac{\sigma^2_{X+YZ}}{\sqrt{1+z^2}}\right).$$

Puisque X et Y sont indépendantes, on a

$$\frac{\sigma^2_{X+YZ}}{\sqrt{1+z^2}} = \frac{1}{1+z^2}(\sigma_X^2 + z^2\sigma_Y^2) = 1;$$

il en résulte que l'on a, pour P_Z -presque tout réel z ,

$$P_U^{Z=z} = \mathcal{N}_{\mathbb{R}}(0, 1),$$

quantité indépendante de z , ceci démontre que **les variables aléatoires U et Z sont indépendantes** et que l'on a $P_U^{Z=z} = P_U = \mathcal{N}_{\mathbb{R}}(0, 1)$, c'est à dire que **la loi de U est la loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$** .

En conclusion, **la variable aléatoire U , transformée non linéaire de la variable aléatoire gaussienne (X, Y, Z) à valeurs dans \mathbb{R}^3 est gaussienne**. On peut même remarquer que **la variable aléatoire (U, Z) est gaussienne** à valeurs dans \mathbb{R}^2 , puisque U et Z sont indépendantes et de loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$.

Exercice 13.3. Caractérisation des lois gaussiennes sur \mathbb{R} . Soient X et Y deux variables aléatoires réelles, admettant un moment d'ordre deux, **indépendantes** et de même loi μ telle que

$$\int_{\mathbb{R}} x d\mu(x) = 0 \quad \text{et} \quad \int_{\mathbb{R}} x^2 d\mu(x) = \sigma^2.$$

Démontrer que si μ est la loi $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$, la variable aléatoire $\frac{X+Y}{\sqrt{2}}$ a pour loi $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$. Démontrer qu'inversement, si la variable aléatoire $(X + Y)/\sqrt{2}$ a pour loi μ , alors μ est la loi $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$. Pour la réciproque, on supposera que $\sigma = 1$. Démontrer que l'on a, pour tout réel t et tout entier n ,

$$\widehat{\mu}(t) = \left[\widehat{\mu}\left(\frac{t}{2^n}\right) \right]^{4^n},$$

puis que $\widehat{\mu}(t) \neq 0$. Poser alors, pour tout $t \neq 0$,

$$h(t) = \frac{\ln |\widehat{\mu}(t)|}{t^2},$$

et démontrer que la fonction h est constante. En déduire $|\widehat{\mu}(t)|$ puis $\widehat{\mu}(t)$.

Solution. Si μ est la loi $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$, X et Y étant indépendantes de loi $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$, la variable aléatoire (X, Y) est gaussienne et donc aussi la variable aléatoire $\frac{X+Y}{\sqrt{2}}$. Puisqu'alors on a

$$E\left(\frac{X+Y}{\sqrt{2}}\right) = 0 \quad \text{et} \quad \sigma_{\frac{X+Y}{\sqrt{2}}}^2 = \frac{1}{2}(\sigma_X^2 + \sigma_Y^2) = \sigma^2,$$

on a bien que $P_{\frac{X+Y}{\sqrt{2}}} = \mu$.

Inversement, on peut supposer, sans perte de généralité, que $\sigma = 1$. Par indépendance des variables aléatoires X et Y , la fonction caractéristique $\varphi_{\frac{X+Y}{\sqrt{2}}}$ vérifie, pour tout réel t ,

$$\varphi_{\frac{X+Y}{\sqrt{2}}}(t) = \varphi_X\left(\frac{t}{\sqrt{2}}\right)\varphi_Y\left(\frac{t}{\sqrt{2}}\right) = \left[\widehat{\mu}\left(\frac{t}{\sqrt{2}}\right)\right]^2;$$

l'hypothèse que la variable aléatoire $\frac{X+Y}{\sqrt{2}}$ a pour loi μ se traduit alors par la relation

$$\forall t \in \mathbb{R} \quad \widehat{\mu}(t) = \left[\widehat{\mu}\left(\frac{t}{\sqrt{2}}\right)\right]^2,$$

ce qui implique que, pour tout $t \in \mathbb{R}$,

$$\widehat{\mu}(t) = \left[\widehat{\mu}\left(\frac{t}{2}\right)\right]^4.$$

Il en résulte que, par itération, on a, pour tout réel t et tout entier n ,

$$\widehat{\mu}(t) = \left[\widehat{\mu}\left(\frac{t}{2^n}\right)\right]^{4^n}. \quad (13.44)$$

Supposons alors qu'il existe t_0 tel que $\widehat{\mu}(t_0) = 0$. Il résulte de (13.44) que, pour tout entier n , on a $\widehat{\mu}\left(\frac{t_0}{2^n}\right) = 0$, et, par continuité de $\widehat{\mu}$, que $\widehat{\mu}(0) = 0$, ce qui est faux, puisque $\widehat{\mu}(0) = 1$. On vient de démontrer que **la transformée de Fourier $\widehat{\mu}$ ne s'annule pas**. Il résulte de (13.44) que, pour tout entier n , on a

$$h(t) = \frac{\ln |\widehat{\mu}(t)|}{t^2} = \frac{4^n \ln \left| \widehat{\mu}\left(\frac{t}{2^n}\right) \right|}{t^2} = \frac{\ln \left| \widehat{\mu}\left(\frac{t}{2^n}\right) \right|}{\left(\frac{t}{2^n}\right)^2} = h\left(\frac{t}{2^n}\right).$$

Mais, la variable aléatoire X ayant une moyenne nulle et une variance 1, la transformée de Fourier $\widehat{\mu}$ admet le développement limité d'ordre deux au voisinage de 0 donné par

$$\widehat{\mu}(t) = 1 - \frac{t^2}{2} + o(t^2):$$

on a donc aussi

$$|\widehat{\mu}(t)| = 1 - \frac{t^2}{2} + o(t^2),$$

ce qui donne

$$\ln |\widehat{\mu}(t)| = -\frac{t^2}{2} + o(t^2) \quad \text{et} \quad h(t) = -\frac{1}{2} + o(1).$$

Il résulte alors de (13.44) que, pour tout réel $t \neq 0$, on a le développement asymptotique en n :

$$h(t) = h\left(\frac{t}{2^n}\right) = -\frac{1}{2} + o(1).$$

On a donc, pour tout réel $t \neq 0$, $h(t) = -\frac{1}{2}$, et par conséquent

$$\boxed{|\widehat{\mu}(t)| = \exp\left(-\frac{t^2}{2}\right)}.$$

Écrivons alors, pour tout réel $t \neq 0$, $\widehat{\mu}(t)$ sous la forme polaire

$$\widehat{\mu}(t) = |\widehat{\mu}(t)| \exp[i g(t)].$$

Il résulte de (13.44) que l'on a, pour tout entier n ,

$$\exp[i g(t)] = \exp\left(\frac{t^2}{2}\right) \widehat{\mu}(t) = \exp\left(\frac{t^2}{2}\right) \left[|\widehat{\mu}\left(\frac{t}{2^n}\right)\right]^{4^n},$$

ce qui donne le développement asymptotique en n

$$\exp[i g(t)] = \exp\left(\frac{t^2}{2}\right) \left[1 - \frac{t^2}{2 \cdot 4^n} + o\left(\frac{t^2}{4^n}\right)\right]^{4^n}.$$

Le membre de droite convergeant vers 1, il en résulte que l'on a, pour tout réel $t \neq 0$, $\exp[i g(t)] = 1$. On a démontré que, pour tout réel t ,

$$\boxed{\widehat{\mu}(t) = \exp\left(-\frac{t^2}{2}\right)},$$

c'est à dire que μ est la loi gaussienne $\mathcal{M}_{\mathbb{R}}(0, 1)$.

On étudie maintenant une autre caractérisation de variables aléatoires gaussiennes. Il s'agit d'une version d'un théorème de **Bernstein** un peu plus générale que celle usuellement énoncée.

Exercice 13.4. Caractérisation des variables aléatoires gaussiennes : théorème de Bernstein. Soient X et Y deux variables aléatoires réelles **indépendantes** et telles que les variables aléatoires $X+Y$ et $X-Y$ soient indépendantes ; l'objet de l'exercice est de démontrer que X et Y sont deux variables aléatoires gaussiennes. Pour cela, on note $\mu = P_X$, $\nu = P_Y$ et $\gamma = \mu * \nu$.

1. Démontrer que la transformée de Fourier $\widehat{\gamma}$ de γ vérifie la relation

$$\forall (u, v) \in \mathbb{R}^2 \quad \widehat{\gamma}(u+v)\widehat{\gamma}(u-v) = [\widehat{\gamma}(u)]^2 |\widehat{\gamma}(v)|^2 \quad (13.45)$$

2. Soit $\overline{\gamma}$ la probabilité définie par, pour tout borélien A , $\overline{\gamma}(A) = \gamma(-A)$ et $\delta = \gamma * \overline{\gamma}$. Démontrer que la transformée de Fourier $\widehat{\delta}$ de δ vérifie la relation

$$\forall (u, v) \in \mathbb{R}^2 \quad \widehat{\delta}(u+v)\widehat{\delta}(u-v) = [\widehat{\delta}(u)]^2 [\widehat{\delta}(v)]^2 \quad (13.46)$$

et que l'ensemble $G = \{t \in \mathbb{R} \mid \widehat{\delta}(t) \neq 0\}$ est un groupe. En déduire que la transformée de Fourier $\widehat{\delta}$ ne s'annule jamais; déterminer alors $\widehat{\delta}$ puis $|\widehat{\gamma}|$.

3. On pose, pour tout réel t , $g(t) = \frac{\widehat{\gamma}(t)}{|\widehat{\gamma}(t)|}$. Démontrer que g vérifie la relation

$$\forall (u, v) \in \mathbb{R}^2 \quad g^2(u+v) = g^2(u)g^2(v) \quad (13.47)$$

4. Soit Φ une application borélienne de \mathbb{R} dans \mathbb{C} telle que

- pour tout réel t , $|\Phi(t)| = 1$,
- pour tous réels s et t ,

$$\Phi(s+t) = \Phi(s)\Phi(t). \quad (13.48)$$

Démontrer qu'il existe un réel c tel que l'on ait, pour tout réel t , $\Phi(t) = \exp(ict)$. Pour cela, on démontrera qu'il existe un réel a tel que $f(a) \neq 0$, où f est la fonction définie, pour tout réel x , par

$$f(x) = \int_0^x \Phi(t) dt,$$

puis, après avoir remarqué que, pour tout réel x , on a

$$\Phi(x) = \frac{f(x+a) - f(x)}{f(a)},$$

ou en déduira que Φ est dérivable, ce qui permet de conclure.

5. En déduire qu'il existe un réel m et un réel $a > 0$ tels que l'on ait, pour tout réel t ,

$$\widehat{\gamma}(t) = \exp\left(imt - a\frac{t^2}{2}\right). \quad (13.49)$$

Démontrer alors que les variables aléatoires X et Y sont gaussiennes.

6. Généraliser ce résultat à des variables aléatoires à valeurs dans \mathbb{R}^d .

Solution.

1. Par indépendance des variables aléatoires $X+Y$ et $X-Y$ d'une part, puis des variables aléatoires X et Y d'autre part, on a, sur les fonctions caractéristiques, la relation : pour tous réels u et v ,

$$\varphi_{(X+Y, X-Y)}(u, v) = \varphi_{X+Y}(u)\varphi_{X-Y}(v) = \varphi_X(u)\varphi_Y(u)\overline{\varphi_X(v)}\varphi_Y(v);$$

puisque l'on a aussi, pour tous réels u et v ,

$$\varphi_{(X+Y, X-Y)}(u, v) = \varphi_{(X, Y)}(u+v, u-v) = \varphi_X(u+v)\varphi_Y(u-v),$$

on en déduit la relation : pour tous réels u et v ,

$$\varphi_X(u+v)\varphi_Y(u-v) = \varphi_X(u)\varphi_Y(u)\overline{\varphi_X(v)}\overline{\varphi_Y(v)};$$

en changeant v en $-v$ dans cette dernière relation, on a pour tous réels u et v ,

$$\varphi_X(u-v)\varphi_Y(u+v) = \varphi_X(u)\varphi_Y(u)\overline{\varphi_X(v)}\overline{\varphi_Y(v)};$$

en multipliant membre à membre ces deux dernières égalités et en remarquant que

$$\widehat{\gamma} = \widehat{\mu} \widehat{\nu} = \varphi_{X+Y} = \varphi_X \varphi_Y,$$

on obtient alors la relation

$$\forall (u, v) \in \mathbb{R}^2 \quad \widehat{\gamma}(u+v)\widehat{\gamma}(u-v) = [\widehat{\gamma}(u)]^2 |\widehat{\gamma}(v)|^2. \quad (13.50)$$

2. Le théorème de transfert permet d'établir l'égalité sur les transformées de Fourier de γ et $\overline{\gamma}$:

$$\forall t \in \mathbb{R} \quad \widehat{\gamma}(t) = \overline{\widehat{\gamma}(t)}. \quad (13.51)$$

Il en résulte que l'on a, pour tout réel t ,

$$\widehat{\delta}(t) = \widehat{\gamma}(t)\widehat{\overline{\gamma}}(t) = \widehat{\gamma}(t)\overline{\widehat{\gamma}(t)} = |\widehat{\gamma}(t)|^2 \geq 0; \quad (13.52)$$

en prenant les modules dans la relation (13.50), on obtient alors que, pour tous réels u et v ,

$$\widehat{\delta}(u+v)\widehat{\delta}(u-v) = [\widehat{\delta}(u)]^2 [\widehat{\delta}(v)]^2. \quad (13.53)$$

Il en résulte que si u et v sont tels que $\widehat{\delta}(u)$ et $\widehat{\delta}(v)$ sont différents de 0, on a aussi $\widehat{\delta}(u+v) \neq 0$ et $\widehat{\delta}(u-v) \neq 0$; puisque de plus $\widehat{\delta}(0) = 1$, **G est un groupe**. De plus, $\widehat{\delta}$ étant continue, **G est ouvert** et est donc identique à \mathbb{R} . Autrement dit, $\widehat{\delta}$ ne s'annule pas et est donc strictement positif. Posons alors, pour tout réel t ,

$$f(t) = -\ln \widehat{\delta}(t);$$

la relation (13.53) donne, pour tous réels u et v ,

$$f(u+v) + f(u-v) = 2[f(u) + f(v)]. \quad (13.54)$$

Puisque f est continue, positive, et que $f(0) = 0$, il en résulte qu'il existe¹⁰ $a > 0$ tel que l'on ait, pour tout réel u , $f(u) = au^2$, soit encore $\widehat{\delta}(u) = \exp(-au^2)$. En utilisant la relation (13.52), on vient de démontrer qu'il existe $a > 0$ tel que l'on ait :

$$\boxed{\forall t \in \mathbb{R} \quad |\widehat{\gamma}(t)| = \exp\left(-\frac{at^2}{2}\right)}. \quad (13.55)$$

10. On utilise l'argument classique : sur les entiers, puis sur les rationnels, puis par continuité, on prolonge aux réels.

3. En prenant les modules dans la relation (13.50), il vient, pour tous réels u et v ,

$$|\widehat{\gamma}(u+v)| |\widehat{\gamma}(u-v)| = |\widehat{\gamma}(u)|^2 |\widehat{\gamma}(v)|^2, \quad (13.56)$$

et en quotientant membre à membre les égalités des relations (13.50) et (13.56), on obtient :

$$\forall (u, v) \in \mathbb{R}^2 \quad g(u+v) g(u-v) = g^2(u);$$

en échangeant u en v , on a aussi

$$\forall (u, v) \in \mathbb{R}^2 \quad g(u+v) g(v-u) = g^2(v),$$

et donc, en multipliant membre à membre ces égalités, en tenant compte de ce que, pour tout t , on a $g(-t) = \overline{g(t)}$, et $|g(t)| = 1$, il vient :

$$\forall (u, v) \in \mathbb{R}^2 \quad g^2(u+v) = g^2(u) g^2(v). \quad (13.57)$$

4. Si on avait, pour tout a , $f(a) = 0$, les fonctions $\Re(\Phi)$ et $\Im(\Phi)$ seraient nulles presque partout, ce qui n'est pas, puisque $|\Phi| = 1$; soit donc a tel que $f(a) \neq 0$. On a

$$f(x+a) - f(x) = \int_x^{x+a} \Phi(t) dt,$$

et, après le changement de variables $s = t - x$ et application de la propriété de semi-groupe (13.48) pour Φ ,

$$f(x+a) - f(x) = \int_0^a \Phi(x+s) ds = \Phi(x) \int_0^a \Phi(s) ds = \Phi(x) f(a),$$

ce qui donne

$$\Phi(x) = \frac{f(x+a) - f(x)}{f(a)}. \quad (13.58)$$

La continuité de f implique alors celle de Φ ; la fonction f , définie comme fonction de la borne supérieure de l'intégrale de Φ , est alors dérivable, ce qui entraîne à son tour, en vertu de (13.58), que Φ est dérivable. En dérivant par rapport à s dans l'égalité $\Phi(s+t) = \Phi(s)\Phi(t)$, on a donc, pour tous s et t ,

$$\Phi'(s+t) = \Phi'(s)\Phi(t).$$

et donc, pour tout t ,

$$\Phi'(t) = \Phi'(0)\Phi(t). \quad (13.59)$$

Si $\Phi'(0) = 0$, on a alors pour tout t , $\Phi'(t) = 0$ et Φ est constant, non nul, puisque on a $|\Phi| = 1$; d'après (13.48), on a alors $\Phi(t) = 1$ pour tout t et $c = 0$ convient. Si $\Phi'(0) \neq 0$, il résulte de (13.59) que $\Phi(0) = 1$. Soit alors $c = -i\Phi'(0)$; on a $\Phi'(t) = ic\Phi(t)$, et donc

$$\frac{d}{dt} [\Phi(t) \exp(-ict)] = \Phi'(t) \exp(-ict) - ic\Phi(t) \exp(-ict) = 0.$$

Il en résulte que l'on a $\Phi(t) \exp(-ict) = \Phi(0) = 1$, soit, pour tout t , $\Phi(t) = \exp(ict)$. Enfin, puisque $|\Phi(1)| = |\exp(ic)| = 1$, c est réel.

5. En vertu de la relation (13.57) et de la définition de g , on peut appliquer le résultat de la question précédente à fonction g^2 ; il existe donc un réel m tel que l'on ait, pour tout t , $g^2(t) = \exp(i2mt)$. Puisque $g(0) = 1$, on obtient par continuité de g que $g(t) = \exp(imt)$. Il résulte alors de (13.55) que l'on a, pour tout t ,

$$\widehat{\gamma}(t) = \exp\left(imt - a\frac{t^2}{2}\right), \quad (13.60)$$

où $a > 0$; autrement dit, γ est la probabilité gaussienne $\mathcal{N}_{\mathbb{R}}(m, a)$. La variable aléatoire $X + Y$ est donc gaussienne; on montrerait de même que $X - Y$ est gaussienne. Ces variables aléatoires étant indépendantes, la variable aléatoire $(X + Y, X - Y)$ est aussi gaussienne; ainsi, les variables aléatoires X et Y sont gaussiennes, comme transformées linéaires de la variable aléatoire gaussienne $(X + Y, X - Y)$.

6. Pour généraliser à \mathbb{R}^d , il suffit d'appliquer, pour tous x et y de \mathbb{R}^d , le résultat précédent aux variables aléatoires réelles $\langle X, x \rangle$ et $\langle Y, y \rangle$.

Exercice 13.5. Une caractérisation de la loi gaussienne en termes de moyenne et variance empirique. Soient X_1, X_2, \dots, X_n des variables aléatoires réelles indépendantes de même loi μ telle que

$$\int_{\mathbb{R}} x^2 d\mu(x) < +\infty.$$

On définit les variables aléatoires, appelées respectivement moyenne et variance empirique, par

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \Sigma_n = \frac{1}{n} \sum_{i=1}^n X_i^2 - M_n^2.$$

On note X la variable aléatoire à valeurs dans \mathbb{R}^n , $X = (X_1, X_2, \dots, X_n)$.

1. On suppose que μ est la loi gaussienne $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$. Quelle est la loi de X ? Soit C une matrice orthogonale $n \times n$ telle que, pour tout $j = 1, 2, \dots, n$ on ait $C_{1j} = 1/\sqrt{n}$. Exprimer M_n et Σ_n à l'aide des composantes de CX et en déduire que M_n et Σ_n sont des variables aléatoires indépendantes.

Dans le cas où $m = 0$ et $\sigma = 1$, préciser les lois des variables aléatoires M_n et Σ_n .

2. On s'intéresse à une réciproque: on suppose que les variables aléatoires M_n et Σ_n sont indépendantes. Pour simplifier, on suppose les variables aléatoires X_i centrées. On note $\sigma^2 = \int_{\mathbb{R}} x^2 d\mu(x)$, φ la transformée de Fourier de μ , puis $S_n = nM_n$ et $V_n = n\Sigma_n$.

(a) Calculer la moyenne $E(V_n)$ en fonction de σ .

(b) Démontrer que l'application de \mathbb{R}^2 dans \mathbb{C} : $(u, v) \mapsto E[\exp i(uS_n + vV_n)]$ est différentiable. Justifier la relation:

$$\forall u \in \mathbb{R} \quad E[V_n \exp i(uS_n)] = [\varphi(u)]^n E(V_n). \quad (13.61)$$

(c) Calculer $E[V_n \exp i(uS_n)]$ à l'aide des dérivées première et seconde de φ .

(d) Dédurre alors de la relation (13.61) que φ est solution de l'équation différentielle

$$\frac{\varphi''}{\varphi} - \left[\frac{\varphi'}{\varphi} \right]^2 = -\sigma^2,$$

et en conclure que μ est la loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$.

(e) Démontrer qu'il en est de même si on ne suppose plus les variables aléatoires X_i centrées.

Solution.

1. Les variables aléatoires X_i sont **indépendantes** et de loi gaussienne $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$, la variable aléatoire X , à valeurs dans \mathbb{R}^n , est donc aussi gaussienne de moyenne (m, m, \dots, m) et de matrice de covariance $\sigma^2 I$, où I est la matrice identité de \mathbb{R}^n . Il résulte de la définition de C que

$$M_n = \frac{1}{\sqrt{n}}(CX)_1,$$

et donc que

$$n\Sigma_n = \|X\|^2 - [(CX)_1]^2;$$

la matrice C , étant orthogonale, conserve la norme, ce qui donne :

$$n\Sigma_n = \sum_{i=2}^n [(CX)_i]^2.$$

La variable aléatoire CX , transformée linéaire de la variable aléatoire gaussienne X , est elle-même gaussienne de moyenne (Cm, Cm, \dots, Cm) et de matrice de covariance $\sigma^2 C I C^* = \sigma^2 I$ (puisque la matrice C est orthogonale). Il en résulte que les composantes de CX sont indépendantes, et donc aussi que **les variables aléatoires M_n et Σ_n sont indépendantes**.

Dans le cas où $m = 0$ et $\sigma = 1$, **la variable aléatoire M_n suit alors la loi $\mathcal{N}_{\mathbb{R}}(0, 1/n)$** ; $n\Sigma_n$ étant somme de $n - 1$ carrés de variables aléatoires gaussiennes centrées réduites et indépendantes, **la variable aléatoire $n\Sigma_n$ suit la loi du chi-deux à $n - 1$ degrés de liberté**.

2. Les variables aléatoires M_n et Σ_n étant indépendantes, les variables S_n et V_n le sont aussi. On suppose dans un premier temps que les variables aléatoires X_i sont **centrées**.

(a) On a

$$V_n = \sum_{i=1}^n X_i^2 - \frac{S_n^2}{n}, \text{ et donc, } EV_n = \sum_{i=1}^n EX_i^2 - \frac{ES_n^2}{n};$$

mais, puisque $ES_n = \sum_{i=1}^n EX_i = 0$, on a, par indépendance des X_i ,

$$ES_n^2 = \sigma_{S_n}^2 = \sum_{i=1}^n \sigma_{X_i}^2 = n\sigma^2.$$

Enfin, les X_i étant centrées, on a $\sigma_{X_i}^2 = EX_i^2$, ce qui donne donc :

$$E(V_n) = (n-1)\sigma^2.$$

(b) Les variables aléatoires S_n et V_n sont indépendantes; on a donc, pour tout $(u, v) \in \mathbb{R}^2$,

$$E[\exp i(us_n + vV_n)] = \varphi_{S_n}(u) \varphi_{V_n}(v),$$

où φ_{S_n} et φ_{V_n} sont les fonctions caractéristiques de S_n et V_n . Ces variables aléatoires admettant un moment d'ordre un, leur fonction caractéristique est dérivable; par conséquent, l'application $(u, v) \mapsto E[\exp i(uS_n + vV_n)]$ est différentiable.

Les variables aléatoires S_n et V_n étant indépendantes, il en est de même de variables aléatoires V_n et $\exp i(uS_n)$; puisqu'elles sont intégrables, on a, pour tout réel u ,

$$E[V_n \exp i(uS_n)] = E[V_n] E[\exp i(uS_n)] = \varphi'_{S_n}(u) E(V_n);$$

les X_i étant indépendantes, il en résulte que l'on a :

$$\forall u \in \mathbb{R} \quad E[V_n \exp i(uS_n)] = [\varphi(u)]^n E(V_n). \quad (13.62)$$

Remarque. On aurait aussi pu déjà exploiter la différentiabilité établie ci-dessus et dire que

$$\left[\frac{\partial}{\partial v} E[\exp i(us_n + vV_n)] \right]_{v=0} = \varphi_{S_n}(u) \varphi'_{V_n}(0);$$

puisque de plus $|\exp i(us_n + vV_n)| \leq 1$, que

$$\left| \frac{\partial}{\partial v} \exp i(us_n + vV_n) \right| = |V_n \exp i(us_n + vV_n)| \leq |V_n|,$$

et que V_n est P-intégrable, il résulte du théorème de dérivation d'une intégrale dépendant d'un paramètre que

$$\frac{\partial}{\partial v} E[\exp i(us_n + vV_n)] = v E[V_n \exp i(us_n + vV_n)].$$

Comme on a

$$\varphi'_{V_n}(0) = i E V_n,$$

en tenant compte de l'indépendance des X_i , on retrouve la relation (13.62).

(c) Puisque $V_n = \sum_{k=1}^n X_k^2 - \frac{S_n^2}{n}$, on a

$$E[V_n \exp i(us_n)] = \sum_{k=1}^n E[X_k^2 \exp i(us_n)] - \frac{1}{n} E[S_n^2 \exp i(us_n)]. \quad (13.63)$$

Il résulte de l'indépendance des X_k que l'on a, pour tout $k \leq n$ fixé,

$$E[X_k^2 \exp i(us_n)] = \left\{ \prod_{\substack{i \neq k \\ 1 \leq i \leq n}} E[\exp i(uX_i)] \right\} E[X_k^2 \exp i(uX_k)].$$

soit, puisque les variables aléatoires X_i sont indépendantes,

$$E[X_k^2 \exp i(uS_n)] = [\varphi(u)]^{n-1} E[X_k^2 \exp i(uX_k)]. \quad (13.64)$$

Puisque les variables aléatoires X_k et S_n admettent un moment d'ordre deux, leur fonction caractéristique est deux fois dérivable et on a

$$E[X_k^2 \exp i(uX_k)] = -\varphi''_{X_k}(u) \quad \text{et} \quad E[S_n^2 \exp i(uS_n)] = -\varphi''_{S_n}(u). \quad (13.65)$$

Les variables aléatoires X_k étant indépendantes et de même loi, on a $\varphi_{S_n}(u) = [\varphi(u)]^n$. Il résulte alors de (13.63), (13.64) et (13.65) que l'on a

$$E[V_n \exp i(uS_n)] = -n [\varphi(u)]^{n-1} \varphi''(u) \\ + \frac{1}{n} [n(n-1) [\varphi(u)]^{n-2} [\varphi'(u)]^2 + n [\varphi(u)]^{n-1} \varphi''(u)]$$

soit

$$E[V_n \exp i(uS_n)] = -(n-1) [\varphi(u)]^{n-1} \varphi''(u) + (n-1) [\varphi(u)]^{n-2} [\varphi'(u)]^2. \quad (13.66)$$

(d) Remarquons d'abord, puisque φ est continue et que $\varphi(0) = 1$, que $\varphi^{-1}(\{0\}^c)$ est un voisinage ouvert de 0. En tenant compte des égalités (13.62) et (13.66) ainsi que de la valeur de EV_n , on voit que, sur l'ouvert $\varphi^{-1}(\{0\}^c)$, φ est solution de l'équation différentielle :

$$\frac{\psi''(u)}{\psi(u)} - \left[\frac{\psi'(u)}{\psi(u)} \right]^2 = -\sigma^2.$$

Puisque, de plus, on a $\varphi(0) = 1$ et $\varphi'(0) = 0$ (car les X_i sont centrées), il existe $a_1 > 0$ tel que l'on ait, pour tout $u \in [-a_1, a_1]$,

$$\varphi(u) = \exp\left(-\frac{\sigma^2 u^2}{2}\right). \quad (13.67)$$

En particulier, $\varphi(a_1) \neq 0$; φ étant continue est non nulle sur un intervalle $[-a_1, a_2]$ où $a_2 > a_1$ et l'égalité (13.67) est encore valable sur cet intervalle. Par récurrence, on montre de même qu'il existe une suite **strictement croissante** de réels $a_n > 0$ tels que l'égalité (13.67) soit encore valable sur l'intervalle $[-a_1, a_n]$. Si cette suite était bornée, elle convergerait vers un réel $a > 0$. On aurait, pour tout n , $\varphi(a_n) = \exp(-\sigma^2 a_n^2/2)$ et aussi, par continuité de φ et de l'exponentielle, $\varphi(a) = \exp(-\sigma^2 a^2/2) > 0$ et il y aurait contradiction. En conclusion, l'égalité (13.67) est valable sur $[-a_1, +\infty[$ et donc en fait sur tout \mathbb{R} , puisque, pour tout réel u , on a $\varphi(-u) = \overline{\varphi(u)}$. On vient de démontrer que μ est la loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$.

(e) Si les variables aléatoires X_i sont de moyenne m , on considère les variables aléatoires centrées $\overset{\circ}{X}_i = X_i - m$, qui sont encore indépendantes, de telle sorte qu'un calcul simple conduit aux égalités :

$$\overset{\circ}{S}_n = \sum_{i=1}^n \overset{\circ}{X}_i = S_n - nm \quad \text{et} \quad \overset{\circ}{V}_n = \sum_{i=1}^n \overset{\circ}{X}_i^2 - \frac{\overset{\circ}{S}_n^2}{n} = V_n.$$

Si les variables aléatoires S_n et V_n sont indépendantes, il en est de même des variables aléatoires \hat{S}_n et \hat{V}_n et ainsi, d'après la question précédente, les variables aléatoires \hat{X}_i sont de loi $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$. Il en résulte que la mesure μ est la loi gaussienne $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$.

Exercice 13.6. Initiation à la théorie du signal. Soient S (le signal) et V (le bruit) deux variables aléatoires réelles indépendantes, gaussiennes de loi respectives $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$ et $\mathcal{N}_{\mathbb{R}}(0, t)$, où m est un réel quelconque, σ et t sont des réels strictement positifs. L'observation est la variable aléatoire

$$R = tS + V.$$

Calculer la meilleure approximation au sens des moindres carrés du signal au vu de l'observation, c'est à dire l'espérance conditionnelle $E^{\sigma(R)}S$, où $\sigma(R)$ désigne la tribu engendrée par R . Pour cela, choisir des constantes a et b telles que les variables aléatoires $aR + bS$ et R soient indépendantes et en déduire l'espérance conditionnelle $E^{\sigma(R)}S$.

Solution. Puisque les variables aléatoires S et V sont indépendantes et gaussiennes, la variable aléatoire (S, V) est gaussienne et donc aussi sa transformée linéaire $(aR + bS, R)$. Ainsi, pour que $aR + bS$ et R soient indépendantes, il faut et il suffit que $\text{cov}(aR + bS, R) = 0$. En vertu de la bilinéarité de la covariance et de l'indépendance des variables aléatoires S et V , on peut écrire les égalités

$$\text{cov}(aR + bS, R) = a\sigma_{\mathbb{R}}^2 + b \text{cov}(S, R) = a(t^2\sigma^2 + t) + bt\sigma^2.$$

On choisit a et b différents de 0 et tels que

$$a = -\frac{b\sigma^2}{t\sigma^2 + 1}.$$

et ainsi les variables aléatoires $aR + bS$ et R sont indépendantes. On a alors

$$E^{\sigma(R)}(aR + bS) = E(aR + bS) = aE(R) + bE(S);$$

on a aussi

$$E^{\sigma(R)}(aR + bS) = aR + bE^{\sigma(R)}(S),$$

ce qui donne

$$bE^{\sigma(R)}(S) = -a(R - E(R)) + bE(S),$$

soit, en tenant compte du choix de a et b ,

$$E^{\sigma(R)}(S) = \frac{\sigma^2}{t\sigma^2 + 1}(R - tm) + m,$$

ou encore :

$$E^{\sigma(R)}(S) = \frac{m + \sigma^2 R}{t\sigma^2 + 1}.$$

Exercice 13.7. Initiation à la théorie du signal (suite). Soient n réels strictement positifs t_1, t_2, \dots, t_n . Soient S (le signal) et, pour $i = 1, 2, \dots, n$, W_{t_i} (le bruit à l'instant t_i), des variables aléatoires réelles indépendantes, gaussiennes de loi respectives $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$ et $\mathcal{N}_{\mathbb{R}}(0, t_i)$, où m est un réel quelconque et σ un réel strictement positif. L'observation à l'instant t_i est la variable aléatoire

$$R_i = t_i S + W_{t_i}.$$

On munit \mathbb{R}^n de sa base canonique et on note t le vecteur de composantes t_1, t_2, \dots, t_n , R et W les variables aléatoires à valeurs dans \mathbb{R}^n de composantes respectives R_1, R_2, \dots, R_n , et $W_{t_1}, W_{t_2}, \dots, W_{t_n}$, de sorte que l'on a :

$$R = S t + W.$$

Calculer la meilleure approximation au sens des moindres carrés du signal au vu des observations R_1, R_2, \dots, R_n , c'est à dire l'espérance conditionnelle $E^{\sigma(R)} S$, où $\sigma(R)$ désigne la tribu engendrée par R . Pour cela, choisir un vecteur u de \mathbb{R}^n et une constante b tels que les variables aléatoires $\langle u, R \rangle + bS$ et R soient indépendantes et en déduire l'espérance conditionnelle $E^{\sigma(R)} S$.

Solution. Puisque les variables aléatoires S et $W_{t_1}, W_{t_2}, \dots, W_{t_n}$ sont indépendantes et gaussiennes, la variable aléatoire (S, W) est gaussienne et donc aussi sa transformée linéaire $(\langle u, R \rangle + bS, R)$. Ainsi, pour que $\langle u, R \rangle + bS$ et R soient indépendantes, il faut et il suffit que la matrice de covariance de $(\langle u, R \rangle + bS, R)$ soit nulle, ce qui s'écrit encore

$$\forall j = 1, 2, \dots, n \quad \text{cov}(\langle u, R \rangle + bS, R_j) = 0,$$

soit

$$\forall j = 1, 2, \dots, n \quad \text{cov}(\langle u, R \rangle, R_j) + b \text{cov}(S, R_j) = 0. \quad (13.68)$$

On a, par indépendance de S et W_{t_j} ,

$$\text{cov}(S, R_j) = \text{cov}(S, t_j S + W_{t_j}) = t_j \sigma_S^2;$$

de plus on a

$$\text{cov}(\langle u, R \rangle, R_j) = \langle u, t \rangle \text{cov}(S, R_j) + \text{cov}(\langle u, W \rangle, R_j),$$

soit

$$\text{cov}(\langle u, R \rangle, R_j) = \langle u, t \rangle [t_j \sigma_S^2] + [\langle u, W \rangle, S] + \text{cov}(\langle u, W \rangle, W_{t_j}),$$

soit encore, puisque les variables aléatoires $\langle u, W \rangle$ et S sont indépendantes

$$\text{cov}(\langle u, R \rangle, R_j) = \langle u, t \rangle [t_j \sigma_S^2] + \text{cov}(\langle u, W \rangle, W_{t_j}).$$

La variable aléatoire W ayant ses composantes indépendantes, on a

$$\text{cov}(\langle u, W \rangle, W_{t_j}) = u_j \sigma_{W_{t_j}}^2 = u_j t_j.$$

Ainsi, la condition d'indépendance (13.68) s'écrit, en simplifiant par t_j ,

$$\forall j = 1, 2, \dots, n \quad \langle u, t \rangle \sigma^2 + u_j + b \sigma^2 = 0. \quad (13.69)$$

Pour u , on prend le vecteur $\mathbf{1}$ dont toutes les composantes valent 1 et on choisit alors b tel que

$$\langle \mathbf{1}, t \rangle \sigma^2 + 1 + b \sigma^2 = 0. \quad (13.70)$$

et ainsi, pour ce choix, les variables aléatoires $\langle \mathbf{1}, R \rangle + bS$ et R sont indépendantes. On a alors

$$E^{\sigma(R)}(\langle \mathbf{1}, R \rangle + bS) = E(\langle \mathbf{1}, R \rangle + bS) = E(\langle \mathbf{1}, R \rangle) + bE(S);$$

on a aussi

$$E^{\sigma(R)}(\langle \mathbf{1}, R \rangle + bS) = \langle \mathbf{1}, R \rangle + bE^{\sigma(R)}(S),$$

ce qui donne, en comparant les membres de droite de ces égalités,

$$bE^{\sigma(R)}(S) = -\langle \mathbf{1}, R \rangle + E(\langle \mathbf{1}, R \rangle) + bm.$$

En tenant compte de la valeur de b donnée par l'égalité (13.70) et de l'égalité

$$E(\langle \mathbf{1}, R \rangle) = \langle \mathbf{1}, ER \rangle = m \langle \mathbf{1}, t \rangle,$$

il vient

$$E^{\sigma(R)}(S) = \frac{\sigma^2}{\langle \mathbf{1}, t \rangle \sigma^2 + 1} \langle \mathbf{1}, R \rangle + m \left(\frac{1}{\langle \mathbf{1}, t \rangle \sigma^2 + 1} \right).$$

ou encore :

$$E^{\sigma(t)}(S) = \frac{m + \sigma^2 \sum_{j=1}^n R_j}{1 + \sigma^2 \sum_{j=1}^n t_j}.$$

Exercice 13.8. Forme quadratique d'une variable aléatoire gaussienne. (Théorème de Cochran.) Soit X une variable aléatoire à valeurs dans un espace euclidien E de dimension d , de loi gaussienne $\mathcal{N}_E(0, \mathbf{I})$, où \mathbf{I} est l'application identique sur E . Soit a un vecteur unitaire de E et U et V les variables aléatoires réelles définies par

$$U = \langle X, a \rangle \quad \text{et} \quad V = \|X\|^2 - \langle X, a \rangle^2.$$

1. Démontrer que les variables aléatoires U et V sont indépendantes et identifier leur loi.

2. Soit Y une variable aléatoire à valeurs dans E de loi gaussienne $\mathcal{N}_E(m, \mathbf{I})$, où $m \in E$. Dédurre de la question précédente que la loi de $\|Y\|^2$ est la convolution d'un chi-deux à $d - 1$ degrés de liberté et de la loi du carré d'une variable aléatoire réelle gaussienne de loi $\mathcal{N}_R(\|m\|, 1)$.

Solution.

1. Soit $(e_i)_{i=1, \dots, d}$ une base **orthonormée** de E de premier vecteur a . On a alors

$$U = \langle X, e_1 \rangle \quad \text{et} \quad V = \sum_{i=2}^d \langle X, e_i \rangle^2,$$

et les variables aléatoires $\langle X, e_i \rangle$, $i = 1, \dots, d$, sont indépendantes. L'indépendance de U et V en résulte. La loi de U est gaussienne; on a

$$EU = \langle EX, e_1 \rangle = 0 \quad \text{et} \quad \sigma_{\langle X, e_1 \rangle}^2 = \langle \Lambda_X a, a \rangle = \|a\|^2 = 1;$$

la loi de U est donc la loi $\mathcal{N}_{\mathbb{R}}(0, 1)$. De même, les variables aléatoires $\langle X, e_i \rangle$, $i = 2, \dots, d$, sont de loi $\mathcal{N}_{\mathbb{R}}(0, 1)$ et sont indépendantes, **la loi de V est donc la loi du chi-deux à $d - 1$ degrés de liberté.**

2. On a

$$\|Y\|^2 = \|Y - m\|^2 + 2 \langle Y - m, m \rangle + \|m\|^2,$$

soit encore

$$\|Y\|^2 = \left[\|Y - m\|^2 - \left\langle Y - m, \frac{m}{\|m\|} \right\rangle^2 \right] + \left[\left\langle Y - m, \frac{m}{\|m\|} \right\rangle + \|m\| \right]^2.$$

La variable aléatoire $Y - m$ suit la loi $\mathcal{N}_E(0, \mathbf{I})$; il résulte de la question précédente que les variables aléatoires $\|Y - m\|^2 - \left\langle Y - m, \frac{m}{\|m\|} \right\rangle^2$ et $\left\langle Y - m, \frac{m}{\|m\|} \right\rangle + \|m\|$ sont indépendantes de lois respectives la loi du chi-deux à $d - 1$ degrés de liberté et la loi $\mathcal{N}_{\mathbb{R}}(\|m\|, 1)$, ce qui démontre que **la loi de $\|Y\|^2$ est la convolution d'un chi-deux à $d - 1$ degrés de liberté et de la loi du carré d'une variable aléatoire réelle gaussienne de loi $\mathcal{N}_{\mathbb{R}}(\|m\|, 1)$.**

Exercice 13.9. Moyenne et variance empiriques. Soient Y_1, Y_2, \dots, Y_n , n variables aléatoires indépendantes à valeurs dans un espace euclidien E de dimension d , de loi gaussienne $\mathcal{N}_E(0, \mathbf{I})$, où \mathbf{I} est l'application identique sur E . On note $Y = (Y_1, Y_2, \dots, Y_n)$ la variable aléatoire à valeurs dans E^n muni de la structure euclidienne produit.

1. Démontrer que les variables aléatoires $\|Y\|^2$ et $\frac{Y}{\|Y\|}$ sont indépendantes (on étudiera d'abord le cas où $E = \mathbb{R}$, en pensant à faire un changement de variables en coordonnées sphériques, et on en déduira le cas général). Préciser leur loi dans le cas où $E = \mathbb{R}$.

2. Soient n variables aléatoires réelles indépendantes X_1, X_2, \dots, X_n , de loi gaussienne $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$, où $m \in \mathbb{R}$ et $\sigma > 0$. On définit les variables aléatoires réelles M et V , et les variables aléatoires X et X' à valeurs dans \mathbb{R}^n par

$$M = \frac{1}{n} \sum_{j=1}^n X_j \quad \text{et} \quad V = \sum_{j=1}^n (X_j - M)^2,$$

$$X = (X_1, X_2, \dots, X_n) \quad \text{et} \quad X' = (X_1 - M, X_2 - M, \dots, X_n - M);$$

on pose $Z = \frac{X'}{\sqrt{V}}$. On remarquera que l'on a $M = \frac{1}{n} \langle X, e \rangle$, où e est le vecteur de \mathbb{R}^n $(1, 1, \dots, 1)$.

- (a) Démontrer que les variables aléatoires M et X' sont indépendantes.
- (b) Calculer l'opérateur d'auto-covariance de X' .
- (c) Soit H l'hyperplan orthogonal à e ; démontrer qu'existe une isométrie B de \mathbb{R}^{n-1} sur H et une variable aléatoire U à valeurs dans \mathbb{R}^{n-1} de loi $\mathcal{N}_{\mathbb{R}^{n-1}}(0, \sigma^2 \mathbf{1}_{\mathbb{R}^{n-1}})$ telles que $X' = BU$ P-p.s.
- (d) En déduire que les variables aléatoires M, V et Z sont indépendantes.
- (e) Déterminer les lois de M et $\frac{1}{\sigma^2} V$.

Solution.

1. Plaçons nous d'abord dans le cas où $E \cong \mathbb{R}$ et soient f et g des fonctions de \mathbb{R} dans lui-même, mesurables **positives**. On a, par le théorème de transfert,

$$E\left[f(\|Y\|^2)g\left(\frac{Y}{\|Y\|}\right)\right] = \int_{\mathbb{R}^n} f(\|y\|^2)g\left(\frac{y}{\|y\|}\right) \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{\|y\|^2}{2}\right) dy.$$

Effectuons le changement de variables en coordonnées sphériques défini par

$$\begin{cases} y_1 &= \rho \cos \varphi_1 \\ y_2 &= \rho \sin \varphi_1 \cos \varphi_2 \\ &\dots\dots\dots \\ y_{n-1} &= \rho \sin \varphi_1 \dots \sin \varphi_{n-2} \cos \varphi_{n-1} \\ y_n &= \rho \sin \varphi_1 \dots \sin \varphi_{n-2} \sin \varphi_{n-1}. \end{cases}$$

ce qui définit un difféomorphisme de $\mathbb{R}^n \setminus (\bigcup_{i=1}^n D_i)$ sur $]0, +\infty[\times]0, \pi[^{n-2} \times]0, 2\pi[$, où D_i est la droite engendrée par le i -ième vecteur de la base canonique de \mathbb{R}^n . Le jacobien de la transformation étant

$$J(\rho, \varphi_1, \varphi_2, \dots, \varphi_{n-1}) = \rho^{n-1} \prod_{i=1}^{n-2} (\sin \varphi_i)^{n-i-1},$$

le changement de variables et une application du théorème de Fubini conduisent à l'égalité

$$E\left[f(\|Y\|^2)g\left(\frac{Y}{\|Y\|}\right)\right] = I_1(f) I_2(g), \quad (13.71)$$

où on a posé

$$I_1(f) = \int_{]0, +\infty[} \frac{1}{(2\pi)^{\frac{n}{2}}} \rho^{n-1} f(\rho^2) \exp\left(-\frac{\rho^2}{2}\right) d\rho.$$

$$I_2(g) =$$

$$\int_{]0, \pi[^{n-2} \times]0, 2\pi[} \prod_{i=1}^{n-2} (\sin \varphi_i)^{n-i-1} g(\Phi(\varphi_1, \varphi_2, \dots, \varphi_{n-1})) d(\varphi_1, \varphi_2, \dots, \varphi_{n-1}), \quad (13.72)$$

et

$$\Phi(\varphi_1, \varphi_2, \dots, \varphi_{n-1}) = (\cos \varphi_1, \sin \varphi_1 \cos \varphi_2, \dots, \sin \varphi_1 \dots \sin \varphi_{n-2} \sin \varphi_{n-1}).$$

En particulier, on obtient

$$E[f(\|Y\|^2)] = I_1(f) I_2(1) \quad \text{et} \quad E\left[g\left(\frac{Y}{\|Y\|}\right)\right] = I_1(1) I_2(g),$$

et

$$E\{1\} = I_1(1) I_2(1) = 1.$$

Il résulte alors de (13.71) que l'on a, pour toutes fonctions f et g mesurables positives,

$$E\left[f(\|Y\|^2)g\left(\frac{Y}{\|Y\|}\right)\right] = E[f(\|Y\|^2)] E\left[g\left(\frac{Y}{\|Y\|}\right)\right]. \quad (13.73)$$

ce qui est nécessaire et suffisant pour assurer l'**indépendance des variables aléatoires** $\|Y\|^2$ et $\frac{Y}{\|Y\|}$.

Si maintenant E est un espace euclidien quelconque, soit $(e_i)_{i=1, \dots, d}$ une base orthonormée de E ; les variables aléatoires réelles $Z_{i,j} = (Y_i, e_j)$, où $i = 1, \dots, n$ et $j = 1, \dots, d$, sont indépendantes de loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$, puisque transformées linéaires de la variable aléatoire gaussienne Y à valeurs dans E^n et puisque, pour deux couples (i, j) et (k, l) différents, on a $\text{cov}((Y_i, e_j), (Y_k, e_l)) = 0$. De plus, on a

$$\|Y\|^2 = \sum_{i=1}^n \sum_{j=1}^d Z_{i,j}^2.$$

Soit Z la variable aléatoire à valeurs dans E^{nd} définie par

$$Z = (Z_{1,1}, \dots, Z_{1,d}, Z_{2,1}, \dots, Z_{2,d}, \dots, Z_{n,1}, \dots, Z_{n,d}),$$

on a alors $\|Z\|^2 = \|Y\|^2$ et, d'après la propriété établie précédemment, les variables aléatoires $\|Z\|^2$ et $\frac{Z}{\|Z\|}$ sont indépendantes. Il en résulte que les variables aléatoires $\|Y\|^2$ et $\frac{1}{\|Z\|}(\sum_{j=1}^d Z_{1,j}e_j, \dots, \sum_{j=1}^d Z_{n,j}e_j)$ sont aussi indépendantes; puisque l'on a

$$\frac{1}{\|Z\|} \left(\sum_{j=1}^d Z_{1,j}e_j, \dots, \sum_{j=1}^d Z_{n,j}e_j \right) = \frac{1}{\|Y\|} (Y_1, \dots, Y_n) = \frac{Y}{\|Y\|}.$$

les variables aléatoires $\|Y\|^2$ et $\frac{Y}{\|Y\|}$ sont indépendantes.

Si $E = \mathbb{R}$, la loi de $\|Y\|^2$ est la loi du chi-deux à n degrés de liberté (somme de n variables aléatoires indépendantes, carrés de variables aléatoires de loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$). On a montré que, pour toute fonction g mesurable positive, on a

$$E\left[g\left(\frac{Y}{\|Y\|}\right)\right] = I_1(1) I_2(g).$$

Soit S_n la sphère de \mathbb{R}^n de centre 0 et de rayon 1 et soit μ la mesure image de la mesure $I_1(1) \left[\prod_{i=1}^{n-2} (\sin \varphi_i)^{n-i-1} \right] d(\varphi_1, \varphi_2, \dots, \varphi_{n-1})$ sur $]0, \pi[\times]0, 2\pi[$ par l'application Φ ; μ peut être appelée **probabilité uniforme sur S_n** et on a $E\left[g\left(\frac{Y}{\|Y\|}\right)\right] = \int_{S_n} g(x) d\mu(x)$. Ainsi, la loi de $\frac{Y}{\|Y\|}$ est la loi uniforme (au sens précédent) sur S_n .

2. Les variables aléatoires X_1, X_2, \dots, X_n étant indépendantes de loi gaussienne $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$, la variable aléatoire X est gaussienne de loi $\mathcal{N}_{\mathbb{R}^n}(me, \sigma^2 \mathbf{1}_{\mathbb{R}^n})$. La variable aléatoire (M, X') , transformée linéaire de X est alors de loi gaussienne dans \mathbb{R}^{n+1} .

(a) Ainsi, pour que M et X' soient indépendantes, il faut et il suffit que leur opérateur d'intercovariance $\Lambda_{M, X'}$ soit nul. On remarque que X' est centrée, puis que l'on a

$$X' = X - Me \quad \text{et} \quad M = \frac{1}{n} \langle X, e \rangle.$$

si bien que

$$EX' = EX - E \frac{1}{n} \langle X, e \rangle e = me - \frac{1}{n} \langle me, e \rangle e = 0.$$

On a, pour tout $u \in \mathbb{R}^n$,

$$\Lambda_{M, X'} u = \text{cov}(M, \langle X', u \rangle) = \text{cov}(M, \langle X, u \rangle - M \langle e, u \rangle),$$

soit

$$\Lambda_{M, X'} u = \text{cov}\left(\frac{1}{n} \langle X, e \rangle, \langle X, u \rangle\right) - \langle e, u \rangle \sigma_M^2 = \frac{1}{n} \langle \Lambda_X e, u \rangle - \langle e, u \rangle \sigma_M^2;$$

puisque l'on a $\Lambda_X = \sigma^2 \mathbf{1}_{\mathbb{R}^n}$, il vient

$$\sigma_M^2 = \frac{1}{n^2} \sigma_{\langle X, e \rangle}^2 = \frac{1}{n^2} \langle \Lambda_X e, e \rangle = \frac{\sigma^2 \|e\|^2}{n^2} = \frac{\sigma^2}{n}, \quad (13.74)$$

et donc aussi

$$\Lambda_{M, X'} u = \frac{\sigma^2}{n} \langle e, u \rangle - \frac{\sigma^2}{n} \langle e, u \rangle = 0;$$

ainsi $\Lambda_{M, X'} = 0$ et **les variables aléatoires M et X' sont indépendantes.**

(b) Puisque l'on a $X = X' + Me$ et que M et X' sont indépendantes, on a

$$\Lambda_X = \Lambda_{X'} + \Lambda_{Me}.$$

Pour tous u et v de \mathbb{R}^n , on a

$$\langle \Lambda_{Me} u, v \rangle = \text{cov}(\langle Me, u \rangle, \langle Me, v \rangle) = \langle e, u \rangle \langle e, v \rangle \sigma_M^2,$$

ce qui donne, d'après (13.74),

$$\Lambda_{X'} = \sigma^2 \left(\mathbf{1}_{\mathbb{R}^n} - \frac{1}{n} ee^* \right),$$

où ee^* est l'endomorphisme défini par, pour tous u et v de \mathbb{R}^n , $\langle ee^* u, v \rangle = \langle e, u \rangle \langle e, v \rangle$ (sa représentation matricielle dans la base canonique est le produit de Kronecker de e par lui-même).

(c) On a $\text{Ker } \Lambda_{X'} = \mathbb{R}e$ et, $\Lambda_{X'}$ étant auto-adjoint, $\text{Im}(\Lambda_{X'}) = (\text{Ker } \Lambda_{X'})^\perp = H$; ainsi X' prend P-p.s. ses valeurs dans H (qui est de dimension $n-1$). Soit alors une isométrie B de \mathbb{R}^{n-1} sur H ; notons i_H l'injection canonique de H dans \mathbb{R}^n . Soit la variable aléatoire $U = B^* i_H^* X' = (i_H B)^* X'$ à valeurs dans \mathbb{R}^{n-1} ; elle est de loi gaussienne $\mathcal{N}_{\mathbb{R}^{n-1}}(E((i_H B)^* X'), (i_H B)^* \Lambda_{X'} (i_H B)) = \mathcal{N}_{\mathbb{R}^{n-1}}(0, \sigma^2 \mathbf{1}_{\mathbb{R}^{n-1}})$, puisque X' est centrée et que l'on a

$$B^* i_H^* \Lambda_{X'} i_H B = \sigma^2 \mathbf{1}_{\mathbb{R}^{n-1}}.$$

Enfin, puisque $BB^* = \mathbf{1}_H$ et que X' prend P-p.s. ses valeurs dans H , on a P-p.s. $X' = BU$.

(d) Il résulte de la première question que les variables aléatoires $\|U\|^2$ et $\frac{U}{\|U\|}$ sont indépendantes; les variables aléatoires $V = \|X'\|^2$ et $Z = \frac{X'}{\|X'\|}$ le sont donc aussi. Puisque les variables aléatoires V et Z sont $\sigma(X')$ -mesurables et que M et X' sont indépendantes, il en résulte que **les variables aléatoires M , V et Z sont indépendantes.**

(e) Puisque

$$EM = \frac{1}{n} \langle EX, e \rangle = \frac{m}{n} \|e\|^2 = m$$

il résulte de (13.74) que M suit la loi $\mathcal{N}_{\mathbb{R}}(m, \frac{\sigma^2}{n})$. Par ailleurs, on a

$$\frac{V}{\sigma^2} = \left\| \frac{X'}{\sigma} \right\|^2 = \left\| \frac{U}{\sigma} \right\|^2 \quad \text{P-p.s. ;}$$

ainsi, U/σ suivant la loi $\mathcal{N}_{\mathbb{R}^{n-1}}(0, \mathbf{1}_{\mathbb{R}^{n-1}})$, V/σ^2 suit la loi du chi-deux à $n-1$ degrés de liberté.

Remarque. Les résultats de ce dernier exercice conduisent au **test de Student**; c'est un test paramétrique. On considère une variable aléatoire réelle X de loi $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$ dont on ne connaît pas les paramètres. On veut tester l'hypothèse que m est inférieur ou égal à une valeur donnée m_0 au vu d'un échantillon (x_1, x_2, \dots, x_n) . Soit (X_1, X_2, \dots, X_n) un échantillon empirique de X , c'est à dire n variables aléatoires indépendantes de même loi que X . On introduit les variables aléatoires centrées normalisées $\hat{X}_i = \frac{X_i - m}{\sigma}$ (elles sont donc de loi $\mathcal{N}_{\mathbb{R}}(0, 1)$) et les moments empiriques associés à cet échantillon, à savoir

$$\begin{aligned} M_n &= \frac{1}{n} \sum_{j=1}^n X_j & \Sigma_n^2 &= \frac{1}{n-1} \sum_{j=1}^n (X_j - M_n)^2, \text{ et} \\ \hat{M}_n &= \frac{1}{n} \sum_{j=1}^n \hat{X}_j & (\hat{\Sigma}_n)^2 &= \frac{1}{n-1} \sum_{j=1}^n (\hat{X}_j - \hat{M}_n)^2; \end{aligned}$$

on a

$$M_n = \sigma \hat{M}_n + m \quad \text{et} \quad \Sigma_n^2 = \sigma^2 (\hat{\Sigma}_n)^2,$$

si bien que, les variables aléatoires $\overset{\circ}{M}_n$ et $(\overset{\circ}{\Sigma}_n)^2$ étant indépendantes, comme on vient de le voir, il en est de même des variables aléatoires M_n et Σ_n^2 ; de plus $\sqrt{n}\overset{\circ}{M}_n$ suit la loi $\mathcal{N}_{\mathbb{R}}(0, 1)$ et $(n-1)(\overset{\circ}{\Sigma}_n)^2$ suit la loi du chi-deux χ_{n-1}^2 .

La variable aléatoire $T_n = \sqrt{n}\overset{\circ}{M}_n / \overset{\circ}{\Sigma}_n$, qui est égale à $\sqrt{n-1} \frac{\sqrt{n}\overset{\circ}{M}_n}{\sqrt{n-1}\overset{\circ}{\Sigma}_n}$, suit alors la loi de Student de paramètre $n-1$ (voir l'exercice du chapitre 9 sur les lois de Student); cette loi est tabulée. Remarquons que

$$T_n = \sqrt{n} \frac{M_n - m}{\Sigma_n}.$$

Le test consiste alors, pour un niveau α donné, à déterminer dans la table la valeur $t_{n-1, \alpha}$ pour laquelle on a $P(T_n \leq t_{n-1, \alpha}) = 1 - \alpha$. Puisque $T_n \leq t_{n-1, \alpha}$ si et seulement si $M_n \in]-\infty, t_{n-1, \alpha} \frac{\Sigma_n}{\sqrt{n}} + m]$, sous l'hypothèse que la vraie valeur (inconnue) de m soit inférieure ou égale à m_0 , on a donc $M_n \in]-\infty, t_{n-1, \alpha} \frac{\Sigma_n}{\sqrt{n}} + m_0]$ avec probabilité supérieure ou égale à $1 - \alpha$. Le test de Student propose d'accepter cette hypothèse, avec probabilité inférieure ou égale à α de se tromper, si l'échantillon est tel que $\bar{x} \in]-\infty, t_{n-1, \alpha} \frac{s}{\sqrt{n}} + m_0]$, où

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j, \quad s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2;$$

on dit qu'on a déterminé une **région de confiance** au seuil ou au niveau α .

Chapitre 14

Convergence de mesures et convergence en loi

Par souci de simplification, nous nous limitons à l'étude de mesures sur \mathbb{R}^d ; tout ce que nous allons dire est en fait valable lorsque l'espace E est métrique, localement compact, dénombrable à l'infini¹, en particulier si E est une partie compacte, une partie ouverte, ou une partie fermée de \mathbb{R}^d , ou d'un espace vectoriel de dimension finie.

La référence fondamentale sur les questions de convergence de mesures est le livre de Billingsley [1]. Les notions topologiques nécessaires peuvent être trouvées, par exemple, dans le livre de J. Dieudonné, *Fondements de l'analyse moderne*².

14.1. Convergence de mesures bornées sur \mathbb{R}^d

On note \mathcal{M} l'ensemble des mesures (positives) **bornées** sur \mathbb{R}^d muni de sa tribu borélienne et, pour $b > 0$, $\mathcal{M}(b)$ le sous-ensemble des mesures μ de **masse** inférieure ou égale à b (c'est-à-dire telles que $\mu(\mathbb{R}^d) \leq b$). Enfin, on note \mathcal{M}^1 l'ensemble des probabilités sur \mathbb{R}^d .

On introduit les trois espaces vectoriels de **fonctions réelles continues** sur \mathbb{R}^d :

- $\mathcal{C}_K(\mathbb{R}^d)$, espace des fonctions continues à **support compact**,
- $\mathcal{C}_0(\mathbb{R}^d)$, espace des fonctions continues **tendant vers 0 à l'infini**³,

1. On dit qu'un espace localement compact est **dénombrable à l'infini** s'il existe une suite $(K_n)_{n \in \mathbb{N}}$ de compacts telle que $K_n \subset \overset{\circ}{K}_{n+1}$ pour tout n , et telle que $\bigcup_{n \in \mathbb{N}} K_n = E$. Les ouverts et les fermés de \mathbb{R}^d sont localement compacts dénombrables à l'infini (pour un ouvert borné de \mathbb{R}^d , on peut prendre pour K_n l'ensemble des points à distance supérieure ou égale à $1/n$ de la frontière de l'ouvert). Si E est compactifié par adjonction d'un point à l'infini (compactifié d'Alexandrov), cela revient à dire que le point à l'infini possède une base dénombrable de voisinages.

2. Dieudonné J. (1965) *Fondements de l'analyse moderne*, Cahiers Scientifiques, fascicule XXVIII, Gauthier-Villars, Éditeur.

3. Dans un espace localement compact E , on dit qu'une fonction réelle f **tend vers 0 à l'infini** si, pour tout $\varepsilon > 0$, il existe un compact K tel que $\sup_{x \in K^c} |f(x)| \leq \varepsilon$. Si de plus E est dénombrable à l'infini, il suffit pour cela que la suite $(f(x_n))_{n \in \mathbb{N}}$ tende vers 0 pour toute suite $(x_n)_{n \in \mathbb{N}}$ tendant vers l'infini (par définition, une suite tend vers l'infini si, pour tout compact K , elle est située dans K^c à partir d'un certain rang).

– $\mathcal{C}_b(\mathbb{R}^d)$, espace des fonctions continues **bornées**.

On a les inclusions

$$\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d) \subset \mathcal{C}_0(\mathbb{R}^d) \subset \mathcal{C}_b(\mathbb{R}^d).$$

Pour la norme, $\|f\| = \sup_{x \in \mathbb{R}^d} |f(x)|$, l'espace $\mathcal{C}_b(\mathbb{R}^d)$ est un espace de Banach. $\mathcal{C}_0(\mathbb{R}^d)$ est un sous-espace fermé de $\mathcal{C}_b(\mathbb{R}^d)$, et $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$ est dense dans $\mathcal{C}_0(\mathbb{R}^d)$. L'espace $\mathcal{C}_0(\mathbb{R}^d)$ est **séparable**⁴, mais $\mathcal{C}_b(\mathbb{R}^d)$ ne l'est pas.

Définition 14.1. Sur \mathcal{M} , on définit respectivement les trois topologies **vague**, **faible et étroite** comme les topologies les moins fines⁵ rendant continues les applications $\mu \mapsto \int f d\mu$ de \mathcal{M} dans \mathbb{R} pour tout $f \in \mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$ (resp. pour tout $f \in \mathcal{C}_0(\mathbb{R}^d)$, resp. pour tout $f \in \mathcal{C}_b(\mathbb{R}^d)$).

En particulier, une suite $(\mu_n)_{n \in \mathbb{N}}$ de mesures bornées **converge** vers la mesure μ ,

$$\left\{ \begin{array}{ll} \text{vaguement} & \text{si } \lim_n \int f d\mu_n = \int f d\mu \quad \text{pour tout } f \in \mathcal{C}_{\mathcal{X}}(\mathbb{R}^d), \\ \text{faiblement} & \text{si } \lim_n \int f d\mu_n = \int f d\mu \quad \text{pour tout } f \in \mathcal{C}_0(\mathbb{R}^d), \\ \text{étroitement} & \text{si } \lim_n \int f d\mu_n = \int f d\mu \quad \text{pour tout } f \in \mathcal{C}_b(\mathbb{R}^d). \end{array} \right.$$

Remarque. Une base de voisinages de μ pour l'une de ces topologies est définie par les ensembles de la forme

$$V_{\varepsilon, f_1, \dots, f_n}(\mu) = \left\{ \nu \in \mathcal{M} \mid \sup_{1 \leq i \leq n} \left| \int f_i d\mu - \int f_i d\nu \right| \leq \varepsilon \right\},$$

où $\varepsilon > 0$ et les f_i appartiennent respectivement à $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$, $\mathcal{C}_0(\mathbb{R}^d)$, $\mathcal{C}_b(\mathbb{R}^d)$.

Il est clair que la topologie vague est moins fine que la topologie faible, laquelle est moins fine que la topologie étroite. En particulier, **une suite qui converge étroitement converge aussi faiblement, une suite qui converge faiblement converge aussi vaguement**.

De plus, la topologie faible sur $\mathcal{M}(1)$ est strictement moins fine que la topologie étroite, comme le montre l'exemple suivant : soit x un vecteur

4. (a) Un espace métrique est **séparable** s'il possède un sous-ensemble **dénombrable** dense.

(b) Un sous-ensemble H d'un espace vectoriel normé est dit **total** si le sous-espace vectoriel engendré par H (c'est-à-dire l'ensemble des combinaisons linéaires finies d'éléments de H) est dense dans E .

(c) Un espace vectoriel normé qui possède un sous-ensemble total H dénombrable est séparable (considérer les combinaisons linéaires finies d'éléments de H à coefficients rationnels).

(d) Il existe dans $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$ et dans $\mathcal{C}_0(\mathbb{R}^d)$ des ensembles **totaux** dénombrables, ce qui n'est pas le cas pour $\mathcal{C}_b(\mathbb{R}^d)$.

5. (a) Ce sont des topologies « **initiales** ».

(b) Une topologie \mathcal{T}_1 sur un ensemble X est **moins fine** qu'une topologie \mathcal{T}_2 sur X si tout ouvert pour \mathcal{T}_1 est un ouvert pour \mathcal{T}_2 (elle possède « moins » d'ouverts que \mathcal{T}_2). Cela revient aussi à dire que l'application identité $(X, \mathcal{T}_2) \rightarrow (X, \mathcal{T}_1)$ est continue.

non nul de \mathbb{R}^d et $\mu_n = \delta_{nx}$; la suite $(\mu_n)_{n \in \mathbb{N}}$ converge faiblement vers la mesure nulle μ puisque, pour tout $f \in \mathcal{C}_0(\mathbb{R}^d)$, on a $\lim_n \int f d\mu_n = \lim_n f(nx) = 0$; toutefois, elle ne converge pas étroitement vers μ , puisque $\lim_n \int 1 d\mu_n = 1$ et que $\int 1 d\mu = 0$. On vient en même temps de montrer que \mathcal{M}^1 n'est pas faiblement fermé (une limite faible de probabilités n'est pas nécessairement une probabilité).

Nous allons **comparer** ces topologies sur $\mathcal{M}(b)$ (et en conséquence, comparer les notions correspondantes de convergence de suites de mesures de masse $\leq b$) et en étudier certaines propriétés.

Proposition 14.2 (Comparaison des topologies). (a) Sur $\mathcal{M}(b)$ les topologies **vague** et **faible** coïncident avec la topologie la moins fine rendant continues les applications $\mu \mapsto \int f d\mu$ lorsque f parcourt un ensemble **total** \mathcal{H} dans $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$ ou $\mathcal{C}_0(\mathbb{R}^d)$.

(b) Sur \mathcal{M}^1 les trois topologies coïncident.

(c) L'espace $\mathcal{M}(b)$ est métrisable et **compact** pour la **topologie faible**.

Démonstration. Remarquons d'abord que si les applications $\mu \mapsto \int f d\mu$ sont continues lorsque f parcourt l'ensemble \mathcal{H} , elles le sont aussi lorsque f parcourt l'espace vectoriel $\widetilde{\mathcal{H}}$ engendré par \mathcal{H} .

(a) Soit une fonction f de $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$ ou $\mathcal{C}_0(\mathbb{R}^d)$ quelconque, et soit $\varepsilon > 0$; il existe une fonction g de $\widetilde{\mathcal{H}}$ telle que $\|f - g\| \leq \frac{\varepsilon}{4b}$. Par l'inégalité triangulaire, on a

$$\begin{aligned} & \left| \int f d\mu - \int f d\nu \right| \\ & \leq \left| \int f d\mu - \int g d\mu \right| + \left| \int g d\mu - \int g d\nu \right| + \left| \int g d\nu - \int f d\nu \right|, \end{aligned}$$

soit

$$\left| \int f d\mu - \int f d\nu \right| \leq 2b \|f - g\| + \left| \int g d\mu - \int g d\nu \right|.$$

Il en résulte que dès que $\nu \in V_{\varepsilon/2, g}(\mu)$, on a $|\int f d\mu - \int f d\nu| \leq \varepsilon$; autrement dit, $V_{\varepsilon/2, g}(\mu) \subset V_{\varepsilon, f}(\mu)$, ce qui démontre que $V_{\varepsilon, f}(\mu)$ est un voisinage de μ pour la topologie initiale associée à $\widetilde{\mathcal{H}}$: puisque de plus, a priori, celle-ci est moins fine que les topologies vague et faible, ceci démontre le résultat.

Remarque. On notera que, par exemple, l'ensemble dénombrable \mathcal{H} des fonctions du type $x^n \exp(-x^2)$, $n \in \mathbb{N}$, est dense dans $\mathcal{C}_b(\mathbb{R})$. Par ailleurs, le même résultat pour la topologie étroite n'a pas d'intérêt: un ensemble \mathcal{H} total dans $\mathcal{C}_b(\mathbb{R}^d)$ n'est pas dénombrable et est très « gros ».

(b) Montrons maintenant que les topologies faible et étroite coïncident sur \mathcal{M}^1 . Pour cela, il suffit de montrer que, si $P \in \mathcal{M}^1$, un voisinage de P du type $V_{\varepsilon, f}(P)$ pour la topologie étroite, où $f \in \mathcal{C}_b(\mathbb{R}^d)$ et $\varepsilon > 0$ sont

quelconques, en est un voisinage faible. Soient donc $f \in \mathcal{C}_b(\mathbb{R}^d)$ et $\varepsilon > 0$ quelconques et soit $(h_p)_{p \in \mathbb{N}}$ une suite de fonctions de $\mathcal{C}_0(\mathbb{R}^d)$ positives qui converge simplement vers 1 en croissant. Pour tout entier p , $f h_p \in \mathcal{C}_0(\mathbb{R}^d)$ et, pour tout $Q \in \mathcal{M}^1$, on a

$$\begin{aligned} & \left| \int f dP - \int f dQ \right| \\ & \leq \left| \int (f - f h_p) dP \right| + \left| \int f h_p dP - \int f h_p dQ \right| + \left| \int (f h_p - f) dQ \right|, \end{aligned}$$

et donc,

$$\begin{aligned} & \left| \int f dP - \int f dQ \right| \\ & \leq \|f\| \int (1 - h_p) dP + \left| \int f h_p dP - \int f h_p dQ \right| + \|f\| \int (1 - h_p) dQ. \end{aligned}$$

Puisque P et Q sont des probabilités, on a

$$\int (1 - h_p) dP + \int (1 - h_p) dQ = 2 \left(1 - \int h_p dP \right) + \int h_p dP - \int h_p dQ,$$

et, a fortiori,

$$0 \leq \int (1 - h_p) dP + \int (1 - h_p) dQ = 2 \left(1 - \int h_p dP \right) + \left| \int h_p dP - \int h_p dQ \right|.$$

Il en résulte que l'on a

$$\begin{aligned} & \left| \int f dP - \int f dQ \right| \\ & \leq \|f\| \left[2 \left(1 - \int h_p dP \right) + \left| \int h_p dP - \int h_p dQ \right| \right] + \left| \int f h_p dP - \int f h_p dQ \right|. \end{aligned}$$

Par convergence monotone décroissante (P est une probabilité), on a

$$\lim_p \int (1 - h_p) dP = 0;$$

on choisit alors p tel que $0 \leq \int (1 - h_p) dP \leq \varepsilon / (4 \|f\|)$. Pour tout $Q \in \mathcal{M}^1 \cap V_{\varepsilon/(4\|f\|), h_p}(P) \cap V_{\varepsilon/2, f h_p}(P)$, on a alors $|\int f dP - \int f dQ| \leq \varepsilon$, et donc, $Q \in \mathcal{M}^1 \cap V_{\varepsilon, f}(P)$; on vient de montrer que $\mathcal{M}^1 \cap V_{\varepsilon/(4\|f\|), h_p}(P) \cap V_{\varepsilon/2, f h_p}(P) \subset \mathcal{M}^1 \cap V_{\varepsilon, f}(P)$, ce qui démontre, puisque h_p et $f h_p$ sont dans $\mathcal{C}_0(\mathbb{R}^d)$, que $\mathcal{M}^1 \cap V_{\varepsilon, f}(P)$ est un voisinage de P pour la topologie faible sur \mathcal{M}^1 . Puisque de plus, a priori, la topologie faible est moins fine que la topologie étroite, ceci démontre leur égalité sur \mathcal{M}^1 .

(c) Choisissons une suite $(f_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$ qui soit dense dans $\mathcal{C}_0(\mathbb{R}^d)$. On définit la distance d sur $\mathcal{M}(b)$ par

$$d(\mu, \nu) = \sum_{n=1}^{+\infty} \frac{1}{2^n \|f_n\|} \left| \int f_n d\mu - \int f_n d\nu \right|.$$

Il s'agit bien d'une distance : en effet $d(\mu, \nu)$ a toujours une valeur finie, la symétrie et l'inégalité triangulaire se vérifient immédiatement. De plus, si $d(\mu, \nu) = 0$, on a, pour tout n , $\int f_n d\mu = \int f_n d\nu$, et donc aussi, par densité, $\int f d\mu = \int f d\nu$ pour tout $f \in \mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$, ce qui démontre que $\mu = \nu$ (cf. chap. 8, corollaire 8.5). D'après (a) (prendre pour \mathcal{H} l'ensemble total constitué par les fonctions $f_n, n \in \mathbb{N}$), la topologie induite par d coïncide avec la topologie faible.

Pour démontrer que $\mathcal{M}(b)$ est faiblement compact, il suffit alors de démontrer que de toute suite, on peut extraire une sous-suite convergente. Soit donc, pour tout $p, \mu_p \in \mathcal{M}(b)$. On va utiliser le **procédé diagonal** pour extraire de la suite $(\mu_p)_{p \in \mathbb{N}}$ une sous-suite convergente. Pour tout n , la suite de réels $(\int f_n d\mu_p)_{p \in \mathbb{N}}$ est **bornée** par $b \|f_n\|$. On peut alors extraire une sous-suite convergente de la suite $(\int f_1 d\mu_p)_{p \in \mathbb{N}}$; notons φ_1 l'injection croissante de \mathbb{N} dans lui-même qui définit la suite extraite. Pour la même raison, on peut extraire une sous-suite convergente de la suite $(\int f_2 d\mu_{\varphi_1(p)})_{p \in \mathbb{N}}$; notons φ_2 l'injection croissante de \mathbb{N} dans lui-même qui définit la suite extraite; les suites $(\int f_1 d\mu_{\varphi_2(p)})_{p \in \mathbb{N}}$ et $(\int f_2 d\mu_{\varphi_2(p)})_{p \in \mathbb{N}}$ sont alors convergentes. Par récurrence, on construit de même pour tout entier k la suite $(\mu_{\varphi_k(p)})_{p \in \mathbb{N}}$, sous-suite de $(\mu_{\varphi_{k-1}(p)})_{p \in \mathbb{N}}$ telle que les suites $(\int f_i d\mu_{\varphi_k(p)})_{p \in \mathbb{N}}$ pour tout $i \leq k$ soient convergentes. Alors, pour tout entier k , la suite $(\int f_k d\mu_{\varphi_p(p)})_{p \in \mathbb{N}}$ est convergente : c'est à partir du rang k une sous-suite de la suite convergente $(\int f_k d\mu_{\varphi_k(p)})_{p \in \mathbb{N}}$.

Par densité, on a aussi que, pour tout $f \in \mathcal{C}_0(\mathbb{R}^d)$, la suite $(\int f d\mu_{\varphi_k(k)})_{k \in \mathbb{N}}$ est convergente de limite $\Psi(f)$. L'application Ψ est alors une forme linéaire **positive** sur $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$; le théorème de **Riesz**⁶ (cf. par exemple, Métivier [18], p. 87) assure qu'il existe une mesure unique μ telle que $\Psi(f) = \int f d\mu$ pour tout $f \in \mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$. Par densité de $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$ dans $\mathcal{C}_0(\mathbb{R}^d)$, on a alors $\lim_k \int f d\mu_{\varphi_k(k)} = \int f d\mu$ pour tout $f \in \mathcal{C}_0(\mathbb{R}^d)$, c'est-à-dire que la sous-suite $(\mu_{\varphi_k(k)})_{k \in \mathbb{N}}$ converge faiblement vers μ . Il reste à vérifier que $\mu \in \mathcal{M}(b)$. Soit $(h_p)_{p \in \mathbb{N}}$ une suite de fonctions de $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$ positives qui converge simplement vers 1 en croissant; on a, pour tout p ,

$$0 \leq \lim_k \int h_p d\mu_{\varphi_k(k)} = \int h_p d\mu \leq b.$$

et donc, par convergence monotone,

6. **Théorème de Riesz.** Soit Φ une forme linéaire positive sur $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$. Il existe une mesure unique sur \mathbb{R}^d , muni de sa tribu borélienne, qui représente Φ , c'est-à-dire telle que l'on ait, pour tout $f \in \mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$, $\Phi(f) = \int_{\mathbb{R}^d} f d\mu$; cette mesure possède les propriétés suivantes :

- (i) μ est finie sur tout compact (on dit que c'est une mesure de Radon),
- (ii) pour tout $B \in \mathcal{B}_{\mathbb{R}^d}$, on a

$$\mu(B) = \inf \{ \mu(O) \mid O \text{ ouvert, } O \supset B \}; \quad \mu(B) = \sup \{ \mu(K) \mid K \text{ compact, } K \subset B \}.$$

$$\lim_p \int h_p d\mu = \mu(\mathbb{R}^d) \leq b. \quad \square$$

Notation. On note traditionnellement $\mu_n \implies \mu$ le fait que la suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ .

Remarque. En conséquence de la proposition précédente, pour qu'une suite $(\mu_n)_{n \in \mathbb{N}}$ de mesures de $\mathcal{M}(b)$ converge **faiblement** vers μ , il faut et il suffit que la suite $(\int f d\mu_n)_{n \in \mathbb{N}}$ converge vers $\int f d\mu$ pour tout f parcourant un ensemble total \mathcal{H} de $\mathcal{C}_b(\mathbb{R}^d)$, par exemple $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$ lui-même. De plus, si les μ_n et μ sont des **probabilités**, pour que la suite $(\mu_n)_{n \in \mathbb{N}}$ converge **étroitement** vers μ , il faut et il suffit que la suite $(\int f d\mu_n)_{n \in \mathbb{N}}$ converge vers $\int f d\mu$ pour tout f parcourant un ensemble total \mathcal{H} de $\mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$.

Proposition 14.3. Soient μ_n , $n \in \mathbb{N}$, et μ des mesures de $\mathcal{M}(b)$ telles que la suite $(\mu_n)_{n \in \mathbb{N}}$ converge vaguement (ou faiblement) vers μ . La suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ si et seulement si $\lim_n \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$.

Démonstration. La condition est évidemment nécessaire. Pour la **condition suffisante**, soient $f \in \mathcal{C}_b(\mathbb{R}^d)$ et $\varepsilon > 0$ quelconques. La mesure μ étant de masse finie, on peut choisir $\varphi \in \mathcal{C}_{\mathcal{X}}^+(\mathbb{R}^d)$ telle

$$\|\varphi\| \leq 1 \quad \text{et} \quad 0 < \int (1 - \varphi) d\mu \leq \frac{\varepsilon}{8\|f\|}.$$

Puisque $f\varphi \in \mathcal{C}_{\mathcal{X}}(\mathbb{R}^d)$, il existe un entier N_1 tel que, pour tout $n \geq N_1$, on ait

$$\left| \int f\varphi d\mu_n - \int f\varphi d\mu \right| \leq \frac{\varepsilon}{2}.$$

Par ailleurs, puisque

$$\int (1 - \varphi) d\mu_n = \mu_n(\mathbb{R}^d) - \int \varphi d\mu_n,$$

il résulte des hypothèses que la suite de terme général $\int (1 - \varphi) d\mu_n$ converge vers $\mu(\mathbb{R}^d) - \int \varphi d\mu = \int (1 - \varphi) d\mu$. Il existe donc un entier N_2 tel que, pour tout $n \geq N_2$, on ait

$$0 \leq \int (1 - \varphi) d\mu_n \leq \frac{\varepsilon}{4\|f\|}.$$

Pour tout $n \geq N = \max(N_1, N_2)$, on a, par l'inégalité triangulaire,

$$\begin{aligned} & \left| \int f d\mu_n - \int f d\mu \right| \\ & \leq \left| \int (f - f\varphi) d\mu_n \right| + \left| \int f\varphi d\mu_n - \int f\varphi d\mu \right| + \left| \int (f\varphi - f) d\mu \right|; \end{aligned}$$

on a, a fortiori,

$$\begin{aligned} & \left| \int f d\mu_n - \int f d\mu \right| \\ & \leq \|f\| \int (1-\varphi) d\mu_n + \left| \int f\varphi d\mu_n - \int f\varphi d\mu \right| + \|f\| \int (1-\varphi) d\mu, \end{aligned}$$

et, par conséquent,

$$\left| \int f d\mu_n - \int f d\mu \right| \leq \varepsilon,$$

ce qui démontre que la suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ . \square

Remarque. Sur l'espace mesurable $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ la suite des mesures de Dirac $\delta_{1/n}$ converge étroitement vers la mesure de Dirac δ_0 ; toutefois, on a, pour tout $n \in \mathbb{N}^*$, $\delta_{1/n}(\{0\}) = 0$ et par conséquent, on a $\lim_n \delta_{1/n}(\{0\}) \neq \delta_0(\{0\})$. Ainsi la **convergence étroite d'une suite de mesures $(\mu_n)_{n \in \mathbb{N}}$ vers une mesure μ n'entraîne pas la convergence, pour tout borélien A , de la suite $(\mu_n(A))_{n \in \mathbb{N}}$** . La proposition 14.5 suivante donne une réponse à ce problème de convergence.

Définition 14.4. Soit μ une mesure sur \mathbb{R}^d . Un borélien A est dit de μ -**continuité** si $\mu(\partial A) = 0$, où ∂A désigne la frontière (ou le bord) de A .

Proposition 14.5 (Critères de convergence étroite). Soient μ_n , $n \in \mathbb{N}$, et μ des mesures de $\mathcal{M}(b)$. Les assertions suivantes sont équivalentes :

- (i) La suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ .
- (ii) Pour tout **fermé** F , on a

$$\limsup_n \mu_n(F) \leq \mu(F).$$

et de plus, on a

$$\lim_n \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d).$$

- (iii) Pour tout **ouvert** O , on a

$$\liminf_n \mu_n(O) \geq \mu(O).$$

et de plus, on a

$$\lim_n \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d).$$

- (iv) Pour tout **borélien** A de μ -**continuité**, on a

$$\lim_n \mu_n(A) = \mu(A).$$

Démonstration. On fait la démonstration suivant le schéma suivant :

$$(i) \Leftrightarrow (ii) \Leftrightarrow (iii) \Rightarrow (iv) \Rightarrow (ii).$$

(i) \Rightarrow (ii). On définit pour tout $j \in \mathbb{N}^*$ la fonction $\varphi_j \in \mathcal{C}_b(\mathbb{R})$ par

$$\varphi_j(u) = \begin{cases} 1 & \text{si } u \leq 0, \\ 1 - ju & \text{si } 0 < u < \frac{1}{j}, \\ 0 & \text{si } u \geq \frac{1}{j}; \end{cases}$$

la suite de fonctions $(\varphi_j)_{j \in \mathbb{N}^*}$ est décroissante et converge simplement vers $\mathbf{1}_{]-\infty, 0]}$. Soit un fermé F ; on définit alors la fonction $f_j \in \mathcal{C}_b(\mathbb{R}^d)$ en posant, pour $x \in \mathbb{R}^d$, $f_j(x) = \varphi_j(d(x, F))$. Alors, pour tout $x \in \mathbb{R}^d$, $\lim_j \searrow f_j(x) = \mathbf{1}_{]-\infty, 0]}(d(x, F)) = \mathbf{1}_F(x)$; les fonctions f_j étant bornées par 1 et la mesure μ étant bornée, on peut appliquer le théorème de convergence monotone décroissante, ce qui donne $\lim_j \int f_j d\mu = \mu(F)$. Soit $\varepsilon > 0$ quelconque; il existe alors j_0 tel que

$$\mu(F) \leq \int f_{j_0} d\mu \leq \mu(F) + \varepsilon.$$

Puisque, par hypothèse, la suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ , on a alors

$$\lim_n \int f_{j_0} d\mu_n = \int f_{j_0} d\mu \leq \mu(F) + \varepsilon.$$

Mais puisque, pour tout $n \in \mathbb{N}$, on a $\mu_n(F) \leq \int f_{j_0} d\mu_n$, il vient

$$\limsup_n \mu_n(F) \leq \limsup_n \int f_{j_0} d\mu_n \leq \mu(F) + \varepsilon,$$

ce qui, étant donné l'arbitraire de $\varepsilon > 0$, démontre que $\limsup_n \mu_n(F) \leq \mu(F)$.

(ii) \Rightarrow (iii). Si O est un ouvert, μ étant bornée, on passe au complémentaire: on a, pour tout n ,

$$\mu_n(O) = \mu_n(\mathbb{R}^d) - \mu_n(O^c),$$

et donc, en prenant les limites inférieures et en utilisant l'hypothèse,

$$\liminf_n \mu_n(O) = \liminf_n \mu_n(\mathbb{R}^d) - \limsup_n \mu_n(O^c) \geq \mu(\mathbb{R}^d) - \mu(O^c) = \mu(O).$$

(iii) \Rightarrow (ii). On raisonne de la même manière en passant au complémentaire.

(iii) \Rightarrow (iv). Soit A un borélien de μ -continuité; puisque les assertions (ii) et (iii) sont équivalentes, en utilisant la croissance des mesures et les inclusions $\overset{\circ}{A} \subset A \subset \overline{A}$, on a successivement les inégalités

$$\begin{aligned} \mu(\overset{\circ}{A}) &\leq \liminf_n \mu_n(\overset{\circ}{A}) \leq \liminf_n \mu_n(A) \\ &\leq \limsup_n \mu_n(A) \leq \limsup_n \mu_n(\overline{A}) \leq \mu(\overline{A}). \end{aligned}$$

Mais, puisque $\partial A = \overline{A} \setminus \overset{\circ}{A}$ et que $\mu(\partial A) = 0$, on a $\mu(\overline{A}) = \mu(\overset{\circ}{A}) = \mu(A)$, ce qui démontre que

$$\liminf_n \mu_n(A) = \limsup_n \mu_n(A) = \mu(A),$$

c'est-à-dire que la suite de terme général $\mu_n(A)$ converge vers $\mu(A)$.

(ii) (et donc (iii)) \Rightarrow (i). Préalablement, on établit une relation du même type que celle donnant la moyenne d'une variable aléatoire positive comme intégrale du complément à 1 de sa fonction de répartition, à savoir : si X est une variable aléatoire positive, on a

$$\int_{\Omega} X dP = \int_{\mathbb{R}} P(X > x) d\lambda(x) = \int_{\mathbb{R}} P(X \geq x) d\lambda(x).$$

Il résulte du théorème de Fubini que, pour tout $f \in \mathcal{C}_b^+(\mathbb{R})$ et tout $\mu \in \mathcal{M}(b)$, on a, en notant λ la mesure de Lebesgue sur \mathbb{R} ,

$$\begin{aligned} \int_{[0, \|f\|]} \mu(f \geq u) d\lambda(u) &= \int_{[0, \|f\|]} \left[\int_{\mathbb{R}^d} \mathbf{1}_{(f(x) \geq u)} d\mu(x) \right] d\lambda(u) \\ &= \int_{\mathbb{R}^d} \left[\int_{[0, \|f\|]} \mathbf{1}_{[0, \|f\|]}(u) \mathbf{1}_{(f(x) \geq u)} d\lambda(u) \right] d\mu(x), \end{aligned}$$

soit

$$\int_{[0, \|f\|]} \mu(f \geq u) d\lambda(u) = \int_{\mathbb{R}^d} f d\mu; \quad (14.1)$$

on a de même

$$\int_{[0, \|f\|]} \mu(f > u) d\lambda(u) = \int_{\mathbb{R}^d} f d\mu. \quad (14.2)$$

Soit alors $f \in \mathcal{C}_b^+(\mathbb{R}^d)$ quelconque. La mesure $\lambda|_{[0, \|f\|]}$ étant bornée, le lemme de Fatou-Lebesgue donne

$$\begin{aligned} \limsup_n \int_{\mathbb{R}^d} f d\mu_n &= \limsup_n \int_{[0, \|f\|]} \mu_n(f \geq u) d\lambda(u) \\ &\leq \int_{[0, \|f\|]} \limsup_n \mu_n(f \geq u) d\lambda(u); \end{aligned}$$

puisque $(f \geq u)$ est **fermé** et qu'on suppose l'assertion (ii) vraie, on obtient

$$\limsup_n \int_{\mathbb{R}^d} f d\mu_n \leq \int_{[0, \|f\|]} \mu(f \geq u) d\lambda(u) = \int_{\mathbb{R}^d} f d\mu.$$

On a de même

$$\begin{aligned} \liminf_n \int_{\mathbb{R}^d} f d\mu_n &= \liminf_n \int_{[0, \|f\|]} \mu_n(f > u) d\lambda(u) \\ &\geq \int_{[0, \|f\|]} \liminf_n \mu_n(f > u) d\lambda(u); \end{aligned}$$

puisque $(f > u)$ est **ouvert** et qu'on suppose l'assertion (iii) vraie, on obtient

$$\liminf_n \int_{\mathbb{R}^d} f d\mu_n \geq \int_{\{0, \|f\|\}} \mu(f > u) d\lambda(u) = \int_{\mathbb{R}^d} f d\mu.$$

Au total, on a démontré que $\lim_n \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$ pour tout $f \in \mathcal{C}_b^+(\mathbb{R})$; par linéarité on a cette relation pour tout $f \in \mathcal{C}_b(\mathbb{R})$ de signe quelconque (il suffit d'appliquer le résultat précédent à la fonction positive $\|f\| - f$).

(iv) \Rightarrow (ii). Soit F un fermé quelconque. Soit, pour tout $\varepsilon > 0$, le fermé $F_\varepsilon = \{x \in \mathbb{R}^d \mid d(x, F) \leq \varepsilon\}$. L'application Φ de $[0, 1]$ dans $[0, b]$ définie par $\Phi(\varepsilon) = \mu(F_\varepsilon)$ est croissante et bornée; elle admet un ensemble I de points de discontinuité au plus dénombrable. Pour tout $\varepsilon \in [0, 1] \setminus I$, F_ε étant fermé, on a

$$\partial F_\varepsilon = F_\varepsilon \setminus \left[\bigcup_{n \in \mathbb{N}^*} F_{\varepsilon - \frac{1}{n}} \right],$$

ce qui donne

$$\mu(\partial F_\varepsilon) = \mu(F_\varepsilon) - \lim_n \mu(F_{\varepsilon - \frac{1}{n}});$$

ε étant point de continuité de Φ , il vient $\mu(\partial F_\varepsilon) = 0$. Ainsi, on peut trouver une suite $(\varepsilon_k)_{k \in \mathbb{N}^*}$ décroissant vers 0 et telle que, pour tout $k \in \mathbb{N}^*$, on ait $\mu(\partial F_{\varepsilon_k}) = 0$. Par hypothèse, on a alors, pour tout $k \in \mathbb{N}^*$,

$$\limsup_n \mu_n(F) \leq \limsup_n \mu_n(F_{\varepsilon_k}) = \lim_n \mu_n(F_{\varepsilon_k}) = \mu(F_{\varepsilon_k}).$$

Comme $F = \bigcap_{k \in \mathbb{N}^*} F_{\varepsilon_k}$ et que la suite des ensembles F_{ε_k} est décroissante, on a $\lim_k \mu(F_{\varepsilon_k}) = \mu(F)$, ce qui donne

$$\limsup_n \mu_n(F) \leq \mu(F).$$

Enfin, \mathbb{R}^d étant un ensemble de μ -continuité, on a bien $\lim_n \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$. \square

Remarque. Sur l'espace mesurable $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, on considère, pour tout $n \in \mathbb{N}^*$, la mesure $\mu_n = \frac{1}{n} \sum_{j=1}^n \delta_{j/n}$. Cette suite de mesure converge étroitement vers $\mathbf{1}_{[0,1]} \cdot \lambda$, où λ est la mesure de Lebesgue sur \mathbb{R} . En effet, pour tout $f \in \mathcal{C}_b(\mathbb{R})$, on a

$$\lim_n \int_{\mathbb{R}} f d\mu_n = \lim_n \frac{1}{n} \sum_{j=1}^n f\left(\frac{j}{n}\right) = \int_0^1 f(x) dx = \int_{\mathbb{R}} f \mathbf{1}_{[0,1]} d\lambda :$$

(la somme $\sum_{j=1}^n \frac{1}{n} f\left(\frac{j}{n}\right)$ est une somme de Riemann associée à f et à la subdivision de $[0, 1]$ déterminée par les points j/n). Toutefois, puisque pour tout n on a $\mu_n(\mathbb{Q} \cap [0, 1]) = 1$, on a $\lim_n \mu_n(\mathbb{Q} \cap [0, 1]) = 1$, tandis que l'on a

$\mathbf{1}_{[0,1]} \cdot \lambda(\mathbb{Q} \cap [0, 1]) = 0$; ce qui n'est pas en contradiction avec la proposition précédente, puisque $\mathbb{Q} \cap [0, 1]$ n'est pas un ensemble de $\mathbf{1}_{[0,1]} \cdot \lambda$ -continuité (on a $\partial(\mathbb{Q} \cap [0, 1]) = [0, 1]$ et donc $\mathbf{1}_{[0,1]} \cdot \lambda[\partial(\mathbb{Q} \cap [0, 1])] = 1$).

Définition 14.6. Une suite $(\mu_n)_{n \in \mathbb{N}}$ de mesures de $\mathcal{M}(b)$ est **tendue** si, pour tout $\varepsilon > 0$, il existe un compact K de \mathbb{R}^d tel que

$$\sup_{n \in \mathbb{N}} \mu_n(K^c) \leq \varepsilon.$$

Corollaire 14.7. Si la suite $(\mu_n)_{n \in \mathbb{N}}$ de mesures de $\mathcal{M}(b)$ converge étroitement vers $\mu \in \mathcal{M}(b)$, elle est **tendue**.

Démonstration. Soient $\varepsilon > 0$ quelconque et une boule **ouverte** O telle que $\mu(O) \geq \mu(\mathbb{R}^d) - \varepsilon/2$. Puisque $\lim_n \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$, il existe un entier N_1 tel que, pour tout $n \geq N_1$, on ait $\mu_n(\mathbb{R}^d) \leq \mu(\mathbb{R}^d) + \varepsilon/2$. Par ailleurs, il résulte de la proposition 14.5 que,

$$\liminf_n \mu_n(O) \geq \mu(O) \geq \mu(\mathbb{R}^d) - \frac{\varepsilon}{2};$$

il existe donc un entier N_2 tel que, pour tout $n \geq N_2$, on ait $\mu_n(O) \geq \mu(\mathbb{R}^d) - \varepsilon/2$. Posons $N = \max(N_1, N_2)$; pour tout $n \geq N$, on a alors

$$\mu_n(O^c) = \mu_n(\mathbb{R}^d) - \mu_n(O) \leq \left[\mu(\mathbb{R}^d) + \frac{\varepsilon}{2} \right] - \mu(\mathbb{R}^d) + \frac{\varepsilon}{2} = \varepsilon.$$

Il en résulte que

$$\sup_{n \geq N} \mu_n(O^c) \leq \varepsilon. \quad (14.3)$$

Il reste à choisir un compact K contenant O tel que l'on ait $\mu_n(K^c) \leq \varepsilon$ dès que $0 \leq n \leq N$, ce qui est possible puisque, pour de tels n , en nombre fini, on a $\lim_p \mu_n(B_f(0, p)^c) = 0$, où $B_f(0, p)$ désigne la boule fermée de centre 0 et de rayon p ; ainsi, puisque $K^c \subset O^c$, il résulte de 14.3 que l'on a $\sup_{n \in \mathbb{N}} \mu_n(K^c) \leq \varepsilon$. \square

Le célèbre et important théorème de **Paul Lévy** donne une **caractérisation de la convergence étroite** d'une suite de mesures **en termes de transformées de Fourier**.

Théorème 14.8 (Théorème de Lévy). Soit, pour tout $n \in \mathbb{N}$, une mesure $\mu_n \in \mathcal{M}(b)$.

(a) Si la suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ , alors la suite $(\widehat{\mu}_n)_{n \in \mathbb{N}}$ des transformées de Fourier de μ_n converge simplement vers $\widehat{\mu}$, transformée de Fourier de μ .

(b) Inversement, si la suite $(\widehat{\mu}_n)_{n \in \mathbb{N}}$ des transformées de Fourier des μ_n converge simplement vers une fonction φ **continue en 0**, alors il existe une unique mesure $\mu \in \mathcal{M}(b)$ telle que $\varphi = \widehat{\mu}$; de plus, la suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ .

(c) En fait, dans l'un quelconque de ces cas, la convergence de la suite $(\widehat{\mu}_n)_{n \in \mathbb{N}}$ est uniforme sur tout compact.

Démonstration. Observons tout d'abord que si une suite $(\mu_n)_{n \in \mathbb{N}}$ de mesures bornées sur \mathbb{R}^d converge faiblement (resp. étroitement) vers une mesure μ , on a $\lim_n \int f d\mu_n = \int f d\mu$ si f est une fonction continue sur \mathbb{R}^d tendant vers zéro à l'infini (resp. une fonction continue bornée sur \mathbb{R}^d) et à valeurs **complexes** : il suffit de remarquer que la convergence a lieu pour $\Re f$ et $\Im f$. Dans cette démonstration, la notation $\mathcal{C}_0(\mathbb{R}^d)$ désignera l'espace des fonctions continues définies sur \mathbb{R}^d et à valeurs complexes.

(a) Pour tout $t \in \mathbb{R}^d$, la fonction $\exp i(\cdot, t)$ est continue bornée, et la suite de terme général $\widehat{\mu}_n(t)$ converge donc vers $\widehat{\mu}(t)$.

(b) Montrons d'abord que la suite $(\mu_n)_{n \in \mathbb{N}}$ est **faiblement convergente**. Puisque $\mathcal{M}(b)$ est métrisable et compact pour la topologie faible, pour qu'il en soit ainsi, il faut et il suffit que cette suite admette au plus une valeur d'adhérence faible.

On sait en effet que dans un espace métrique compact toute suite possède *au moins* un point adhérent, et qu'une suite qui n'admet qu'un seul point adhérent converge vers ce point.

Soit donc μ une valeur d'adhérence faible de la suite $(\mu_n)_{n \in \mathbb{N}}$ et soit $(\mu_{\psi(n)})_{n \in \mathbb{N}}$ une sous-suite qui converge faiblement vers μ (ψ est l'injection de \mathbb{N} dans \mathbb{N} qui définit la sous-suite). On va démontrer que $\mu_{\psi(n)}$ tend **étroitement** vers μ quand n tend vers l'infini, ce qui d'après (a) assurera que la suite de terme général $\widehat{\mu_{\psi(n)}}$ converge simplement vers $\widehat{\mu}$. Puisque par hypothèse $\widehat{\mu}_n$ tend simplement vers φ quand n tend vers l'infini, il en est de même pour toute sous-suite, et on aura $\widehat{\mu} = \varphi$. L'unicité d'une valeur d'adhérence faible μ résultera alors de l'injectivité de la transformation de Fourier et on aura démontré la convergence faible de la suite $(\mu_n)_{n \in \mathbb{N}}$ vers μ .

Il nous faut donc montrer que la suite $(\mu_{\psi(n)})_{n \in \mathbb{N}}$ converge étroitement vers μ , et pour cela il suffit, puisqu'on a déjà la convergence faible, de montrer que

$$\lim_n \mu_{\psi(n)}(\mathbb{R}^d) = \mu(\mathbb{R}^d)$$

(voir prop. 14.3). Mais on sait par hypothèse que

$$\lim_n \mu_{\psi(n)}(\mathbb{R}^d) = \lim_n \mu_n(\mathbb{R}^d) = \lim_n \widehat{\mu}_n(0) = \varphi(0).$$

Puisque $\mu(\mathbb{R}^d) = \widehat{\mu}(0)$, il nous suffit donc de montrer que $\widehat{\mu}(0) = \varphi(0)$;

Pour cela, on observe d'abord que pour $\varepsilon > 0$ on a

$$\lim_n \int_{[0, \varepsilon]^d} \widehat{\mu}_n(t) dt = \int_{[0, \varepsilon]^d} \varphi(t) dt. \quad (14.4)$$

En effet, puisque $\widehat{\mu}_n$ tend simplement vers φ quand n tend vers l'infini et que les fonctions $\widehat{\mu}_n$ sont bornées en module par b , cela résulte du théorème de convergence dominée.

On utilise alors le lemme suivant :

Lemme. Soit $\varepsilon > 0$. Il existe une fonction $f_\varepsilon \in \mathcal{C}_0(\mathbb{R}^d)$ telle que pour toute mesure bornée ν sur \mathbb{R}^d on ait

$$\int_{[0, \varepsilon]^d} \widehat{\nu}(t) dt = \int_{\mathbb{R}^d} f_\varepsilon d\nu. \quad (14.5)$$

Démonstration du lemme. On a, par le théorème de Fubini,

$$\begin{aligned} \int_{[0, \varepsilon]^d} \widehat{\nu}(t) dt &= \int_{[0, \varepsilon]^d} \left[\int_{\mathbb{R}^d} \exp(i \langle x, t \rangle) d\nu(x) \right] dt \\ &= \int_{\mathbb{R}^d} \left[\int_{[0, \varepsilon]^d} \exp(i \langle x, t \rangle) dt \right] d\nu(x); \end{aligned}$$

or, toujours par le théorème de Fubini, on a

$$\int_{[0, \varepsilon]^d} \exp(i \langle x, t \rangle) dt = \prod_{j=1}^d \left[\int_0^\varepsilon \exp(ix_j t_j) dt_j \right].$$

On obtient donc (14.5) en posant, pour $u \in \mathbb{R}$,

$$g_\varepsilon(u) = \int_{[0, \varepsilon]} \exp(iut) dt = \begin{cases} \frac{\exp(i\varepsilon u) - 1}{iu} & \text{si } u \neq 0, \\ \varepsilon & \text{si } u = 0. \end{cases}$$

et, pour $x \in \mathbb{R}^d$, $f_\varepsilon(x) = \prod_{j=1}^d g_\varepsilon(x_j)$. Il est clair que $f_\varepsilon \in \mathcal{C}_0(\mathbb{R}^d)$.

Suite de la démonstration de (b). Puisque la suite de terme général $\widehat{\mu}_{\psi(n)}$ converge faiblement vers μ et que $f_\varepsilon \in \mathcal{C}_0(\mathbb{R}^d)$, on a

$$\lim_n \int f_\varepsilon d\widehat{\mu}_{\psi(n)} = \int f_\varepsilon d\mu,$$

d'où, d'après le dernier lemme,

7. La relation (14.5) est un cas particulier de la relation $\int \widehat{\mu} d\nu = \int \widehat{\nu} d\mu$, valable pour des mesures bornées μ, ν quelconques sur \mathbb{R}^d . Lorsque μ est une mesure de densité h par rapport à la mesure de Lebesgue ($h \in L^1(\mathbb{R}^d)$), on pose $\widehat{h} = \widehat{\mu}$ et on dit que \widehat{h} est la transformée de Fourier de la fonction h ; on a $\int \widehat{h} d\nu = \int \widehat{\nu}(t)h(t) dt$. On obtient (14.5) en prenant $h = 1_{[0, \varepsilon]^d}$. $\widehat{h} = f_\varepsilon$. Le fait que $\widehat{h} \in \mathcal{C}_0(\mathbb{R}^d)$ est un fait général (lemme de Riemann-Lebesgue).

$$\lim_n \int_{[0, \varepsilon]^d} \widehat{\mu_{\psi(n)}}(t) dt = \int_{[0, \varepsilon]^d} \widehat{\mu}(t) dt.$$

D'après (14.4) appliqué à la sous-suite $(\mu_{\psi(n)})_{n \in \mathbb{N}}$, on a

$$\frac{1}{\varepsilon^d} \int_{[0, \varepsilon]^d} \widehat{\mu}(t) dt = \frac{1}{\varepsilon^d} \int_{[0, \varepsilon]^d} \varphi(t) dt.$$

Grâce à la continuité de $\widehat{\mu}$ et de φ , on obtient en prenant la limite pour ε tendant vers zéro des deux membres de l'égalité précédente, la relation

$$\widehat{\mu}(0) = \varphi(0). \quad (14.6)$$

On a montré la convergence faible de la suite $(\mu_n)_{n \in \mathbb{N}}$. Enfin, il résulte de (14.6) et de la convergence simple de la suite $(\widehat{\mu}_n)_{n \in \mathbb{N}}$ vers φ que

$$\lim_n \widehat{\mu}_n(0) = \widehat{\mu}(0)$$

ou, autrement dit, que

$$\lim_n \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d).$$

Ceci achève de montrer la convergence **étroite** de la suite $(\mu_n)_{n \in \mathbb{N}}$ vers sa limite faible μ .

(c) D'après le corollaire 14.7, la suite $(\mu_n)_{n \in \mathbb{N}}$, qui converge étroitement, est tendue. Soit $\varepsilon > 0$ quelconque; on choisit alors un compact K_ε tel que

$$\sup_{n \in \mathbb{N}} \mu_n(K_\varepsilon^c) \leq \frac{\varepsilon}{3}.$$

On a, pour tout $n \in \mathbb{N}$, et tous t, t' de \mathbb{R}^d ,

$$\begin{aligned} |\widehat{\mu}_n(t) - \widehat{\mu}_n(t')| &= \left| \int_{\mathbb{R}^d} [\exp(i \langle x, t \rangle) - \exp(i \langle x, t' \rangle)] d\mu_n(x) \right| \\ &\leq \int_{K_\varepsilon} |\exp(i \langle x, t \rangle) - \exp(i \langle x, t' \rangle)| d\mu_n(x) + 2\mu_n(K_\varepsilon^c), \end{aligned}$$

soit, par l'inégalité des accroissements finis,

$$|\widehat{\mu}_n(t) - \widehat{\mu}_n(t')| \leq \int_{K_\varepsilon} |\langle x, t - t' \rangle| d\mu_n(x) + 2\mu_n(K_\varepsilon^c);$$

on a donc, pour tous t, t' de \mathbb{R}^d ,

$$\sup_{n \in \mathbb{N}} |\widehat{\mu}_n(t) - \widehat{\mu}_n(t')| \leq \|t - t'\| \left[b \sup_{x \in K_\varepsilon} \|x\| \right] + \frac{2\varepsilon}{3},$$

et, pour tous t, t' de \mathbb{R}^d tels que $\|t - t'\| \leq \frac{\varepsilon}{3b \sup_{x \in K_\varepsilon} \|x\|}$,

$$\sup_{n \in \mathbb{N}} |\widehat{\mu}_n(t) - \widehat{\mu}_n(t')| \leq \varepsilon.$$

Autrement dit la suite des fonctions $\widehat{\mu}_n$ est **équi-continue** (uniformément en t) ; puisqu'elle converge simplement, elle converge uniformément sur tout compact. \square

14.2. Convergence en loi

Toujours par souci de simplification, nous nous limitons à l'étude de variables aléatoires à valeurs dans \mathbb{R}^d ; tout ce que nous allons dire est en fait valable lorsque les variables aléatoires sont à valeurs dans un espace métrique E localement compact et dénombrable à l'infini.

Définition 14.9. Soit pour tout $n \in \mathbb{N}$ une variable aléatoire X_n définie sur un espace probabilisé $(\Omega_n, \mathcal{A}_n, \mathbb{P}^n)$, à valeurs dans \mathbb{R}^d , et soit X une **variable aléatoire** définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans \mathbb{R}^d . La suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ **converge en loi vers** X si la suite $(\mathbb{P}_{X_n}^n)_{n \in \mathbb{N}}$ des lois des X_n converge étroitement vers la loi \mathbb{P}_X de X .

Notation : $X_n \xrightarrow{\mathcal{L}} X$.

Remarque. Cette notion de convergence n'est pas relative aux variables aléatoires en tant qu'applications, mais concerne les **lois** des variables aléatoires ; elle fournit en particulier une notion d'**approximation pour les lois** de variables aléatoires. Il est à remarquer que les variables aléatoires ne sont pas nécessairement définies sur le même espace probabilisé. En pratique, il n'y a pas toujours de variable aléatoire limite X naturelle et cela conduit à une deuxième définition de la convergence en loi d'une suite de variables aléatoires utilisée couramment et simultanément avec la précédente.

Définition 14.10. Soit pour tout $n \in \mathbb{N}$ une variable aléatoire X_n définie sur un espace probabilisé $(\Omega_n, \mathcal{A}_n, \mathbb{P}^n)$, à valeurs dans \mathbb{R}^d , et soit μ une **probabilité** sur \mathbb{R}^d . La suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ **converge en loi vers** μ si la suite $(\mathbb{P}_{X_n}^n)_{n \in \mathbb{N}}$ des lois des X_n , converge étroitement vers la loi μ .

Notation : $X_n \xrightarrow{\mathcal{L}} \mu$.

Il faut remarquer que dans ce cas, les objets mathématiques figurant de chaque côté de la flèche sont de **nature différente**.

Exemple : on démontrera ultérieurement que si pour tout $n \in \mathbb{N}$, X_n est une variable aléatoire réelle de loi binomiale $\mathcal{B}(n, \frac{\lambda}{n})$, où $\lambda > 0$, alors la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la loi de Poisson $\mathcal{P}(\lambda)$.

Conceptuellement, il n'y a donc rien de nouveau par rapport à la notion de convergence étroite d'une suite de probabilités et les critères de convergence en loi sont ceux de la convergence étroite d'une suite de **probabilités**. On donne toutefois une formulation du théorème de Lévy en termes de convergence en loi.

Théorème 14.11 (Théorème de Lévy; convergence en loi). Soit pour tout $n \in \mathbb{N}$ une variable aléatoire X_n définie sur un espace probabilisé $(\Omega_n, \mathcal{A}_n, \mathbb{P}^n)$, à valeurs dans \mathbb{R}^d , de fonction caractéristique φ_{X_n} .

(a) Si la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X , où X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans \mathbb{R}^d , alors la suite $(\varphi_{X_n})_{n \in \mathbb{N}}$ des fonctions caractéristiques converge simplement (et même uniformément sur tout compact de \mathbb{R}^d) vers la fonction caractéristique φ_X de X .

(b) Inversement, si la suite $(\varphi_{X_n})_{n \in \mathbb{N}}$ des fonctions caractéristiques converge simplement vers une fonction φ continue en 0, alors φ est la transformée de Fourier d'une probabilité μ sur \mathbb{R}^d , et la suite des variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X .

De plus il existe une variable aléatoire (non unique) X définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans \mathbb{R}^d , telle que la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X .

Démonstration. Ce n'est qu'une reformulation du théorème de Lévy pour la convergence étroite de mesures bornées, une fois rappelé que φ_{X_n} est, par définition, la transformée de Fourier de la loi de X_n . Seul le dernier point de la réciproque nécessite un éclaircissement : d'après le théorème de Lévy (th. 14.8), la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la probabilité μ telle que $\hat{\mu} = \varphi$ (μ est bien une probabilité, puisque $\lim_n \varphi_{X_n}(0) \equiv 1 = \varphi(0) = \hat{\mu}(0)$); on considère alors l'application identique X de \mathbb{R}^d sur lui-même; c'est une variable aléatoire définie sur l'espace probabilisé $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, \mu)$ à valeurs dans \mathbb{R}^d de loi μ et telle que la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X . \square

Exemple 14.1. Soit pour tout $n \in \mathbb{N}$ une variable aléatoire définie sur un espace probabilisé $(\Omega_n, \mathcal{A}_n, \mathbb{P}^n)$, à valeurs dans \mathbb{R}^d , et soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans \mathbb{R}^d . On a l'équivalence :

$$X_n \xrightarrow{\mathcal{L}} X \iff \forall t \in \mathbb{R}^d, \langle X_n, t \rangle \xrightarrow{\mathcal{L}} \langle X, t \rangle.$$

En effet, pour tout $t \in \mathbb{R}^d$ et tout réel α , on a $\varphi_{X_n}(\alpha t) = \varphi_{\langle X_n, t \rangle}(\alpha)$ et $\varphi_X(\alpha t) = \varphi_{\langle X, t \rangle}(\alpha)$. Il suffit alors d'appliquer le théorème 14.11.

Remarque. Avec les mêmes notations qu'au théorème 14.11, il est clair que si f est une application continue de \mathbb{R}^d dans \mathbb{R}^k et si la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X , alors la suite $(f(X_n))_{n \in \mathbb{N}}$ converge en loi vers $f(X)$, puisque, pour toute fonction $g \in \mathcal{C}_b(\mathbb{R}^k)$, on a, par le théorème de transfert,

$$\int_{\mathbb{R}^k} g d\mathbb{P}_{f(X_n)}^n = \int_{\mathbb{R}^d} f \circ g d\mathbb{P}_{X_n}^n \quad \text{et} \quad \int_{\mathbb{R}^k} g d\mathbb{P}_X = \int_{\mathbb{R}^d} f \circ g d\mathbb{P}_X.$$

La proposition suivante étend la classe de fonctions pour lesquelles on a cette propriété.

Proposition 14.12. *Soit pour tout $n \in \mathbb{N}$, une variable aléatoire X_n définie sur un espace probabilisé $(\Omega_n, \mathcal{A}_n, P^n)$, à valeurs dans \mathbb{R}^d , et soit X une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{R}^d . Soit f une application borélienne de \mathbb{R}^d dans \mathbb{R}^k telle que f soit P_X -p.s. continue. Si la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X , alors la suite $(f(X_n))_{n \in \mathbb{N}}$ converge en loi vers $f(X)$.*

Démonstration. Soit $C \in \mathcal{B}_{\mathbb{R}^d}$ tel que $P_X(C) = 1$ et tel que f soit continue sur C . Soit un fermé F quelconque de \mathbb{R}^k . On a, pour tout n ,

$$P_{f(X_n)}^n(F) = P_{X_n}^n[f^{-1}(F)] \leq P_{X_n}^n[\overline{f^{-1}(F)}],$$

et donc, d'après la proposition 14.5,

$$\limsup_n P_{f(X_n)}^n(F) \leq \limsup_n P_{X_n}^n[\overline{f^{-1}(F)}] \leq P_X(\overline{f^{-1}(F)}). \quad (14.7)$$

De plus, on a les inclusions

$$f^{-1}(F) \subset \overline{f^{-1}(F)} \subset C^c \cup f^{-1}(F);$$

en effet, soit $x \in \overline{f^{-1}(F)}$ quelconque. Si $x \in C^c$ on a a fortiori $x \in C^c \cup f^{-1}(F)$. Si $x \in C$, puisque $x \in \overline{f^{-1}(F)}$, il existe une suite $(x_n)_{n \in \mathbb{N}}$ de points de $f^{-1}(F)$ (c'est-à-dire tels que $f(x_n) \in F$) qui converge vers x , point de continuité de f ; on a alors $\lim_n f(x_n) = f(x)$, et puisque F est fermé, $f(x) \in F$ et on a encore $x \in C^c \cup f^{-1}(F)$. Puisque $P_X(C^c) = 0$, il en résulte que l'on a

$$P_X[\overline{f^{-1}(F)}] = P_X[f^{-1}(F)] = P_{f(X)}(F),$$

et, en reportant dans (14.7),

$$\limsup_n P_{f(X_n)}^n(F) \leq P_{f(X)}(F).$$

ce qui démontre que la suite $(f(X_n))_{n \in \mathbb{N}}$ converge en loi vers $f(X)$. \square

Exemple 14.2. Soient, sur l'espace probabilisé (Ω, \mathcal{A}, P) , des variables aléatoires (X_n, Y_n) , $n \in \mathbb{N}$, et (X, Y) à valeurs dans \mathbb{R}^2 telles que $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$; alors, par exemple, $X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$ et $X_n Y_n \xrightarrow{\mathcal{L}} XY$. Supposons, pour simplifier, que $Y_n \neq 0$ partout, pour tout n . Si $P_Y(\{0\}) = 0$, (c'est-à-dire $P_{(X,Y)}(\mathbb{R} \times \{0\}) = 0$), on peut dire que $\frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X}{Y}$.

On compare, lorsque cela a un sens, la convergence en loi et la convergence en probabilité d'une suite de variables aléatoires.

Proposition 14.13. *Si une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{R}^d converge en probabilité vers une variable aléatoire X (définie sur (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{R}^d) elle converge aussi en loi vers X .*

Démonstration. Soit $f \in \mathcal{C}_b(\mathbb{R}^d)$ quelconque. Pour tout $\varepsilon > 0$, on a, par le théorème de transfert,

$$\left| \int_{\mathbb{R}^d} f dP_{X_n} - \int_{\mathbb{R}^d} f dP_X \right| = |E f(X_n) - E f(X)| \\ \leq \varepsilon + 2 \|f\| P(|f(X_n) - f(X)| > \varepsilon);$$

puisque f est continue, la suite $(f(X_n))_{n \in \mathbb{N}}$ converge en probabilité vers $f(X)$ et il en résulte que

$$0 \leq \limsup_n \left| \int_{\mathbb{R}^d} f dP_{X_n} - \int_{\mathbb{R}^d} f dP_X \right| \leq \varepsilon,$$

ce qui, étant donné l'arbitraire de ε , démontre que $\lim_n \int_{\mathbb{R}^d} f dP_{X_n} = \int_{\mathbb{R}^d} f dP_X$. \square

Remarque. **La réciproque est fautive** et, comme le montre le contre-exemple suivant, elle n'est même pas vraie avec une suite stationnaire ! On considère sur l'espace probabilisé (Ω, \mathcal{A}, P) une variable aléatoire de Bernoulli X de paramètre $1/2$ et on pose, pour tout $n \in \mathbb{N}$, $X_n = X$; trivialement, on a $X_n \xrightarrow{\mathcal{L}} X$. La variable aléatoire $Y = 1 - X$ est encore de loi de Bernoulli de paramètre $1/2$, si bien que l'on a aussi $X_n \xrightarrow{\mathcal{L}} Y$. Par contre, puisque $|X_n - Y| = |2X - 1| = 1$ P-p.s., pour tout $\varepsilon \in]0, 1[$, on a $P(|X_n - Y| > \varepsilon) = 1$ et la suite $(X_n)_{n \in \mathbb{N}}$ ne converge pas en probabilité vers Y .

Toutefois, on a une réciproque partielle :

Proposition 14.14. *Si une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{R}^d converge en loi vers une variable aléatoire P-p.s. constante a , elle converge aussi en probabilité vers a .*

Démonstration. Pour tout $\varepsilon > 0$, on a $\delta_a(\partial B_f(a, \varepsilon)) = 0$, où $B_f(a, \varepsilon)$ est la boule fermée de centre a et de rayon ε , si bien que d'après la proposition 14.5, on a $\lim_n P_{X_n}(B_f(a, \varepsilon)) = \delta_a(B_f(a, \varepsilon)) = 1$. Il en résulte que $\lim_n P(\|X_n - a\| > \varepsilon) = 0$. \square

Le lemme de **Scheffé** donne une condition suffisante de convergence en loi dans le cas où les variables aléatoires admettent une **densité**.

Lemme 14.15 (Lemme de Scheffé). Soit, pour tout $n \in \mathbb{N}$, une variable aléatoire X_n définie sur un espace probabilisé $(\Omega_n, \mathcal{A}_n, \mathbb{P}^n)$, à valeurs dans \mathbb{R}^d et admettant une densité f_{X_n} . Si la suite $(f_{X_n})_{n \in \mathbb{N}}$ converge λ_d -p.p. vers une fonction f telle que $\int_{\mathbb{R}^d} f \, d\lambda_d = 1$, alors la suite $(X_n)_{n \in \mathbb{N}}$ **converge en loi** vers la loi $f \cdot \lambda_d$. De plus, on a

$$\lim_n \sup_{A \in \mathcal{B}_{\mathbb{R}^d}} \left| \mathbb{P}_{X_n}(A) - \int_A f \, d\lambda_d \right| = 0.$$

Démonstration. Pour tout $A \in \mathcal{B}_{\mathbb{R}^d}$, on a

$$\left| \mathbb{P}_{X_n}(A) - \int_A f \, d\lambda_d \right| = \left| \int_A (f_{X_n} - f) \, d\lambda_d \right| \leq \int_{\mathbb{R}^d} |f_{X_n} - f| \, d\lambda_d :$$

on a donc

$$\sup_{A \in \mathcal{B}_{\mathbb{R}^d}} \left| \mathbb{P}_{X_n}(A) - \int_A f \, d\lambda_d \right| \leq \int_{\mathbb{R}^d} |f_{X_n} - f| \, d\lambda_d. \quad (14.8)$$

On rappelle l'égalité très utile

$$|a - b| = a + b - 2 \min(a, b) \quad (a, b \in \mathbb{R});$$

si bien que l'on a

$$\int_{\mathbb{R}^d} |f_{X_n} - f| \, d\lambda_d = \int_{\mathbb{R}^d} f_{X_n} \, d\lambda_d + \int_{\mathbb{R}^d} f \, d\lambda_d - 2 \int_{\mathbb{R}^d} \min(f_{X_n}, f) \, d\lambda_d.$$

soit, en tenant compte de ce que f_{X_n} et f sont des densités de probabilité,

$$\int_{\mathbb{R}^d} |f_{X_n} - f| \, d\lambda_d = 2 - 2 \int_{\mathbb{R}^d} \min(f_{X_n}, f) \, d\lambda_d.$$

Mais, puisque l'on a

$$\forall n \in \mathbb{N} \quad 0 \leq \min(f_{X_n}, f) \leq f \quad \text{et} \quad \lim_n \min(f_{X_n}, f) = f \quad \lambda_d\text{-p.p.},$$

il résulte du théorème de convergence dominée que

$$\lim_n \int_{\mathbb{R}^d} |f_{X_n} - f| \, d\lambda_d = 2 - 2 \lim_n \int_{\mathbb{R}^d} \min(f_{X_n}, f) \, d\lambda_d = 2 - 2 \int_{\mathbb{R}^d} f \, d\lambda_d = 0,$$

ce qui, en tenant compte de (14.8), donne le résultat. \square

La proposition suivante donne un **critère de convergence en loi pour une suite de variables aléatoires discrètes** à valeurs dans \mathbb{Z} .

Proposition 14.16. Soient $X_n, n \in \mathbb{N}$ et X des variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans \mathbb{Z} . On a l'équivalence

$$X_n \xrightarrow{\mathcal{L}} X \quad \iff \quad \forall r \in \mathbb{Z} \quad \lim_n \mathbb{P}(X_n = r) = \mathbb{P}(X = r).$$

Démonstration. Si la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X , pour r fixé, on choisit $f \in \mathcal{C}_X(\mathbb{R})$ à support dans l'intervalle $]r - 1/2, r + 1/2[$ telle que $f(r) \neq 0$. Puisque l'on a

$$\int_{\mathbb{R}} f dP_{X_n} = f(r)P(X_n = r) \quad \text{ct} \quad \int_{\mathbb{R}} f dP_X = f(r)P(X = r).$$

et que $\lim_n \int_{\mathbb{R}} f dP_{X_n} = \int_{\mathbb{R}} f dP_X$, il vient $\lim_n P(X_n = r) = P(X = r)$.

Inversement, pour tout $f \in \mathcal{C}_X(\mathbb{R})$ de support compact K , on a

$$\int_{\mathbb{R}} f dP_{X_n} = \sum_{r \in K} f(r)P(X_n = r),$$

somme ne comportant qu'un nombre fini de termes: si, pour tout $r \in \mathbb{Z}$, on a $\lim_n P(X_n = r) = P(X = r)$, il en résulte que $\lim_n \int_{\mathbb{R}} f dP_{X_n} = \int_{\mathbb{R}} f dP_X$, ce qui démontre la convergence étroite de la suite des probabilités P_{X_n} vers la **probabilité** P_X . \square

Historiquement, la convergence en loi a été définie en termes de convergence de suites de fonctions de répartition. Mais, comme le montre la proposition suivante, la définition n'était pas extrêmement simple...

Proposition 14.17 (Convergence en loi et fonctions de répartition). *Soit pour tout $n \in \mathbb{N}$ une variable aléatoire réelle X_n définie sur un espace probabilisé $(\Omega_n, \mathcal{A}_n, P^n)$, de fonction de répartition F_{X_n} et soit X une variable aléatoire réelle définie sur un espace probabilisé (Ω, \mathcal{A}, P) , de fonction de répartition F_X . La suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si la suite $(F_{X_n}(x))_{n \in \mathbb{N}}$ converge vers $F_X(x)$ en tout point x de continuité de F_X .*

Démonstration. Supposons que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X . Soit x un point de continuité de F_X . Puisque $\partial(]-\infty, x]) = \{x\}$ et que $P_X(\{x\}) = F_X(x) - F_X(x-0) = 0$, la demi-droite $]-\infty, x]$ est un ensemble de P_X -continuité et on a, d'après la proposition 14.5,

$$\lim_n P_{X_n}(]-\infty, x]) = P_X(]-\infty, x]) = F_X(x),$$

ce qui démontre la condition nécessaire.

Inversement, supposons que la suite $(F_{X_n}(x))_{n \in \mathbb{N}}$ converge vers $F_X(x)$ en tout point x de continuité de F_X . Soient $f \in \mathcal{C}_0(\mathbb{R})$ et $\varepsilon > 0$ quelconques. Puisque l'ensemble des points de discontinuité de F_X est dénombrable (peut-être vide), il existe une fonction en escalier du type $g = \sum_{j=1}^k \alpha_j \mathbf{1}_{]a_j, b_j]}$, avec $a_j < b_j \leq a_{j+1} < b_{j+1}$, et où les a_j et b_j sont points de continuité de F_X , telle que $\|f - g\| \leq \varepsilon$. On a alors, par hypothèse,

$$\int_{\mathbb{R}} g dP_{X_n} = \sum_{j=1}^k \alpha_j (F_{X_n}(b_j) - F_{X_n}(a_j)) \xrightarrow{n \rightarrow +\infty} \sum_{j=1}^k \alpha_j (F_X(b_j) - F_X(a_j)),$$

c'est-à-dire $\lim_n \int_{\mathbb{R}} g dP_{X_n} = \int_{\mathbb{R}} g dP_X$. Par l'inégalité triangulaire, on a

$$\begin{aligned} & \left| \int_{\mathbb{R}} f dP_{X_n} - \int_{\mathbb{R}} f dP_X \right| \\ & \leq \left| \int_{\mathbb{R}} (f - g) dP_{X_n} \right| + \left| \int_{\mathbb{R}} g dP_{X_n} - \int_{\mathbb{R}} g dP_X \right| + \left| \int_{\mathbb{R}} (g - f) dP_X \right|, \end{aligned}$$

ce qui donne

$$\left| \int_{\mathbb{R}} f dP_{X_n} - \int_{\mathbb{R}} f dP_X \right| \leq 2 \|f - g\| + \left| \int_{\mathbb{R}} g dP_{X_n} - \int_{\mathbb{R}} g dP_X \right|;$$

on a donc

$$0 \leq \limsup_n \left| \int_{\mathbb{R}} f dP_{X_n} - \int_{\mathbb{R}} f dP_X \right| \leq 2\varepsilon,$$

ce qui, étant donné l'arbitraire de ε , montre que $\lim_n \int_{\mathbb{R}} f dP_{X_n} = \int_{\mathbb{R}} f dP_X$. La suite des probabilités P_{X_n} converge donc étroitement vers P_X . \square

Remarque. Comme le montre l'exemple suivant, on ne peut pas s'attendre à la convergence simple (partout) de la suite des fonctions de répartition F_{X_n} : si $X_n = 1/n$ et si $X = 0$, on a $X_n \xrightarrow{\mathcal{L}} X$, tandis que, pour tout n , $F_{X_n}(0) = 0$, et que $F_X(0) = 1$. On a par ailleurs un résultat analogue pour des variables aléatoires à valeurs dans \mathbb{R}^d , mais il est peu utilisable.

Exemple 14.3. La réciproque du lemme de Scheffé est fautive. Soit, pour tout $n \in \mathbb{N}^*$, une variable aléatoire réelle X_n définie sur l'espace probabilisé (Ω, \mathcal{A}, P) , admettant une densité f_{X_n} définie pour tout réel x par

$$f_{X_n}(x) = \mathbf{1}_{]0,1[}(x)(1 - \cos(2\pi nx)).$$

La suite $(f_{X_n})_{n \in \mathbb{N}^*}$ ne converge pas λ -p.p. (elle diverge en tout point de $]0, 1[$, elle converge ailleurs), alors que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers la loi $\mathbf{1}_{]0,1[} \cdot \lambda$. En effet, la fonction de répartition de X_n est donnée par

$$F_{X_n}(x) = \begin{cases} 0 & \text{si } x < 0, \\ x - \frac{\sin(2\pi nx)}{2\pi n} & \text{si } 0 < x \leq 1, \\ 1 & \text{si } x > 1, \end{cases}$$

si bien que l'on a

$$\lim_n F_{X_n}(x) = \begin{cases} 0 & \text{si } x < 0, \\ x & \text{si } 0 < x \leq 1, \\ 1 & \text{si } x > 1, \end{cases}$$

ce qui, en vertu de la proposition 14.17, démontre la convergence en loi de la suite $(X_n)_{n \in \mathbb{N}^*}$ vers la probabilité $\mathbf{1}_{]0,1[} \cdot \lambda_d$, loi uniforme sur $]0, 1[$.

Voici maintenant deux théorèmes de convergence en loi liés à la loi de **Poisson**. Pour leur démonstration, on utilise le lemme classique suivant.

Lemme 14.18. *Pour tout nombre complexe z et tout $n \in \mathbb{N}^*$, on a*

$$|\exp(z) - (1 + \frac{z}{n})^n| \leq \exp(|z|) - (1 + \frac{|z|}{n})^n. \quad (14.9)$$

Il en résulte que, pour tout $z \in \mathbb{C}$, la suite de terme général $(1 + \frac{z}{n})^n$ est convergente et que

$$\lim_n (1 + \frac{z}{n})^n = \exp(z).$$

Démonstration. La formule du binôme donne, pour tout $z \in \mathbb{C}$,

$$\exp(z) - (1 + \frac{z}{n})^n = \sum_{j=0}^{+\infty} \frac{z^j}{j!} - \sum_{j=0}^n \binom{n}{j} \frac{z^j}{n^j};$$

en tenant compte de l'égalité

$$\frac{\binom{n}{j}}{n^j} = \frac{1}{j!} \prod_{k=0}^{j-1} (1 - \frac{k}{n}),$$

il vient

$$\exp(z) - (1 + \frac{z}{n})^n = \sum_{j=n+1}^{+\infty} \frac{z^j}{j!} + \sum_{j=0}^n \frac{z^j}{j!} \left[1 - \prod_{k=0}^{j-1} (1 - \frac{k}{n}) \right]. \quad (14.10)$$

Puisque $1 - \prod_{k=0}^{j-1} (1 - \frac{k}{n}) \geq 0$, il en résulte que

$$|\exp(z) - (1 + \frac{z}{n})^n| \leq \sum_{j=n+1}^{+\infty} \frac{|z|^j}{j!} + \sum_{j=0}^n \frac{|z|^j}{j!} \left[1 - \prod_{k=0}^{j-1} (1 - \frac{k}{n}) \right],$$

ce qui donne (14.9), en réutilisant (14.10) pour $|z|$.

Enfin, puisque $\ln(1 + \frac{z}{n})^n = |z| + o(1)$, on a $\lim_n (1 + \frac{|z|}{n})^n = \exp(|z|)$, ce qui entraîne, d'après (14.9), $\lim_n (1 + \frac{z}{n})^n = \exp(z)$. \square

On démontre maintenant le **théorème de Poisson** (dont on a déjà donné une démonstration élémentaire dans le premier tome) en utilisant le théorème de Lévy.

Théorème 14.19 (Théorème de Poisson). *Soit, pour tout $n \in \mathbb{N}^*$, une variable aléatoire de loi binomiale $\mathcal{B}(n, p_n)$. On suppose que $\lim_n n p_n = \lambda$, où $\lambda > 0$. Alors la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers la loi de Poisson $\mathcal{P}(\lambda)$.*

Démonstration. Pour tout $n \in \mathbb{N}^*$, la fonction caractéristique de X_n est donnée par

$$\forall t \in \mathbb{R} \quad \varphi_{X_n}(t) = [p_n \exp(it) + (1 - p_n)]^n = [1 + p_n(\exp(it) - 1)]^n.$$

Il résulte de (14.9) que, pour tout $z \in \mathbb{C}$, on a

$$|\exp(np_n z) - (1 + p_n z)^n| \leq \exp(np_n |z|) - (1 + p_n |z|)^n;$$

puisque, par hypothèse,

$$\ln(1 + p_n |z|)^n = n \left[\frac{\lambda |z|}{n} + o\left(\frac{1}{n}\right) \right] \xrightarrow{n \rightarrow +\infty} \lambda |z|.$$

on a alors

$$\lim_n [\exp(np_n |z|) - (1 + p_n |z|)^n] = 0,$$

et donc

$$\lim_n (1 + p_n z)^n = \exp(\lambda z).$$

En prenant $z = [\exp(it) - 1]$, il vient

$$\forall t \in \mathbb{R} \quad \lim_n \varphi_{X_n}(t) = \exp[\lambda(\exp(it) - 1)],$$

ce qui, en vertu du théorème de Lévy, démontre le résultat. \square

Ce premier théorème de Poisson se généralise de la manière suivante.

Théorème 14.20 (Théorème des événements rares, théorème de Poisson). Soit, pour tout $n \in \mathbb{N}^*$, une famille finie $\{A_{n,j} \mid 1 \leq j \leq M_n\}$ d'événements indépendants définis sur un espace probabilisé (Ω, \mathcal{A}, P) . On pose $P(A_{n,j}) = p_{n,j}$ et on note

$$S_n = \sum_{j=1}^{M_n} \mathbf{1}_{A_{n,j}}.$$

On suppose que la suite de terme général M_n tend en croissant vers $+\infty$, que

$$\max_{1 \leq j \leq M_n} p_{n,j} \xrightarrow{n \rightarrow +\infty} 0 \quad \text{et que} \quad \sum_{j=1}^{M_n} p_{n,j} \xrightarrow{n \rightarrow +\infty} \lambda, \quad (14.11)$$

où $\lambda > 0$. Alors la suite $(S_n)_{n \in \mathbb{N}^*}$ converge en loi vers la loi de Poisson $\mathcal{P}(\lambda)$.

Démonstration. On utilise encore le théorème de Lévy. Par indépendance des $A_{n,j}$, $1 \leq j \leq M_n$, on a, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} \varphi_{S_n}(t) &= \prod_{j=1}^{M_n} \varphi_{\mathbf{1}_{A_{n,j}}}(t) = \prod_{j=1}^{M_n} [p_{n,j} \exp(it) + (1 - p_{n,j})] \\ &= \prod_{j=1}^{M_n} [1 + p_{n,j}(\exp(it) - 1)]. \end{aligned}$$

Si Log est la détermination principale du logarithme complexe, il résulte de la formule de Taylor avec reste intégral que, pour tout z tel que $|z| < 1$, on a

$$\text{Log}(1+z) = z - z^2 \int_0^1 (1-u) \frac{1}{(1+uz)^2} du.$$

Notons $z = \exp(it) - 1$; puisque $\max_{1 \leq j \leq M_n} p_{n,j} \xrightarrow{n \rightarrow +\infty} 0$, il existe N tel que, pour tout $n \geq N$, on ait $\max_{1 \leq j \leq M_n} |p_{n,j} z| < 1/2$. Pour tout $n \geq N$, on a alors

$$\text{Log} \varphi_{S_n}(t) = z \sum_{j=1}^{M_n} p_{n,j} - z^2 \sum_{j=1}^{M_n} p_{n,j}^2 \int_0^1 \frac{1-u}{(1+u p_{n,j} z)^2} du.$$

D'après l'inégalité triangulaire, on a, pour tout $n \geq N$ et tout $u \in [0, 1]$,

$$|1 + u p_{n,j} z| \geq 1 - p_{n,j} |z| \geq \frac{1}{2};$$

on a donc, pour tout $n \geq N$,

$$\left| \sum_{j=1}^{M_n} p_{n,j}^2 \int_0^1 \frac{1-u}{(1+u p_{n,j} z)^2} du \right| \leq 2 \left[\max_{1 \leq j \leq M_n} p_{n,j} \right] \left[\sum_{j=1}^{M_n} p_{n,j} \right].$$

Il résulte alors des hypothèses que $\lim_n \text{Log} \varphi_{S_n}(t) = \lambda z$; autrement dit, on a

$$\forall t \in \mathbb{R} \quad \lim_n \varphi_{S_n}(t) = \exp[\lambda(\exp(it) - 1)],$$

ce qui, en vertu du théorème de Lévy, démontre le résultat. \square

Remarque. Le théorème 14.20 tire son nom du fait qu'il montre qu'un phénomène aléatoire qui peut se représenter comme une **superposition d'événements rares** (c'est-à-dire d'événements de « petite » probabilité, au sens des conditions (14.11)) et indépendants, suit approximativement une loi de Poisson. Par ailleurs, ce théorème est une généralisation du théorème de Poisson 14.19. En effet, supposons (avec les notations du théorème 14.20) que $M_n = n$, et que, pour tout $n \in \mathbb{N}^*$, la famille finie d'événements indépendants $\{A_{n,j} \mid 1 \leq j \leq n\}$ soit telle que $P(A_{n,j}) = p_n$, indépendamment de j vérifiant $1 \leq j \leq n$, la suite $(p_n)_{n \in \mathbb{N}}$ vérifiant de surcroît la condition $\lim_n n p_n = \lambda$ ($\lambda > 0$). La variable aléatoire S_n suit alors la **loi binomiale** $\mathcal{B}(n, p_n)$ et les conditions (14.11) sont bien satisfaites puisque

$$p_n = \max_{1 \leq j \leq n} p_{n,j} \xrightarrow{n \rightarrow +\infty} 0 \quad \text{et que} \quad \sum_{j=1}^n p_{n,j} = n p_n \xrightarrow{n \rightarrow +\infty} \lambda.$$

Les deux théorèmes affirment que la suite $(S_n)_{n \in \mathbb{N}^*}$ converge en loi vers la loi de Poisson $\mathcal{P}(\lambda)$.

14.3. Théorème limite central

Le **théorème limite central** du calcul des probabilités montre que, sous des conditions plus ou moins générales, la loi de la somme d'un grand nombre de variables aléatoires indépendantes est « proche » d'une **loi normale**. Il existe de nombreuses versions de ce théorème (en particulier avec des hypothèses du type Lindeberg⁸); nous n'en donnerons qu'une version élémentaire. La démonstration moderne de ces différentes versions repose sur le théorème de Lévy et consiste à faire un développement asymptotique de la fonction caractéristique de la somme (centrée et réduite) de n variables aléatoires indépendantes.

Il résulte de la proposition 12.13 du chapitre 12 que, si une variable aléatoire admet un moment d'ordre $2k$, sa fonction caractéristique admet un développement limité d'ordre $2k$. Le lemme suivant donne (dans le cas d'un développement limité d'ordre 2) une majoration du « reste » qui peut parfois être utile.

Lemme 14.21. *Si la variable aléatoire réelle X admet un moment d'ordre deux, sa fonction caractéristique φ_X admet un développement limité d'ordre deux en 0 donné par, pour tout réel t ,*

$$\varphi_X(t) = 1 + it EX - \frac{t^2}{2} EX^2 + o(t^2).$$

Plus précisément, on a l'inégalité, pour tout réel t ,

$$\left| \varphi_X(t) - \left(1 + it EX - \frac{t^2}{2} EX^2 \right) \right| \leq t^2 E \left[\min \left(X^2, |t| \frac{|X|^3}{6} \right) \right]. \quad (14.12)$$

Démonstration. La formule de Taylor avec reste intégral écrite à l'ordre 2 donne, pour tout réel x ,

$$\exp(ix) = 1 + ix - x^2 \int_0^1 (1-u) \exp(iux) du,$$

soit, puisque $\int_0^1 (1-u) du = \frac{1}{2}$,

$$\exp(ix) - \left(1 + ix - \frac{x^2}{2} \right) = -x^2 \int_0^1 (1-u) [\exp(iux) - 1] du;$$

il en résulte que

$$\left| \exp(ix) - \left(1 + ix - \frac{x^2}{2} \right) \right| \leq x^2.$$

La même formule de Taylor à l'ordre 3 donne, pour tout réel x ,

8. Pour un théorème limite central avec une condition du type **Lindeberg**, voir par exemple Rényi A. [18], p. 415.

$$\exp(ix) = 1 + ix - \frac{x^2}{2} - i \frac{x^3}{2} \int_0^1 (1-u)^2 \exp(iux) du ;$$

il en résulte que

$$\left| \exp(ix) - \left(1 + ix - \frac{x^2}{2} \right) \right| \leq \frac{|x|^3}{6}.$$

Au total, on a, pour tout réel x ,

$$\left| \exp(ix) - \left(1 + ix - \frac{x^2}{2} \right) \right| \leq \min\left(x^2, \frac{|x|^3}{6}\right).$$

La majoration (14.12) en résulte immédiatement. Mais, par convergence dominée (prendre une suite quelconque qui tend vers 0), on a

$$\lim_{t \rightarrow 0} E \left[\min\left(X^2, |t| \frac{|X|^3}{6}\right) \right] = 0,$$

ce qui démontre le résultat. \square

Remarque. Si on n'a pas besoin de cette majoration précise du reste, pour établir le développement limité, il suffit d'appliquer la formule de Taylor-Young à φ_X qui est ici deux fois dérivable (cf. prop. 12.13, chapitre 12).

Théorème 14.22 (Théorème limite central). Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{R}^d , indépendantes, de même loi et admettant un moment d'ordre deux⁹. La suite de terme général Y_n , défini pour tout $n \in \mathbb{N}^*$ par

$$Y_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - EX_j),$$

converge en loi vers la loi gaussienne $\mathcal{N}_{\mathbb{R}^d}(0, C_{X_1})$, où C_{X_1} est la matrice de covariance des X_j .

En particulier, si $d = 1$ et si $Z_n = \frac{Y_n}{\sigma_{X_1}}$, où σ_{X_1} est l'écart-type des X_j , la suite des fonctions de répartition F_{Z_n} des Z_n converge simplement vers Φ , fonction de répartition de la loi $\mathcal{N}_{\mathbb{R}}(0, 1)$, donnée, pour tout réel z , par

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Démonstration. Les variables aléatoires X_j étant indépendantes et de même loi, la fonction caractéristique de Y_n est donnée par, pour tout $t \in \mathbb{R}^d$,

$$\varphi_{Y_n}(t) = \prod_{j=1}^n \varphi_{(X_j - EX_j)}\left(\frac{t}{\sqrt{n}}\right) = \left[\varphi_{(X_1 - EX_1, t)}\left(\frac{1}{\sqrt{n}}\right) \right]^n.$$

9. C'est à dire de carré (de norme) intégrable.

Le lemme 14.21 appliqué à la variable aléatoire réelle centrée $\langle X_1 - EX_1, t \rangle$ donne le développement asymptotique

$$\varphi_{Y_n}(t) = \left[1 - \frac{1}{2n} E(\langle X_1 - EX_1, t \rangle^2) + o\left(\frac{1}{n}\right) \right]^n.$$

Le lemme 14.18 assure que la suite de terme général $\varphi_{Y_n}(t)$ converge et que

$$\lim_n \varphi_{Y_n}(t) = \exp \left[-\frac{1}{2} E(\langle X_1 - EX_1, t \rangle^2) \right],$$

soit, puisque $E(\langle X_1 - EX_1, t \rangle^2) = \langle C_{X_1} t, t \rangle$,

$$\lim_n \varphi_{Y_n}(t) = \exp \left[-\frac{1}{2} \langle C_{X_1} t, t \rangle \right].$$

Le théorème de Lévy assure alors que : $Y_n \xrightarrow{\mathcal{L}} \mathcal{N}_{\mathbb{R}^d}(0, C_{X_1})$.

Si $d = 1$, on a, pour tout réel t ,

$$\varphi_{Z_n}(t) = \varphi_{Y_n} \left(\frac{t}{\sigma_{X_1}} \right) \quad \text{et donc} \quad \lim_n \varphi_{Z_n}(t) = \exp \left[-\frac{t^2}{2} \right].$$

c'est-à-dire que la suite des probabilités P_{Z_n} converge étroitement vers la loi $\mathcal{N}_{\mathbb{R}}(0, 1)$, ce qui est encore équivalent, puisque Φ est continue, à la convergence simple de la suite des fonctions de répartition des Z_n vers la fonction de répartition Φ de la loi limite. \square

Remarque. La fonction de répartition Φ de la loi $\mathcal{N}_{\mathbb{R}}(0, 1)$ est tabulée¹⁰. Nous rappelons (cf. tome 1, p. 230) trois valeurs d'accroissements de la fonction Φ d'usage courant :

$\Phi(1,64) - \Phi(-1,64) = 0,9$	$\Phi(1,96) - \Phi(-1,96) = 0,95$
$\Phi(3,09) - \Phi(-3,09) = 0,99.$	

Une application du théorème limite central est de démontrer le **théorème de Karl Pearson**, théorème qui est à la base du **test du chi-deux**.

Théorème 14.23 (Théorème de Karl Pearson). *Soit $k \in \mathbb{N}^*$ fixé. Pour tout $n \in \mathbb{N}^*$, on considère une partition $(A_j^n)_{1 \leq j \leq k}$ de Ω par des ensembles \mathcal{A} -mesurables. On suppose que ces partitions sont indépendantes, c'est-à-dire que les familles, indexées sur n , constituées par les éléments de ces partitions sont indépendantes. On suppose de plus que,*

$$\forall n \in \mathbb{N}^*, \quad \forall j = 1, 2, \dots, k, \quad P(A_j^n) = p_j,$$

10. Une table de la fonction de répartition Φ de la loi de Gauss centrée réduite figure, par exemple, dans Rényi A. [18], p. 585-586.

où $p_j > 0$ et $\sum_{j=1}^k p_j = 1$. On définit, pour tout $j = 1, 2, \dots, k$, les variables aléatoires réelles

$$N_j^n = \sum_{l=1}^n \mathbf{1}_{N_j^l},$$

puis la variable aléatoire

$$\chi_{k,n}^2 = \sum_{j=1}^k \frac{(N_j^n - np_j)^2}{np_j} = n \sum_{j=1}^k \frac{(\frac{1}{n}N_j^n - p_j)^2}{p_j}.$$

Alors, la suite des lois $\mathbf{P}_{\chi_{k,n}^2}$ converge étroitement vers la loi χ_{k-1}^2 du **chi-deux** à $k-1$ degrés de liberté; autrement dit, la suite des variables aléatoires $\chi_{k,n}^2$ converge en loi vers la loi du **chi-deux** à $k-1$ degrés de liberté.

Démonstration. Pour tout $n \in \mathbb{N}^*$, on définit les variables aléatoires X^n et N^n , à valeurs dans \mathbb{R}^k , par

$$X^n = \begin{pmatrix} \mathbf{1}_{N_1^n} \\ \vdots \\ \mathbf{1}_{N_k^n} \end{pmatrix} \quad \text{et} \quad N^n = \sum_{j=1}^n X^j = \begin{pmatrix} N_1^n \\ \vdots \\ N_k^n \end{pmatrix}.$$

On rappelle que la loi de N^n est la **loi multinomiale** $M(n; p_1, p_2, \dots, p_{k-1})$ et que les variables aléatoires X^n sont de même loi, de moyenne et matrice de covariance C_{X^n} données par

$$EX^n = \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} \quad \text{et} \quad (C_{X^n})_{i,j} = \begin{cases} p_i(1-p_i) & \text{si } i = j \\ -p_i p_j & \text{si } i \neq j \end{cases},$$

ce qui peut s'écrire, si on note p le vecteur de composantes p_j , $j = 1, 2, \dots, k$,

$$EX^n = p \quad \text{et} \quad C_{X^n} = \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & p_k \end{pmatrix} - pp^*.$$

Les variables aléatoires X^j étant de plus indépendantes, il résulte du théorème limite central que la suite de terme général Y_n , défini pour tout $n \in \mathbb{N}^*$ par

$$Y_n = \frac{1}{\sqrt{n}} \left[\sum_{j=1}^n X_j - np \right],$$

converge en loi vers la loi gaussienne $\mathcal{N}_{\mathbb{R},d}(0, C_{X_1})$, ce qui est équivalent, par le lemme de Lévy, à

$$\forall t \in \mathbb{R}^k \quad \lim_n \varphi_{Y_n}(t) = \exp\left(-\frac{\langle C_{X_1} t, t \rangle}{2}\right). \quad (14.13)$$

Par ailleurs, si on note M la matrice diagonale définie par

$$\forall j = 1, 2, \dots, k \quad M_{j,j} = \frac{1}{p_j} \quad \text{et} \quad M_{i,j} = 0 \text{ si } i \neq j.$$

on a $\chi_{k,n}^2 = \langle MY_n, Y_n \rangle = \|M^{\frac{1}{2}} Y_n\|^2$. Puisque, pour tout $t \in \mathbb{R}^k$, on a $\varphi_{M^{\frac{1}{2}} Y_n}(t) = \varphi_{Y_n}(M^{\frac{1}{2}} t)$, il résulte de (14.13) que

$$\forall t \in \mathbb{R}^k \quad \lim_n \varphi_{M^{\frac{1}{2}} Y_n}(t) = \exp\left(-\frac{\langle M^{\frac{1}{2}} C_{X_1} M^{\frac{1}{2}} t, t \rangle}{2}\right).$$

et donc, encore par le théorème de Lévy, que la suite de terme général $M^{\frac{1}{2}} Y_n$ converge en loi vers la loi gaussienne $\mathcal{N}_{\mathbb{R}^d}(0, M^{\frac{1}{2}} C_{X_1} M^{\frac{1}{2}})$. Mais on a

$$M^{\frac{1}{2}} C_{X_1} M^{\frac{1}{2}} = M^{\frac{1}{2}} [M^{-1} - \rho \rho^*] M^{\frac{1}{2}} = I - (M^{\frac{1}{2}} \rho)(M^{\frac{1}{2}} \rho)^*$$

et

$$\|M^{\frac{1}{2}} \rho\|^2 = \sum_{j=1}^k p_j = 1.$$

autrement dit, $M^{\frac{1}{2}} \rho$ est un vecteur unitaire; si on choisit une transformation orthogonale O telle que $O(M^{\frac{1}{2}} \rho) = e_1$, on a alors

$$O[M^{\frac{1}{2}} C_{X_1} M^{\frac{1}{2}}] O^* = I - (e_1)(e_1)^* = \begin{pmatrix} 0 & 0 \\ 0 & I_{\mathbb{R}^{k-1}} \end{pmatrix}.$$

si bien que, toujours par le théorème de Lévy, la suite de terme général $OM^{\frac{1}{2}} Y_n$ converge en loi vers la loi gaussienne $\mathcal{N}_{\mathbb{R}^d}(0, OM^{\frac{1}{2}} C_{X_1} M^{\frac{1}{2}} O) = \mathcal{N}_{\mathbb{R}^d}(0, I - (e_1)(e_1)^*)$. Puisque O est orthogonale, on a $\chi_{k,n}^2 = \|OM^{\frac{1}{2}} Y_n\|^2$; il en résulte que, si U est une variable aléatoire de loi $\mathcal{N}_{\mathbb{R}^d}(0, I - (e_1)(e_1)^*)$, la suite des lois des $\chi_{k,n}^2$ converge étroitement vers la loi de $\|U\|^2$, c'est-à-dire la loi du chi-deux à $k - 1$ degrés de liberté. \square

Remarque. Soit $q = (q_1, \dots, q_k)^*$ un vecteur de \mathbb{R}^k distinct de p . On note

$$\kappa_{k,n}^2 \equiv n \sum_{j=1}^k \frac{\left(\frac{1}{n} N_j^n - q_j\right)^2}{q_j};$$

la loi forte des grands nombres assure que, pour tout $j = 1, 2, \dots, k$, on a

$$\lim_n \frac{1}{n} N_j^n = p_j \text{ P-p.s.}$$

Puisque $q \neq p$, il en résulte que la suite de terme général

$$\sum_{j=1}^k \left(\frac{1}{n} N_j^n - q_j\right)^2 / q_j$$

converge P-p.s. vers un nombre $\alpha > 0$; la suite de terme général $\kappa_{k,n}^2$ converge alors P-p.s. vers $+\infty$ avec n .

Exemple 14.4. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d , indépendantes, de même loi μ . Soit $(D_j)_{1 \leq j \leq k}$ une partition de \mathbb{R}^d par des boréliens telle que $\mu(D_j) = p_j > 0$, pour tout $j = 1, 2, \dots, k$. Pour chaque n , les $A_j^n = X_n^{-1}(D_j)$, $j = 1, 2, \dots, k$, forment une partition de Ω et ces partitions sont indépendantes. De plus, on a, pour tout $n \in \mathbb{N}^*$, $P(A_j^n) = \mu(D_j) = p_j$. On définit, pour tout $j = 1, 2, \dots, k$, les variables aléatoires réelles

$$N_j^n = \sum_{l=1}^n \mathbf{1}_{D_j}(X_l),$$

puis les variables aléatoires

$$\chi_{k,n}^2 = n \sum_{j=1}^k \frac{\left(\frac{1}{n} N_j^n - p_j\right)^2}{p_j} \quad \text{et} \quad \kappa_{k,n}^2 \equiv n \sum_{j=1}^k \frac{\left(\frac{1}{n} N_j^n - q_j\right)^2}{q_j}.$$

Alors, la suite des lois $P_{\chi_{k,n}^2}$ converge étroitement vers la loi χ_{k-1}^2 du **chi-deux** à $k-1$ degrés de liberté, tandis que, pour tout vecteur $q \neq p$, la suite de terme général $\kappa_{k,n}^2$ converge vers $+\infty$ P-p.s.

Cet exemple est à la base du **test du chi-deux** dont on donne maintenant la problématique; suit un exemple de mise en pratique.

Le problème. Suite à la modélisation probabiliste d'un phénomène aléatoire, on s'intéresse à une variable aléatoire X à valeurs dans \mathbb{R}^d , censée représenter une « grandeur vectorielle » liée à ce phénomène. La loi μ de X est **inconnue** de l'expérimentateur; toutefois, celui-ci, à l'issue de calculs et raisonnements, est conduit à formuler des **hypothèses** sur cette loi. Il s'agit de « tester » l'hypothèse H que X est de loi μ au vu d'un « échantillon » de taille n , $\underline{x}_n = (x_1, x_2, \dots, x_n)$, obtenu en observant n réalisations « indépendantes » de ce phénomène. Cet échantillon est censé être la réalisation de n variables aléatoires X_1, X_2, \dots, X_n , indépendantes de même loi¹¹ que X .

On garde les notations de l'exemple 14.4, les D_j sont appelés **classes**. Pour tout vecteur \underline{x}_n de \mathbb{R}^{dn} , on définit $f_j(\underline{x}_n) = \frac{1}{n} \sum_{l=1}^n \mathbf{1}_{D_j}(x_l)$, fréquence du nombre de points x_l situés dans D_j . L'**effectif observé** de points x_l situés dans D_j pour l'échantillon \underline{x}_n est alors $n_j(\underline{x}_n) = n f_j(\underline{x}_n)$, l'**effectif théorique** dans D_j , sous l'hypothèse H , est $n p_j$. La « distance du chi-deux » entre ces effectifs observés et théoriques est

11. On rappelle que $\underline{X}_n = (X_1, X_2, \dots, X_n)$ est appelé « échantillon empirique » de taille n de la variable aléatoire X .

$$\Delta(\underline{x}_n) = \sum_{j=1}^k \frac{(n_j(\underline{x}_n) - np_j)^2}{np_j} \equiv n \sum_{j=1}^k \frac{(f_j(\underline{x}_n) - p_j)^2}{p_j},$$

qui est donc une réalisation de la variable aléatoire χ_{k-1}^2 . Pour tout réel $c > 0$, on considère l'ensemble $R_c^n = \{\underline{x}_n = (x_1, x_2, \dots, x_n) \mid \Delta(\underline{x}_n) > c\}$, appelé **zone de rejet** de l'hypothèse H. On a, en regard de l'exemple précédent,

$$P(\underline{X}_n \in R_c^n) = P(\Delta(\underline{X}_n) > c) \xrightarrow[n \rightarrow +\infty]{} \chi_{k-1}^2(\{c, +\infty\}),$$

et ceci indépendamment de la loi μ de X. Cela conduit à adopter la **règle du test du chi-deux** suivante : accepter l'hypothèse H si $\Delta(\underline{x}_n) \leq c$, la rejeter sinon. Le **risque d'erreur** est la probabilité de rejeter H alors que H est vraie ; il vaut, lorsque n est grand, $\chi_{k-1}^2(\{c, +\infty\})$ (en pratique, on considère que l'approximation est bonne dès que $np_j \geq 5$, pour tout $j = 1, \dots, k$). Pour un risque d'erreur α donné (en général 0,1 ou 0,05 ou 0,01), on détermine dans la table du chi-deux le réel c_α tel que l'on ait $\chi_{k-1}^2(\{c_\alpha, +\infty\}) = \alpha$, et on accepte ou rejette l'hypothèse H au vu de \underline{x}_n selon que $\Delta(\underline{x}_n) \leq c_\alpha$ ou non.

Exemple. On lance un dé n fois; on obtient n_j fois le chiffre j , pour $j = 1, 2, \dots, 6$. On se demande si ce dé est équilibré au vu des deux échantillons suivants :

$$\begin{array}{l} \mathbf{n = 60} \quad n_1 = 11 \quad n_2 = 8 \quad n_3 = 12 \quad n_4 = 9 \quad n_5 = 8 \quad n_6 = 12 \\ \mathbf{n = 600} \quad n_1 = 110 \quad n_2 = 80 \quad n_3 = 120 \quad n_4 = 90 \quad n_5 = 80 \quad n_6 = 120 \end{array}$$

Au lancer de dé est associé une variable aléatoire X de loi μ portée par l'ensemble $\{1, 2, \dots, 6\}$. Les classes sont les singletons $\{j\}$, $j = 1, 2, \dots, 6$ et on a $k = 6$. L'hypothèse H est que la loi μ est uniforme. On choisit $\alpha = 0,05$, de sorte que la table donne $c_\alpha = 11,1$, c'est-à-dire que l'on a $\chi_5^2(\{11,1, +\infty\}) = 0,05$. On calcule alors $\Delta(\underline{x}_n)$ pour ces deux échantillons :

– pour $n = 60$,

$$\Delta(\underline{x}_n) = \frac{1}{60} \left[\frac{(11 - 10)^2 + (8 - 10)^2 + (12 - 10)^2 + (9 - 10)^2 + (8 - 10)^2 + (11 - 10)^2}{6} \right],$$

soit

$$\Delta(\underline{x}_n) = 1,8 \leq 11,1, \text{ ce qui conduit à accepter H au vu de } \underline{x}_{60}.$$

– pour $n = 600$,

$$\Delta(\underline{x}_n) = \frac{1}{600} \left[\frac{(110 - 100)^2 + (80 - 100)^2 + (120 - 100)^2 + (90 - 100)^2 + (80 - 100)^2 + (110 - 100)^2}{6} \right],$$

soit

$\Delta(\underline{x}_n) = 18 > 11,1$, ce qui conduit à rejeter H au vu de \underline{x}_{600} .

14.4. Estimation

On présente succinctement le problème de l'estimation de la loi d'une variable aléatoire et on donne une méthode de construction d'estimateur connu sous le nom d'**estimateur du maximum de vraisemblance**. Sous des hypothèses de régularité de la densité, on montre que cet estimateur a des **propriétés asymptotiques** intéressantes. Pour une étude détaillée de la théorie de l'estimation, on pourra par exemple consulter les livres de Fourgeaud et Fuchs [11] ou de Dacunha-Castelle et Duflo [8].

Le problème d'estimation paramétrique. Suite à la modélisation probabiliste d'un phénomène aléatoire, on s'intéresse à une variable aléatoire X à valeurs dans \mathbb{R} , censée représenter une « grandeur réelle » liée à ce phénomène. La loi μ de X est **inconnue** de l'expérimentateur ; toutefois, celui-ci, à l'issue de calculs et raisonnements, est conduit à supposer que cette loi appartient à une famille de lois dépendant d'un paramètre $\theta \in \Theta$, ouvert de \mathbb{R}^p . Il s'agit d'**estimer** la « vraie » valeur θ_0 du paramètre au vu d'un **échantillon** de taille n , $\underline{x}_n = (x_1, x_2, \dots, x_n)$, obtenu en observant n réalisations indépendantes de ce phénomène. Cet échantillon est censé être la réalisation de n variables aléatoires X_1, X_2, \dots, X_n , indépendantes de même loi que X .

Modélisation statistique du problème. On considère une **structure statistique**, c'est-à-dire une famille d'espaces probabilisés $[(\Omega, \mathcal{A}, P_\theta)]_{\theta \in \Theta}$, où Θ est un ouvert de \mathbb{R}^p , sur lesquels on définit la variable aléatoire réelle X et un **échantillon** de taille infinie, c'est-à-dire une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires P_θ -indépendantes de même loi μ_θ que X (mesure image de P_θ par X), pour tout $\theta \in \Theta$. On suppose que l'application $\theta \mapsto P_\theta$ est injective. Soit g une application de Θ dans un ouvert Θ' de \mathbb{R}^k avec $k \leq p$; un **estimateur** de $g(\theta)$ au vu d'un **échantillon** de taille n , sera une variable aléatoire (appelée « statistique » par les statisticiens) $T_n = \varphi_n(\underline{X}_n)$, où φ_n est une fonction mesurable de \mathbb{R}^n dans Θ' . Cet estimateur est dit **sans biais** si $E_\theta(T_n) = g(\theta)$, où E_θ désigne l'intégration par rapport à la probabilité P_θ . Si l'échantillon est infini, la suite $T = (T_n)_{n \in \mathbb{N}^*}$ est appelée estimateur de $g(\theta)$; il sera intéressant s'il est **consistant (en probabilité, resp. presque sûrement)**, c'est-à-dire si la suite $(T_n)_{n \in \mathbb{N}^*}$ converge en P_θ -probabilité, resp. P_θ -p.s.

La **méthode du maximum de vraisemblance** permet souvent de trouver de tels estimateurs. Elle n'a qu'un fondement empirique issu de l'expérience

suivante : on tire au hasard une boule dans une des deux urnes U_1 et U_2 , sans savoir dans laquelle (mais tout en sachant la composition de ces urnes) ; U_1 contient une boule rouge et 9 noires, U_2 contient 9 boules rouges et une noire. Si le tirage donne une boule rouge, on est tenté de dire que l'on a tiré dans U_2 qui donne la plus grande probabilité de tirage d'une boule rouge.

Dans la suite, on fait l'hypothèse qu'existe une mesure σ -finie μ sur \mathbb{P} (souvent la mesure de Lebesgue) telle que, pour tout $\theta \in \Theta$, la loi μ_θ de X soit de densité la fonction $f(\cdot, \theta)$ par rapport à μ . La variable aléatoire \underline{X}_n admet alors pour densité (par rapport à la mesure $\mu^{\otimes n}$) la fonction L_n , appelée **fonction de vraisemblance** (relative à l'échantillon de taille n) définie par, pour tout $\underline{x}_n \in \mathbb{R}^n$,

$$L_n(\underline{x}_n, \theta) = \prod_{j=1}^n f(x_j, \theta).$$

Un estimateur T_n de θ est appelé **estimateur du maximum de vraisemblance** de θ s'il s'écrit sous la forme $\hat{\varphi}_n(\underline{X}_n)$, où $\hat{\varphi}_n$ est une fonction mesurable satisfaisant à la condition :

$$\forall \underline{x}_n \in \mathbb{R}^n \quad L_n(\underline{x}_n, \hat{\varphi}_n(\underline{x}_n)) = \sup_{\theta \in \Theta} L_n(\underline{x}_n, \theta). \quad (14.14)$$

Si de plus $f(x, \cdot)$ est différentiable, $\hat{\varphi}_n(\underline{x}_n)$ est solution de l'**équation de vraisemblance** :

$$\forall \underline{x}_n \in \mathbb{R}^n \quad \frac{\partial}{\partial \theta} L_n(\underline{x}_n, \hat{\varphi}_n(\underline{x}_n)) = 0. \quad (14.15)$$

Il faut noter qu'alors $\hat{\varphi}_n(\underline{x}_n)$ n'est a priori qu'un point stationnaire ; il faudra aller voir plus avant pour s'assurer de l'existence de maximum.

Dans le cas où $f(x, \theta) > 0$ pour tout (x, θ) , l'équation (14.14) est équivalente à l'équation, plus maniable, obtenue en prenant les logarithmes et $\hat{\varphi}_n$ est solution de l'équation :

$$\forall \underline{x}_n \in \mathbb{R}^n \quad \ln L_n(\underline{x}_n, \hat{\varphi}_n(\underline{x}_n)) = \sup_{\theta \in \Theta} \ln L_n(\underline{x}_n, \theta).$$

Si de plus $f(x, \cdot)$ est différentiable, $\hat{\varphi}_n(\underline{x}_n)$ est solution de l'**équation de log-vraisemblance** :

$$\forall \underline{x}_n \in \mathbb{R}^n \quad \frac{\partial}{\partial \theta} \ln L_n(\underline{x}_n, \hat{\varphi}_n(\underline{x}_n)) = 0.$$

De tels estimateurs, quand ils existent ne sont en général pas uniques ; ils sont souvent consistants et jouissent de propriétés de **normalité asymptotique**. Le problème d'existence est un problème de maximum : les conditions

d'existence sont souvent de type différentiabilité, mais ce n'est pas toujours le cas. Voici un exemple de chaque cas présenté avec les notations ci-dessus :

Cas gaussien. $\Theta = \mathbb{R} \times \mathbb{R}^{+*}$, et pour tout $\theta = (m, \sigma^2)$, $\mu_\theta = \mathcal{N}_{\mathbb{R}}(m, \sigma^2)$. On a, pour tout $\underline{x}_n \in \mathbb{R}^n$,

$$\ln L_n(\underline{x}_n, \theta) = \sum_{j=1}^n \ln f(x_j, \theta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - m)^2,$$

et donc

$$\begin{aligned} \frac{\partial}{\partial m} \ln L_n(\underline{x}_n, (m, \sigma^2)) &= \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - m), \\ \frac{\partial}{\partial \sigma^2} \ln L_n(\underline{x}_n, (m, \sigma^2)) &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{j=1}^n (x_j - m)^2. \end{aligned}$$

Les solutions de l'équation de **maximum de vraisemblance** sont à chercher parmi les points stationnaires, c'est-à-dire les solutions des équations de log-vraisemblance

$$\frac{\partial}{\partial m} \ln L_n(\underline{x}_n, (m, \sigma^2)) = 0 \quad \text{et} \quad \frac{\partial}{\partial \sigma^2} \ln L_n(\underline{x}_n, (m, \sigma^2)) = 0,$$

soit ici

$$\widehat{m}_n = \frac{1}{n} \sum_{j=1}^n x_j \quad \text{et} \quad \widehat{\sigma}_n^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \widehat{m}_n)^2.$$

Il reste à vérifier que ce point stationnaire correspond à un **maximum**. On pourrait étudier les dérivées secondes, mais ici, on le voit directement par le calcul suivant :

$$\begin{aligned} \ln L_n(\underline{x}_n, (\widehat{m}_n, \widehat{\sigma}_n^2)) - \ln L_n(\underline{x}_n, (m, \sigma^2)) \\ &= -\frac{n}{2} \ln\left(\frac{\widehat{\sigma}_n^2}{\sigma^2}\right) - \frac{1}{2\widehat{\sigma}_n^2} \sum_{j=1}^n (x_j - \widehat{m}_n)^2 + \frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - m)^2 \\ &= \frac{n}{2} \left[\left(\frac{\widehat{\sigma}_n^2}{\sigma^2} - 1\right) - \ln\left(\frac{\widehat{\sigma}_n^2}{\sigma^2}\right) \right] + n \frac{(\widehat{m}_n - m)^2}{2\sigma^2} \geq 0, \end{aligned}$$

car on a, pour tout $x > 0$, $(x - 1) > \ln x$. Ainsi, il existe un unique estimateur du maximum de vraisemblance (M_n, Σ_n) , où $M_n = \frac{1}{n} \sum_{j=1}^n x_j$ et $\Sigma_n = \frac{1}{n} \sum_{j=1}^n (X_j - M_n)^2$ sont les moyenne et variance empirique de l'échantillon.

Remarque. On a $E_\theta[\Sigma_n] = \frac{n-1}{n} \sigma^2$; on dit que Σ_n est un estimateur **biaisé** de σ^2 .

Cas uniforme. $\Theta = \mathbb{R}^+$, et pour tout $\theta > 0$, μ_θ est loi uniforme $\mathcal{U}([0, \theta])$ sur l'intervalle $[0, \theta]$. On a, pour tout $\underline{x}_n \in \mathbb{R}^n$,

$$L_n(\underline{x}_n, \theta) = \frac{1}{\theta^n} \prod_{j=1}^n \mathbf{1}_{[0, \theta]}(x_j).$$

La fonction $L_n(\underline{x}_n, \cdot)$ n'est pas différentiable ; toutefois $L_n(\underline{x}_n, \cdot)$ sera maximum pour θ le plus petit possible, c'est-à-dire en $\hat{\varphi}_n(\underline{x}_n) = \max_{1 \leq j \leq n} \underline{x}_j$. Dans ce cas, il existe un unique estimateur du maximum de vraisemblance de θ , $T_n = \max_{1 \leq j \leq n} (X_j)$.

En fait, les propriétés asymptotiques sont des propriétés des solutions de l'équation de vraisemblance (qui donnent des points stationnaires) et non des propriétés des solutions donnant les maximum de $L_n(\underline{x}_n, \cdot)$. À titre d'exemple, le théorème suivant établit, sous des hypothèses très fortes, un résultat de normalité asymptotique dans le cas où le paramètre est réel (ce théorème se généralise au cas d'un paramètre multidimensionnel).

Théorème 14.24. Soit $\theta_0 \in \Theta$ la vraie valeur du paramètre. On suppose que $f(x, \theta) > 0$ pour tout $(x, \theta) \in \mathbb{R} \times \Theta$, où Θ est un ouvert de \mathbb{R} , que, pour tout réel x , $f(x, \cdot)$ est deux fois continûment dérivable et que la fonction $\frac{\partial^2}{\partial \theta^2} \ln f(x, \cdot)$ est continue en θ uniformément en x . On suppose de plus qu'il existe une fonction g μ -intégrable telle que l'on ait la majoration

$$\forall (x, \theta) \in \mathbb{R} \times \Theta \quad \left| \frac{\partial^2}{\partial \theta^2} f(x, \theta) \right| \leq g(x), \quad (14.16)$$

et une fonction h telle que la fonction $x \mapsto h(x) f(x, \theta_0)$ soit μ -intégrable et telle que l'on ait la majoration

$$\forall (x, \theta) \in \mathbb{R} \times \Theta \quad \left| \frac{\partial^2}{\partial \theta^2} \ln f(x, \theta) \right| \leq h(x). \quad (14.17)$$

Soit $I(\theta_0) = - \int_{\mathbb{R}} \left(\frac{\partial^2}{\partial \theta^2} \ln f(x, \theta_0) \right) f(x, \theta_0) d\mu(x)$ ($I(\theta_0)$ est appelée **quantité d'information de Fisher**). On a $0 \leq I(\theta_0) < +\infty$. On suppose que $I(\theta_0) > 0$.

Soit une suite $(\hat{\varphi}_n)_{n \in \mathbb{N}^*}$ de solution de l'équation de vraisemblance (14.15) et soit $T_n = \hat{\varphi}_n(\underline{X}_n)$. Si cette suite d'estimateurs $(T_n)_{n \in \mathbb{N}^*}$ converge P_{θ_0} -p.s. vers θ_0 , alors la suite de terme général $Y_n = \sqrt{n I(\theta_0)}(T_n - \theta_0)$ converge en loi vers la loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$.

Démonstration. Soit φ la fonction définie par, pour tout (x, θ) , $\varphi(x, \theta) = \frac{\partial}{\partial \theta} \ln f(x, \theta)$ et U_θ la variable aléatoire $U_\theta = \varphi(X, \theta)$. La condition (14.17) implique que

$$\int_{\mathbb{R}} \left| \frac{\partial^2}{\partial \theta^2} \ln f(x, \theta_0) \right| f(x, \theta_0) d\mu(x) < +\infty \quad (14.18)$$

(on dit que le modèle est **régulier** en θ_0). Le théorème de transfert et (14.18) assurent alors que U_{θ_0} est de carré P_{θ_0} -intégrable et que $E_{\theta_0}(U_{\theta_0}) = 0$. En

effet, on a

$$\begin{aligned} \int_{\Omega} (U_{\theta_0})^2 dP_{\theta_0} &= \int_{\mathbb{R}} \left(\frac{\partial}{\partial \theta} \ln f(x, \theta_0) \right)^2 f(x, \theta_0) d\mu(x) \\ &= \int_{\mathbb{R}} \left[\frac{\frac{\partial}{\partial \theta} f(x, \theta_0)}{f(x, \theta_0)} \right]^2 f(x, \theta_0) d\mu(x); \end{aligned}$$

et, puisque

$$\frac{\partial^2}{\partial \theta^2} \ln f(x, \theta) = \frac{\frac{\partial^2}{\partial \theta^2} f(x, \theta) f(x, \theta) - \left[\frac{\partial}{\partial \theta} f(x, \theta) \right]^2}{[f(x, \theta)]^2},$$

il vient, en tenant compte des conditions (14.16) et (14.17),

$$\begin{aligned} \int_{\Omega} (U_{\theta_0})^2 dP_{\theta_0} &= - \int_{\mathbb{R}} \left[\frac{\partial^2}{\partial \theta^2} \ln f(x, \theta_0) \right] f(x, \theta_0) d\mu(x) + \int_{\mathbb{R}} \frac{\partial^2}{\partial \theta^2} f(x, \theta_0) d\mu(x) < +\infty. \end{aligned} \quad (14.19)$$

Il en résulte que $\int_{\Omega} |U_{\theta_0}| dP_{\theta_0} < +\infty$ et, par le théorème de transfert, que

$$\int_{\Omega} |U_{\theta_0}| dP_{\theta_0} = \int_{\mathbb{R}} \left| \frac{\partial}{\partial \theta} \ln f(x, \theta_0) \right| f(x, \theta_0) d\mu(x) < +\infty.$$

De plus, on a

$$E_{\theta_0}(U_{\theta_0}) = \int_{\mathbb{R}} \frac{\frac{\partial}{\partial \theta} f(x, \theta_0)}{f(x, \theta_0)} f(x, \theta_0) d\mu(x) = \int_{\mathbb{R}} \frac{\partial}{\partial \theta} f(x, \theta_0) d\mu(x).$$

Soit K un voisinage compact de θ_0 contenu dans Θ ; le théorème des accroissements finis et la condition (14.16) assurent que l'on a, pour tout $\theta \in K$,

$$\left| \frac{\partial}{\partial \theta} f(x, \theta) \right| \leq \left| \frac{\partial}{\partial \theta} f(x, \theta_0) \right| + c g(x),$$

où $c > 0$ est une constante qui dépend de K . On peut donc appliquer le théorème de dérivation d'une intégrale dépendant d'un paramètre, ce qui donne

$$E_{\theta_0}(U_{\theta_0}) = \left[\frac{\partial}{\partial \theta} \int_{\mathbb{R}} f(x, \theta) d\mu(x) \right]_{\theta=\theta_0}.$$

soit, puisque, pour tout θ , on a $\int_{\mathbb{R}} f(x, \theta) d\mu(x) = 1$, $E_{\theta_0}(U_{\theta_0}) = 0$.

De même, on a, d'après la condition (14.16),

$$\int_{\mathbb{R}} \frac{\partial^2}{\partial \theta^2} f(x, \theta) d\mu(x) = \frac{\partial^2}{\partial \theta^2} \int_{\mathbb{R}} f(x, \theta) d\mu(x) = 0.$$

Il en résulte, en reportant dans (14.19), que

$$E_{\theta_0}(U_{\theta_0}^2) = - \int_{\mathbb{R}} \left[\frac{\partial^2}{\partial \theta^2} \ln f(x, \theta_0) \right] f(x, \theta_0) d\mu(x) = I(\theta_0),$$

ce qui montre en particulier que $0 \leq I(\theta_0) < +\infty$.

Soit, pour tout $n \in \mathbb{N}^*$, une solution $\hat{\varphi}_n(\underline{x}_n)$ de l'équation de log-vraisemblance

$$\frac{\partial}{\partial \theta} \ln L_n(\underline{x}_n, \theta) = \sum_{j=1}^n \varphi(x_j, \theta) = 0.$$

La formule de Taylor à l'ordre 1 avec reste intégral appliquée en $\hat{\varphi}_n(\underline{x}_n)$ implique que

$$\sum_{j=1}^n \varphi(x_j, \theta_0) = (\theta_0 - \hat{\varphi}_n(\underline{x}_n)) \int_0^1 \sum_{j=1}^n \frac{\partial}{\partial \theta} \varphi \left[x_j, \hat{\varphi}_n(\underline{x}_n) + t(\theta_0 - \hat{\varphi}_n(\underline{x}_n)) \right] dt,$$

ce qui donne l'égalité

$$\sum_{j=1}^n \varphi(X_j, \theta_0) = (\theta_0 - T_n) \int_0^1 \sum_{j=1}^n \frac{\partial}{\partial \theta} \varphi \left[X_j, T_n + t(\theta_0 - T_n) \right] dt.$$

soit encore,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(X_j, \theta_0) = [\sqrt{n}(\theta_0 - T_n)] \int_0^1 \frac{1}{n} \sum_{j=1}^n \frac{\partial}{\partial \theta} \varphi \left[X_j, T_n + t(\theta_0 - T_n) \right] dt. \quad (14.20)$$

Les variables aléatoires $\varphi(X_j, \theta_0)$ sont indépendantes de même loi (sous P_{θ_0}) que U_{θ_0} (elles admettent donc un moment d'ordre deux): le théorème de la limite central montre alors que

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(X_j, \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}_{\mathbb{R}}(0, I(\theta_0)). \quad (14.21)$$

Il reste à étudier la suite de terme général

$$\int_0^1 \frac{1}{n} \sum_{j=1}^n \frac{\partial}{\partial \theta} \varphi \left[X_j, T_n + t(\theta_0 - T_n) \right] dt.$$

D'après la condition (14.17), les variables aléatoires $\frac{\partial}{\partial \theta} \varphi(X_j, \theta_0)$ admettent une moyenne sous P_{θ_0} égale à $-I(\theta_0)$; de plus, elles sont indépendantes. Il résulte de la **loi forte des grands nombres** que

$$\lim_n \frac{1}{n} \sum_{j=1}^n \frac{\partial}{\partial \theta} \varphi(X_j, \theta_0) = -I(\theta_0) \quad P_{\theta_0}\text{-p.s.} \quad (14.22)$$

Soit alors

$$A_n(t) = \frac{1}{n} \sum_{j=1}^n \left[\frac{\partial}{\partial \theta} \varphi [X_j, T_n + t(\theta_0 - T_n)] - \frac{\partial}{\partial \theta} \varphi (X_j, \theta_0) \right];$$

démontrons que P_{θ_0} -p.s., pour tout $t \in [0, 1]$, $\lim_n A_n(t) = 0$ (attention à l'ordre dans lequel les assertions ont été énoncées).

Soit $\varepsilon > 0$ quelconque. La fonction $\theta \mapsto \frac{\partial}{\partial \theta} \varphi(x, \theta)$ étant continue, uniformément en x , il existe un intervalle V centré en θ_0 et contenu dans Θ tel que, dès que $\theta \in V$ on ait,

$$\sup_{x \in \mathbb{R}} \left| \frac{\partial}{\partial \theta} \varphi(x, \theta) - \frac{\partial}{\partial \theta} \varphi(x, \theta_0) \right| \leq \varepsilon. \quad (14.23)$$

Par hypothèse, il existe $N \in \mathcal{A}$ tel que $P_{\theta_0}(N) = 0$ et tel que, pour tout $\omega \notin N$, on ait $\lim_n T_n(\omega) = \theta_0$; soit un tel ω et soit $K(\omega)$ tel que l'on ait, pour tout $n \geq K(\omega)$, $T_n(\omega) \in V$. On a alors, si $0 < t < 1$,

$$\begin{aligned} |A_n(t)(\omega)| &\leq \frac{1}{n} \sum_{j=1}^{K(\omega)} \left| \frac{\partial}{\partial \theta} \varphi [X_j(\omega), T_n(\omega) + t(\theta_0 - T_n(\omega))] - \frac{\partial}{\partial \theta} \varphi (X_j(\omega), \theta_0) \right| \\ &\quad + \frac{1}{n} \sum_{j=K(\omega)+1}^n \left| \frac{\partial}{\partial \theta} \varphi [X_j(\omega), T_n(\omega) + t(\theta_0 - T_n(\omega))] - \frac{\partial}{\partial \theta} \varphi (X_j(\omega), \theta_0) \right|. \end{aligned}$$

soit, en tenant compte de (14.23),

$$|A_n(t)(\omega)| \leq \frac{2}{n} \sum_{j=1}^{K(\omega)} g(X_j(\omega)) + \varepsilon.$$

Il en résulte que $\limsup_n |A_n(t)(\omega)| \leq \varepsilon$, ce qui, étant donné l'arbitraire de ε , démontre que, pour tout $t \in [0, 1]$, $\lim_n A_n(t)(\omega) = 0$. Puisque l'on a

$$|A_n(t)(\omega)| \leq 2g(x),$$

il résulte du théorème de convergence dominée que la suite de terme général

$$\int_0^1 \frac{1}{n} \sum_{j=1}^n \left[\frac{\partial}{\partial \theta} \varphi [X_j(\omega), T_n(\omega) + t(\theta_0 - T_n(\omega))] - \frac{\partial}{\partial \theta} \varphi (X_j(\omega), \theta_0) \right] dt$$

converge vers 0; puisque ceci est vrai pour tout $\omega \notin N$, il résulte alors de (14.22) que

$$\begin{aligned} \lim_n \int_0^1 \frac{1}{n} \sum_{j=1}^n \frac{\partial}{\partial \theta} \varphi [X_j, T_n + t(\theta_0 - T_n)] dt &= \lim_n \frac{1}{n} \sum_{j=1}^n \frac{\partial}{\partial \theta} \varphi (X_j, \theta_0) \\ &= -I(\theta_0) \quad P_{\theta_0}\text{-p.s.} \end{aligned}$$

Puisque $I(\theta_0) \neq 0$, on a aussi

$$\lim_n \left[\frac{1}{n} \sum_{j=1}^n \frac{\partial}{\partial \theta} \varphi(X_j, \theta_0) \right]^{-1} = -[I(\theta_0)]^{-1} \quad \mathbb{P}_{\theta_0}\text{-p.s.},$$

et la convergence a lieu aussi en loi. L'égalité (14.20) et la convergence en loi démontrée en (14.22) impliquent alors, par le **lemme de Slutsky** (voir exercice 8), que

$$\sqrt{n} I(\theta_0) (T_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}_{\mathbb{R}}(0, 1). \quad \square$$

Exercices

Exercice 14.1. Convergence étroite d'une suite de probabilités portées par \mathbb{Z} . Soit, pour tout $n \in \mathbb{N}$, une probabilité μ_n sur l'espace mesurable $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, portée par \mathbb{Z} , c'est-à-dire de la forme $\mu_n = \sum_{r \in \mathbb{Z}} a_r^n \delta_r$, où pour tous $r \in \mathbb{Z}$, $a_r^n \geq 0$. Démontrer que la suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers une probabilité μ si et seulement si, pour tout $r \in \mathbb{Z}$, la suite $(a_r^n)_{n \in \mathbb{N}}$ converge vers un réel $a_r \geq 0$ et si on a $\sum_{r \in \mathbb{Z}} a_r = 1$ et $\mu = \sum_{r \in \mathbb{Z}} a_r \delta_r$ (μ est portée par \mathbb{Z}).

Solution. Si la suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers une probabilité μ , on a, pour tous les intervalles **ouverts** $]r - 1, r[$, où $r \in \mathbb{Z}$,

$$0 \leq \mu(]r - 1, r[) \leq \liminf_n \mu_n(]r - 1, r[);$$

puisque, pour tout $n \in \mathbb{N}$, on a $\mu_n(]r - 1, r[) = 0$, il en résulte que $\mu(]r - 1, r[) = 0$. La probabilité μ est donc portée par une partie de \mathbb{Z} et est de la forme $\mu = \sum_{r \in \mathbb{Z}} a_r \delta_r$, où $a_r \geq 0$. Pour toute fonction $f \in \mathcal{C}_{\mathcal{K}}(\mathbb{R})$ à support dans l'intervalle $]r - 1/2, r + 1/2[$ telle que $f(r) \neq 0$, on a alors

$$\int_{\mathbb{R}} f d\mu_n = f(r) \mu_n(\{r\}) = f(r) a_r^n \quad \text{et} \quad \int_{\mathbb{R}} f d\mu = f(r) \mu(\{r\}).$$

Puisque $\lim_n \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\mu$, il vient $\lim_n a_r^n = \mu(\{r\}) \geq 0$. Enfin, μ étant par hypothèse une probabilité, on a

$$\sum_{r \in \mathbb{Z}} \mu(\{r\}) = \sum_{r \in \mathbb{Z}} a_r = 1.$$

Inversement, supposons que, pour tout $r \in \mathbb{Z}$, la suite $(a_r^n)_{n \in \mathbb{N}}$ converge vers un réel $a_r \geq 0$ et que l'on ait $\sum_{r \in \mathbb{Z}} a_r = 1$ et $\mu = \sum_{r \in \mathbb{Z}} a_r \delta_r$. Pour tout $f \in \mathcal{C}_{\mathcal{K}}(\mathbb{R})$ de support compact K , on a

$$\int_{\mathbb{R}} f d\mu_n = \sum_{r \in K} f(r) a_r^n \quad \text{et} \quad \int_{\mathbb{R}} f d\mu = \sum_{r \in K} f(r) a_r.$$

les sommes ne comportant qu'un nombre fini de termes; il en résulte que $\lim_n \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\mu$, ce qui démontre la convergence vague, et donc étroite de la suite des probabilités μ_n vers la probabilité μ .

Exercice 14.2. Approximation binomiale de la loi hypergéométrique. Soit, pour $j \in \mathbb{N}^*$ fixé, $U^j = U_1^j \sqcup U_2^j$ un ensemble fini partitionné en deux sous ensembles non vides U_1^j et U_2^j ; on note $|U^j| = r^j$ et $|U_1^j| = r_j^1$ (et donc $|U_2^j| = r^j - r_j^1 \geq 1$). Soit un entier n tel que $1 \leq n < r^j$. On extrait « au hasard », c'est-à-dire de manière uniforme, n éléments de U^j ; déterminer, pour tout k tel que $0 \leq k \leq n$, la probabilité d'obtenir exactement k éléments de U_1^j (et donc $n - k$ éléments de U_2^j).

On suppose de plus que les deux suites d'entiers $(r_j^1)_{j \in \mathbb{N}^*}$ et $(r_j)_{j \in \mathbb{N}^*}$ sont croissantes et tendent vers l'infini avec j de telle sorte que $\frac{r_j^1}{r_j} \xrightarrow{j \rightarrow +\infty} p$, où $p \in]0, 1[$. Soit un couple d'entiers tels que $0 \leq k \leq n$; démontrer qu'il existe un entier j_0 tel que, pour tout $j \geq j_0$, on ait $n \leq r^j - r_j^1$ et $n \leq r_j^1$; si $j \geq j_0$, on pose

$$P_{k,n}(r_j^1, r_j) = \frac{\binom{r_j^1}{k} \binom{r^j - r_j^1}{n-k}}{\binom{r^j}{n}}.$$

Démontrer que l'on a la convergence suivante :

$$P_{k,n}(r_j^1, r_j) \xrightarrow{j \rightarrow +\infty} \binom{n}{k} p^k (1-p)^{n-k}.$$

Interpréter ce résultat en termes de convergence étroite d'une suite de probabilités (on utilisera l'exercice précédent).

Solution. Une réalisation est une partie de U^j à n éléments; on choisit pour ensemble des réalisations l'ensemble $\Omega^j = \{A \in \mathcal{P}(U^j) \mid |A| = n\}$. L'événement étudié est la partie A_k de Ω^j :

$$A_k = \left\{ A \in \Omega^j \mid |A \cap U_1^j| = k \right\}.$$

Sur l'espace probabilisable $(\Omega^j, \mathcal{A}^j)$ où $\mathcal{A}^j = \mathcal{P}(\Omega^j)$, on met la probabilité uniforme P^j (c'est la traduction de l'usage courant de l'expression « au hasard »). On cherche en fait la probabilité $P^j(A_k)$. L'ensemble A_k est vide si et seulement si $r_j^1 < k \leq n$ ou si $0 \leq k < n - (r^j - r_j^1)$. Sinon, c'est-à-dire si

$$\max(0, n - (r^j - r_j^1)) \leq k \leq \min(n, r_j^1),$$

on a : $|A_k| = \binom{r_j^1}{k} \binom{r^j - r_j^1}{n-k}$; on a de plus $|\Omega^j| = \binom{r^j}{n}$. Il en résulte que l'on a alors

$$P^j(A_k) = \frac{\binom{r_j^1}{k} \binom{r^j - r_j^1}{n-k}}{\binom{r^j}{n}}.$$

Les événements A_k , $0 \leq k \leq n$ forment une partition de Ω^j ; la mesure

$$\mu^j = \sum_{k=0}^n P^j(A_k) \delta_k = \sum_{k=\max(0, n-(r^j-r_j^1))}^{\min(n, r_j^1)} \frac{\binom{r_j^1}{k} \binom{r^j - r_j^1}{n-k}}{\binom{r^j}{n}} \delta_k$$

est donc une probabilité, appelée **loi hypergéométrique**.

Soit $\varepsilon > 0$ tel que $\varepsilon < \min(p, 1-p)$. Il existe j_1 tel que, pour tout $j \geq j_1$, on ait $p - \varepsilon \leq \frac{r_j^1}{r_j} \leq p + \varepsilon$, ce qui implique $r^j(1-p-\varepsilon) \leq r_j - r_j^1 \leq r^j(1-p+\varepsilon)$; puisque les suites $(r_j^1)_{j \in \mathbb{N}^*}$ et $(r_j)_{j \in \mathbb{N}^*}$ tendent vers l'infini avec j , il existe donc j_0 tel que, pour tout $j \geq j_0$, on ait $n \leq r^j - r_j^1$ et $n \leq r_j^1$. Pour un tel j , on a alors

$$\mu^j = \sum_{k=0}^n \mathbb{P}^j(\mathbf{A}_k) \delta_k = \sum_{k=0}^n \mathbb{P}_{k,n}(r_j^1, r_j) \delta_k \quad (14.24)$$

et, après simplification des coefficients binomiaux,

$$\mathbb{P}_{k,n}(r_j^1, r_j) = \binom{n}{k} \prod_{l=0}^{k-1} \left(\frac{r_j^1 - l}{r_j - l} \right) \prod_{l=0}^{(n-k)-1} \left(\frac{r_j - r_j^1 - l}{r_j - l} \right),$$

soit

$$\mathbb{P}_{k,n}(r_j^1, r_j) = \binom{n}{k} \prod_{l=0}^{k-1} \left(\frac{\frac{r_j^1}{r_j} - \frac{l}{r_j}}{1 - \frac{l}{r_j}} \right) \prod_{l=0}^{(n-k)-1} \left(\frac{1 - \frac{r_j^1}{r_j} - \frac{l}{r_j}}{1 - \frac{l}{r_j}} \right).$$

En vertu des hypothèses faites, cela démontre la convergence

$$\mathbb{P}_{k,n}(r_j^1, r_j) \xrightarrow{j \rightarrow +\infty} \binom{n}{k} p^k (1-p)^{n-k}.$$

L'égalité (14.24) et l'exercice précédent montrent alors la convergence étroite de la suite des probabilités μ^j vers la loi binomiale $\mathcal{B}(n, p) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k$.

Exercice 14.3. Loïs géométriques et exponentielles. Soit X une variable aléatoire réelle positive; pour tout $a > 0$, on définit les variables aléatoires $V_a = \left\lfloor \frac{X}{a} \right\rfloor$ et $X_a = a \left\lfloor \frac{X}{a} \right\rfloor$.

1. Si X est de loi exponentielle $\exp(\lambda)$, où $\lambda > 0$, déterminer la loi de V_a .
2. Si, pour tout $a > 0$, la loi de V_a est la loi géométrique sur \mathbb{N} de paramètre $1 - \exp(-\lambda a)$, calculer la fonction de répartition de X_a . Étudier la convergence étroite quand a tend vers 0 de la famille des lois des X_a .

Solution.

1. La variable aléatoire V_a est à valeurs dans \mathbb{N} . Si X est de loi exponentielle $\exp(\lambda)$, on a, pour tout $n \in \mathbb{N}$,

$$\mathbb{P}(V_a = n) = \mathbb{P}(na \leq X < (n+1)a) = \int_{na}^{(n+1)a} \lambda \exp(-\lambda x) dx,$$

soit

$$\mathbb{P}(V_a = n) = \exp(-\lambda na) [1 - \exp(-\lambda a)].$$

c'est-à-dire que **la loi de V_a est la loi géométrique sur \mathbb{N} de paramètre $1 - \exp(-\lambda a)$.**

2. Inversement, si, pour tout $a > 0$, la loi de V_a est la loi géométrique sur \mathbb{N} de paramètre $1 - \exp(-\lambda a)$, X_a étant à valeurs dans $a\mathbb{N}$, on a, pour tout $n \in \mathbb{N}$,

$$P(X_a = na) = P(V_a = n) = \exp(-\lambda na) [1 - \exp(-\lambda a)].$$

Pour tout réel $x \geq 0$, on a alors

$$P(X_a > x) = \sum_{n: na > x} \exp(-\lambda na) [1 - \exp(-\lambda a)] :$$

si on pose

$$n_0(x) = \inf(n \in \mathbb{N} \mid na > x) = \left[\frac{x}{a} \right] + 1,$$

on a

$$P(X_a > x) = [1 - \exp(-\lambda a)] \sum_{n=n_0(x)}^{+\infty} \exp(-\lambda na),$$

soit

$$P(X_a > x) = \exp(-\lambda n_0(x)a) = \exp(-\lambda a) \exp\left(-\lambda a \left[\frac{x}{a} \right]\right).$$

Autrement dit, puisque de plus, pour tout $x < 0$, on a $P(X_a > x) = 1$, la fonction de répartition F_{X_a} de X_a est donnée par

$$F_{X_a}(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - \exp(-\lambda a) \exp\left(-\lambda a \left[\frac{x}{a} \right]\right) & \text{si } x \geq 0. \end{cases}$$

En remarquant que, pour tout réel $x \geq 0$, on a $x - a < a \left[\frac{x}{a} \right] \leq x$, il vient que $\lim_{a \rightarrow 0} a \left[\frac{x}{a} \right] = x$; on a donc

$$\lim_{a \rightarrow 0} F_{X_a}(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - \exp(-\lambda x) & \text{si } x \geq 0, \end{cases}$$

ce qui démontre que **la famille des lois des variables aléatoires X_a tend étroitement vers la loi exponentielle $\exp(\lambda)$ quand a tend vers 0** (on dit encore que la famille des variables aléatoires X_a converge en loi vers la loi $\exp(\lambda)$ quand a tend vers 0).

Exercice 14.4. Convergence étroite de suite de probabilités gaussiennes sur \mathbb{R} .

Sur l'espace mesurable $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ on considère la suite des mesures gaussiennes $\mu_n = f_n \cdot \lambda$, $n \in \mathbb{N}$, où λ est la mesure de Lebesgue et f_n est la densité définie par, pour tout réel x ,

$$f_n(x) = \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left(-\frac{(x - m_n)^2}{2\sigma_n^2}\right),$$

m_n étant un réel quelconque et σ_n un réel strictement positif.

1. Si les suites $(m_n)_{n \in \mathbb{N}}$ et $(\sigma_n)_{n \in \mathbb{N}}$ sont convergentes respectivement vers m et σ , étudier la convergence étroite de la suite $(\mu_n)_{n \in \mathbb{N}}$ directement à partir de la définition de cette notion de convergence. Que dit de plus le lemme de Scheffé dans le cas où $\sigma > 0$?

2. Si la suite $(m_n)_{n \in \mathbb{N}}$ est bornée et si la suite $(\sigma_n)_{n \in \mathbb{N}}$ tend vers $+\infty$ avec n , étudier les convergences faible et étroite de la suite $(\mu_n)_{n \in \mathbb{N}}$.

Solution. Pour tout $f \in \mathcal{C}_b(\mathbb{R})$, on a

$$\int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f(x) \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left(-\frac{(x-m_n)^2}{2\sigma_n^2}\right) dx,$$

soit, par le changement de variables $y = \frac{x-m_n}{\sigma_n}$,

$$\int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f(y\sigma_n + m_n) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy. \quad (14.25)$$

1. Si les suites $(m_n)_{n \in \mathbb{N}}$ et $(\sigma_n)_{n \in \mathbb{N}}$ sont convergentes, f étant continue, on a

$$\lim_n f(y\sigma_n + m_n) = f(y\sigma + m),$$

puisque de plus, on a, pour tout n ,

$$|f(y\sigma_n + m_n)| \leq \|f\|_{\infty},$$

fonction intégrable par rapport à la probabilité gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$, il résulte du théorème de convergence dominée que l'on a

$$\lim_n \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f(y\sigma + m) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy. \quad (14.26)$$

Si $\sigma > 0$, on peut faire le changement de variables défini par $x = y\sigma + m$, ce qui donne

$$\lim_n \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx,$$

et ainsi on a établi la **convergence étroite de la suite $(\mu_n)_{n \in \mathbb{N}}$ vers la probabilité gaussienne $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$.**

Puisqu'ici on a, pour tout réel x , $\lim_n f_n(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$, le lemme de Scheffé s'applique et donne une convergence uniforme en les boréliens, à savoir que la suite de terme général

$$\sup_{A \in \mathcal{B}_{\mathbb{R}}} \left| \int_A \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-m_n)^2}{2\sigma_n^2}\right) dx - \int_A \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \right|$$

converge vers 0.

Si $\sigma = 0$, puisque $\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = 1$, la relation (14.26) donne

$$\lim_n \int_{\mathbb{R}} f d\mu_n = f(m) = \int_{\mathbb{R}} f d\delta_m.$$

Ainsi on a établi la **convergence étroite de la suite $(\mu_n)_{n \in \mathbb{N}}$ vers la mesure de Dirac en m .**

2. Pour tout $f \in \mathcal{C}_0(\mathbb{R})$, la relation (14.25) est en particulier vraie; si la suite $(m_n)_{n \in \mathbb{N}}$ est bornée et si la suite $(\sigma_n)_{n \in \mathbb{N}}$ tend vers $+\infty$ avec n , pour tout $y \neq 0$, on a $\lim_n |y\sigma_n + m_n| = 0$, et donc $\lim_n f(y\sigma_n + m_n) = 0$. Le théorème de convergence dominée conduit à $\lim_n \int_{\mathbb{R}} f d\mu_n = 0$. Autrement dit, **la suite $(\mu_n)_{n \in \mathbb{N}}$ converge faiblement vers la mesure nulle 0**. Il n'y a bien sûr pas convergence étroite puisque $\lim_n \mu_n(\mathbb{R}) = 1$ et $0(\mathbb{R}) = 0$.

Remarque. Sous ces dernières hypothèses, pour tout $f \in \mathcal{C}_b(\mathbb{R})$, on a, après le changement de variables défini par $y = x - m_n$,

$$\int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f(y + m_n) \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma_n^2}\right) dy :$$

on a de plus

$$\lim_n f(y + m_n) \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma_n^2}\right) = 0.$$

Cela donne un **exemple où le théorème de convergence dominée ne s'applique pas**.

Exercice 14.5. Variables aléatoires gaussiennes et convergence en loi. (On peut utiliser les résultats de l'exercice précédent). Soit $\{X_0; Z_n, n \in \mathbb{N}\}$ une famille de variables aléatoires réelles gaussiennes définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , indépendantes. On suppose que les Z_n sont de même loi $\mathcal{N}_{\mathbb{R}}(0, \sigma^2)$ où $\sigma > 0$. Soit un réel ρ non nul; pour tout $n \in \mathbb{N}^*$, on définit la variable aléatoire

$$X_n = \rho X_{n-1} + Z_n.$$

Démontrer que X_n admet un moment d'ordre deux et calculer ses moyenne et variance. Étudier la convergence en loi de la suite $(X_n)_{n \in \mathbb{N}}$.

Solution. La variable aléatoire X_0 admet un moment d'ordre deux; supposons qu'il en soit de même pour X_n . Puisque Z_{n+1} est gaussienne, elle admet un moment d'ordre deux et donc aussi X_{n+1} . Par linéarité, on a, pour tout $n \in \mathbb{N}^*$, $EX_n = \rho EX_{n-1}$, et donc

$$EX_n = \rho^n EX_0.$$

La variable aléatoire X_{n-1} est fonction linéaire de $(X_0, Z_1, \dots, Z_{n-1})$; puisque les variables aléatoires X_0, Z_1, \dots, Z_n sont indépendantes, les variables aléatoires X_{n-1} et Z_n le sont aussi. Il en résulte que l'on a, pour tout $n \in \mathbb{N}^*$,

$$\sigma_{X_n}^2 = \rho^2 \sigma_{X_{n-1}}^2 + \sigma_{Z_n}^2 = \rho^2 \sigma_{X_{n-1}}^2 + \sigma^2;$$

un calcul simple conduit à

$$\sigma_{X_n}^2 = \begin{cases} \rho^{2n} \left(\sigma_{X_0}^2 - \frac{\sigma^2}{1 - \rho^2} \right) + \frac{\sigma^2}{1 - \rho^2} & \text{si } |\rho| \neq 1, \\ \sigma_{X_0}^2 + n\sigma^2 & \text{si } |\rho| = 1. \end{cases}$$

Les variables aléatoires X_0, Z_1, \dots, Z_n sont indépendantes et gaussiennes; la variable aléatoire vectorielle (X_0, Z_1, \dots, Z_n) est alors gaussienne. La variable aléatoire réelle X_n , fonction linéaire de (X_0, Z_1, \dots, Z_n) est alors gaussienne. Pour étudier la convergence en loi de la suite $(X_n)_{n \in \mathbb{N}}$, on applique les résultats de l'exercice précédent :

• si $|\rho| < 1$, on a $\lim_n EX_n = 0$ et $\lim_n \sigma_{X_n}^2 = \frac{\sigma^2}{1-\rho^2}$; **la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la loi $\mathcal{N}_{\mathbb{R}}(0, \frac{\sigma^2}{1-\rho^2})$** ;

• si $|\rho| \geq 1$, on a $\lim_n \sigma_{X_n}^2 = +\infty$;

- si $EX_0 = 0$ ou si $|\rho| = 1$, la suite $(EX_n)_{n \in \mathbb{N}}$ est bornée et, d'après l'exercice précédent, **la suite $(X_n)_{n \in \mathbb{N}}$ ne converge pas en loi**.

- si $EX_0 \neq 0$, et $|\rho| > 1$, on a $\lim_n |EX_n| = +\infty$; ce cas n'a pas été étudié dans l'exercice précédent. Posons $m_n = EX_n$ et $\sigma_n = \sigma_{X_n}$; pour tout $f \in \mathcal{C}_0(\mathbb{R})$ on a

$$\int_{\mathbb{R}} f dP_{X_n} = \int_{\mathbb{R}} f(x) \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left(-\frac{(x-m_n)^2}{2\sigma_n^2}\right) dx,$$

soit, par le changement de variables défini par $y = \frac{x-m_n}{\sigma_n}$,

$$\int_{\mathbb{R}} f dP_{X_n} = \int_{\mathbb{R}} f(y\sigma_n + m_n) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy. \quad (14.27)$$

De plus, dans ce cas, on a

$$y\sigma_n + m_n = \rho^n \left[y \left(\sigma_{X_0}^2 - \frac{\sigma^2}{1-\rho^2} + \frac{\sigma^2}{\rho^{2n}(1-\rho^2)} \right)^{\frac{1}{2}} + EX_0 \right].$$

Il en résulte que $\lim_n f(y\sigma_n + m_n) = 0$ pour λ -presque tout y et le théorème de convergence dominée conduit à $\lim_n \int_{\mathbb{R}} f dP_{X_n} = 0$. Autrement dit, **la suite $(P_{X_n})_{n \in \mathbb{N}}$ converge faiblement vers la mesure nulle \emptyset** . Il n'y a bien sûr pas convergence étroite.

Remarque. En résumé, **la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi si et seulement si $|\rho| < 1$** . Il faut noter que l'on aurait pu traiter cet exercice à l'aide du théorème de Lévy.

Exercice 14.6. Convergence en loi. Sur l'espace probabilisé (Ω, \mathcal{A}, P) , on considère, pour tout $n \in \mathbb{N}^*$, des variables aléatoires X_n et Y_n . On suppose que les X_n , $n \in \mathbb{N}^*$, sont de même loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$ et que la loi de Y_n est $P_{Y_n} = (1 - \frac{1}{n})\delta_1 + \frac{1}{n}\delta_0$. Étudier la convergence en loi de la suite $(X_n Y_n)_{n \in \mathbb{N}^*}$.

Solution. La suite $(Y_n)_{n \in \mathbb{N}^*}$ converge en loi vers 1 et on a $\lim_n P(Y_n = 0) = 0$. Pour tout $f \in \mathcal{C}_b(\mathbb{R})$, on a pour tout $n \in \mathbb{N}^*$,

$$\int_{\Omega} f(X_n) dP = \int_{\mathbb{R}} f d\mathcal{N}_{\mathbb{R}}(0, 1),$$

et donc

$$\left| \int_{\Omega} f(X_n Y_n) dP - \int_{\mathbb{R}} f d\mathcal{N}_{\mathbb{R}}(0, 1) \right| = \left| \int_{\Omega} f(X_n Y_n) dP - \int_{\Omega} f(X_n) dP \right|.$$

En tenant compte de ce que les ensembles $(Y_n = 1)$ et $(Y_n = 0)$ forment une partition de Ω , à un ensemble de probabilité nulle près, et que, sur l'ensemble $(Y_n = 1)$, on a $X_n = X_n Y_n$, il vient

$$\begin{aligned} & \left| \int_{\Omega} f(X_n Y_n) dP - \int_{\mathbb{R}} f d\mathcal{N}_{\mathbb{R}}(0, 1) \right| \\ &= \left| \int_{\Omega} f(X_n Y_n) dP - \int_{(Y_n=1)} f(X_n Y_n) dP - \int_{(Y_n=0)} f(X_n) dP \right|, \end{aligned}$$

ce qui donne encore

$$\left| \int_{\Omega} f(X_n Y_n) dP - \int_{\mathbb{R}} f d\mathcal{N}_{\mathbb{R}}(0, 1) \right| = \left| \int_{(Y_n=0)} f(X_n Y_n) dP - \int_{(Y_n=0)} f(X_n) dP \right|.$$

Il en résulte que

$$\left| \int_{\Omega} f(X_n Y_n) dP - \int_{\mathbb{R}} f d\mathcal{N}_{\mathbb{R}}(0, 1) \right| \leq 2 \|f\|_{\infty} P(Y_n = 0),$$

ce qui démontre que

$$\lim_n \int_{\Omega} f(X_n Y_n) dP = \int_{\mathbb{R}} f d\mathcal{N}_{\mathbb{R}}(0, 1),$$

et donc que la suite $(X_n Y_n)_{n \in \mathbb{N}}$ converge en loi vers la loi $\mathcal{N}_{\mathbb{R}}(0, 1)$.

Exercice 14.7. Convergence en loi d'une suite de variables aléatoires à valeurs dans \mathbb{R}^2 et de la suite de ses marginales. Soient, sur le même espace probabilisé (Ω, \mathcal{A}, P) , deux suites $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles qui convergent en loi respectivement vers les variables aléatoires indépendantes X et Y .

1. Si, pour tout $n \in \mathbb{N}$, X_n et Y_n sont indépendantes, démontrer que la suite des variables aléatoires (X_n, Y_n) , $n \in \mathbb{N}$, converge en loi vers (X, Y) . En conclure en particulier que la suite des variables aléatoires $X_n + Y_n$, $n \in \mathbb{N}$, converge en loi vers $X + Y$.

2. On étudie un contre-exemple, dans le cas où on supprime l'hypothèse « pour tout $n \in \mathbb{N}$, X_n et Y_n sont indépendantes ». Soient X et Y deux variables aléatoires réelles indépendantes de même loi de Bernoulli $(\delta_0 + \delta_1)/2$. On pose, pour tout $n \in \mathbb{N}^*$,

$$X_n = X + \frac{1}{n} \quad \text{et} \quad Y_n = (1 - X) - \frac{1}{n}.$$

Étudier la convergence en loi des trois suites $(X_n)_{n \in \mathbb{N}^*}$, $(Y_n)_{n \in \mathbb{N}^*}$ et $(X_n + Y_n)_{n \in \mathbb{N}^*}$. En conclure que la suite des variables aléatoires (X_n, Y_n) , $n \in \mathbb{N}^*$, ne converge pas en loi vers (X, Y) .

Solution.

1. Puisque pour tout $n \in \mathbb{N}$ les variables aléatoires X_n et Y_n sont indépendantes, la fonction caractéristique $\varphi_{(X_n, Y_n)}$ de (X_n, Y_n) est donnée par, pour tout $(u, v) \in \mathbb{R}^2$,

$$\varphi_{(X_n, Y_n)}(u, v) = \varphi_{X_n}(u) \varphi_{Y_n}(v).$$

Les deux suites $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ convergeant en loi respectivement vers les variables aléatoires X et Y , le théorème de Lévy assure que

$$\lim_n \varphi_{X_n}(u) = \varphi_X(u) \quad \text{et} \quad \lim_n \varphi_{Y_n}(v) = \varphi_Y(v),$$

ce qui implique que

$$\lim_n \varphi_{(X_n, Y_n)}(u, v) = \varphi_X(u) \varphi_Y(v);$$

les variables aléatoires X et Y étant indépendantes, on a encore

$$\lim_n \varphi_{(X_n, Y_n)}(u, v) = \varphi_{(X, Y)}(u, v).$$

La partie réciproque (b) du théorème de Lévy (th. 14.11) montre alors que **la suite des variables aléatoires** (X_n, Y_n) , $n \in \mathbb{N}$, **converge en loi vers** (X, Y) . La variable aléatoire $X_n + Y_n$ étant une fonction **continue** de (X_n, Y_n) , **la suite des variables aléatoires** $X_n + Y_n$, $n \in \mathbb{N}$, **converge alors en loi vers** $X + Y$.

2. Les suites $(X_n)_{n \in \mathbb{N}^*}$ et $(Y_n)_{n \in \mathbb{N}^*}$ convergent P-p.s., et donc en loi, respectivement vers X et $1 - X$. Puisque les variables aléatoires X , $1 - X$ et Y ont même loi on a

$$X_n \xrightarrow{\mathcal{L}} X \quad \text{et} \quad Y_n \xrightarrow{\mathcal{L}} Y.$$

Par contre, on a, pour tout $n \in \mathbb{N}^*$, $X_n + Y_n = 1$; il en résulte que **la suite des variables aléatoires** $X_n + Y_n$, $n \in \mathbb{N}$, **converge en loi vers** δ_1 alors que, les variables aléatoires X et Y étant indépendantes, on a

$$P_{X+Y} = \frac{1}{4}(\delta_0 + \delta_2) + \frac{1}{2}\delta_1;$$

la suite des variables aléatoires $X_n + Y_n$, $n \in \mathbb{N}$, **ne converge pas en loi vers** $X + Y$. A fortiori, **la suite des variables aléatoires** (X_n, Y_n) , $n \in \mathbb{N}$, **ne converge pas en loi vers** (X, Y) .

Le lemme de **Slutsky** donne une hypothèse alternative à l'indépendance pour assurer la propriété de convergence en loi étudiée à la première question de l'exercice précédent.

Exercice 14.8. Lemme de Slutsky. Soient, sur le même espace probabilisé (Ω, \mathcal{A}, P) , deux suites $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles qui convergent en loi respectivement vers une variable aléatoire X et une **constante** y_0 . Démontrer que la suite des variables aléatoires (X_n, Y_n) , $n \in \mathbb{N}$, converge en loi vers (X, y_0) (on admettra que l'ensemble $\mathcal{H} = \{(x, y) \mapsto f(x)g(y) \mid f, g \in \mathcal{C}_0(\mathbb{R})\}$ est total dans $\mathcal{C}_0(\mathbb{R}^2)$ ou, alternativement, on utilisera le théorème de Lévy). En conclure en particulier que la suite des variables aléatoires $X_n + Y_n$, $n \in \mathbb{N}$, converge en loi vers $X + y_0$.

Démontrer que si la suite $(X_n)_{n \in \mathbb{N}}$ converge **en loi** vers une variable aléatoire X et si la suite $(X_n - Y_n)_{n \in \mathbb{N}}$ converge **en probabilité** vers 0, alors la suite $(Y_n)_{n \in \mathbb{N}}$ converge en loi vers X .

Solution. Soient f et g appartenant à $\mathcal{C}_0(\mathbb{R})$ quelconques. La suite $(Y_n)_{n \in \mathbb{N}}$ convergant en loi vers une **constante** y_0 , converge en probabilité vers y_0 ; la fonction g étant continue, la suite $(g(Y_n))_{n \in \mathbb{N}}$ converge alors en probabilité vers $g(y_0)$. Soit $\varepsilon > 0$ quelconque; on a alors

$$\lim_n P(|g(Y_n) - g(y_0)| > \varepsilon) = 0. \quad (14.28)$$

Par le théorème de transfert, on a

$$\begin{aligned} & \left| \int_{\mathbb{R}^2} f(x)g(y) dP_{(X_n, Y_n)}(x, y) - \int_{\mathbb{R}^2} f(x)g(y) dP_X \otimes \delta_{y_0}(x, y) \right| \\ &= \left| \int_{\Omega} f(X_n)g(Y_n) dP - \int_{\Omega} f(X)g(y_0) dP \right|. \end{aligned}$$

Il résulte alors de l'inégalité triangulaire que

$$\begin{aligned} & \left| \int_{\mathbb{R}^2} f(x)g(y) dP_{(X_n, Y_n)}(x, y) - \int_{\mathbb{R}^2} f(x)g(y) dP_X \otimes \delta_{y_0}(x, y) \right| \\ & \leq \left| \int_{\Omega} f(X_n)g(Y_n) dP - \int_{\Omega} f(X_n)g(y_0) dP \right| \\ & \quad + \left| \int_{\Omega} f(X_n)g(y_0) dP - \int_{\Omega} f(X)g(y_0) dP \right|, \end{aligned}$$

et donc que

$$\begin{aligned} & \left| \int_{\mathbb{R}^2} f(x)g(y) dP_{(X_n, Y_n)}(x, y) - \int_{\mathbb{R}^2} f(x)g(y) dP_X \otimes \delta_{y_0}(x, y) \right| \\ & \leq \|f\|_{\infty} \left[\int_{\Omega} |g(Y_n) - g(y_0)| dP \right] + |g(y_0)| \left| \int_{\Omega} f(X_n) - f(X) dP \right|, \end{aligned}$$

soit, après avoir partitionné dans la première intégrale par l'ensemble $(|g(Y_n) - g(y_0)| > \varepsilon)$ et son complémentaire,

$$\begin{aligned} & \left| \int_{\mathbb{R}^2} f(x)g(y) dP_{(X_n, Y_n)}(x, y) - \int_{\mathbb{R}^2} f(x)g(y) dP_X \otimes \delta_{y_0}(x, y) \right| \\ & \leq \|f\|_{\infty} \left[\varepsilon + \int_{(|g(Y_n) - g(y_0)| > \varepsilon)} |g(Y_n) - g(y_0)| dP \right] + \|g\|_{\infty} \left| \int_{\Omega} f(X_n) - f(X) dP \right| \\ & \leq \|f\|_{\infty} [\varepsilon + 2\|g\|_{\infty} P(|g(Y_n) - g(y_0)| > \varepsilon)] + \|g\|_{\infty} \left| \int_{\mathbb{R}} f dP_{X_n} - \int_{\mathbb{R}} f dP_X \right|. \end{aligned}$$

Il résulte de la convergence en loi vers X de la suite $(X_n)_{n \in \mathbb{N}}$, puis de la relation (14.28), que l'on a

$$\limsup_n \left| \int_{\mathbb{R}^2} f(x)g(y) dP_{(X_n, Y_n)}(x, y) - \int_{\mathbb{R}^2} f(x)g(y) dP_X \otimes \delta_{y_0}(x, y) \right| \leq \|f\|_{\infty} \varepsilon;$$

le membre de droite de cette inégalité étant positif, l'arbitraire de ε assure alors que

$$\lim_n \int_{\mathbb{R}^2} f(x)g(y) dP_{(X_n, Y_n)}(x, y) = \int_{\mathbb{R}^2} f(x)g(y) dP_X \otimes \delta_{y_0}(x, y);$$

l'ensemble \mathcal{H} étant total dans $\mathcal{C}_0(\mathbb{R}^2)$, cela démontre que **la suite des variables aléatoires (X_n, Y_n) , $n \in \mathbb{N}$, converge en loi vers (X, y_0)** . La somme étant une application continue, il en résulte que **la suite des variables aléatoires $X_n + Y_n$, $n \in \mathbb{N}$, converge en loi vers $X + y_0$** .

Remarque. On peut faire une démonstration en presque tout point analogue en utilisant le théorème de Lévy (ce n'est pas surprenant : dans les deux points de vue est présent le même argument de densité, celui-ci étant un ingrédient de la démonstration du théorème de Lévy). En voici une présentation.

La suite $(Y_n)_{n \in \mathbb{N}}$ convergeant en loi vers une **constante** y_0 , converge en probabilité vers y_0 ; la fonction $y \mapsto \exp(i v y)$ étant continue, la suite $(\exp(i v Y_n))_{n \in \mathbb{N}}$ converge alors en probabilité vers $\exp(i v y_0)$. Soit $\varepsilon > 0$ quelconque; on a alors

$$\lim_n P(|\exp(i v Y_n) - \exp(i v y_0)| > \varepsilon) = 0. \quad (14.29)$$

Soient $\varphi_{(X_n, Y_n)}$ la fonction caractéristique de (X_n, Y_n) et $\widehat{P_X \otimes \delta_{y_0}}$ la transformée de Fourier de la probabilité $P_X \otimes \delta_{y_0}$. Il résulte alors de l'inégalité triangulaire que

$$\begin{aligned} & |\varphi_{(X_n, Y_n)}(u, v) - \widehat{P_X \otimes \delta_{y_0}}(u, v)| \\ & \leq \left| \int_{\Omega} \exp(i u X_n) \exp(i v Y_n) dP - \int_{\Omega} \exp(i u X_n) \exp(i v y_0) dP \right| \\ & \quad + \left| \int_{\Omega} \exp(i u X_n) \exp(i v y_0) dP - \int_{\Omega} \exp(i u X) \exp(i v y_0) dP \right|, \end{aligned}$$

et donc que

$$\begin{aligned} & |\varphi_{(X_n, Y_n)}(u, v) - \widehat{P_X \otimes \delta_{y_0}}(u, v)| \\ & \leq \int_{\Omega} |\exp(i v Y_n) - \exp(i v y_0)| dP + \left| \int_{\Omega} \exp(i u X_n) - \exp(i u X) dP \right| \end{aligned}$$

soit, après avoir partitionné par l'ensemble $(|\exp(i v Y_n) - \exp(i v y_0)| > \varepsilon)$ et son complémentaire dans la première intégrale,

$$\begin{aligned} & |\varphi_{(X_n, Y_n)}(u, v) - \widehat{P_X \otimes \delta_{y_0}}(u, v)| \\ & \leq \varepsilon + \int_{(|\exp(i v Y_n) - \exp(i v y_0)| > \varepsilon)} |\exp(i v Y_n) - \exp(i v y_0)| dP + |\varphi_{X_n}(u) - \varphi_X(u)| \\ & \leq \varepsilon + 2P(|\exp(i v Y_n) - \exp(i v y_0)| > \varepsilon) + |\varphi_{X_n}(u) - \varphi_X(u)|. \end{aligned}$$

Il résulte de la convergence en loi vers X de la suite $(X_n)_{n \in \mathbb{N}}$, traduite à l'aide du théorème de Lévy, puis de la relation (14.29) que l'on a

$$0 \leq \limsup_n |\varphi_{(X_n, Y_n)}(u, v) - \widehat{P_X \otimes \delta_{y_0}}(u, v)| \leq \varepsilon,$$

ce qui, étant donné l'arbitraire de ε montre que

$$\lim_n \varphi_{(X_n, Y_n)}(u, v) = \widehat{P_X \otimes \delta_{y_0}}(u, v);$$

la partie réciproque du théorème de Lévy démontre alors que **la suite des variables aléatoires** (X_n, Y_n) , $n \in \mathbb{N}$, **converge en loi vers** (X, y_0) .

Si la suite $(X_n)_{n \in \mathbb{N}}$ converge **en loi** vers une variable aléatoire X et si la suite $(X_n - Y_n)_{n \in \mathbb{N}}$ converge **en probabilité** vers 0, la suite $(X_n - Y_n)_{n \in \mathbb{N}}$ converge **en loi** vers 0; donc, par le lemme de Slutsky précédemment démontré, on a $(X_n, Y_n - X_n) \xrightarrow{\mathcal{L}} (X, 0)$. Il en résulte que, puisque $Y_n = (Y_n - X_n) + X_n$, la suite $(Y_n)_{n \in \mathbb{N}}$ converge en loi vers X .

Exercice 14.9. Développement décimal, convergence en loi et théorème de Lévy.

Soit une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{A}, P) , indépendantes, de même loi uniforme sur l'ensemble des entiers $\{0, 1, 2, \dots, 9\}$. On définit, pour tout $n \in \mathbb{N}$, la variable aléatoire $Y_n = \sum_{j=0}^n \frac{X_j}{10^j}$. Démontrer que la suite $(Y_n)_{n \in \mathbb{N}}$ converge P.-p.s. vers une variable aléatoire Y dont on déterminera la loi.

Solution. On a P.-p.s., pour tout $n \in \mathbb{N}$, $0 \leq \frac{X_n}{10^n} \leq \frac{1}{10^{n-1}}$, ce qui montre que la série de terme général $\frac{X_n}{10^n}$ est P.-p.s. convergente, c'est-à-dire que **la suite $(Y_n)_{n \in \mathbb{N}}$ converge P.-p.s. vers une variable aléatoire Y** ; il y a donc aussi convergence en probabilité et donc aussi en loi. Le théorème de Lévy va permettre d'identifier la loi de Y . Les variables aléatoires X_n étant indépendantes et de même loi, donc de même fonction caractéristique, la fonction caractéristique de Y_n est donnée en tout réel t par

$$\varphi_{Y_n}(t) = \prod_{j=0}^n \varphi_{X_j}\left(\frac{t}{10^j}\right) = \prod_{j=0}^n \varphi_{X_0}\left(\frac{t}{10^j}\right).$$

La fonction caractéristique de X_0 est donnée en tout réel t par

$$\varphi_{X_0}(t) = \frac{1}{10} \sum_{j=0}^9 \exp(ijt) = \begin{cases} \frac{1}{10} \frac{1 - \exp(i10t)}{1 - \exp(it)} & \text{si } \exp(it) \neq 1, \\ 1 & \text{sinon.} \end{cases}$$

Si $\exp(it) \neq 1$, c'est-à-dire si $t \notin 2\pi\mathbb{Z}$, on a, pour tout $j \in \mathbb{N}$, $\exp(i \frac{t}{10^j}) \neq 1$ (car s'il existe $j \in \mathbb{N}$ tel que $\exp(i \frac{t}{10^j}) = 1$, on a aussi $\exp(it) = 1$), si bien que l'on a, en simplifiant,

$$\varphi_{Y_n}(t) = \frac{1}{10^{n+1}} \prod_{j=0}^n \frac{1 - \exp\left(i \frac{t}{10^{j-1}}\right)}{1 - \exp\left(i \frac{t}{10^j}\right)} = \frac{1}{10^{n+1}} \frac{1 - \exp(i10t)}{1 - \exp\left(i \frac{t}{10^n}\right)}.$$

Dans ce cas, on a

$$\varphi_{Y_n}(t) = \frac{1 - \exp(i10t)}{10^{n+1}} \left[\frac{1}{1 - \left(1 + i \frac{t}{10^n} + o\left(\frac{t}{10^n}\right)\right)} \right].$$

On a donc, pour tout réel $t \neq 2\pi\mathbb{Z}$,

$$\lim_n \varphi_{Y_n}(t) = \frac{\exp(i10t) - 1}{10it} = \widehat{\mathcal{U}([0, 10])}(t). \quad (14.30)$$

où $\widehat{\mathcal{U}([0, 10])}$ est la transformée de Fourier de la loi uniforme sur l'intervalle $[0, 10]$. Puisque la suite $(Y_n)_{n \in \mathbb{N}}$ converge en loi vers Y , il résulte du théorème de Lévy que la suite $(\varphi_{Y_n})_{n \in \mathbb{N}}$ converge simplement vers la fonction caractéristique φ_Y de Y . La relation (14.30) implique alors que l'on a, pour tout réel $t \neq 2\pi\mathbb{Z}$, $\varphi_Y(t) = \widehat{\mathcal{U}([0, 10])}(t)$; les fonctions φ_Y et $\widehat{\mathcal{U}([0, 10])}$ étant continues sont alors

égales, ce qui, en vertu de l'injectivité de la transformée de Fourier, prouve que **la loi de Y est la loi uniforme sur l'intervalle** $[0, 10]$.

Exercice 14.10. Convergence en loi et fonctions de répartition. Soit une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{A}, P) , indépendantes, de même loi, de fonction de répartition F . On définit, pour tout $n \in \mathbb{N}^*$, les variables aléatoires I_n et M_n par

$$I_n = \min_{1 \leq i \leq n} X_i \quad \text{et} \quad M_n = \max_{1 \leq i \leq n} X_i.$$

1. Étudier la convergence en loi des suites $(I_n)_{n \in \mathbb{N}^*}$ et $(M_n)_{n \in \mathbb{N}^*}$.

2. On suppose que les X_n , $n \in \mathbb{N}^*$, sont de même loi exponentielle $\exp(\lambda)$ où $\lambda > 0$. On pose, pour tout $n \in \mathbb{N}^*$, $Z_n = \frac{M_n}{\ln n}$; étudier la convergence en loi de la suite $(Z_n)_{n \in \mathbb{N}^*}$.

Solution. Les variables aléatoires I_n et M_n étant définies par des opérations relatives à la structure d'ordre, il est judicieux ici d'employer le critère de convergence en loi en termes de fonctions de répartition.

1. Pour tout réel x , on a, en tenant compte de l'indépendance des variables aléatoires X_n ,

$$P(I_n > x) = P\left[\bigcap_{1 \leq i \leq n} (X_i > x)\right] = \prod_{1 \leq i \leq n} P(X_i > x);$$

les X_n ayant même fonction de répartition F , la fonction de répartition de I_n est alors donnée par, pour tout réel x ,

$$F_{I_n}(x) = 1 - (1 - F(x))^n.$$

Il en résulte que

$$\lim_n F_{I_n}(x) = \begin{cases} 0 & \text{si } F(x) = 0, \\ 1 & \text{si } 0 < F(x) \leq 1. \end{cases}$$

— Si $x_i = \inf\{x \mid F(x) > 0\} > -\infty$, on a alors

$$\lim_n F_{I_n}(x) = \begin{cases} 0 & \text{si } x < x_i, \\ 1 & \text{si } x > x_i, \end{cases}$$

ce qui démontre que : $I_n \xrightarrow{\mathcal{L}} x_i$.

— Si $x_i = \inf\{x \mid F(x) > 0\} = -\infty$, on a alors, pour tout réel x , $\lim_n F_{I_n}(x) = 1$; la fonction limite n'est pas une fonction de répartition : **il n'y a pas convergence en loi de la suite** $(I_n)_{n \in \mathbb{N}^*}$.

De même, pour tout réel x , on a, en tenant compte de l'indépendance des variables aléatoires X_n ,

$$P(M_n \leq x) = P\left[\bigcap_{1 \leq i \leq n} (X_i \leq x)\right] = \prod_{1 \leq i \leq n} P(X_i \leq x);$$

les X_n ayant même fonction de répartition F , la fonction de répartition de M_n est alors donnée par

$$F_{M_n}(x) = [F(x)]^n.$$

Il en résulte que

$$\lim_n F_{M_n}(x) = \begin{cases} 0 & \text{si } F(x) < 1, \\ 1 & \text{si } F(x) = 1. \end{cases}$$

— Si $x_s = \inf\{x \mid F(x) = 1\} < +\infty$, on a alors

$$\lim_n F_{M_n}(x) = \begin{cases} 0 & \text{si } x < x_s, \\ 1 & \text{si } x > x_s, \end{cases}$$

ce qui démontre que : $\boxed{M_n \xrightarrow{\mathcal{L}} x_s}$.

— Si $x_s = \inf\{x \mid F(x) = 1\} = +\infty$, on a alors, pour tout réel x , $\lim_n F_{M_n}(x) = 0$: la fonction limite n'est pas une fonction de répartition ; **il n'y a pas convergence en loi de la suite $(M_n)_{n \in \mathbb{N}^*}$** .

— On a dans ce cas, pour tout réel x ,

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - \exp(-\lambda x) & \text{si } x > 0, \end{cases}$$

si bien que la fonction de répartition de Z_n est donnée par, pour tout $x > 0$,

$$F_{Z_n}(x) = F_{M_n}(x \ln n) = \begin{cases} 0 & \text{si } x \leq 0, \\ [1 - \exp(-\lambda x \ln n)]^n & \text{si } x > 0. \end{cases}$$

Si $x > 0$, on a alors

$$\ln F_{Z_n}(x) = n \ln \left[1 - \frac{1}{n^{\lambda x}} \right] = n \left[-\frac{1}{n^{\lambda x}} + o\left(\frac{1}{n^{\lambda x}}\right) \right].$$

Il en résulte que

$$\lim_n \ln F_{Z_n}(x) = \begin{cases} 0 & \text{si } \lambda x > 1, \\ -\infty & \text{si } 0 < \lambda x < 1; \end{cases}$$

puisque de plus, pour tout $x \leq 0$, on a $\lim_n F_{Z_n}(x) = 0$, il vient

$$\lim_n F_{Z_n}(x) = \begin{cases} 1 & \text{si } x > \frac{1}{\lambda}, \\ 0 & \text{si } x < \frac{1}{\lambda}. \end{cases}$$

ce qui, en remarquant que $\frac{1}{\lambda} = EX_1$, démontre que : $\boxed{Z_n \xrightarrow{\mathcal{L}} EX_1}$.

Exercice 14.11. Inégalité intégrale pour la partie réelle d'une fonction caractéristique : convergence en loi d'une série de variables aléatoires indépendantes (théorème de Lévy). Toutes les variables aléatoires sont définies sur le même espace probabilisé (Ω, \mathcal{A}, P) .

Soit X une variable aléatoire réelle de fonction caractéristique φ_X . Soit g la fonction réelle définie sur \mathbb{R} par

$$g(x) = \begin{cases} 1 - \frac{\sin x}{x} & \text{si } x \neq 0, \\ 0 & \text{si } x = 0. \end{cases}$$

1. Vérifier que $g \in \mathcal{C}_b(\mathbb{R})$, est positive, et que $g(x) = 0$ si et seulement si $x = 0$. Soit $\delta > 0$ quelconque ; démontrer l'égalité :

$$\frac{1}{\delta} \int_0^\delta (1 - \Re \varphi_X(t)) dt = \int_\Omega g(\delta X) dP. \quad (14.31)$$

2. Pour tout $\varepsilon > 0$, on note $I_\varepsilon = \inf_{|x| > \varepsilon} g(x) > 0$. Démontrer que l'on a

$$P(|X| > \varepsilon) \leq \frac{1}{I_\varepsilon \delta} \int_0^\delta (1 - \Re \varphi_X(t)) dt = \frac{1}{2I_\varepsilon \delta} \int_{-\delta}^\delta (1 - \varphi_X(t)) dt. \quad (14.32)$$

Soit une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires réelles. On note $S_n = \sum_{j=1}^n X_j$.

3. Démontrer que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers 0 (et donc en probabilité vers 0) si et seulement si il existe $\delta > 0$ tel que la suite $(\varphi_{X_n}(t))_{n \in \mathbb{N}^*}$ converge vers 1 pour tout $t \in [-\delta, \delta]$.

4. On suppose les variables aléatoires $X_n, n \in \mathbb{N}^*$, indépendantes. Démontrer que la suite $(S_n)_{n \in \mathbb{N}^*}$ converge en loi si et seulement si elle converge en probabilité (théorème de Lévy).

Solution.

1. La continuité en 0 résulte de ce que $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$; g est de plus continue en tout autre point, paire et positive (car $|\sin x| \leq |x|$). De plus $\lim_{x \rightarrow +\infty} g(x) = 1$, ce qui montre $g \in \mathcal{C}_b(\mathbb{R})$. Enfin, pour tout $x > 0$, on a

$$xg(x) = \int_0^x (1 - \sin u) du > 0,$$

ce qui démontre que $g(x) = 0$ si et seulement si $x = 0$. On a

$$\frac{1}{\delta} \int_0^\delta (1 - \Re \varphi_X(t)) dt = \frac{1}{\delta} \int_0^\delta (1 - E \cos(tX)) dt;$$

puisque $0 \leq 1 - \cos(tX)$, on peut appliquer le théorème de Fubini, ce qui donne :

$$\frac{1}{\delta} \int_0^\delta (1 - \Re \varphi_X(t)) dt = \frac{1}{\delta} \int_\Omega \left[\int_0^\delta (1 - \cos(tX)) dt \right] dP.$$

Il en résulte que

$$\begin{aligned} \frac{1}{\delta} \int_0^\delta (1 - \Re \varphi_X(t)) dt &= \int_{(X \neq 0)} \left[\frac{1}{\delta} \int_0^\delta (1 - \cos(tX)) dt \right] dP \\ &= \int_{(X \neq 0)} \left[1 - \frac{\sin \delta X}{\delta X} \right] dP, \end{aligned}$$

ce qui démontre (14.31), puisque $g(0) = 0$.

2. Puisque $\varphi_X(-t) = \overline{\varphi_X(t)}$, il vient, après le changement de variable défini par $-t = u$,

$$\int_{-\delta}^0 (1 - \varphi_X(t)) dt = \int_0^{\delta} (1 - \varphi_X(-u)) du = \int_0^{\delta} (1 - \overline{\varphi_X(u)}) du,$$

ce qui implique que

$$\int_{-\delta}^{\delta} (1 - \varphi_X(t)) dt = 2 \int_0^{\delta} (1 - \Re \varphi_X(t)) dt;$$

on obtient ainsi l'égalité dans la relation (14.32). Il résulte de (14.31), de la positivité de g et de la définition de I_ε que l'on a

$$\begin{aligned} \frac{1}{\delta} \int_0^{\delta} (1 - \Re \varphi_X(t)) dt &= \int_{\Omega} g(\delta X) dP \\ &\geq \int_{(|X| > \varepsilon)} g(\delta X) dP \\ &\geq I_\varepsilon P(|X| > \varepsilon), \end{aligned}$$

ce qui achève de démontrer (14.32).

3. Si la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers 0, le théorème de Lévy assure la convergence simple de la suite $(\varphi_{X_n}(t))_{n \in \mathbb{N}^*}$ vers 1, et donc a fortiori sur tout intervalle $[-\delta, \delta]$. **Inversement**, supposons qu'il existe $\delta > 0$ tel que la suite $(\varphi_{X_n}(t))_{n \in \mathbb{N}^*}$ converge vers 1 pour tout $t \in [-\delta, \delta]$. Il résulte de (14.32) que, pour tout $\varepsilon > 0$ et pour tout $n \in \mathbb{N}^*$, on a

$$P(|X_n| > \varepsilon) \leq \frac{1}{2I_\varepsilon \delta} \int_{-\delta}^{\delta} |1 - \varphi_{X_n}(t)| dt.$$

Puisque $\lim_n |1 - \varphi_{X_n}(t)| = 0$ sur $[-\delta, \delta]$ et que $|1 - \varphi_{X_n}(t)| \leq 2$, il résulte du théorème de convergence dominée que

$$\lim_n P(|X_n| > \varepsilon) = 0.$$

ce qui démontre que **la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en probabilité, et donc en loi, vers 0.**

4. **Supposons que la suite $(S_n)_{n \in \mathbb{N}^*}$ converge en loi.** Soit $\delta > 0$ quelconque fixé et soient des entiers m et n quelconques tels que $m < n$ il résulte de (14.32) que, pour tout $\varepsilon > 0$, on a

$$P(|S_n - S_m| > \varepsilon) \leq \frac{1}{2I_\varepsilon \delta} \int_{-\delta}^{\delta} |1 - \varphi_{S_n - S_m}(t)| dt. \quad (14.33)$$

Les variables aléatoires S_m et $S_n - S_m$ sont indépendantes, ce qui donne, pour tout réel t , l'égalité (en termes de fonctions caractéristiques)

$$\varphi_{S_n}(t) = \varphi_{S_m}(t) \varphi_{S_n - S_m}(t).$$

et donc l'égalité

$$\varphi_{S_n}(t) - \varphi_{S_m}(t) = \varphi_{S_m}(t) [1 - \varphi_{S_n - S_m}(t)].$$

Puisque la suite $(S_n)_{n \in \mathbb{N}^*}$ converge en loi, il résulte du théorème de Lévy que la suite $(\varphi_{S_n})_{n \in \mathbb{N}^*}$ converge simplement vers une fonction φ qui vaut 1 en 0 et que la convergence est **uniforme** sur $[-\delta, \delta]$. Il existe donc N tel que l'on ait $|\varphi_{S_m}(t)| \geq 1/2$ dès que $m \geq N$. Si on a $n > m \geq N$, on a alors

$$|\varphi_{S_n}(t) - \varphi_{S_m}(t)| = |\varphi_{S_m}(t)| |1 - \varphi_{S_n - S_m}(t)| \geq \frac{1}{2} |1 - \varphi_{S_n - S_m}(t)|,$$

ce qui, en reportant dans (14.33), donne la majoration

$$P(|S_n - S_m| > \varepsilon) \leq \frac{1}{1_\varepsilon \delta} \int_{-\delta}^{\delta} |\varphi_{S_n}(t) - \varphi_{S_m}(t)| dt.$$

Puisque $\lim_{n,m} |\varphi_{S_n}(t) - \varphi_{S_m}(t)| = 0$ et que $|\varphi_{S_n}(t) - \varphi_{S_m}(t)| \leq 2$, il résulte d'une double application du théorème de convergence dominée que

$$\lim_{n,m} P(|S_n - S_m| > \varepsilon) = 0.$$

c'est-à-dire que la suite $(S_n)_{n \in \mathbb{N}^*}$ est de **Cauchy** pour la convergence en probabilité. Ainsi, **la suite $(S_n)_{n \in \mathbb{N}^*}$ converge en probabilité**. La convergence en probabilité impliquant la convergence en loi, la réciproque est vraie.

Remarque. En conséquence de l'**inégalité d'Ottaviani**, on a démontré en exercice au chapitre 10 (ex. 10, chap. 10) l'autre partie de ce **théorème de Lévy**, à savoir que, pour une **série** de variables aléatoires **indépendantes**, les convergences en probabilité et P-p.s. sont équivalentes.

Exercice 14.12. Variables aléatoires gaussiennes, lois conditionnelles, fonctions caractéristiques et convergence en loi.

Notation. Un vecteur (x_1, x_2, \dots, x_n) de \mathbb{R}^n est noté \underline{x}_n . Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{A}, P) . On suppose que X_1 est de loi gaussienne $\mathcal{N}_{\mathbb{R}}(0, 1)$ et que, pour tout $n > 1$, une loi conditionnelle $P_{\underline{X}_{n+1}}^{\underline{X}_n = \underline{x}_n}$ de X_{n+1} sachant \underline{X}_n est, pour tout $\underline{x}_n \in \mathbb{R}^n$, la loi gaussienne $\mathcal{N}_{\mathbb{R}}(x_n, 1)$.

1. Quelle est la loi de (X_1, X_2) ? Trouver, à un facteur multiplicatif près, une combinaison linéaire de X_1 et X_2 qui soit indépendante de X_1 .
2. Soit \mathcal{B}_n la tribu engendrée par \underline{X}_n . Calculer les espérances conditionnelles $E^{\mathcal{B}_n} X_{n+1}$ et $E^{\mathcal{B}_n} X_{n+1}^2$. En déduire les moyenne et variance de X_n . Montrer que la suite $(X_n)_{n \in \mathbb{N}^*}$ ne converge pas dans L^2 .
3. Justifier l'existence d'une densité $f_{\underline{X}_n}$ pour la variable aléatoire \underline{X}_n et la calculer (on traitera d'abord le cas $n = 3$). Quelle est la fonction caractéristique de \underline{X}_n ?

4. Soit $j < k$. Quelle est la loi de la variable aléatoire (X_j, X_k) ? Quel est le coefficient de corrélation de X_j et X_k ? Étudier la convergence en loi de la suite de variables aléatoires $(X_j, \frac{X_k}{\sqrt{k}})_{k \in \mathbb{N}^*}$; que peut-on dire de la loi limite?
5. Soit, pour tout $n \in \mathbb{N}^*$, la variable aléatoire $Z_n = \frac{1}{n\sqrt{n}} \sum_{j=1}^n X_j$. Étudier la convergence en loi de la suite variables aléatoires $(Z_n)_{n \in \mathbb{N}^*}$.

Solution.

1. Puisque X_1 admet une densité, et du fait de l'existence d'une densité conditionnelle de X_2 sachant X_1 , la variable aléatoire (X_1, X_2) admet une densité $f_{(X_1, X_2)}$ donnée par, pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$\begin{aligned} f_{(X_1, X_2)}(x_1, x_2) &= f_{X_1}(x_1) f_{X_2}^{X_1=x_1}(x_2) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_2 - x_1)^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right), \end{aligned}$$

ce qui montre que (X_1, X_2) est **gaussienne** de densité donnée par, pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$f_{(X_1, X_2)}(x_1, x_2) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(x_1^2 + (x_2 - x_1)^2)\right].$$

La variable aléatoire $(Y, Z) = (aX_1 + bX_2, X_1)$ est gaussienne comme transformée linéaire de la variable aléatoire gaussienne (X_1, X_2) . Donc, pour que Y et Z soient indépendantes, il faut et il suffit que $\text{cov}(Y, Z) = 0$, ou encore, puisque Z est centrée, que $E(YZ) = 0$. On a

$$E(YZ) = aEX_1^2 + bE(X_1X_2) = a + b \int_{\mathbb{R}} x_1 m_{X_2}^{X_1=x_1} f_{X_1}(x_1) dx_1.$$

où $m_{X_2}^{X_1=x_1}$ est la moyenne conditionnelle de X_2 sachant X_1 , soit

$$E(YZ) = a + b E(X_1^2) = a + b;$$

ainsi Y et Z sont indépendantes si et seulement si $a + b = 0$.

2. Un représentant (ou version) de l'espérance conditionnelle $E^{\mathcal{B}_n} X_{n+1}$ s'obtient en composant la moyenne conditionnelle $m_{X_{n+1}}^{X_n=x_n}$ avec \underline{X}_n , ce qui donne

$$E^{\mathcal{B}_n} X_{n+1} = X_n.$$

On dit alors que la suite $(X_n)_{n \in \mathbb{N}}$ est une **martingale**¹² relativement à la suite croissante (pour l'inclusion) de sous-tribus $(\mathcal{B}_n)_{n \in \mathbb{N}}$, appelée elle-même **filtration**. De même, par le théorème de transfert conditionnel, on a

$$E^{\mathcal{B}_n}(X_{n+1}^2) = \left[\int_{\mathbb{R}} u^2 dP_{X_{n+1}}^{X_n=u} \right] \circ \underline{X}_n = 1 + X_n^2.$$

12. La théorie des martingales est étudiée au chapitre 15.

Il en résulte que l'on a

$$E(X_{n+1}) = E\left[E^{\mathcal{B}_n} X_{n+1}\right] = E(X_n),$$

et donc que

$$E(X_n) = E(X_1) = 0.$$

De même, on a

$$E(X_n^2) = E\left[E^{\mathcal{B}_n}(X_{n+1}^2)\right] = 1 + E(X_n^2),$$

et, puisque $E(X_1^2) = 1$,

$$E(X_n^2) = n.$$

La suite $(X_n)_{n \in \mathbb{J}^*}$ n'est pas bornée dans L^2 et ne converge donc pas dans L^2 .

3. Le même raisonnement qu'à la première question montre que la variable aléatoire (X_1, X_2, X_3) admet une densité $f_{(X_1, X_2, X_3)}$ donnée par, pour tout $(x_1, x_2, x_3) \in \mathbb{R}^3$,

$$f_{(X_1, X_2, X_3)}(x_1, x_2, x_3) = f_{X_1}(x_1) f_{X_2}^{X_1=x_1}(x_2) f_{X_3}^{X_2=x_2}(x_3),$$

ce qui montre que (X_1, X_2, X_3) est **gaussienne** de densité donnée par, pour tout $(x_1, x_2, x_3) \in \mathbb{R}^3$,

$$f_{(X_1, X_2, X_3)}(x_1, x_2, x_3) = \frac{1}{(2\pi)^{3/2}} \exp\left[-\frac{1}{2}(x_1^2 + (x_2 - x_1)^2 + (x_3 - x_2)^2)\right].$$

On remarque que

$$x_1^2 + (x_2 - x_1)^2 + (x_3 - x_2)^2 = \langle \underline{\underline{A}} \underline{\underline{x}}_3, \underline{\underline{x}}_3 \rangle.$$

où

$$\underline{\underline{A}} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}; \text{ un calcul simple donne } \underline{\underline{A}}^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$

Le même raisonnement montre alors que la variable aléatoire $\underline{\underline{X}}_n$ admet une densité $f_{\underline{\underline{X}}_n}$ donnée par, pour tout $\underline{\underline{x}}_n \in \mathbb{R}^n$,

$$f_{\underline{\underline{X}}_n}(\underline{\underline{x}}_n) = f_{X_1}(x_1) \prod_{j=1}^{n-1} f_{X_{j+1}}^{X_j=x_j}(x_{j+1}).$$

ce qui montre que $\underline{\underline{X}}_n$ est **gaussienne** de densité donnée par, pour tout $\underline{\underline{x}}_n \in \mathbb{R}^n$,

$$f_{\underline{\underline{X}}_n}(\underline{\underline{x}}_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left[-\frac{1}{2}(x_1^2 + (x_2 - x_1)^2 + \cdots + (x_n - x_{n-1})^2)\right].$$

On remarque que

$$x_1^2 + (x_2 - x_1)^2 + \cdots + (x_n - x_{n-1})^2 = \langle \underline{A}_n \underline{x}_n, \underline{x}_n \rangle,$$

où

$$\underline{A}_n = \begin{pmatrix} 2 & -1 & 0 & & 0 \\ -1 & 2 & -1 & 0 & \\ 0 & -1 & 2 & -1 & 0 \\ & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & & 0 & -1 & 2 & -1 \\ 0 & & & 0 & -1 & 1 \end{pmatrix};$$

l'inversion de \underline{A}_n (par exemple, par résolution du système linéaire associé) donne

$$\underline{A}_n^{-1} = \begin{pmatrix} 1 & 1 & 1 & \cdots & \cdots & 1 \\ 1 & 2 & 2 & \cdots & \cdots & 2 \\ 1 & 2 & 3 & \cdots & \cdots & 3 \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \\ 1 & 2 & 3 & \cdots & n-1 & n-1 \\ 1 & 2 & 3 & \cdots & n-1 & n \end{pmatrix}.$$

On a vu à la question précédente que la variable aléatoire gaussienne \underline{X}_n est centrée; sa fonction caractéristique est alors donnée par, pour tout $\underline{t}_n \in \mathbb{R}^n$,

$$\varphi_{\underline{X}_n}(\underline{t}_n) = \exp \left[-\frac{1}{2} \langle \underline{A}_n^{-1} \underline{t}_n, \underline{t}_n \rangle \right]. \quad (14.34)$$

4. Soit $j < k$. La variable aléatoire (X_j, X_k) , marginale de \underline{X}_n , est encore gaussienne centrée et sa matrice de covariance vaut

$$C_{(X_j, X_k)} = \begin{pmatrix} j & j \\ j & k \end{pmatrix}.$$

Le coefficient de corrélation ρ_{X_j, X_k} de X_j et X_k est alors

$$\rho_{X_j, X_k} = \frac{\text{cov}(X_j, X_k)}{\sigma_{X_j} \sigma_{X_k}} = \frac{j}{\sqrt{jk}} = \sqrt{\frac{j}{k}}.$$

La fonction caractéristique de (X_j, X_k) est donnée, pour tout $(u, v) \in \mathbb{R}^2$, par

$$\varphi_{(X_j, X_k)}(u, v) = \exp \left[-\frac{1}{2} \left\langle C_{(X_j, X_k)} \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \right\rangle \right],$$

soit

$$\varphi_{(X_j, X_k)}(u, v) = \exp \left[-\frac{1}{2} (ju^2 + 2juv + kv^2) \right].$$

Il en résulte que

$$\varphi_{(X_j, \frac{X_k}{\sqrt{k}})}(u, v) = \varphi_{(X_j, X_k)}(u, \frac{v}{\sqrt{k}}) = \exp \left[-\frac{1}{2}(ju^2 + 2\frac{j}{\sqrt{k}}uv + v^2) \right],$$

ce qui implique que

$$\lim_k \varphi_{(X_j, \frac{X_k}{\sqrt{k}})}(u, v) = \exp \left[-\frac{1}{2}(ju^2 + v^2) \right];$$

il résulte du théorème de Lévy que

$$\left(X_j, \frac{X_k}{\sqrt{k}} \right) \xrightarrow[k \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_{\mathbb{R}^2} \left(0, \begin{pmatrix} j & 0 \\ 0 & 1 \end{pmatrix} \right).$$

La loi limite est la loi produit $\mathcal{N}_{\mathbb{R}}(0, j) \otimes \mathcal{N}_{\mathbb{R}}(0, 1)$; on dit, qu'à j fixé, les variables aléatoires X_j et $\frac{X_k}{\sqrt{k}}$ sont asymptotiquement indépendantes.

5. Soit $\underline{1}_n$ le vecteur de \mathbb{R}^n dont toutes les composantes valent 1. On a alors $Z_n = n^{-\frac{3}{2}} \langle \underline{X}_n, \underline{1}_n \rangle$, si bien que la fonction caractéristique de Z_n est donnée par, pour tout réel t ,

$$\varphi_{Z_n}(t) = \varphi_{(\underline{X}_n, \underline{1}_n)}(n^{-\frac{3}{2}}t) = \varphi_{\underline{X}_n}(n^{-\frac{3}{2}}t \underline{1}_n).$$

Il résulte de (14.34) que l'on a

$$\varphi_{Z_n}(t) = \exp \left[-\frac{t^2}{2n^3} \langle A_n^{-1} \underline{1}_n, \underline{1}_n \rangle \right] = \exp \left[-\frac{t^2 S_n}{2n^3} \right],$$

où on note S_n la somme des termes de A_n^{-1} . Pour calculer cette somme, on peut sommer parallèlement à la première diagonale, ce qui donne

$$S_n = (1 + 2 + \dots + n) + 2[(1 + 2 + \dots + (n-1)) + (1 + 2 + \dots + (n-2)) + \dots + 1],$$

soit

$$\begin{aligned} S_n &= \frac{n(n+1)}{2} + 2 \left[\frac{(n-1)n}{2} + \frac{(n-2)(n-1)}{2} + \dots + 1 \right] \\ &= \frac{n(n+1)}{2} + 2 \left[\binom{n}{2} + \binom{n-1}{2} + \dots + \binom{3}{2} + \binom{3}{3} \right]. \end{aligned}$$

Or, d'après les relations du triangle de Pascal, on a

$$\binom{n+1}{3} = \binom{n}{2} + \binom{n-1}{2} + \dots + \binom{3}{2} + \binom{3}{3}.$$

ce qui implique que

$$S_n = \frac{n(n+1)(2n+1)}{6}.$$

et donc que

$$\varphi_{Z_n}(t) = \exp \left[-\frac{t^2}{12} \frac{(n+1)(2n+1)}{n^2} \right].$$

Il en résulte que

$$\lim_n \varphi_{Z_n}(t) = \exp \left[-\frac{t^2}{6} \right],$$

ce qui, en vertu du théorème de Lévy (th. 14.11), montre que

$$\boxed{Z_n \xrightarrow{\mathcal{L}} \mathcal{N}_{\frac{1}{3}} \left(0, \frac{1}{3} \right)}.$$

Chapitre 15

Processus et martingales discrets

On introduit d'abord, sur des exemples, quelques notions relatives aux processus. On s'attache ensuite à l'étude des martingales bornées dans L^2 , et en particulier aux résultats de convergence presque sûre.

15.1. Quelques exemples de processus

Mouvement d'une particule dans un fluide. Notons (X_t, V_t) le couple position-vitesse à l'instant t d'une particule dans un fluide : cette particule est soumise à de nombreuses collisions avec d'autres particules, si bien que la meilleure façon de modéliser le phénomène consiste à considérer ce couple (X_t, V_t) comme une variable aléatoire. La famille $\{(X_t, V_t)\}_{t \in \mathbb{R}^+}$ est un **processus stochastique à temps continu** ; on suppose bien entendu que toutes ces variables aléatoires sont définies sur un même espace probabilisé (Ω, \mathcal{A}, P) . Du point de vue probabiliste, comme d'ailleurs du point de vue de l'étude physique du phénomène, on s'intéresse à certaines grandeurs. Les grandeurs observables, ou mesurables, à l'instant t sont celles qui ne dépendent que de l'**histoire** passé du processus jusqu'à l'instant t – autrement dit celles qui sont « fonctions » des valeurs de X_s et V_s pour $s \leq t$. Un théorème classique de théorie de la mesure affirme qu'une variable aléatoire Y est « fonction » d'une variable aléatoire X , c'est-à-dire s'écrit $Y = f(X)$, où f est une fonction mesurable si et seulement si Y est mesurable par rapport à la tribu $\sigma(X)$ engendrée par X . La généralisation de ce résultat à une famille non dénombrable de variables aléatoires (ici $X_s, V_s, s \leq t$) n'est pas sans poser quelques problèmes (d'ailleurs, qu'est-ce qu'une fonction mesurable de tous les X_s, V_s , pour $s \leq t$?), mais il est raisonnable de considérer que les grandeurs observables à l'instant t sont mesurables par rapport à la tribu $\mathcal{A}_t = \sigma(\{X_s, V_s \mid s \leq t\})$. En ce sens, on dit que l'**histoire** du processus à l'instant t est résumée par la tribu \mathcal{A}_t .

Notons $f(x, v)$ la valeur d'une grandeur liée à cette particule fournie par un appareil de mesure lorsque la particule a pour position-vitesse le couple (x, v) . Si un observateur fait des mesures en une suite croissante d'instantanés $t_1, t_2, \dots, t_n, \dots$ ce qui est connu de l'observateur est le processus d'observation $\{f(X_{t_n}, V_{t_n})\}_{n \in \mathbb{N}^*}$, processus **discret** dont l'**histoire** à l'instant t_n est résumée par la tribu $\mathcal{B}_n = \sigma(f(X_{t_i}, V_{t_i}) \mid i \leq n)$. On peut

envisager l'étude de ce dernier processus avec son histoire propre, c'est-à-dire la **filtration** $(\mathcal{B}_n)_{n \in \mathbb{N}^*}$, ou une histoire plus riche, par exemple celle du processus position-vitesse lui-même, à savoir la **filtration** $(\mathcal{A}_{1,n})_{n \in \mathbb{N}^*}$; des situations intermédiaires sont aussi envisageables.

Marches aléatoires dans \mathbb{R}^n . Soit $X = (X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^n , *indépendantes*, et telle que les variables aléatoires X_n , $n \in \mathbb{N}^*$, soient *de même loi*. On note, pour tout $n \in \mathbb{N}$, $S_n = \sum_{j=0}^n X_j$; la famille de variables aléatoires $S = (S_n)_{n \in \mathbb{N}}$ est un processus discret appelée **marche aléatoire sur \mathbb{R}^n** , issue du point (éventuellement aléatoire) X_0 .

Processus de renouvellement. Une marche aléatoire à valeurs dans \mathbb{R} , où les X_n sont ≥ 0 (on conserve les notations ci-dessus) est appelée **processus de renouvellement**. Voici l'exemple qui est à l'origine de ce nom.

Imaginons une machine qui marche en continu et dont une pièce peut devenir défective ; lorsque c'est le cas, on la remplace instantanément par une pièce identique. La variable aléatoire X_n modélise le **temps de vie** de la n -ième pièce et si on pose $S_0 = 0$, $S_n = \sum_{j=0}^n X_j$ est la **date** de renouvellement de cette pièce.

Un autre exemple classique de processus de renouvellement est celui d'une **file d'attente** de clients, S_n représentant la **date d'arrivée** au guichet du n -ième client.

Processus de saut. Dans la situation décrite ci-dessus d'un processus de renouvellement $S = (S_n)_{n \in \mathbb{N}}$, on peut considérer pour $t \geq 0$ donné le nombre N_t d'indices n tels que $S_n \leq t$ (c'est un nombre qui dépend du hasard, autrement dit une variable aléatoire). Dans les exemples considérés, il s'agit du nombre de remplacements de pièces avant l'instant t ou du nombre de clients arrivés entre 0 et t . La famille $(N_t)_{t \in \mathbb{R}^+}$ est un processus à temps continu dont les réalisations $\omega \mapsto N_t(\omega)$ sont des fonctions croissantes à valeurs dans \mathbb{N} . On parle à ce propos de processus de saut. Dans le cas particulier où les variables aléatoires X_n , $n \in \mathbb{N}^*$, sont de même **loi exponentielle**, le processus $(N_t)_{t \in \mathbb{R}^+}$ est un processus de Poisson (cf. ex. 11.3). Autre exemple : le nombre N_t d'impulsions enregistrées par un compteur Geiger pendant un intervalle de temps $[0, t]$.

L'indice n'a pas toujours une interprétation temporelle. Pour étudier la répartition des molécules d'un gaz par unité de volume à un instant donné, on partitionne l'espace en cubes numérotés. On considère alors le **processus discret** $(X_n)_{n \in \mathbb{N}^*}$, où X_n est la variable aléatoire donnant le nombre de molécules situées dans le n -ième cube.

Il n'est d'ailleurs pas nécessaire de discrétiser l'espace, et on peut définir la notion de processus indexé par \mathbb{R}^3 : dans l'exemple des molécules d'un

gaz, cela revient à considérer pour chaque borélien A de \mathbb{R}^3 la variable aléatoire X_A donnant le nombre de molécules de gaz situées dans A .

15.2. Processus et martingales : définitions

Par souci de simplification, on n'envisagera que des processus à valeurs dans \mathbb{R} ou $\overline{\mathbb{R}}$, les définitions suivantes se généralisant facilement à des processus à valeurs dans \mathbb{R}^n .

Définition 15.1. *Un processus indexé par l'ensemble d'indices partiellement ordonné I est la donnée d'une famille $(X_i)_{i \in I}$ de variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{A}, P) . On parle de **processus discret** si l'ensemble d'indices I est de plus dénombrable infini (I sera en général égal à \mathbb{N} , \mathbb{N}^* ou $\overline{\mathbb{N}}$). Une famille $(\mathcal{A}_i)_{i \in I}$ de sous-tribus de \mathcal{A} , croissante pour l'inclusion, est appelée **filtration**. L'objet $(\Omega, \mathcal{A}, P, (\mathcal{A}_i)_{i \in I})$ est alors appelé **base de processus**. Si $X = (X_i)_{i \in I}$ est un processus discret, sa **filtration naturelle** est la famille de sous tribus \mathcal{A}_i , $i \in I$, où \mathcal{A}_i est la tribu $\sigma(X_j \mid j \leq i)$. Le processus discret $X = (X_i)_{i \in I}$ est **adapté** (sous-entendu, relativement à la filtration $(\mathcal{A}_i)_{i \in I}$) si, pour tout $i \in I$, X_i est \mathcal{A}_i -mesurable.*

Exemple 15.1. Pour une marche aléatoire $S = (S_n)_{n \in \mathbb{I}^*}$, il est facile de voir que sa filtration naturelle est la même que la filtration naturelle du processus $X = (X_n)_{n \in \mathbb{N}^*}$ qui permet de la définir.

Une classe importante de processus est celle des **martingales discrètes** à valeurs **réelles**. L'étude de la convergence presque sûre de suites de variables aléatoires est souvent facilitée par l'introduction de martingales associées, pour lesquelles on dispose de bons théorèmes de convergence.

Définition 15.2. *Soit $(\Omega, \mathcal{A}, P, (\mathcal{A}_i)_{i \in I})$ une base de processus, où I est dénombrable et partiellement ordonné. Soit $X = (X_i)_{i \in I}$ un processus **adapté** tel que, pour tout $i \in I$, X_i soit P -intégrable (resp. positive). Le processus X est une **sous-martingale intégrable** (resp. **sous-martingale positive**) si, pour tout i et j tels que $i \leq j$, on a*

$$E^{\mathcal{A}_i} X_j \geq X_i,$$

une **snrmartingale intégrable** (resp. **surmartingale positive**) si, pour tout i et j tels que $i \leq j$, on a

$$E^{\mathcal{A}_i} X_j \leq X_i.$$

Le processus X est une **martingale intégrable** s'il est à la fois une **sous- et snrmartingale intégrable**, ce qui est équivalent à dire que, pour tout i et j tels que $i \leq j$, on a

$$E^{\mathcal{A}_i} X_j = X_i.$$

Si $I = \mathbb{N}$ ou \mathbb{N}^* on parle de sous-martingale, surmartingale ou martingale **discrète**; si $I = \overline{\mathbb{N}}$ et si $\mathcal{A}_\infty = \bigvee_{n \in \mathbb{N}} \mathcal{A}_n$, tribu engendrée par la réunion des tribus \mathcal{A}_n , $n \in \mathbb{N}$, on parle de **martingale (discrète) fermée**. Avec la même définition de \mathcal{A}_∞ , une martingale discrète intégrable $X = (X_n)_{n \in \mathbb{N}}$ est **fermable** s'il existe une variable aléatoire X_∞ , \mathcal{A}_∞ -mesurable, telle que l'on ait, pour tout $n \in \mathbb{N}$, $X_n = E^{\mathcal{A}_n} X_\infty$; le processus $X = (X_n)_{n \in \overline{\mathbb{N}}}$ est alors une martingale fermée.

Une martingale discrète intégrable $X = (X_n)_{n \in \mathbb{N}}$ est dite **bornée dans L^1** si $\sup_{n \in \mathbb{N}} E|X_n| < +\infty$.

Une martingale discrète $X = (X_n)_{n \in \mathbb{N}}$ est dite **de carré intégrable** ou dans L^2 (resp. **bornée dans L^2**) si, pour tout $n \in \mathbb{N}$, X_n est de carré intégrable (resp. si $\sup_{n \in \mathbb{N}} EX_n^2 < +\infty$).

Remarque. Une sous-martingale (resp. surmartingale) croît (resp. décroît) en espérance conditionnelle; une martingale est constante en espérance conditionnelle.

Voici des exemples simples de tels processus.

Exemple 15.2. 1. Un processus de renouvellement est une sous-martingale par rapport à sa filtration naturelle.

2. Considérons la marche aléatoire dans \mathbb{Z} définie, avec les notations ci-dessus, par $S_0 = a$ et $S_n = \sum_{j=1}^n X_j$, $n \in \mathbb{N}^*$, où les X_n , $n \in \mathbb{N}^*$, forment une suite de variables aléatoires *indépendantes* de même loi $p\delta_1 + (1-p)\delta_{-1}$ (avec $0 < p < 1$). Cette marche peut modéliser par exemple la fortune d'un joueur qui joue à **pile on face** et qui, à chaque jet, gagne ou perd une unité respectivement avec probabilité p et $1-p$. Si $(\mathcal{A}_n)_{i \in \mathbb{N}}$ est la filtration du processus S , on remarque que, si $n \in \mathbb{N}^*$, $\mathcal{A}_n = \sigma(X_j \mid 1 \leq j \leq n)$. On a alors

$$E^{\mathcal{A}_n}(S_{n+1}) = S_n + E^{\mathcal{A}_n}(X_{n+1});$$

puisque les X_n sont indépendantes, on a

$$E^{\mathcal{A}_n}(X_{n+1}) = E(X_{n+1}) = p - (1-p) = 2p - 1.$$

Il en résulte que

$$\left\{ \begin{array}{ll} S \text{ est une sous-martingale} & \text{si } p > \frac{1}{2}, \\ S \text{ est une surmartingale} & \text{si } p < \frac{1}{2}, \\ S \text{ est une martingale} & \text{si } p = \frac{1}{2}. \end{array} \right.$$

Remarque.

1. Pour que le processus X soit une sous-martingale il faut et il suffit que le processus $-X$ soit une surmartingale.

2. Si X et Y sont des sous-martingales, pour tout réels positifs a et b , le processus $aX + bY = (aX_i + bY_i)_{i \in I}$ est encore une sous-martingale.

3. Si X et Y sont des sous-martingales, le processus $X \vee Y = (X_i \vee Y_i)_{i \in I}$ est une sous-martingale; de même, si X et Y sont des surmartingales, le processus $X \wedge Y = (X_i \wedge Y_i)_{i \in I}$ est une surmartingale¹. La première assertion résulte de ce que, si $i \leq j$, on a $E^{\mathcal{A}_i}(X_j \vee Y_j) \geq E^{\mathcal{A}_i} X_j \geq X_i$ et $E^{\mathcal{A}_i}(X_j \vee Y_j) \geq E^{\mathcal{A}_i} Y_j \geq Y_i$, la seconde, de ce que $E^{\mathcal{A}_i}(X_j \wedge Y_j) \leq E^{\mathcal{A}_i} X_j \leq X_i$ et $E^{\mathcal{A}_i}(X_j \wedge Y_j) \leq E^{\mathcal{A}_i} Y_j \leq Y_i$.

4. Si X est une sous-martingale (resp. une surmartingale) intégrable et si $i \leq j$, on a $EX_j \geq EX_i$ (resp. $EX_j \leq EX_i$); en particulier, si X est une martingale intégrable on a $EX_j = EX_i$.

5. Pour que le processus X soit une sous-martingale (resp. surmartingale ou martingale) **discrète** il faut et il suffit que l'on ait, pour tout $n \in \mathbb{N}$,

$$E^{\mathcal{A}_n} X_{n+1} \geq X_n, \text{ (resp. } E^{\mathcal{A}_n} X_{n+1} \leq X_n \text{ ou } E^{\mathcal{A}_n} X_{n+1} = X_n \text{)}.$$

6. Soit $X = (X_n)_{n \in \mathbb{N}}$ une martingale dans L^2 ; le processus $X^2 = (X_n^2)_{n \in \mathbb{N}}$ est une sous-martingale et, en conséquence, la suite $(EX_n^2)_{n \in \mathbb{N}}$ est croissante. En effet, puisque X_n est \mathcal{A}_n -mesurable, on a, pour tout n ,

$$E^{\mathcal{A}_n}(X_{n+1} - X_n)^2 = E^{\mathcal{A}_n} X_{n+1}^2 + E^{\mathcal{A}_n} X_n^2 - 2X_n E^{\mathcal{A}_n} X_{n+1} = E^{\mathcal{A}_n} X_{n+1}^2 - E^{\mathcal{A}_n} X_n^2,$$

ce qui démontre que $E^{\mathcal{A}_n} X_{n+1}^2 \geq E^{\mathcal{A}_n} X_n^2$. En intégrant, il en résulte que $EX_{n+1}^2 \geq EX_n^2$.

Dans la suite, sauf mention du contraire, les processus introduits sont définis sur la même base de processus $(\Omega, \mathcal{A}, P, (\mathcal{A}_n)_{n \in \mathbb{N}})$. Lorsqu'il n'y aura pas d'ambiguïté, on supprimera l'adjectif « discret ».

Exemple 15.3. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires intégrables définies sur l'espace probabilisé (Ω, \mathcal{A}, P) ; notons, pour tout $n \in \mathbb{N}$, \mathcal{B}_n la tribu $\sigma(X_i \mid i \leq n)$ et Y_n les variables aléatoires définies par $Y_0 = 0$ et $Y_n = \sum_{i=1}^n (X_i - E^{\mathcal{B}_{i-1}} X_i)$ si $n \geq 1$. Le processus $(Y_n)_{n \in \mathbb{N}}$ est une martingale sur la base de processus $(\Omega, \mathcal{A}, P, (\mathcal{B}_n)_{n \in \mathbb{N}})$; on dit aussi que $(Y_n)_{n \in \mathbb{N}}$ est une martingale relativement à la filtration $(\mathcal{B}_n)_{n \in \mathbb{N}}$, ou plus brièvement, quand il n'y a pas d'ambiguïté, une martingale.

Un cas particulier important est celui où les variables aléatoires X_n sont *indépendantes* et, dans ce cas, on a $Y_n = \sum_{i=1}^n (X_i - EX_i)$ si $n \geq 1$.

Exemple 15.4. Soit U une variable aléatoire intégrable et soit $(\mathcal{A}_n)_{n \in \mathbb{N}}$ une filtration sur l'espace probabilisé (Ω, \mathcal{A}, P) ; soit $X_n = E^{\mathcal{A}_n} U$. Le processus $(X_n)_{n \in \mathbb{N}}$ est une martingale bornée dans L^1 .

1. Rappel de notations : pour tous réels a et b , $a \vee b = \max(a, b)$ et $a \wedge b = \min(a, b)$, ce qui se lit respectivement « a sup b » et « a inf b ».

Exemple 15.5. Soient $X = (X_i)_{i \in I}$ une sous-martingale positive et f une fonction de \mathbb{R}^+ dans lui-même, convexe croissante et telle que $f(X_i)$ soit intégrable pour tout i ; il résulte de l'inégalité de Jensen que le processus $f(X) = (f(X_i))_{i \in I}$ est une sous-martingale positive: en particulier il en est ainsi pour les processus $X^p = (X_i^p)_{i \in I}$ avec $p \geq 1$ et $X^+ = (X_i^+)_{i \in I}$.

Exemple 15.6. Soit $X = (X_i)_{i \in I}$ une martingale; le processus $(|X_i|)_{i \in I}$ est une sous-martingale puisque, si $i \leq j$, on a $|X_i| = |E^{\mathcal{A}_i} X_j| \leq E^{\mathcal{A}_i} |X_j|$. Plus généralement, si f est une fonction convexe continue telle que $f(X_i)$ soit intégrable pour tout i , il résulte de l'inégalité de Jensen que le processus $f(X) = (f(X_i))_{i \in I}$ est une sous-martingale. En particulier, si X est une martingale L^2 , le processus $X^2 = (X_i^2)_{i \in I}$ est une sous-martingale.

15.3. Temps d'arrêt

La notion de temps introduite pour la modélisation d'un processus aléatoire est en fait relative à l'horloge de l'observateur et le phénomène aléatoire étudié n'a aucune raison a priori d'évoluer simplement suivant cette horloge. On est alors amené à introduire des temps aléatoires, appelés temps d'arrêt, qui tiennent lieu d'horloge interne du processus.

Soit une base de processus $(\Omega, \mathcal{A}, P, (\mathcal{A}_n)_{n \in \mathbb{N}})$. On note $\mathcal{A}_\infty = \bigvee_{n \in \mathbb{N}} \mathcal{A}_n$.

Définition 15.3. Une application T de Ω dans $\overline{\mathbb{N}}$ est un temps d'arrêt si, pour tout $n \in \mathbb{N}$, on a $(T = n) \in \mathcal{A}_n$.

Remarque. Si $(\mathcal{A}_n)_{n \in \mathbb{N}}$ est la filtration naturelle d'un processus $X = (X_n)_{n \in \mathbb{N}}$ à valeurs dans un espace mesurable (E, \mathcal{E}) , une application T de Ω dans $\overline{\mathbb{N}}$ est un temps d'arrêt si, pour tout $n \in \mathbb{N}$, il existe une application mesurable f_n de $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ à valeurs 0 ou 1 telle que l'on ait $\mathbf{1}_{(T=n)} = f_n(X_0, X_1, \dots, X_n)$.

Si T est un temps d'arrêt, on a $(T = +\infty) \in \mathcal{A}_\infty$; en effet, $(T < +\infty) = \bigcup_{n \in \mathbb{N}} (T = n)$ et, pour tout $n \in \mathbb{N}$, on a $(T = n) \in \mathcal{A}_\infty$, puisque $\mathcal{A}_n \subset \mathcal{A}_\infty$.

Exemple 15.7. Toute application constante T de Ω dans $\overline{\mathbb{N}}$ est un temps d'arrêt.

Exemple 15.8. Le temps d'entrée T_A d'un processus adapté X dans un borélien A est un temps d'arrêt. Il est défini par

$$T_A = \inf(n \in \mathbb{N} \mid X_n \in A), \text{ avec la convention } \inf \emptyset = +\infty.$$

En effet, il résulte de l'adaptation de X et de la croissance de la suite des sous-tribus de la filtration que l'on a $(T_A = 0) = (X_0 \in A) \in \mathcal{A}_0$ et, pour tout $n \in \mathbb{N}^*$,

$$(\tau_A = n) = \left[\bigcap_{k=0}^{n-1} (X_k \notin A) \right] \cap (X_n \in A) \in \mathcal{A}_n.$$

Exemple 15.9. Le temps de dernier passage τ_A d'un processus adapté X dans un borélien A n'est pas un temps d'arrêt. Il est défini par

$$\tau_A = \sup(n \in \mathbb{N}^* \mid X_n \in A), \text{ avec la convention } \sup \emptyset = 0.$$

En effet, on a $(\tau_A = 0) = \bigcap_{n \in \mathbb{N}^*} (X_n \notin A) \in \mathcal{A}_\infty$, mais, en général, $(\tau_A = 0) \notin \mathcal{A}_0$, et de plus on a, pour tout $n \in \mathbb{N}^*$,

$$(\tau_A = n) = (X_n \in A) \cap \left[\bigcap_{k \geq n+1} (X_k \notin A) \right] \notin \mathcal{A}_n.$$

Notation. On note \mathcal{T} (resp. \mathcal{T}_b) l'ensemble des temps d'arrêt (resp. temps d'arrêt bornés), relativement à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$.

Définition 15.4. Soit T un temps d'arrêt; la famille d'événements \mathcal{A}_T définie par

$$\mathcal{A}_T = \{A \in \mathcal{A}_\infty \mid \forall n \in \mathbb{N} \quad A \cap (T = n) \in \mathcal{A}_n\}$$

est une tribu; elle est appelée **tribu des événements antérieurs à T** .

Remarque. T est \mathcal{A}_T -mesurable.

Lemme 15.5. (a) Soient T un temps d'arrêt et $A \in \mathcal{A}_\infty$; $A \in \mathcal{A}_1$ si et seulement si² pour tout $n \in \mathbb{N}$, $A \cap (T \leq n) \in \mathcal{A}_n$.

(b) Une application T de Ω dans $\overline{\mathbb{N}}$ est un temps d'arrêt si et seulement, pour tout $n \in \mathbb{N}$, $(T \leq n) \in \mathcal{A}_n$.

(c) Si T_1, T_2, \dots, T_k sont des temps d'arrêt, il en est de même des applications $\inf_{1 \leq i \leq k} T_i$ et $\sup_{1 \leq i \leq k} T_i$. En particulier, si T est un temps d'arrêt, pour tout entier k , $T \wedge k$ est un temps d'arrêt borné.

Démonstration. (a) Supposons que T soit un temps d'arrêt et soit $A \in \mathcal{A}_T$; on a, si $k \leq n$, $A \cap (T = k) \in \mathcal{A}_k \subset \mathcal{A}_n$ et donc

$$A \cap (T \leq n) = A \cap \left[\bigcup_{k \leq n} (T = k) \right] \in \mathcal{A}_n.$$

Inversement supposons que, pour tout $n \in \mathbb{N}$, $A \cap (T \leq n) \in \mathcal{A}_n$; on a, si $n \geq 1$, $A \cap (T \leq n-1) \in \mathcal{A}_{n-1} \subset \mathcal{A}_n$ et, par conséquent,

$$A \cap (T = n) = [A \cap (T \leq n)] \setminus [A \cap (T \leq n-1)] \in \mathcal{A}_n;$$

de plus $A \cap (T = 0) = A \cap (T \leq 0) \in \mathcal{A}_0$.

2. Cette caractérisation justifie le nom de **tribu des événements antérieurs à T** .

(b) Cela résulte de la propriété précédente, en remarquant que T est un temps d'arrêt si et seulement si $\Omega \in \mathcal{A}_T$.

(c) Les T_i sont des temps d'arrêt; d'après la caractérisation des temps d'arrêt précédemment démontrée, on a, pour tout $n \in \mathbb{N}$, $(\inf_{1 \leq i \leq k} T_i \leq n) = \bigcup_{1 \leq i \leq k} (T_i \leq n) \in \mathcal{A}_n$ et $(\sup_{1 \leq i \leq k} T_i \leq n) = \bigcap_{1 \leq i \leq k} (T_i \leq n) \in \mathcal{A}_n$, ce qui démontre le résultat en vertu de cette même caractérisation. \square

Le lemme suivant caractérise les fonctions \mathcal{A}_T -mesurables et donne l'expression de l'espérance conditionnelle d'une variable aléatoire par rapport à la tribu \mathcal{A}_T .

Convention. Une application X définie sur une partie Ω' de Ω sera dite \mathcal{A} -mesurable si elle est mesurable relativement à l'espace trace $(\Omega', \Omega' \cap \mathcal{A})$.

Lemme 15.6. (a) Soient T un temps d'arrêt et X une application de Ω dans $\overline{\mathbb{R}}$, \mathcal{A}_∞ -mesurable. L'application X est \mathcal{A}_T -mesurable si et seulement si, pour tout $n \in \mathbb{N}$, sa restriction $X_{(T=n)}$ à l'ensemble $(T = n)$ est \mathcal{A}_n -mesurable.

(b) Soit X une variable aléatoire numérique définie sur (Ω, \mathcal{A}, P) , positive ou intégrable. On a, pour tout $n \in \overline{\mathbb{N}}$,

$$E^{\mathcal{A}_T} X = E^{\mathcal{A}_n} X \quad \text{sur } (T = n),$$

autrement dit,

$$E^{\mathcal{A}_T} X = \sum_{n \in \overline{\mathbb{N}}} \mathbf{1}_{(T=n)} E^{\mathcal{A}_n} X.$$

Démonstration. (a) Supposons que $X = \mathbf{1}_B$, où $B \in \mathcal{A}_\infty$; pour tout $n \in \mathbb{N}$, sa restriction $X_{(T=n)}$ à l'ensemble $(T = n)$ est $\mathbf{1}_{B \cap (T=n)}$. L'équivalence annoncée résulte alors de la définition de la tribu \mathcal{A}_T ; elle est encore alors valable, par linéarité, lorsque X est étagée, puis lorsque X est positive (X est alors limite simple d'une suite croissante de fonctions étagées). Le cas général s'obtient alors en décomposant X en parties positive et négative.

(b) Soit X une variable aléatoire numérique positive et soit, pour tout $n \in \overline{\mathbb{N}}$, Y_n un représentant de $E^{\mathcal{A}_n} X$; Y_n est \mathcal{A}_n -mesurable et positive. La variable aléatoire positive $Y = \sum_{n \in \overline{\mathbb{N}}} \mathbf{1}_{(T=n)} Y_n$ est alors, d'après la propriété précédente, \mathcal{A}_T -mesurable. De plus, puisque X est positive, on a, pour tout $A \in \mathcal{A}_T$,

$$\int_A X \, dP = \sum_{n \in \overline{\mathbb{N}}} \int_{A \cap (T=n)} X \, dP,$$

soit, puisque $A \cap (T = n)$ appartient à \mathcal{A}_n , et que les intégrales sont positives,

$$\int_A X \, dP = \sum_{n \in \overline{\mathbb{N}}} \int_{A \cap (T=n)} E^{\mathcal{A}_n} X \, dP = \int_A \left[\sum_{n \in \overline{\mathbb{N}}} \mathbf{1}_{(T=n)} Y_n \right] dP = \int_A Y \, dP.$$

ce qui démontre le résultat lorsque X est positive. Le cas où X est de signe quelconque et intégrable se démontre alors en décomposant X en parties positive et négative. \square

On obtient alors immédiatement le corollaire suivant :

Corollaire 15.7. *Soient X un processus adapté et T un temps d'arrêt. L'application X_T définie sur l'ensemble $(T < +\infty)$ par $X_T = X_n$ sur $(T = n)$, pour tout $n \in \mathbb{N}$, est \mathcal{A}_T -mesurable.*

De même, si $(X_n)_{n \in \mathbb{N}}$ est un processus adapté et T un temps d'arrêt l'application X_T définie par $X_T = X_n$ sur $(T = n)$, pour tout $n \in \mathbb{N}$, est \mathcal{A}_T -mesurable.

Proposition 15.8. *Soient S et T des temps d'arrêt.*

(a) *Les événements $(S < T)$, $(S = T)$ et $(S \leq T)$ appartiennent à \mathcal{A}_S et à \mathcal{A}_T .*

(b) *Si $B \in \mathcal{A}_S$ alors $B \cap (S \leq T) \in \mathcal{A}_T$.*

(c) *En conséquence, si les temps d'arrêt S et T sont tels que $S \leq T$, on a $\mathcal{A}_S \subset \mathcal{A}_T$. Il en résulte que les familles de sous-tribus $(\mathcal{A}_T)_{T \in \mathcal{T}_b}$ et $(\mathcal{A}_T)_{T \in \mathcal{T}}$ sont des filtrations.*

Démonstration. (a) On a, pour tout $n \in \mathbb{N}$,

$$(S < T) \cap (S = n) = (n < T) \cap (S = n) \in \mathcal{A}_n,$$

puisque $(n < T) = (T \leq n)^c$ et que S et T sont des temps d'arrêt. Il en résulte que $(S < T)$ appartient à \mathcal{A}_S . Par ailleurs, on a, pour tout $n \in \mathbb{N}^*$,

$$(S < T) \cap (T = n) = (S < n - 1) \cap (T = n) \in \mathcal{A}_n,$$

puisque $(S < n - 1) \in \mathcal{A}_{n-1} \subset \mathcal{A}_n$ et que S et T sont des temps d'arrêt; enfin, on a

$$(S < T) \cap (T = 0) = \emptyset \in \mathcal{A}_0.$$

Il en résulte que $(S < T)$ appartient à \mathcal{A}_T . Ainsi $(S < T) \in \mathcal{A}_S \cap \mathcal{A}_T$.

On a, pour tout $n \in \mathbb{N}$,

$$(S = T) \cap (S = n) = (S = n) \cap (T = n) \in \mathcal{A}_n,$$

ce qui démontre que $(S = T) \in \mathcal{A}_S$; on obtient de même que $(S = T) \in \mathcal{A}_T$.

Tenant compte de ces résultats, il vient

$$(S \leq T) = (S < T) \cup (S = T) \in \mathcal{A}_S \cap \mathcal{A}_T.$$

(b) Si $B \in \mathcal{A}_S$, pour tout $n \in \mathbb{N}$, on a $B \cap (S \leq n) \in \mathcal{A}_n$ et donc

$$[B \cap (S \leq T)] \cap (T = n) = [B \cap (S \leq n)] \cap (T = n) \in \mathcal{A}_n,$$

ce qui démontre que $B \cap (S \leq T) \in \mathcal{A}_T$.

(c) Si les temps d'arrêt S et T sont tels que $S \leq T$, on a $(S \leq T) = \Omega$, et l'assertion précédente assure que $\mathcal{A}_S \subset \mathcal{A}_T$. \square

Remarque. Si $B \in \mathcal{A}_S$, on a aussi $B \cap (S = T) \in \mathcal{A}_T$ et $B \cap (S < T) \in \mathcal{A}_T$ car, d'après la proposition précédente, on a

$$B \cap (S = T) = [B \cap (S \leq T)] \cap (S = T) \in \mathcal{A}_T,$$

et

$$B \cap (S < T) = [B \cap (S \leq T)] \cap (S < T) \in \mathcal{A}_T.$$

15.4. Premier théorème d'arrêt

On démontre un théorème de **caractérisation des martingales** en termes de temps d'arrêt **bornés** : on en déduit le premier théorème d'arrêt de Doob, d'utilisation fréquente.

Théorème 15.9. *Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus adapté. Il y a équivalence entre les propriétés suivantes :*

- (i) X est une martingale intégrable,
- (ii) pour tout $T \in \mathcal{T}_b$, $X_T \in L^1(\Omega, \mathcal{A}_T, P)$ et $EX_T = EX_0$,
- (iii) le processus $(X_T)_{T \in \mathcal{T}_b}$ est une martingale par rapport à la filtration $(\mathcal{A}_T)_{T \in \mathcal{T}_b}$.

Démonstration. (i) \Rightarrow (ii). Supposons que X est une martingale intégrable. Si $T \in \mathcal{T}_b$ est borné par k , on a

$$X_T = \sum_{j=0}^k \mathbf{1}_{(T=j)} X_j,$$

si bien que $X_T \in L^1(\Omega, \mathcal{A}_T, P)$. Soit $A \in \mathcal{A}_T$; on a

$$A = \bigcup_{j=0}^k [A \cap (T = j)] ;$$

et donc

$$\int_A X_T dP = \sum_{j=0}^k \int_{A \cap (T=j)} X_j dP.$$

Puisque X est une martingale et que, pour tout $j \in \mathbb{N}$, $A \cap (T = j) \in \mathcal{A}_j$, il vient alors

$$\int_A X_T dP = \sum_{j=0}^k \int_{A \cap (T=j)} X_j dP = \int_A X_0 dP;$$

X_T étant \mathcal{A}_T -mesurable, on a démontré que $X_T = E^{\mathcal{A}_T} X_k$. En prenant les espérances et en tenant compte de ce que X est une martingale, il vient :

$$EX_T = EX_k = EX_0.$$

(ii) \Rightarrow (iii). Supposons la propriété (ii) vérifiée. Soient S et T deux temps d'arrêt bornés par k tels que $S \leq T \leq k$; on a alors $\mathcal{A}_S \subset \mathcal{A}_T \subset \mathcal{A}_k$. Soit $A \in \mathcal{A}_S$. L'application R définie par $R = S \mathbf{1}_A + k \mathbf{1}_{A^c}$ est évidemment bornée par k et est un temps d'arrêt. En effet, on a

$$(R = n) = \begin{cases} (S = n) \cap A \in \mathcal{A}_n & \text{si } n < k, \\ [(S = k) \cap A] \cup A^c & \text{si } n = k, \\ \emptyset & \text{si } n > k; \end{cases}$$

or, si $n = k$, puisque $(S = k) \cap A \in \mathcal{A}_k$ et que $\mathcal{A}_S \subset \mathcal{A}_k$, on a $(R = n) \in \mathcal{A}_k$, si bien qu'en définitive on a, pour tout $n \in \mathbb{N}$, $(R = n) \in \mathcal{A}_n$.

Appliquant l'hypothèse aux temps d'arrêt bornés R et k , il vient $EX_R = EX_0 = EX_k$, ce qui donne l'égalité

$$E(\mathbf{1}_A X_S + \mathbf{1}_{A^c} X_k) = EX_k,$$

soit encore,

$$E(\mathbf{1}_A X_S) = E(\mathbf{1}_A X_k):$$

X_S étant \mathcal{A}_S -mesurable, il en résulte que $X_S = E^{\mathcal{A}_S} X_k$. On a évidemment de même $X_T = E^{\mathcal{A}_T} X_k$. Puisque l'on a $\mathcal{A}_S \subset \mathcal{A}_T$, on a alors la succession d'égalités

$$E^{\mathcal{A}_S} X_T = E^{\mathcal{A}_S} \left[E^{\mathcal{A}_T} X_k \right] = E^{\mathcal{A}_S} X_k = X_S,$$

ce qui démontre que la propriété (iii) est vérifiée.

(iii) \Rightarrow (i). Il suffit de prendre des temps d'arrêt constants. \square

Corollaire 15.10. Soit $X = (X_n)_{n \in \mathbb{N}}$ une martingale; pour tout temps d'arrêt T , le processus $X^T = (X_{T \wedge n})_{n \in \mathbb{N}}$ est une martingale. Elle est appelée **martingale arrêtée au temps T** .

Démonstration. Pour tout temps d'arrêt borné S , on a $X_S^T = X_{T \wedge S}$ et, $T \wedge S$ étant un temps d'arrêt borné, il résulte du théorème 15.9 appliqué à la martingale X que l'on a

$$E(X_S^T) = E(X_{T \wedge S}) = EX_0 = E(X_0^T),$$

ce qui implique, toujours d'après ce théorème et du fait de l'arbitraire de S , que X^T est une martingale. \square

Une explicitation de l'implication (i) \Rightarrow (iii) du théorème 15.9 donne alors le premier théorème d'arrêt de Doob.

Théorème 15.11 (Premier théorème d'arrêt de Doob). Soit $X = (X_n)_{n \in \mathbb{N}}$ une martingale ; pour tous temps d'arrêt bornés S et T tels que $S \leq T$, on a

$$E^{A_S} X_T = X_S.$$

15.5. Lemme maximal et martingales dans L^2

Il s'agit de démontrer un théorème de convergence presque sûre pour les martingales bornées dans L^2 . Auparavant, on donne le **lemme maximal**, ou **inégalité maximale de Doob**, pour les sous-martingales. C'est en fait une généralisation de l'inégalité de Kolmogorov pour les sommes de variables aléatoires indépendantes.

Lemme 15.12 (Lemme maximal ou inégalité maximale de Doob). (a) Soit X une sous-martingale positive ou intégrable. Pour tout entier N et tout $\varepsilon > 0$, on a l'inégalité

$$P\left(\sup_{0 \leq n \leq N} X_n > \varepsilon\right) \leq \frac{1}{\varepsilon} \left(\int_{\left(\sup_{0 \leq n \leq N} X_n > \varepsilon\right)} X_N dP \right), \quad (15.1)$$

et a fortiori

$$P\left(\sup_{0 \leq n \leq N} X_n > \varepsilon\right) \leq \frac{1}{\varepsilon} E|X_N|. \quad (15.2)$$

Pour tout $\varepsilon > 0$, on a alors l'inégalité

$$P\left(\sup_{n \in \mathbb{N}} X_n > \varepsilon\right) \leq \frac{1}{\varepsilon} \sup_{n \in \mathbb{N}} E|X_n|. \quad (15.3)$$

(b) En particulier, si X est une martingale intégrable bornée dans L^1 (c'est-à-dire telle que $\sup_{n \in \mathbb{N}} E|X_n| < +\infty$), la variable aléatoire $X^* = \sup_{n \in \mathbb{N}} |X_n|$ est finie P-p.s.

Démonstration. (a) Il s'agit de majorer la probabilité de l'ensemble $E = (\sup_{0 \leq n \leq N} X_n > \varepsilon)$; s'il est vide, l'inégalité est triviale et on se place dans le cas où il ne l'est pas. Faisons apparaître l'indice k pour lequel X_k dépasse pour la première fois le seuil ε ; on introduit pour cela les ensembles

$$E_0 = (X_0 > \varepsilon) \quad \text{et, si } 1 \leq k \leq N, \quad E_k = (X_k > \varepsilon) \cap \left[\bigcap_{i=0}^{k-1} (X_i \leq \varepsilon) \right].$$

Ces ensembles forment une partition de E ; par conséquent on a

$$\int_E X_N dP = \sum_{k=0}^N \int_{E_k} X_N dP.$$

Puisque, pour tout k , on a $E_k \in \mathcal{A}_k$ et que X est une sous-martingale, il en résulte que

$$\int_{E_k} X_N dP \geq \sum_{k=0}^N \int_{E_k} X_k dP;$$

par définition de E_k , on a alors

$$\int_E X_N dP \geq \varepsilon \sum_{k=0}^N P(E_k) = \varepsilon P(E),$$

ce qui démontre l'inégalité (15.1); l'inégalité (15.2) en résulte immédiatement.

Enfin, pour tout $\varepsilon > 0$, la suite d'ensembles $(\sup_{0 \leq n \leq N} X_n > \varepsilon)$ étant croissante en N et de réunion $(\sup_{n \in \mathbb{N}} X_n > \varepsilon)$, on a

$$P\left(\sup_{n \in \mathbb{N}} X_n > \varepsilon\right) = \lim_N P\left(\sup_{0 \leq n \leq N} X_n > \varepsilon\right) \leq \frac{1}{\varepsilon} \sup_{n \in \mathbb{N}} E|X_n|.$$

(b) Le processus $|X|$ étant alors une sous-martingale, il résulte de l'inégalité (15.3) que l'on a, pour tout $k \in \mathbb{N}^*$,

$$P(\sup_{n \in \mathbb{N}^*} |X_n| > k) \leq \frac{1}{k} \sup_{n \in \mathbb{N}^*} E|X_n|,$$

ce qui, en passant à la limite en k , donne

$$P\left(\sup_{n \in \mathbb{N}} |X_n| = +\infty\right) = \lim_k P\left(\sup_{n \in \mathbb{N}} |X_n| > k\right) = 0. \quad \square$$

On obtient alors en corollaire l'inégalité de Doob pour les martingales bornées dans L^2 .

Théorème 15.13 (Inégalité de Doob). *Soit X une martingale bornée dans L^2 . La variable aléatoire $X^* = \sup_{n \in \mathbb{N}} |X_n|$ est dans L^2 et on a l'inégalité de Doob*

$$\|X^*\|_{L^2} \leq 2 \sup_{n \in \mathbb{H}} \|X_n\|_{L^2}. \quad (15.4)$$

Démonstration. Soit $M_n = \sup_{0 \leq k \leq n} |X_k|$; il résulte de l'inégalité $M_n \leq \sum_{k=0}^n |X_k|$ que $M_n \in L^2$. Le processus $|X|$ étant une sous-martingale positive et intégrable, il résulte du lemme maximal que, pour tout $a > 0$, on a

$$a E[\mathbf{1}_{(M_n > a)}] \leq E[|X_n| \mathbf{1}_{(M_n > a)}].$$

En intégrant par rapport à la mesure de Lebesgue λ sur \mathbb{R}^+ , on obtient l'inégalité

$$\int_{\mathbb{R}^+} a E[\mathbf{1}_{(M_n > a)}] d\lambda(a) \leq \int_{\mathbb{R}^+} E[|X_n| \mathbf{1}_{(M_n > a)}] d\lambda(a),$$

soit, par application du théorème de Fubini et intégration,

$$E \int_{[0, M_n[} a \, d\lambda(a) = \frac{1}{2} EM_n^2 \leq E [|X_n| M_n] .$$

L'inégalité de Schwarz appliquée au second membre donne

$$\frac{1}{2} EM_n^2 \leq [EX_n^2]^{\frac{1}{2}} [EM_n^2]^{\frac{1}{2}} ,$$

ce qui implique l'inégalité

$$[EM_n^2]^{\frac{1}{2}} \leq 2 [EX_n^2]^{\frac{1}{2}}$$

et a fortiori l'inégalité

$$[EM_n^2]^{\frac{1}{2}} \leq 2 \sup_{n \in \mathbb{N}} \|X_n\|_{L^2} .$$

La suite $(M_n)_{n \in \mathbb{N}}$ convergeant en croissant vers X^* , un passage à la limite et la propriété de Beppo Levi donnent l'inégalité (15.4). \square

Remarque. Sous les hypothèses du théorème 15.13, on a donc en fait la double inégalité

$$\sup_{n \in \mathbb{N}} \|X_n\|_{L^2} \leq \|X^*\|_{L^2} \leq 2 \sup_{n \in \mathbb{N}} \|X_n\|_{L^2} .$$

En corollaire, on obtient un **théorème de convergence pour les martingales bornées dans L^2** .

Théorème 15.14 (Théorème de convergence L^2). *Soit X une martingale bornée dans L^2 . La suite $(X_n)_{n \in \mathbb{N}}$ converge P-p.s. et dans L^2 vers une variable aléatoire X_∞ . On a, pour tout $n \in \mathbb{N}$, $X_n = E^{\mathcal{A}_n} X_\infty$.*

De plus, si la filtration est complète, au sens où la tribu \mathcal{A}_0 contient tous les ensembles \mathcal{A} -négligeables, X_∞ est \mathcal{A}_∞ -mesurable et la martingale X est fermable.

Démonstration. On démontre d'abord la convergence P-p.s.; on note classiquement $\{X \rightarrow\}$ l'ensemble des ω pour lesquels la suite $(X_n(\omega))_{n \in \mathbb{N}}$ converge dans \mathbb{R} . Il résulte du critère de Cauchy que l'on a

$$\{X \rightarrow\} = \bigcap_{\varepsilon \in \mathbb{Q}^+} \bigcup_{N \in \mathbb{N}^*} \bigcap_{m, n \geq N} \{|X_n - X_m| \leq \varepsilon\} ,$$

et donc

$$\{X \rightarrow\}^c = \bigcup_{\varepsilon \in \mathbb{Q}^+} \bigcap_{N \in \mathbb{N}^*} \bigcup_{m, n \geq N} \{|X_n - X_m| > \varepsilon\} ;$$

or, pour $N \in \mathbb{N}^*$, on a

$$\bigcup_{m, n \geq N} \{|X_m - X_n| > \varepsilon\} \subset \left\{ \sup_{m, n \geq N} |X_m - X_n| > \varepsilon \right\} \subset \left\{ \sup_{n \geq N} |X_n - X_N| > \frac{\varepsilon}{3} \right\},$$

la dernière inclusion étant obtenue en prenant la contraposée de la suite d'implications

$$\begin{aligned} \sup_{n \geq N} |X_n - X_N| \leq \frac{\varepsilon}{3} &\implies \forall m, n \geq N \quad |X_m - X_n| \leq \frac{\varepsilon}{3} \text{ et } |X_n - X_N| \leq \frac{\varepsilon}{3} \\ &\implies \forall m, n \geq N \quad |X_m - X_n| \leq \frac{2\varepsilon}{3} < \varepsilon \implies \sup_{m, n \geq N} |X_m - X_n| \leq \varepsilon. \end{aligned}$$

Par l'inégalité de Markov, on a

$$P\left(\sup_{n \geq N} |X_n - X_N| > \frac{\varepsilon}{3}\right) \leq \frac{9}{\varepsilon^2} \left\| \sup_{n \geq N} |X_n - X_N| \right\|_{L^2}^2;$$

l'inégalité de Doob appliquée à la martingale $(X_n - X_N)_{n \geq N}$ ou, ce qui est équivalent, à la martingale $(Y_n)_{n \in \mathbb{N}}$ définie par $Y_n = 0$ si $0 < n \leq N - 1$ et $Y_n = X_n - X_N$ si $n \geq N$, conduit alors à l'inégalité

$$P\left(\sup_{n \geq N} |X_n - X_N| > \frac{\varepsilon}{3}\right) \leq \frac{36}{\varepsilon^2} \sup_{n \geq N} \|X_n - X_N\|_{L^2}^2. \quad (15.5)$$

Mais, puisque X est une martingale dans L^2 , on a

$$\begin{aligned} \|X_n - X_N\|_{L^2}^2 &= E[X_n - X_N]^2 = EX_n^2 + EX_N^2 - 2E[X_n X_N] \\ &= EX_n^2 + EX_N^2 - 2E[X_N(E^{A_n} X_n)] = EX_n^2 - EX_N^2; \end{aligned}$$

ainsi, la suite $(EX_n^2)_{n \in \mathbb{N}}$ est croissante et donc convergente, puisque X est une martingale bornée dans L^2 ; on a donc

$$\sup_{n \geq N} \|X_n - X_N\|_{L^2}^2 = \sup_{n \geq N} EX_n^2 - EX_N^2 = \lim_n EX_n^2 - EX_N^2.$$

En reportant dans l'inégalité (15.5), il vient

$$P\left(\sup_{n \geq N} |X_n - X_N| > \frac{\varepsilon}{3}\right) \leq \frac{36}{\varepsilon^2} \left[\lim_n EX_n^2 - EX_N^2 \right];$$

la suite des ensembles $(\sup_{n \geq N} |X_n - X_N| > \varepsilon/3)$ étant décroissante en N , il en résulte que

$$P\left[\bigcap_{N \in \mathbb{N}^*} \left(\sup_{n \geq N} |X_n - X_N| > \frac{\varepsilon}{3}\right)\right] \leq \frac{36}{\varepsilon^2} \lim_N \left[\lim_n EX_n^2 - EX_N^2 \right] = 0.$$

ce qui implique que

$$P[\{X \text{ n'est pas } \}^c] \leq \sum_{\varepsilon \in \mathbb{Q}^+} P\left[\bigcap_{N \in \mathbb{N}^*} \bigcup_{m, n \geq N} \{|X_m - X_n| > \varepsilon\}\right] = 0.$$

Autrement dit la suite $(X_n)_{n \in \mathbb{N}}$ converge P-p.s. vers une variable aléatoire X_∞ .

Par le lemme de Fatou, on a alors

$$\int_{\Omega} X_\infty^2 dP \leq \liminf_n EX_n^2 \leq \sup_{n \in \mathbb{N}} EX_n^2 < +\infty,$$

ce qui montre que $X_\infty \in L^2$.

Puisque $E[X_n - X_m]^2 = EX_n^2 - EX_m^2$ et que la suite $(EX_n^2)_{n \in \mathbb{N}}$ est convergente, la suite $(X_n)_{n \in \mathbb{N}}$ est de Cauchy dans L^2 ; il en résulte qu'elle converge dans L^2 vers X_∞ . Puisque pour tous m et n tels que $m \geq n$ on a $X_n = E^{\mathcal{A}_n} X_m$, par continuité de l'espérance conditionnelle pour la norme L^2 on a

$$X_n = \lim_m E^{\mathcal{A}_n} X_m = E^{\mathcal{A}_n} X_\infty.$$

En particulier, si la filtration est complète, X_∞ est \mathcal{A}_∞ -mesurable et la martingale X est fermable. \square

Donnons deux exemples de martingale L^2 , l'une bornée dans L^2 , l'autre non bornée dans L^2 .

Exemple 15.10. Soit une suite $(a_n)_{n \in \mathbb{N}}$ de nombres réels. Considérons le processus $S = (S_n)_{n \in \mathbb{N}}$ défini par, pour tout $n \in \mathbb{N}$,

$$S_n = \sum_{j=0}^n a_j X_j.$$

où les X_n , $n \in \mathbb{N}$, forment une suite de variables aléatoires *indépendantes*, de même loi $\frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$. Pour $n \in \mathbb{N}$, on note $\mathcal{A}_n = \sigma(X_j \mid 0 \leq j \leq n)$. Le processus $S = (S_n)_{n \in \mathbb{N}}$ est une martingale pour la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$. En effet, on a

$$E^{\mathcal{A}_n}(S_{n+1}) = S_n + a_{n+1} E^{\mathcal{A}_n}(X_{n+1}),$$

et, puisque les X_n sont indépendantes et centrées, on a

$$E^{\mathcal{A}_n}(X_{n+1}) = E(X_{n+1}) = 0.$$

Il en résulte que

$$E^{\mathcal{A}_n}(S_{n+1}) = S_n.$$

Bien sûr, S_n est dans L^2 , et on a, puisque les X_n sont indépendantes, centrées, de variance 1,

$$E(S_n^2) = \sigma_{S_n}^2 = \sum_{j=0}^n a_j^2.$$

Si on suppose que $\sum_{j=0}^{+\infty} a_j^2 < +\infty$, la martingale S est **bornée dans L^2** et converge donc P-p.s. et dans L^2 .

Si on choisit tous les a_n égaux à 1, la martingale S est dans L^2 mais **n'est pas bornée dans L^2** . Démontrons que P-p.s., la suite $(S_n)_{n \in \mathbb{N}}$ ne converge pas. Il suffit pour cela de démontrer que

$$P\left(\limsup_n \frac{S_n}{\sqrt{n}} = +\infty\right) = 1. \quad (15.6)$$

Pour tout $c > 0$, on a, en conséquence du lemme de Fatou pour les ensembles,

$$\limsup_n P\left[\frac{S_n}{\sqrt{n}} > c\right] \leq P\left[\limsup_n \left(\frac{S_n}{\sqrt{n}} > c\right)\right],$$

et donc, a fortiori,

$$\limsup_n P\left[\frac{S_n}{\sqrt{n}} > c\right] \leq P\left(\limsup_n \frac{S_n}{\sqrt{n}} \geq c\right).$$

Mais, d'après le théorème limite central, on sait que

$$\lim_n P\left[\frac{S_n}{\sqrt{n}} > c\right] = \int_c^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx > 0.$$

Ainsi, on a, pour tout $c > 0$,

$$P\left(\limsup_n \frac{S_n}{\sqrt{n}} \geq c\right) > 0,$$

et donc, par la loi du tout ou rien de Kolmogorov,

$$P\left(\limsup_n \frac{S_n}{\sqrt{n}} \geq c\right) = 1,$$

puisque l'événement $(\limsup_n S_n/\sqrt{n} \geq c)$ est asymptotique. L'égalité (15.6) en résulte immédiatement en écrivant, par exemple, que

$$\left(\limsup_n \frac{S_n}{\sqrt{n}} = +\infty\right) = \bigcup_{p \in \mathbb{N}^+} \left(\limsup_n \frac{S_n}{\sqrt{n}} \geq p\right),$$

ce qui achève la démonstration.

15.6. Décomposition de Doob

À un processus $X = (X_n)_{n \in \mathbb{N}}$ on associe le **processus des accroissements** $\Delta X = (\Delta X_n)_{n \in \mathbb{N}}$ défini par

$$\Delta X_0 = X_0 \quad \text{et, pour tout } n \in \mathbb{N}^*, \quad \Delta X_n = X_n - X_{n-1}.$$

On a alors, pour tout $n \in \mathbb{N}$, $X_n = \sum_{j=0}^n \Delta X_j$.

Définition 15.15. (a) Un processus $X = (X_n)_{n \in \mathbb{N}}$ est **prévisible** si X_0 est \mathcal{A}_0 -mesurable et si, pour tout $n \in \mathbb{N}^*$, X_n est \mathcal{A}_{n-1} -mesurable.

(b) Un processus $A = (A_n)_{n \in \mathbb{N}}$ est **croissant prévisible** s'il est prévisible, si $A_0 = 0$ et s'il vérifie, pour tout $n \in \mathbb{N}^*$,

$$0 \leq A_n \leq A_{n+1} < +\infty \quad \text{P-p.s.}$$

On note alors A_∞ la limite dans $\overline{\mathbb{R}}^+$ de la suite $(A_n)_{n \in \mathbb{N}}$.

Théorème 15.16 (Décomposition de Doob). Soit X une sous-martingale intégrable.

(a) Il existe une martingale intégrable M et un processus croissant prévisible A uniques tels que $X = M + A$.

(b) On a l'équivalence :

$$\sup_{n \in \mathbb{N}} EX_n^+ < +\infty \iff \sup_{n \in \mathbb{N}} E|M_n| < +\infty \quad \text{et } A_\infty \in \mathcal{L}^1.$$

Démonstration. (a) *Existence.* On définit M et A par les processus de leurs accroissements en posant

$$M_0 = X_0 \quad \text{et, pour tout } n \in \mathbb{N}^*, \quad \Delta M_n = X_n - E^{\mathcal{A}_{n-1}} X_n.$$

$$A_0 = 0 \quad \text{et, pour tout } n \in \mathbb{N}^*, \quad \Delta A_n = E^{\mathcal{A}_{n-1}} X_n - X_{n-1}.$$

On a $E^{\mathcal{A}_{n-1}} \Delta M_n = 0$ et M est bien une martingale intégrable ; de plus, X étant une sous-martingale, on a $\Delta A_n \geq 0$. Enfin, par construction, on a $X = M + A$.

Unicité. Soit une deuxième décomposition $X = M' + A'$, où M' est une martingale intégrable et A' un processus croissant prévisible. On a alors

$$\Delta A'_n = \Delta X_n - \Delta M'_n,$$

si bien que, M' étant une martingale et A' un processus croissant prévisible, on a

$$\Delta A'_n = E^{\mathcal{A}_{n-1}}(\Delta X_n) = \Delta A_n.$$

Il en résulte que $A = A'$ et donc aussi $M = M'$.

(b) Supposons que $\sup_{n \in \mathbb{N}} E|M_n| < +\infty$ et $A_\infty \in \mathcal{L}^1$. On a

$$X_n^+ = (M_n + A_n)^+ \leq M_n^+ + A_n,$$

et donc

$$\sup_{n \in \mathbb{N}} EX_n^+ \leq \sup_{n \in \mathbb{N}} EM_n^+ + EA_\infty < +\infty.$$

Inversement, supposons que $\sup_{n \in \mathbb{N}} EX_n^+ < +\infty$; on a $M_n \leq X_n$ et donc $M_n^+ \leq X_n^+$. Il en résulte que $\sup_{n \in \mathbb{N}} EM_n^+ < +\infty$.

Alors, puisque $A_n = X_n - M_n \leq X_n^+ - M_n$ et que $EM_n = EM_0$, on a

$$EA_n \leq \sup_{n \in \mathbb{N}} EX_n^+ - EM_0 ;$$

A étant croissant positif, le lemme de Beppo Levi assure que

$$EA_{\infty} \leq \sup_{n \in \mathbb{N}} EX_n^+ - EM_0 < +\infty.$$

Il reste à remarquer que si M est une martingale intégrable, on a l'équivalence

$$\sup_{n \in \mathbb{N}} EM_n^+ < +\infty \iff \sup_{n \in \mathbb{N}} E|M_n| < +\infty.$$

En effet, l'implication de droite à gauche résulte de l'inégalité $M_n^+ \leq |M_n|$; l'implication inverse résulte de ce que l'on a $|M_n| = 2M_n^+ - M_n$ et donc, puisque M est une martingale, $E|M_n| = 2EM_n^+ - EM_0$. \square

Définition 15.17. Soit X une martingale de L^2 . Le processus croissant prévisible de la décomposition de Doob de la sous-martingale intégrable X^2 est appelé processus croissant prévisible de la martingale X et noté $\langle X \rangle$. C'est l'unique processus croissant prévisible tel que $X^2 - \langle X \rangle$ soit une martingale.

Remarque. Une martingale X de L^2 est bornée dans L^2 si et seulement si $\langle X \rangle_{\infty}$ est intégrable et on a $\sup_{n \in \mathbb{N}} EX_n^2 = EX_0^2 + E\langle X \rangle_{\infty}$.

On donne une loi forte des grands nombres pour une martingale de L^2 .

Théorème 15.18 (Loi forte des grands nombres). Soit X une martingale de L^2 . Sur l'ensemble $\{\langle X \rangle_{\infty} = +\infty\}$, la suite de terme général $\langle X \rangle_n$ est non nulle à partir d'un certain rang (aléatoire) et la suite de terme général $\frac{X_n}{\langle X \rangle_n}$ converge P-p.s. vers 0.

Démonstration. Sur l'ensemble $\{\langle X \rangle_{\infty} = +\infty\}$, la suite de terme général $\langle X \rangle_n$ tend vers $+\infty$ en croissant; elle est donc non nulle à partir d'un certain rang.

Le processus Y défini par le processus de ses accroissements

$$Y_0 = X_0 \quad \text{et, pour tout } n \in \mathbb{N}^*, \quad \Delta Y_n = \frac{\Delta X_n}{1 + \langle X \rangle_n}$$

est une martingale bornée dans L^2 . C'est une martingale, puisque $\langle X \rangle_n$ étant \mathcal{A}_{n-1} -mesurable, on a, pour tout $n \in \mathbb{N}^*$,

$$E^{\mathcal{A}_{n-1}}(\Delta Y_n) = \frac{1}{1 + \langle X \rangle_n} E^{\mathcal{A}_{n-1}}(\Delta X_n) = 0;$$

elle est dans L^2 puisque $(\Delta Y_n)^2 \leq (\Delta X_n)^2$. De plus, pour tout $k \in \mathbb{N}^*$, on a $E(Y_k - Y_{k-1})^2 = EY_k^2 - EY_{k-1}^2$, ce qui implique que

$$EY_n^2 = EY_0^2 + E\left[\sum_{k=1}^n (\Delta Y_k)^2\right];$$

or on a, puisque $\frac{1}{(1+\langle X \rangle_k)^2}$ est \mathcal{A}_{k-1} -mesurable,

$$\mathbb{E}\left[\sum_{k=1}^n (\Delta Y_k)^2\right] = \sum_{k=1}^n \mathbb{E}\left[\frac{1}{(1+\langle X \rangle_k)^2} \mathbb{E}^{\mathcal{A}_{k-1}}(\Delta X_k)^2\right],$$

soit, par définition puis croissance du processus $\langle X \rangle$,

$$\mathbb{E}\left[\sum_{k=1}^n (\Delta Y_k)^2\right] = \sum_{k=1}^n \mathbb{E}\left[\frac{\Delta \langle X \rangle_k}{(1+\langle X \rangle_k)^2}\right] \leq \mathbb{E}\left[\sum_{k=1}^n \int_{\langle X \rangle_{k-1}}^{\langle X \rangle_k} \frac{1}{(1+x)^2} dx\right].$$

Il en résulte que

$$\mathbb{E}\left[\sum_{k=1}^n (\Delta Y_k)^2\right] \leq \int_0^{+\infty} \frac{1}{(1+x)^2} dx < +\infty,$$

et donc que $\sup_{n \in \mathbb{N}} \mathbb{E} Y_n^2 < +\infty$. D'après le théorème 15.14, la suite de terme général Y_n converge P-p.s. et dans L^2 . Il résulte alors du lemme de Kronecker que, sur l'ensemble $\{\langle X \rangle_\infty = +\infty\}$, la suite de terme général $\frac{1}{1+\langle X \rangle_n} (\sum_{k=1}^n \Delta X_k)$ converge vers 0, ce qui donne le résultat. \square

Remarque. Ce dernier théorème est bien une généralisation des théorèmes de loi forte des grands nombres (dans le contexte L^2) pour les variables aléatoires indépendantes. On peut d'ailleurs les redémontrer à l'aide du théorème 15.18. Faisons-le dans un contexte simple.

Supposons donnée, par exemple, des variables aléatoires X_n , $n \in \mathbb{N}$, qui forment une suite de variables aléatoires *indépendantes*, de même loi, *centrées*, et admettant un moment d'ordre 2. Considérons le processus $S = (S_n)_{n \in \mathbb{N}}$ défini par

$$S_n = \sum_{j=0}^n X_j.$$

Pour $n \in \mathbb{N}$, on note $\mathcal{A}_n = \sigma(X_j \mid 0 \leq j \leq n)$. Comme on l'a déjà vu, le processus $S = (S_n)_{n \in \mathbb{N}}$ est une martingale pour la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$; elle est dans L^2 . Calculons son processus croissant prévisible $\langle S \rangle$. On rappelle que

$$\mathbb{E}^{\mathcal{A}_n} [S_{n+1}^2 - S_n^2] = \mathbb{E}^{\mathcal{A}_n} [(\Delta S_{n+1})^2],$$

et, puisque les X_n sont indépendantes, de même loi,

$$\mathbb{E}^{\mathcal{A}_n} [(\Delta S_{n+1})^2] = \mathbb{E}^{\mathcal{A}_n} (X_{n+1}^2) = \mathbb{E}(X_{n+1}^2) = \sigma^2,$$

où σ^2 désigne la variance commune des X_n . Il en résulte que

$$\langle S \rangle_n = n\sigma^2.$$

On a alors, par le théorème de loi forte des martingales,

$$\frac{1}{n} \sum_{j=1}^n X_j = \sigma^2 \frac{S_n - X_0}{\langle S \rangle_n} \xrightarrow{\text{P-p.s.}} 0.$$

15.7. Convergence de martingales intégrables

Définition 15.19. À un processus $X = (X_n)_{n \in \mathbb{N}}$ on associe son processus de variation quadratique $[X] = ([X]_n)_{n \in \mathbb{N}}$ défini par, pour tout $n \in \mathbb{N}$, $[X]_n = \sum_{j=0}^n (\Delta X_j)^2$. On note $[X]_\infty$ la limite dans $\overline{\mathbb{R}}^+$ de la suite croissante $([X]_n)_{n \in \mathbb{N}}$.

Remarque. Soit $x = (x_n)_{n \in \mathbb{N}}$ une suite de réels. Avec les notations ci-dessus, la condition $\sum_{j=0}^{+\infty} |\Delta x_j| < +\infty$ assure la convergence de la suite x . Par contre, comme le montre l'exemple suivant, la suite x peut être divergente alors que sa variation quadratique est finie ; tel est le cas de la suite définie par ses accroissements : $\Delta x_0 = 0$ et, pour tout $n \in \mathbb{N}^*$, $\Delta x_n = 1/n$.

Par ailleurs, la suite x peut être convergente alors que sa variation quadratique est infinie ; tel est le cas de la suite définie par ses accroissements : $\Delta x_0 = 0$ et, pour tout $n \in \mathbb{N}^*$, $\Delta x_n = (-1)^n / \sqrt{n}$, la convergence de la suite x résultant du critère de Leibniz des séries alternées. Le lemme suivant montre que cette situation ne peut pas se produire pour des martingales bornées dans L^1 .

Lemme 15.20. Si X est une martingale bornée dans L^1 , on a

$$[X]_\infty < +\infty \quad \text{P-p.s.}$$

Démonstration. (a) Pour tout $n \in \mathbb{N}^*$, on a

$$\begin{aligned} [X]_n &= X_0^2 + \sum_{j=1}^n (X_j^2 + X_{j-1}^2 - 2X_j X_{j-1}) \\ &= \sum_{j=0}^n X_j^2 + \sum_{j=0}^{n-1} X_j^2 - 2 \sum_{j=1}^n X_{j-1} (X_j - X_{j-1}) - 2 \sum_{j=1}^n X_{j-1}^2, \end{aligned}$$

ce qui s'écrit

$$[X]_n = X_n^2 - 2 \sum_{j=1}^n X_{j-1} \Delta X_j.$$

Il en résulte que, pour tout $n \geq 2$, on a

$$[X]_{n-1} + X_{n-1}^2 = 2X_n X_{n-1} - 2 \sum_{j=1}^n X_{j-1} \Delta X_j ;$$

on vérifie que cette égalité est encore vraie pour $n = 1$, si bien que l'on a l'inégalité, pour tout $n \in \mathbb{N}^*$,

$$[X]_{n-1} \leq 2X_n X_{n-1} - 2 \sum_{j=1}^n X_{j-1} \Delta X_j; \quad (15.7)$$

(b) Soit $\lambda > 0$ et soit T_λ le temps d'arrêt défini par $T_\lambda = \inf(n \in \mathbb{N} \mid |X_n| > \lambda)$, où on pose $\inf \emptyset = +\infty$. Pour tout entier $k \geq 2$, on définit le temps d'arrêt borné $S_k = T_\lambda \wedge k$.

Notons $\|X\|_1 = \sup_{n \in \mathbb{N}} E|X_n|$. On a l'inégalité

$$E[\mathbf{1}_{(S_k \geq 1)} [X]_{S_k-1}] \leq 2\lambda \|X\|_1. \quad (15.8)$$

En effet, sur l'ensemble $(S_k \geq 1)$,

$$\sum_{j=1}^{S_k} X_{j-1} \Delta X_j = \sum_{j=1}^k X_{j-1} \mathbf{1}_{(j \leq T_\lambda)} \Delta X_j; \quad (15.9)$$

remarquons que, si $j \in \mathbb{N}^*$, on a $(j \leq T_\lambda) = (T_\lambda \leq j-1)^c \in \mathcal{A}_{j-1}$ et $(S_k \geq 1)^c = (T_\lambda = 0) \in \mathcal{A}_0$, si bien que la variable aléatoire $\mathbf{1}_{(S_k \geq 1)} X_{j-1} \mathbf{1}_{(j \leq T_\lambda)}$ est \mathcal{A}_{j-1} -mesurable; de plus, par définition de T_λ , on a $|X_{j-1}| \mathbf{1}_{(j \leq T_\lambda)} \leq \lambda$, si bien que, ΔX_j étant intégrable, il en est de même de la variable aléatoire $\mathbf{1}_{(S_k \geq 1)} \sum_{j=1}^k X_{j-1} \mathbf{1}_{(j \leq T_\lambda)} \Delta X_j$. En intégrant les deux membres de (15.9) sur $(S_k \geq 1)$, il vient alors, puisque X est une martingale,

$$E\left[\mathbf{1}_{(S_k \geq 1)} \sum_{j=1}^{S_k} X_{j-1} \Delta X_j\right] = \sum_{j=1}^k E\left[\mathbf{1}_{(S_k \geq 1)} X_{j-1} \mathbf{1}_{(j \leq T_\lambda)} E^{\mathcal{A}_{j-1}} \Delta X_j\right] = 0. \quad (15.10)$$

Toujours par définition de T_λ , et donc de S_k , on a

$$\mathbf{1}_{(S_k \geq 1)} |X_{S_k} X_{S_k-1}| \leq \mathbf{1}_{(S_k \geq 1)} \lambda |X_{S_k}|; \quad (15.11)$$

mais, S_k étant borné, le premier théorème d'arrêt de Doob montre que X_{S_k} est intégrable et que l'on a $X_{S_k} = E^{\mathcal{A}_{S_k}}(X_k)$ d'où il résulte que

$$E|X_{S_k}| \leq E|X_k| \leq \|X\|_1. \quad (15.12)$$

Il résulte alors de (15.11) que

$$E[\mathbf{1}_{(S_k \geq 1)} |X_{S_k} X_{S_k-1}|] \leq \lambda \|X\|_1; \quad (15.13)$$

Enfin, de (15.7) il résulte que

$$\mathbf{1}_{(S_k \geq 1)} [X]_{S_k-1} \leq \mathbf{1}_{(S_k \geq 1)} \left[2|X_{S_k} X_{S_k-1}| - 2 \sum_{j=1}^{S_k} X_{j-1} \Delta X_j \right]; \quad (15.14)$$

en intégrant les deux membres de cette inégalité, on obtient alors de suite l'inégalité annoncée (15.8), en utilisant les relations (15.10) et (15.13).

(c) La suite de terme général S_k convergeant en croissant vers T_λ , la suite de terme général $\mathbf{1}_{(S_k \geq 1)}[X]_{S_k-1}$ converge en croissant vers $\mathbf{1}_{(T_\lambda \geq 1)}[X]_{T_\lambda-1}$, puisque $[X]_n$ est somme de carrés. Un passage à la limite dans (15.8) et la propriété de Beppo Levi donnent donc

$$E[\mathbf{1}_{(T_\lambda \geq 1)}[X]_{T_\lambda-1}] \leq 2\lambda \|X\|_1. \quad (15.15)$$

(d) Remarquant que $(T_\lambda < +\infty) \subset (X^* > \lambda)$, le lemme maximal donne, pour tous $\alpha > 0$ et $\lambda > 0$, les inégalités

$$P[(|X|_\infty \geq \alpha^2) \cap (T_\lambda < +\infty)] \leq P(T_\lambda < +\infty) \leq \frac{\|X\|_1}{\lambda}. \quad (15.16)$$

Puisque, de plus, on a

$$P[(|X|_\infty \geq \alpha^2) \cap (T_\lambda = +\infty)] \leq P[(\mathbf{1}_{(T_\lambda \geq 1)}[X]_{T_\lambda-1} \geq \alpha^2) \cap (T_\lambda = +\infty)],$$

et donc

$$P[(|X|_\infty \geq \alpha^2) \cap (T_\lambda = +\infty)] \leq P(\mathbf{1}_{(T_\lambda \geq 1)}[X]_{T_\lambda-1} \geq \alpha^2), \quad (15.17)$$

il résulte de l'inégalité de Markov et de (15.15) que

$$P[(|X|_\infty \geq \alpha^2) \cap (T_\lambda = +\infty)] \leq \frac{2\lambda}{\alpha^2} \|X\|_1; \quad (15.18)$$

prenant $\lambda = \alpha$ dans cette dernière inégalité, et additionnant membre à membre les inégalités (15.16) et (15.18), il vient

$$P(|X|_\infty \geq \alpha^2) \leq \frac{3}{\alpha} \|X\|_1;$$

l'arbitraire de α démontre alors que $P(|X|_\infty = +\infty) = 0$, ce qui est le résultat annoncé. \square

Le théorème suivant, relatif à la convergence de martingales et démontré pour la première fois par D.L. Burkholder, est alors une conséquence du théorème de convergence pour les martingales bornées dans L^2 , de ce dernier lemme sur la variation quadratique, et du lemme maximal : sa démonstration suit un article de Louis H.Y. Chen paru dans les *Proceedings of the AMS* en 1981.

Notation. Si $X = (X_n)_{n \in \mathbb{N}}$ est un processus, on note $X^* = \sup_{n \in \mathbb{N}} |X_n|$.

Théorème 15.21. Soient M et N deux martingales sur la même base de processus. On suppose que M est bornée dans L^1 . Si les processus de variation quadratique de M et N sont tels que $[N] \leq [M]$, la suite $(N_n)_{n \in \mathbb{N}}$ converge P-p.s.

En particulier, toute martingale bornée dans L^1 converge P-p.s.

Démonstration. On note $\|M\|_1 = \sup_{n \in \mathbb{N}} E|M_n|$.

- Soit $\lambda > 0$ et soit T_λ le temps d'arrêt défini par

$$T_\lambda = \inf(n \in \mathbb{N} \mid |M_n| > \lambda \text{ ou } [M]_n > \lambda^2),$$

où on pose $\inf \emptyset = +\infty$. On va démontrer l'inégalité relative au processus des accroissements de la martingale arrêtée N^{T_λ}

$$E[(\Delta N^{T_\lambda})^*] \leq 2\lambda + \|M\|_1 < +\infty. \quad (15.19)$$

On a, pour tout $n \in \mathbb{N}$,

$$(\Delta N_n^{T_\lambda})^2 \leq [N^{T_\lambda}]_n = [N]_{n \wedge T_\lambda} \leq [M]_{n \wedge T_\lambda} = \mathbf{1}_{(T_\lambda > n)} [M]_n + \mathbf{1}_{(T_\lambda \leq n)} [M]_{T_\lambda},$$

et donc

$$|\Delta N_n^{T_\lambda}| \leq \mathbf{1}_{(T_\lambda > n)} [M]_n^{\frac{1}{2}} + \mathbf{1}_{(T_\lambda \leq n)} [M]_{T_\lambda}^{\frac{1}{2}} \leq \lambda \mathbf{1}_{(T_\lambda > n)} + \mathbf{1}_{(T_\lambda \leq n)} [M]_{T_\lambda}^{\frac{1}{2}}; \quad (15.20)$$

or, par définition de T_λ , on a, sur $(T_\lambda < +\infty)$,

$$[M]_{T_\lambda}^{\frac{1}{2}} = [M]_{T_\lambda-1} + (\Delta M_{T_\lambda})^2]^{\frac{1}{2}} \leq \lambda + |\Delta M_{T_\lambda}|.$$

ce qui, en reportant dans (15.20), implique l'inégalité

$$|\Delta N_n^{T_\lambda}| \leq \lambda + |\Delta M_{T_\lambda}| \mathbf{1}_{(T_\lambda \leq n)}. \quad (15.21)$$

Il en résulte que

$$(\Delta N^{T_\lambda})^* \leq \lambda + |\Delta M_{T_\lambda}| \mathbf{1}_{(T_\lambda < +\infty)}. \quad (15.22)$$

De plus, par l'inégalité triangulaire, on a, sur $(T_\lambda < +\infty)$,

$$|\Delta M_{T_\lambda}| \leq |M_{T_\lambda-1}| + |M_{T_\lambda}| \leq \lambda + |M_{T_\lambda}|;$$

il en résulte que

$$(\Delta N^{T_\lambda})^* \leq 2\lambda + |M_{T_\lambda}| \mathbf{1}_{(T_\lambda < +\infty)}, \quad (15.23)$$

soit, en intégrant

$$E[(\Delta N^{T_\lambda})^*] \leq 2\lambda + E[|M_{T_\lambda}| \mathbf{1}_{(T_\lambda < +\infty)}]. \quad (15.24)$$

Il reste à majorer le membre de droite : or, puisque l'on a

$$\lim_n \mathbf{1}_{(T_\lambda < +\infty)} |M_{T_\lambda \wedge n}| = \mathbf{1}_{(T_\lambda < +\infty)} |M_{T_\lambda}|.$$

il résulte du lemme de Fatou et de l'inégalité (15.24) que l'on a

$$E[(\Delta N^{T_\lambda})^*] \leq 2\lambda + \liminf_n E[\mathbf{1}_{(T_\lambda < +\infty)} |M_{T_\lambda \wedge n}|] \leq 2\lambda + \sup_{n \in \mathbb{N}} E|M_{T_\lambda \wedge n}|;$$

mais, $T_\lambda \wedge n$ étant un temps d'arrêt borné, le premier théorème d'arrêt de Doob montre que l'on a $M_{T_\lambda \wedge n} = E^{\mathcal{A}_{T_\lambda \wedge n}}(M_n)$, d'où il résulte que

$$E|M_{T_\lambda \wedge n}| \leq E|M_n| \leq \|M\|_1,$$

ce qui achève de démontrer l'inégalité (15.19).

• Soit le processus $U = 1 + [N^{T_\lambda}]$: on introduit le processus Y^λ , somme normalisée par U du processus des accroissements de N^{T_λ} défini par ses accroissements en tout $n \in \mathbb{N}$:

$$\Delta Y_n^\lambda = \frac{\Delta N_n^{T_\lambda}}{U_n}.$$

On a, pour tout $n \in \mathbb{N}^*$,

$$\frac{(\Delta N_n^{T_\lambda})^2}{U_n^2} = \frac{\Delta U_n}{U_n^2} \leq \int_{U_{n-1}}^{U_n} \frac{dx}{x^2},$$

et donc

$$\sum_{n=1}^{+\infty} \frac{(\Delta N_n^{T_\lambda})^2}{U_n^2} \leq \sum_{n=1}^{+\infty} \int_{U_{n-1}}^{U_n} \frac{dx}{x^2} \leq \int_1^{+\infty} \frac{dx}{x^2} = 1,$$

ce qui implique l'inégalité

$$\sum_{n=1}^{+\infty} E[(\Delta Y_n^\lambda)^2] = E\left[\sum_{n=1}^{+\infty} \frac{(\Delta N_n^{T_\lambda})^2}{U_n^2}\right] \leq 1. \quad (15.25)$$

• On va démontrer que la suite $(Y_n^\lambda)_{n \in \mathbb{N}}$ converge P-p.s. ; pour cela, on introduit la martingale Z dans L^2 définie par ses accroissements

$$\Delta Z_0 = 0 \text{ et, si } n \in \mathbb{N}^*, \quad \Delta Z_n = \Delta Y_n^\lambda - E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda.$$

C'est une **martingale bornée dans L^2** . En effet, suivant un calcul classique sur la covariance conditionnelle, on a, pour tout $n \geq 1$,

$$\begin{aligned} E[(\Delta Z_n)^2] &= E[(\Delta Y_n^\lambda)^2] - 2E\left[\Delta Y_n^\lambda (E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda)\right] + E\left[(E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda)^2\right] \\ &= E[(\Delta Y_n^\lambda)^2] - E\left[(E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda)^2\right] \leq E[(\Delta Y_n^\lambda)^2]; \end{aligned}$$

mais, Z étant une martingale dans L^2 , on a (calcul déjà fait)

$$E[Z_n^2] - E[Z_{n-1}^2] = E[(\Delta Z_n)^2],$$

si bien que

$$E[Z_n^2] = E[Z_0^2] + \sum_{j=1}^n E[(\Delta Z_j)^2] \leq E[Z_0^2] + \sum_{j=1}^n E[(\Delta Y_j^\lambda)^2];$$

il résulte alors de l'inégalité (15.25) que

$$\sup_{n \in \mathbb{N}^*} E[Z_n^2] \leq E[Z_0^2] + 1.$$

La suite $(Z_n)_{n \in \mathbb{N}}$ converge donc P-p.s..

Nous allons démontrer que la série de terme général $E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda$ est P-p.s. absolument convergente. Remarquons que, U_{n-1} étant \mathcal{A}_{n-1} -mesurable et $N^{T\lambda}$ étant une martingale, on a

$$E^{\mathcal{A}_{n-1}} \left(\frac{\Delta N_n^{T\lambda}}{U_{n-1}} \right) = \frac{1}{U_{n-1}} E^{\mathcal{A}_{n-1}} (\Delta N_n^{T\lambda}) = 0.$$

On peut donc écrire que

$$\begin{aligned} E \left[\sum_{n=1}^{+\infty} |E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda| \right] &= E \left[\sum_{n=1}^{+\infty} \left| E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda - E^{\mathcal{A}_{n-1}} \left(\frac{\Delta N_n^{T\lambda}}{U_{n-1}} \right) \right| \right] \\ &= \sum_{n=1}^{+\infty} E \left[\left| E^{\mathcal{A}_{n-1}} \left(\Delta N_n^{T\lambda} \left(\frac{1}{U_{n-1}} - \frac{1}{U_n} \right) \right) \right| \right], \end{aligned}$$

ce qui, en remarquant que $\frac{1}{U_{n-1}} - \frac{1}{U_n}$ est positif, donne la majoration

$$E \left[\sum_{n=1}^{+\infty} |E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda| \right] \leq \sum_{n=1}^{+\infty} E \left[|\Delta N_n^{T\lambda}| \left(\frac{1}{U_{n-1}} - \frac{1}{U_n} \right) \right],$$

et encore

$$E \left[\sum_{n=1}^{+\infty} |E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda| \right] \leq E \left[(\Delta N^{T\lambda})^* \sum_{n=1}^{+\infty} \left(\frac{1}{U_{n-1}} - \frac{1}{U_n} \right) \right]; \quad (15.26)$$

mais on a, pour $n \geq 1$,

$$\sum_{j=1}^n \left(\frac{1}{U_{j-1}} - \frac{1}{U_j} \right) = \frac{1}{U_0} - \frac{1}{U_n} \leq 1,$$

ce qui implique, puisque U_n est positif, que $\sum_{n=1}^{+\infty} \left(\frac{1}{U_{n-1}} - \frac{1}{U_n} \right) \leq 1$. Les inégalités (15.26) et (15.19) donnent alors

$$E \left[\sum_{n=1}^{+\infty} |E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda| \right] \leq E \left[(\Delta N^{T\lambda})^* \right] < +\infty,$$

et il en résulte que **la série de terme général $E^{\mathcal{A}_{n-1}} \Delta Y_n^\lambda$ est P-p.s. absolument convergente**. Ainsi, puisque pour tout $n \in \mathbb{N}^*$, on a

$$Y_n^\lambda = Y_0^\lambda + Z_n + \sum_{j=1}^n E^{\mathcal{A}_{j-1}} \Delta Y_j^\lambda$$

et que la suite $(Z_n)_{n \in \mathbb{N}}$ converge donc P-p.s., **la suite $(Y_n^\lambda)_{n \in \mathbb{N}}$ converge P-p.s.**

• Définissons le processus Y , en tout $n \in \mathbb{N}$, par

$$Y_n = \sum_{j=0}^n \frac{\Delta N_j}{U_j};$$

pour tout $\lambda > 0$, il coïncide sur $(T_\lambda = +\infty)$ avec le processus Y^λ , ce qui montre que la suite $(Y_n)_{n \in \mathbb{N}}$ converge P-p.s. sur $\bigcup_{\lambda \in \mathbb{Q}^+} (T_\lambda = +\infty)$. Mais on a

$$\begin{aligned} \bigcup_{\lambda \in \mathbb{Q}^+} (T_\lambda = +\infty) &= \bigcup_{\lambda \in \mathbb{Q}^+} [([M]_\infty \leq \lambda^2) \cap (M^* \leq \lambda)] \\ &= ([M]_\infty < +\infty) \cap (M^* < +\infty); \end{aligned}$$

la martingale M étant bornée dans L^1 , le lemme sur la variation quadratique et lemme maximal assurent alors que

$$P\left[\bigcup_{\lambda \in \mathbb{Q}^+} (T_\lambda = +\infty)\right] = 1;$$

ainsi, la suite $(Y_n)_{n \in \mathbb{N}}$ converge P-p.s.. Enfin, puisque on a l'inclusion des ensembles $([M]_\infty < +\infty) \subset ([N]_\infty < +\infty)$, la suite $(U_n)_{n \in \mathbb{N}}$ converge P-p.s. vers une limite finie; un résultat simple d'analyse montre alors que la suite $(\sum_{j=0}^n \Delta N_j)_{n \in \mathbb{N}}$ converge P-p.s., c'est-à-dire que la suite $(N_n)_{n \in \mathbb{N}}$ converge P-p.s., ce qui achève la démonstration. \square

Corollaire 15.22. Soient X une martingale bornée dans L^1 et T un temps d'arrêt; la martingale arrêtée X^T converge P-p.s.

Démonstration. Il suffit de montrer que la martingale X^T est bornée dans L^1 ; or il résulte du premier théorème d'arrêt que l'on a, pour tout $n \in \mathbb{N}$, $X_{T \wedge n} = E^{\mathcal{A}_{T \wedge n}} X_n$, ce qui permet d'écrire la suite d'inégalités

$$\begin{aligned} E|X_n^T| &= E|X_{T \wedge n}| = E|E^{\mathcal{A}_{T \wedge n}} X_n| \leq E[E^{\mathcal{A}_{T \wedge n}} |X_n|] \\ &= E|X_n| \leq \sup_{n \in \mathbb{N}} E|X_n| < +\infty, \end{aligned}$$

d'où il résulte que $\sup_{n \in \mathbb{N}} E|X_n^T| < +\infty$. \square

Remarque. Le contre-exemple suivant montre qu'une martingale bornée dans L^1 peut ne pas converger dans L^1 . Soit une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires indépendantes de même loi $(\delta_0 + \delta_2)/2$; on définit, pour tout $n \in \mathbb{N}$, $Y_n = \prod_{j=0}^n X_j$ et $\mathcal{A}_n = \sigma(X_j \mid 0 \leq j \leq n)$. On a $E^{\mathcal{A}_n} Y_{n+1} = Y_n E^{\mathcal{A}_n} X_{n+1}$, soit par indépendance de X_{n+1} et \mathcal{A}_n , $E^{\mathcal{A}_n} Y_{n+1} = Y_n E X_{n+1} = Y_n$; ainsi le processus Y est une martingale; elle est bornée dans L^1 , puisque, par indépendance, on a, pour tout $n \in \mathbb{N}$, $E|Y_n| = \prod_{j=0}^n E|X_j| = 1$. La suite $(Y_n)_{n \in \mathbb{N}}$ converge donc P-p.s. vers une variable aléatoire Y_∞ ; en remarquant que Y_n ne prend P-p.s. que les valeurs 0 et 2^{n+1} et que $P(Y_n = 2^{n+1}) \approx 2^{-(n+1)}$, on voit que la suite $(Y_n)_{n \in \mathbb{N}}$ converge en probabilité vers 0 et donc que $Y_\infty = 0$; ainsi, la convergence ne peut avoir lieu dans L^1 , puisque $EY_n = 1$. Il est d'ailleurs facile de voir directement que cette suite n'est pas équi-intégrable. La convergence L^1 d'une martingale intégrable est caractérisée par la proposition suivante :

Proposition 15.23. Soit $X = (X_n)_{n \in \mathbb{N}}$ une martingale intégrable. La suite $(X_n)_{n \in \mathbb{I}}$ est équi-intégrable si et seulement si la martingale est fermable. Dans ce cas, la suite $(X_n)_{n \in \mathbb{N}}$ converge P-p.s. et dans L^1 vers une variable aléatoire X_∞ \mathcal{A}_∞ -mesurable.

Démonstration. La suite est équi-intégrable, elle est en particulier bornée dans L^1 et la martingale X est convergente P-p.s. et donc aussi dans L^1 , par équi-intégrabilité. Si on pose $X_\infty = \limsup_n X_n$, X_∞ est \mathcal{A}_∞ -mesurable comme limite de la suite adaptée $(X_n)_{n \in \mathbb{N}}$ et, puisque, si $n \leq p$, on a $X_n = E^{\mathcal{A}_n} X_p$, par convergence L^1 on a aussi $X_n = E^{\mathcal{A}_n} X_\infty$.

Inversement, soit X_∞ \mathcal{A}_∞ -mesurable telle que l'on ait, pour tout $n \in \mathbb{N}$, $X_n = E^{\mathcal{A}_n} X_\infty$; l'équi-intégrabilité de la suite $(X_n)_{n \in \mathbb{N}}$ résulte du lemme général suivant. \square

Lemme 15.24. Soient $X \in L^1(\Omega, \mathcal{A}, P)$ et une famille $(\mathcal{A}_i)_{i \in I}$ de sous-tribus de \mathcal{A} . Notons $X_i = E^{\mathcal{A}_i} X$; la famille $(X_i)_{i \in I}$ est équi-intégrable.

Démonstration. Soit $a > 0$ quelconque : puisque $|X_i| \leq E^{\mathcal{A}_i} |X|$ et que $(|X_i| > a) \in \mathcal{A}_i$, on a, pour tout $i \in I$,

$$P(|X_i| > a) \leq \frac{1}{a} \int |X_i| dP \leq \frac{1}{a} \int |X| dP,$$

et

$$\sup_{i \in I} \int_{(|X_i| > a)} |X_i| dP \leq \sup_{i \in I} \int_{(|X_i| > a)} |X| dP. \quad (15.27)$$

Donc, pour tout $\eta > 0$, il existe $A(\eta) > 0$ tel que $\sup_{i \in I} P(|X_i| > a) \leq \eta$ dès que $a \geq A(\eta)$. Mais, pour tout $\varepsilon > 0$, il existe $\eta(\varepsilon) > 0$ tel que $\int_A |X| dP \leq \varepsilon$ dès que $P(A) \leq \eta(\varepsilon)$. Il résulte alors de (15.27) que, si $a \geq A(\eta(\varepsilon))$, on a $\sup_{i \in I} \int_{(|X_i| > a)} |X_i| dP \leq \varepsilon$, ce qui est le résultat annoncé. \square

15.8. Deuxième théorème d'arrêt

On démontre un théorème de caractérisation des martingales **fermées** en termes de temps d'arrêt quelconques et on en déduit le deuxième théorème d'arrêt de Doob.

Théorème 15.25. On suppose que $\mathcal{A}_\infty = \bigvee_{n \in \mathbb{I}} \mathcal{A}_n$. Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus adapté. Il y a équivalence entre les propriétés suivantes :

- (i) X est une martingale fermée,
- (ii) pour tout $T \in \mathcal{T}$, $X_T \in L^1(\Omega, \mathcal{A}_T, P)$ et $EX_T = EX_0$,
- (iii) le processus $(X_T)_{T \in \mathcal{T}}$ est une martingale par rapport à la filtration $(\mathcal{A}_T)_{T \in \mathcal{T}}$.

Démonstration. (i) \Rightarrow (ii). Supposons que X est une martingale fermée. Pour tout temps d'arrêt T borne par un entier k , le premier théorème d'arrêt montre que $X_T = E^{\mathcal{A}_T} X_k$; mais X étant fermée, on a $X_k = E^{\mathcal{A}_k} X_\infty$; puisque $\mathcal{A}_T \subset \mathcal{A}_k$, il en résulte que

$$X_T = E^{\mathcal{A}_T} \left[E^{\mathcal{A}_k} X_\infty \right] = E^{\mathcal{A}_T} [X_\infty].$$

Le lemme 15.24 montre alors que la famille de variables aléatoires

$$\{X_T \mid T \in \mathcal{T}_b\}$$

est **équi-intégrable**; de plus, on a $EX_T = EX_0$ pour tout $T \in \mathcal{T}_b$. Soit maintenant un temps d'arrêt T quelconque; la famille de variables aléatoires $\{X_{T \wedge n} \mid n \in \mathbb{N}\}$ est alors équi-intégrable, la martingale arrêtée X^T converge P-p.s. et on a $EX_{T \wedge n} = EX_0$ pour tout $n \in \mathbb{N}$. Mais on a $T = \lim_n T \wedge n$, si bien que, sur $(T < +\infty)$, on a $\lim_n X_{T \wedge n} = X_T$; de plus, sur $(T = +\infty)$, on a $X_{T \wedge n} = X_n$; puisque, d'après la proposition 15.23, on a $\lim_n X_n = X_\infty$ P-p.s., on a alors, P-p.s. sur $(T = +\infty)$, $\lim_n X_{T \wedge n} = X_\infty$. En définitive, on a $\lim_n X_{T \wedge n} = X_\infty$ P-p.s. et, par équi-intégrabilité, $\lim_n EX_{T \wedge n} = EX_\infty$. Il en résulte que $EX_T = EX_0$.

(ii) \Rightarrow (iii). Soit $S \in \mathcal{T}$. On démontre d'abord que $X_S = E^{\mathcal{A}_S} [X_\infty]$. Soit $A \in \mathcal{A}_S$. L'application R définie par $R = S \mathbf{1}_A + (+\infty) \mathbf{1}_{A^c}$ est un temps d'arrêt. En effet, on a, pour tout $n \in \mathbb{N}$,

$$(R = n) = (S = n) \cap A \in \mathcal{A}_n.$$

Appliquant l'hypothèse aux temps d'arrêt R et $+\infty$, il vient $EX_R = EX_\infty$, ce qui donne l'égalité

$$E(\mathbf{1}_A X_S + \mathbf{1}_{A^c} X_\infty) = EX_\infty,$$

soit encore,

$$E(\mathbf{1}_A X_S) = E(\mathbf{1}_A X_\infty);$$

X_S étant \mathcal{A}_S -mesurable, il en résulte que $X_S = E^{\mathcal{A}_S} X_\infty$. Maintenant, si S et T sont deux temps d'arrêt tels que $S \leq T$, on a de même $X_T = E^{\mathcal{A}_T} X_\infty$ et, puisque $\mathcal{A}_S \subset \mathcal{A}_T$, on a la succession d'égalités

$$E^{\mathcal{A}_S} X_T = E^{\mathcal{A}_S} \left[E^{\mathcal{A}_T} X_\infty \right] = E^{\mathcal{A}_S} X_\infty = X_S,$$

ce qui démontre que la propriété (iii) est vérifiée.

(iii) \Rightarrow (i). Il suffit de prendre des temps d'arrêt constants, éventuellement égaux à $+\infty$. \square

Une explicitation de l'implication (i) \Rightarrow (iii) du théorème 15.25 donne alors le deuxième théorème d'arrêt de Doob.

Théorème 15.26 (Deuxième théorème d'arrêt de Doob). Soit $X = (X_n)_{n \in \mathbb{N}}$ une martingale fermée; pour tous temps d'arrêt S et T tels que $S \leq T$, on a

$$E^{A_S} X_T = X_S.$$

15.9. Convergence de sous- et surmartingales

On déduit les théorèmes de convergence des sous- et surmartingales du théorème de convergence des martingales bornées dans L^1 et de la décomposition de Doob des sous-martingales.

Théorème 15.27. Soit $X = (X_n)_{n \in \mathbb{N}}$ une sous-martingale telle que

$$\sup_{n \in \mathbb{N}} EX_n^+ < +\infty$$

alors la suite $(X_n)_{n \in \mathbb{N}}$ converge P-p.s.

Démonstration. D'après le théorème 15.16 et l'hypothèse, X admet la décomposition de Doob $X = M + A$, où M est une martingale bornée dans L^1 , donc convergente P-p.s., et A un processus croissant prévisible tel que $A_\infty \in \mathcal{L}^1$, donc fini P-p.s.; la suite $(A_n)_{n \in \mathbb{N}}$ et donc aussi la suite $(X_n)_{n \in \mathbb{N}}$ converge alors P-p.s. \square

Remarque. Soit $X = (X_n)_{n \in \mathbb{N}}$ une sous-martingale de décomposition de Doob $X = M + A$. La suite $(X_n)_{n \in \mathbb{N}}$ converge dans L^1 si et seulement si la suite $(M_n)_{n \in \mathbb{N}}$ est équi-intégrable et si $A_\infty \in \mathcal{L}^1$. En effet, si la suite $(X_n)_{n \in \mathbb{N}}$ converge dans L^1 vers X_∞ , on a

$$\begin{aligned} EA_\infty &= \lim_n \nearrow EA_n = \lim_n [EX_n - EM_n] \\ &= \lim_n [EX_n - EM_0] = EX_\infty - EX_0 < +\infty, \end{aligned}$$

ce qui montre que $A_\infty \in \mathcal{L}^1$. La suite $(A_n)_{n \in \mathbb{N}}$ convergeant en croissant vers A_∞ , converge alors aussi dans L^1 ; la convergence de la suite $(M_n)_{n \in \mathbb{N}}$ dans L^1 en résulte et cette suite est donc équi-intégrable.

Inversement, si la martingale M est équi-intégrable et si $A_\infty \in \mathcal{L}^1$, alors les suites $(M_n)_{n \in \mathbb{N}}$ et $(A_n)_{n \in \mathbb{N}}$ convergent dans L^1 respectivement vers M_∞ et A_∞ ; la suite $(X_n)_{n \in \mathbb{N}}$ converge alors dans L^1 vers $M_\infty + A_\infty$.

Corollaire 15.28. Soit $X = (X_n)_{n \in \mathbb{N}}$ une surmartingale positive; alors la suite $(X_n)_{n \in \mathbb{N}}$ converge P-p.s. vers une variable aléatoire X_∞ à valeurs dans $\overline{\mathbb{R}}^+$ et on a, pour tout $n \in \mathbb{N}$, $X_n \geq E^{A_n} X_\infty$.

Démonstration. S'il existe $c > 0$ tel que l'on ait, pour tout $n \in \mathbb{N}$, $0 \leq X_n \leq c$, alors $-X$ est une sous-martingale intégrable telle que $\sup_{n \in \mathbb{N}} E(-X_n)^+ < +\infty$; d'après le théorème précédent, la suite $(X_n)_{n \in \mathbb{N}}$ converge P-p.s. vers

une variable aléatoire X_∞ . Sous cette hypothèse supplémentaire, cette suite est équi-intégrable et, puisque, si $p \geq n$, on a $X_n \geq E^{\mathcal{A}_n} X_p$, un passage à la limite dans L^1 en p montre que $X_n \geq E^{\mathcal{A}_n} X_\infty$.

Cas général : pour tout $q \in \mathbb{Q}^+$, le processus $X \wedge q = (X_n \wedge q)_{n \in \mathbb{N}}$, minimum de deux surmartingales, est une surmartingale ; de plus, elle est positive et bornée par q ; d'après ce que l'on vient de voir, la suite $(X_n \wedge q)_{n \in \mathbb{N}}$ converge donc P-p.s. ; par un raisonnement classique, on a donc aussi que, P-p.s., pour tout $q \in \mathbb{Q}^+$, la suite $(X_n \wedge q)_{n \in \mathbb{N}}$ converge. Il résulte alors du lemme déterministe 15.29 ci-dessous que P-p.s., la suite $(X_n)_{n \in \mathbb{N}}$ converge dans $\overline{\mathbb{R}}^+$. Enfin, puisque, si $p \geq n$, on a $X_n \geq E^{\mathcal{A}_n} X_p$, le lemme de Fatou conditionnel montre que $X_n \geq E^{\mathcal{A}_n} X_\infty$. \square

Lemme 15.29. Soit $(x_n)_{n \in \mathbb{N}}$ une suite de réels positifs telle que, pour tout $q \in \mathbb{Q}^+$, la suite $(x_n \wedge q)_{n \in \mathbb{N}}$ converge, alors la suite $(x_n)_{n \in \mathbb{N}}$ converge dans $\overline{\mathbb{R}}^+$.

Démonstration. Si $\sup_{n \in \mathbb{N}} x_n < +\infty$, il suffit de prendre $q > \sup_{n \in \mathbb{N}} x_n$ pour obtenir que, pour tout n , $x_n \wedge q = x_n$; par conséquent, par hypothèse, la suite $(x_n)_{n \in \mathbb{N}}$ converge dans \mathbb{R}^+ .

Si $\sup_{n \in \mathbb{N}} x_n = +\infty$, pour tout $q \in \mathbb{Q}^+$, notons l_q la limite de la suite $(x_n \wedge q)_{n \in \mathbb{N}}$; on a $0 \leq l_q \leq q$ et, pour tout $\varepsilon > 0$, il existe $N(\varepsilon, q)$ tel que l'on ait $l_q - \varepsilon \leq x_n \wedge q \leq l_q + \varepsilon$ dès que $n \geq N(\varepsilon, q)$. Soit $B > 0$ quelconque et prenons $\varepsilon = B/4$; choisissons $q > B$.

Si on avait $l_q \leq B - \varepsilon$, on aurait, pour tout $n \geq N(B/4, q)$, $x_n \wedge q \leq B$ et donc $x_n \leq B$; il en résulterait que $\sup_{n \in \mathbb{N}} x_n < +\infty$, et il y aurait contradiction. Ainsi, on a $l_q > B - \varepsilon$; alors, dès que $n \geq N(B/4, q)$, on a $x_n \geq x_n \wedge q \geq l_q - \varepsilon > B - 2\varepsilon = B/2$. Étant donné l'arbitraire de B , on a $\lim_n x_n = +\infty$.

Dans tous les cas, la suite $(x_n)_{n \in \mathbb{N}}$ converge dans $\overline{\mathbb{R}}^+$. \square

Exercices

Exercice 15.1. Tribu des événements antérieurs à un temps d'arrêt. Soit $(\mathcal{A}_n)_{n \in \mathbb{N}}$ une filtration sur l'espace probabilisé (Ω, \mathcal{A}, P) et $\mathcal{A}_\infty = \bigvee_{n \in \mathbb{N}} \mathcal{A}_n$, tribu engendrée par la réunion des tribus \mathcal{A}_n , $n \in \mathbb{N}$. Soient S et T deux temps d'arrêt. Démontrer que, pour tout $Y \in L^1(\Omega, \mathcal{A}, P)$, on a

$$E^{\mathcal{A}_S} [E^{\mathcal{A}_T} Y] = E^{\mathcal{A}_T} [E^{\mathcal{A}_S} Y] = E^{\mathcal{A}_{S \wedge T}} Y.$$

Solution. Soit, pour tout $n \in \mathbb{N}$, $X_n = E^{\mathcal{A}_n} Y$; le processus $(X_n)_{n \in \mathbb{N}}$ est une martingale équi-intégrable, donc convergente P-p.s. et dans L^1 ; si on pose $X_\infty =$

$\limsup_n X_n$, on a $Y = X_\infty$ P.-p.s. et le processus $(X_n)_{n \in \overline{\mathbb{N}}}$ est une martingale fermée. Le deuxième théorème d'arrêt montre alors que $X_T = E^{\mathcal{A}_T} Y$ et, pour tout $n \in \mathbb{N}$,

$$X_n^T = X_{T \wedge n} = E^{\mathcal{A}_{T \wedge n}} X_T,$$

où X^T désigne la martingale X arrêtée en T . Soit n fixé et évaluons $E^{\mathcal{A}_{T \wedge n}} X_T$; pour tout $j \in \overline{\mathbb{N}}$, on a, sur $(T = j)$,

$$E^{\mathcal{A}_{T \wedge n}} X_T = E^{\mathcal{A}_{j \wedge n}} X_T = \begin{cases} E^{\mathcal{A}_j} X_T & \text{si } j < n, \\ E^{\mathcal{A}_n} X_T & \text{si } j \geq n: \end{cases}$$

mais, si $j < n$, puisque $(T = j) \in \mathcal{A}_j$, que $\mathcal{A}_j \subset \mathcal{A}_n$ et que X_j est \mathcal{A}_j -mesurable, donc \mathcal{A}_n -mesurable, on a

$$\mathbf{1}_{(T=j)} E^{\mathcal{A}_j} X_T = E^{\mathcal{A}_j} [\mathbf{1}_{(T=j)} X_T] = \mathbf{1}_{(T=j)} X_j = E^{\mathcal{A}_n} [\mathbf{1}_{(T=j)} X_T] = \mathbf{1}_{(T=j)} E^{\mathcal{A}_n} X_T.$$

Il en résulte que, pour tout $j \in \overline{\mathbb{N}}$, on a dans tous les cas, sur $(T = j)$, $E^{\mathcal{A}_{T \wedge n}} X_T = E^{\mathcal{A}_n} X_T$. On a donc, pour tout $n \in \mathbb{N}$,

$$X_n^T = X_{T \wedge n} = E^{\mathcal{A}_n} X_T. \quad (15.28)$$

Puisque $X_T = E^{\mathcal{A}_T} Y$, X_T est intégrable et la martingale arrêtée X^T est fermable. Appliquons lui le deuxième théorème d'arrêt avec le temps d'arrêt S ; il vient alors d'après (15.28)

$$X_S^T = X_{T \wedge S} = E^{\mathcal{A}_S} X_T = E^{\mathcal{A}_S} [E^{\mathcal{A}_T} Y],$$

et en particulier $X_{T \wedge S} = E^{\mathcal{A}_S} [E^{\mathcal{A}_T} Y]$. En échangeant les rôles de S et T , on a aussi $X_{T \wedge S} = E^{\mathcal{A}_T} [E^{\mathcal{A}_S} Y]$, ce qui implique l'égalité $E^{\mathcal{A}_S} [E^{\mathcal{A}_T} Y] = E^{\mathcal{A}_T} [E^{\mathcal{A}_S} Y]$. Enfin, toujours par le deuxième théorème d'arrêt appliqué avec le temps d'arrêt $S \wedge T$ et la martingale fermée $(X_n)_{n \in \overline{\mathbb{N}}}$, on a $X_{T \wedge S} = E^{\mathcal{A}_{S \wedge T}} Y$, ce qui démontre la dernière égalité.

Exercice 15.2. Problème de la ruine du joueur. Un joueur joue à **pile ou face** avec une pièce non nécessairement équilibrée : on note p la probabilité d'obtenir pile lors d'un jet. Il reçoit un euro de la banque s'il obtient pile et en donne un à la banque s'il obtient face. Sa fortune initiale est de $a \in \mathbb{N}^*$ euros et celle de la banque de $b \in \mathbb{N}^*$ euros. Le joueur joue jusqu'à sa ruine ou celle de la banque. On modélise ce jeu de la manière suivante : $(Y_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires définies sur un espace probabilisé (Ω, \mathcal{A}, P) , indépendantes de même loi $p\delta_1 + q\delta_{-1}$, où $q = 1 - p$. On note S_n la fortune du joueur après n parties, pour un jeu qui ne s'arrêterait pas ; on pose

$$S_0 = a \quad \text{et} \quad S_n = a + \sum_{j=1}^n Y_j.$$

En posant $Y_0 = a$, les filtrations naturelles $(\mathcal{A}_n)_{n \in \mathbb{N}}$ des processus Y et S sont les mêmes. On note T le temps d'arrêt du jeu, c'est-à-dire

$$T = \inf(n \in \mathbb{N}^* \mid S_n = 0 \text{ ou } a + b).$$

On se pose les trois questions : quelle est la probabilité $P(T < +\infty)$ que le jeu s'arrête, quelle est la probabilité $\rho = P(S_T = a + b)$ que le joueur gagne, quel est le temps moyen ET d'arrêt du jeu.

- Déterminer la nature du processus $S = (S_n)_{n \in \mathbb{N}}$ suivant les valeurs de p .
- Étude du cas $p \neq q$** ; on supposera que $p > q$. Écrire la décomposition de Doob de la sous-martingale S et préciser son processus croissant prévisible A . En déduire que $ET < +\infty$; préciser alors la valeur de $P(T < +\infty)$ et donner une expression de ET en fonction de ρ . On définit, pour $s > 0$, le processus U par, pour tout $n \in \mathbb{N}$, $U_n = s^{S_n}$. Déterminer s pour que U soit une martingale non constante ; vérifier qu'alors la martingale arrêtée U^T converge P-p.s. et dans L^1 vers U_T . En déduire les valeurs de ρ puis ET .
- Étude du cas $p = \frac{1}{2}$** . Vérifier que S est une martingale de carré intégrable et déterminer son processus croissant prévisible B . En déduire que $ET < +\infty$; préciser alors la valeur de $P(T < +\infty)$. Vérifier que la martingale arrêtée S^T converge P-p.s., dans L^1 et L^2 vers S_T . En déduire les valeurs de ES_T , ρ et ET .

Solution.

1. On a, si $n \in \mathbb{N}^*$, $E^{\mathcal{A}_{n-1}} \Delta S_n = E^{\mathcal{A}_{n-1}} Y_n$; les variables aléatoires Y_n étant indépendantes, il vient

$$E^{\mathcal{A}_{n-1}} \Delta S_n = EY_n = p - q,$$

ce qui donne la classification suivante pour le processus S :

$$S \text{ est une } \begin{cases} \text{sous-martingale} & \text{si } p > q. \\ \text{martingale} & \text{si } p = q = \frac{1}{2}, \\ \text{surmartingale} & \text{si } p < q. \end{cases}$$

2. Cas $p > q$. La sous-martingale S admet la décomposition de Doob $S = M + A$ où le processus croissant prévisible A est défini par $A_0 = 0$ et $\Delta A_n = E^{\mathcal{A}_{n-1}} \Delta S_n$, ce qui donne

$$A_0 = 0 \quad \text{et} \quad \text{si } n \geq 1, \quad A_n = n(p - q).$$

Le premier théorème d'arrêt appliqué à la martingale $M = (S_n - n(p - q))_{n \in \mathbb{N}}$ et au temps d'arrêt borné $T \wedge n$ donne alors

$$a = ES_0 = E[S_{T \wedge n} - T \wedge n(p - q)],$$

d'où

$$(p - q)E[T \wedge n] = ES_{T \wedge n} - a; \quad (15.29)$$

mais, puisque par définition de T , pour tout $n \in \mathbb{N}$, $0 \leq S_{T \wedge n} \leq a + b$, on a

$$0 \leq (p - q)E[T \wedge n] \leq b.$$

Par la propriété de Beppo Levi, on a $ET = \lim_n \nearrow E[T \wedge n]$. Il en résulte alors que T est intégrable et en particulier que $P(T < +\infty) = 1$. La suite $(S_{T \wedge n})_{n \in \mathbb{N}}$ converge

alors P-p.s. vers S_T ; en passant à la limite dans (15.29), on obtient par convergence dominée

$$(p - q)E[T] = ES_T - a. \quad (15.30)$$

Puisque, par définition de T , on a

$$ES_T = (a + b)P(S_T = a + b),$$

il vient

$$ET = \frac{(a + b)\rho - a}{p - q}. \quad (15.31)$$

Soit $s > 0$; puisque $s^{S_{n-1}}$ est \mathcal{A}_{n-1} -mesurable et indépendante de s^{Y_n} , on a, pour tout $n \in \mathbb{N}^*$,

$$E^{s^{\mathcal{A}_{n-1}}} U_n = E^{s^{\mathcal{A}_{n-1}}} [s^{S_{n-1}} s^{Y_n}] = s^{S_{n-1}} E^{s^{\mathcal{A}_{n-1}}} [s^{Y_n}] = s^{S_{n-1}} E[s^{Y_n}],$$

soit

$$E^{s^{\mathcal{A}_{n-1}}} U_n = U_{n-1} \left(sp + \frac{1}{s} q \right).$$

On a $sp + \frac{1}{s}q = 1$ si et seulement si $s^2 p - s + q = 0$, équation dont une racine évidente est 1 (puisque $p + q = 1$) et l'autre est q/p . Ainsi **pour** $s = q/p$, **U est une martingale non constante**. Par définition de T et puisque $q/p < 1$, on a, pour tout $n \in \mathbb{N}$, $0 \leq U_{T \wedge n} \leq 1$: la martingale arrêtée U^T est donc équi-intégrable et converge P-p.s. et dans L^1 vers U_T .

Par définition de T , U_T prend P-p.s. les valeurs 1 ou $(\frac{q}{p})^{a+b}$; sa moyenne vaut donc

$$EU_T = P(S_T = 0) + \left(\frac{q}{p}\right)^{a+b} P(S_T = a + b).$$

soit

$$EU_T = 1 - \rho + \left(\frac{q}{p}\right)^{a+b} \rho. \quad (15.32)$$

Par ailleurs, d'après le premier théorème d'arrêt, on a, pour tout $n \in \mathbb{N}$,

$$EU_{T \wedge n} = EU_0 = \left(\frac{q}{p}\right)^a,$$

ce qui donne, par convergence dominée,

$$EU_T = \lim_n EU_{T \wedge n} = \left(\frac{q}{p}\right)^a;$$

en reportant dans les égalités (15.32) puis (15.31), il vient

$$\rho = \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^{a+b}} \quad \text{et} \quad ET = \frac{1}{p - q} \left[\frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^{a+b}} - a \right].$$

3. Cas $p = q = 1/2$. Le processus S est alors une martingale dans L^2 (puisque les Y_n sont bornées). Son processus croissant prévisible B est défini par $B_0 = 0$ et ses accroissements donnés, si $n \geq 1$, par $\Delta B_n = E^{\mathcal{A}_{n-1}}(\Delta S_n)^2$, soit, par indépendance de \mathcal{A}_{n-1} et Y_n^2 ,

$$\Delta B_n = E^{\mathcal{A}_{n-1}} Y_n^2 = E Y_n^2 = 1.$$

On a donc

$$B_0 = 0 \quad \text{et, si } n \geq 1, \quad B_n = n;$$

Le premier théorème d'arrêt appliqué à la martingale $(S_n^2 - n)_{n \in \mathbb{N}}$ et au temps d'arrêt borné $T \wedge n$ donne alors, puisque $S_0 = a$,

$$ES_0^2 = E[S_{T \wedge n}^2 - T \wedge n] = a^2; \quad (15.33)$$

puisque $S_{T \wedge n}^2 \leq (a+b)^2$, on alors

$$ES_{T \wedge n}^2 = E[T \wedge n] \leq (a+b)^2 - a^2,$$

ce qui, par la propriété de Beppo Levi, donne

$$ET = \lim_n E[T \wedge n] \leq (a+b)^2 - a^2.$$

Il en résulte que T est intégrable et en particulier que $P(T < +\infty) = 1$. La suite $(S_{T \wedge n})_{n \in \mathbb{N}}$ converge alors P.-p.s. vers S_T ; puisque l'on a, pour tout $n \in \mathbb{N}^*$, $0 \leq S_{T \wedge n} \leq a+b$, le théorème de convergence dominée montre qu'il y a aussi convergence L^1 et L^2 vers S_T . Le premier théorème d'arrêt appliqué à la martingale S et au temps d'arrêt borné $T \wedge n$ donne alors

$$ES_{T \wedge n} = ES_0 = a,$$

et donc, par convergence L^1 ,

$$\boxed{ES_T = a};$$

Puisque, par définition de T ,

$$ES_T = (a+b)P(S_T = a+b) = \rho(a+b),$$

on a

$$\boxed{\rho = \frac{a}{a+b}}.$$

La relation (15.33) donne l'égalité, pour tout $n \in \mathbb{N}^*$,

$$E[T \wedge n] = E[S_{T \wedge n}^2] - a^2, \quad (15.34)$$

et, puisque la suite $(S_{T \wedge n})_{n \in \mathbb{N}}$ converge vers S_T dans L^2 , un passage à la limite dans (15.34) donne

$$E[T] = E[S_T^2] - a^2.$$

Puisque, par définition de T , on a

$$ES_T^2 = (a+b)^2 P(S_T = a+b) = \rho(a+b)^2 = a(a+b),$$

on obtient finalement

$$\boxed{ET = ab}.$$

Exercice 15.3. Jeu de pile ou face avec deux pièces non équilibrées et stratégie d'apprentissage. Un joueur dispose de deux pièces A et B ; la probabilité, inconnue du joueur, d'obtenir pile lors d'un jet avec la pièce A (resp. B) est p^A (resp. p^B). Le joueur gagne une unité à chaque fois qu'il obtient pile. À l'issue de chaque tirage, il choisit la pièce avec laquelle il va faire le tirage suivant en tenant compte des résultats des tirages antérieurs. La stratégie du joueur est de cerner la pièce qui a le plus grand p pour pouvoir « maximiser » son gain. On modélise ce jeu de manière précise comme suit.

On se donne une famille $\{(X_n^A)_{n \in \mathbb{N}}, (X_n^B)_{n \in \mathbb{N}}\}$ de variables aléatoires définies sur un espace probabilisé (Ω, \mathcal{A}, P) , indépendantes. Pour tout $n \in \mathbb{N}$, X_n^A (resp. X_n^B) suit la loi de Bernoulli de paramètre p^A (resp. p^B). On note \mathcal{A}_n la tribu $\sigma(X_j^A, X_j^B \mid 0 \leq j \leq n)$ et $U = (U_n)_{n \in \mathbb{N}}$ un processus adapté à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$ à valeurs dans l'ensemble $\{A, B\}$. On considère le processus $(X_n)_{n \in \mathbb{N}}$ des résultats à chaque tirage ; il vérifie, pour tout $n \in \mathbb{N}$,

$$X_{n+1} = \mathbf{1}_{(U_n=A)} X_{n+1}^A + \mathbf{1}_{(U_n=B)} X_{n+1}^B = X_{n+1}^{U_n}.$$

On pose

$$G_0 = 0 \quad \text{et, si } n \geq 1, \quad G_n = \sum_{j=1}^n X_j.$$

1. Calculer l'espérance conditionnelle $E^{\mathcal{A}_n} X_{n+1}$.

2. On définit le processus M par $M_0 = 0$ et, si $n \in \mathbb{N}^*$, $M_n = \sum_{j=1}^n (X_j - p^{U_{j-1}})$. Vérifier que c'est une martingale de carré intégrable et calculer son processus croissant prévisible. En déduire la convergence P-p.s. vers 0 de la suite de terme général $\frac{G_n}{n} - \frac{1}{n} \sum_{j=1}^n p^{U_{j-1}}$.

3. Pour $J \in \{A, B\}$, on définit les processus N^J, M^J et \tilde{p}^J par

$$N_n^J = \sum_{j=0}^n \mathbf{1}_{(U_j=J)}, \quad M_0^J = 0, \quad \text{et } M_n^J = \sum_{j=1}^n \left[\mathbf{1}_{(U_{j-1}=J, X_j=1)} - \mathbf{1}_{(U_{j-1}=J)} p^J \right],$$

et, si $n \in \mathbb{N}^*$,

$$\tilde{p}_n^J = \frac{1}{N_{n-1}^J} \sum_{j=0}^{n-1} \mathbf{1}_{(U_j=J, X_{j+1}=1)} \quad \text{avec la convention } \frac{0}{0} = 0.$$

Démontrer que M^J est une martingale de carré intégrable et calculer son processus croissant prévisible $\langle M^J \rangle$. En déduire que, sur l'ensemble $\{N_n^J \rightarrow +\infty\}$, la suite de terme général \tilde{p}_n^J converge P-p.s. vers p^J .

4. Soit une suite strictement croissante d'entiers positifs $(v_n)_{n \in \mathbb{N}}$ telle que $\frac{v_n}{n} \rightarrow +\infty$. Le joueur adopte le processus de choix $U = (U_n)_{n \in \mathbb{N}}$ comme suit :

$$\text{si } n \notin \{v_j \mid j \in \mathbb{N}\} \quad U_n = A \mathbf{1}_{(\tilde{p}_n^A \geq \tilde{p}_n^B)} + B \mathbf{1}_{(\tilde{p}_n^A < \tilde{p}_n^B)}, \\ \text{et } U_{v_{2n}} = A, \quad U_{v_{2n+1}} = B;$$

autrement dit, il choisit la pièce qui est apparue le plus fréquemment lorsque n n'est pas dans le support de la suite v , et choisit alternativement les pièces A et B le long de la suite v . Pour ce choix, on a, pour $J \in \{A, B\}$, $N_n^J \rightarrow +\infty$. On suppose par exemple que $p^A > p^B$ (ce qui, rappelons le, est inconnu du joueur).

Étudier, pour $J \in \{A, B\}$, la convergence P-p.s. des suites de terme général $\frac{N_n^J}{n}$.
Étudier alors, pour ce choix, la convergence P-p.s. de la suite de terme général $\frac{U_n}{n}$.

Solution.

1. Le processus U étant adapté, on a

$$E^{\mathcal{A}_n} X_{n+1} = \mathbf{1}_{(U_n=A)} E^{\mathcal{A}_n} X_{n+1}^A + \mathbf{1}_{(U_n=B)} E^{\mathcal{A}_n} X_{n+1}^B,$$

soit, par indépendance des X_{n+1}^J , pour $J \in \{A, B\}$, et de \mathcal{A}_n ,

$$E^{\mathcal{A}_n} X_{n+1} = \mathbf{1}_{(U_n=A)} E X_{n+1}^A + \mathbf{1}_{(U_n=B)} E X_{n+1}^B = \mathbf{1}_{(U_n=A)} p^A + \mathbf{1}_{(U_n=B)} p^B,$$

soit encore

$$E^{\mathcal{A}_n} X_{n+1} = p^{U_n}.$$

2. Il en résulte que

$$E^{\mathcal{A}_{n-1}} [\Delta M_n] = E^{\mathcal{A}_{n-1}} [X_n - p^{U_{n-1}}] = 0,$$

c'est-à-dire que M est une martingale, de carré intégrable, puisque on a $0 \leq X_n \leq 1$ et $0 \leq p^{U_{n-1}} \leq 1$. Son processus croissant prévisible $\langle M \rangle$ est défini par $\langle M \rangle_0 = 0$ et ses accroissements sont donnés, si $n \geq 1$, par

$$\begin{aligned} \Delta \langle M \rangle_n &= E^{\mathcal{A}_{n-1}} (X_n - p^{U_{n-1}})^2 \\ &= \mathbf{1}_{(U_{n-1}=A)} E^{\mathcal{A}_{n-1}} (X_n^A - p^A)^2 + \mathbf{1}_{(U_{n-1}=B)} E^{\mathcal{A}_{n-1}} (X_n^B - p^B)^2 \end{aligned}$$

soit, puisque U_{n-1} est \mathcal{A}_{n-1} -mesurable et que \mathcal{A}_{n-1} et X_n^J sont indépendantes,

$$\begin{aligned} \Delta \langle M \rangle_n &= \mathbf{1}_{(U_{n-1}=A)} E (X_n^A - p^A)^2 + \mathbf{1}_{(U_{n-1}=B)} E (X_n^B - p^B)^2 \\ &= \mathbf{1}_{(U_{n-1}=A)} p^A (1 - p^A) + \mathbf{1}_{(U_{n-1}=B)} p^B (1 - p^B). \end{aligned}$$

ce qui donne

$$\langle M \rangle_n = \sum_{j=1}^n p^{U_{j-1}} (1 - p^{U_{j-1}}).$$

Soit

$$m = \min(p^A(1 - p^A), p^B(1 - p^B)) \quad \text{et} \quad s = \max(p^A(1 - p^A), p^B(1 - p^B));$$

on a alors, pour tout $n \in \mathbb{N}^*$,

$$0 \leq n m \leq \langle M \rangle_n \leq n s.$$

si bien que la suite de terme général $\langle M \rangle_n$ tend vers $+\infty$. Il résulte alors de la loi forte des grands nombres pour les martingales L^2 que

$$\frac{M_n}{\langle M \rangle_n} \xrightarrow{P\text{-p.s.}} 0,$$

et donc aussi, par l'encadrement ci-dessus,

$$\frac{M_n}{n} \xrightarrow{P\text{-p.s.}} 0,$$

ce qui démontre que

$$\boxed{\frac{G_n}{n} - \frac{1}{n} \sum_{j=1}^n p^{U_{j-1}} \xrightarrow{P\text{-p.s.}} 0.} \quad (15.35)$$

3. Soit J fixé. Le processus U étant adapté, on a

$$E^{\mathcal{A}_{n-1}} [\Delta M_n^J] = \mathbf{1}_{(U_{n-1}=J)} E^{\mathcal{A}_{n-1}} [\mathbf{1}_{(X_n^J=1)} - p^J],$$

soit, par indépendance

$$E^{\mathcal{A}_{n-1}} [\Delta M_n^J] = \mathbf{1}_{(U_{n-1}=J)} E [\mathbf{1}_{(X_n^J=1)} - p^J] = 0;$$

le processus M^J est donc une martingale, visiblement de carré intégrable. Son processus croissant prévisible $\langle M^J \rangle$ est défini par $\langle M^J \rangle_0 = 0$ et ses accroissements sont donnés, si $n \geq 1$, par

$$\Delta \langle M^J \rangle_n = E^{\mathcal{A}_{n-1}} (\Delta M_n^J)^2 = E^{\mathcal{A}_{n-1}} [\mathbf{1}_{(U_{n-1}=J)} (\mathbf{1}_{(X_n^J=1)} - p^J)^2],$$

soit, puisque U_{n-1} est \mathcal{A}_{n-1} -mesurable et que \mathcal{A}_{n-1} et X_n^J sont indépendantes,

$$\Delta \langle M \rangle_n = \mathbf{1}_{(U_{n-1}=J)} E [(\mathbf{1}_{(X_n^J=1)} - p^J)^2] = \mathbf{1}_{(U_{n-1}=J)} p^J(1 - p^J),$$

ce qui donne, pour tout $n \in \mathbb{N}^*$,

$$\boxed{\langle M \rangle_n = \mathbf{N}_{n-1}^J p^J(1 - p^J).}$$

Il résulte alors de la loi forte des grands nombres pour les martingales L^2 que, sur l'ensemble $\{\mathbf{N}_n^J \rightarrow +\infty\}$, on a

$$\frac{M_n^J}{\mathbf{N}_n^J} \xrightarrow{P\text{-p.s.}} 0;$$

mais on a $\widetilde{p}_n^J = \frac{M_n^J}{\mathbf{N}_{n-1}^J} + p^J$, ce qui démontre que

$$\boxed{\text{sur l'ensemble } \{\mathbf{N}_n^J \rightarrow +\infty\}, \quad \widetilde{p}_n^J \xrightarrow{P\text{-p.s.}} p^J.}$$

4. Pour ce choix, on a, pour $J \in \{A, B\}$, $N_n^J \rightarrow +\infty$. Supposons par exemple que $p^A > p^B$. Soit ω tel que la suite de terme général $\widetilde{p}_n^J(\omega)$ converge vers p^J pour $J = A$ et B . Il existe un entier $N(\omega)$ tel que $\widetilde{p}_n^A(\omega) \geq \widetilde{p}_n^B(\omega)$ dès que $n \geq N(\omega)$; on a donc $\lim_n \frac{1}{n} \sum_{j=0}^n \mathbf{1}_{(\widetilde{p}_j^A \geq \widetilde{p}_j^B)}(\omega) = 1$. De plus, par définition du processus de choix U , on a, pour $n > N(\omega)$,

$$N_n^A(\omega) - N_{N(\omega)+1}^A(\omega) - \sum_{j=N(\omega)+1}^n \mathbf{1}_{(\widetilde{p}_j^A \geq \widetilde{p}_j^B)}(\omega) = \sum_{j=1}^{+\infty} \mathbf{1}_{(N(\omega) < v_{2j} \leq n)} :$$

mais puisque $v_n/n \rightarrow +\infty$, il existe un entier N' tel que $v_n > n$ dès que $n \geq N'$ et on a donc

$$L_n \equiv \sum_{j=1}^n \mathbf{1}_{(v_{2j} \leq n)} \leq \text{card} \{j \mid v_{2j} \leq n\} ,$$

ce qui implique que $L_n/n \rightarrow 0$. Il en résulte que $N_n^A(\omega)/n \rightarrow 1$. Par un raisonnement identique, on obtiendrait que $N_n^B(\omega)/n \rightarrow 0$. On a ainsi démontré que

$$\boxed{\frac{N_n^A(\omega)}{n} \xrightarrow{P\text{-p.s.}} 1 \quad \text{et} \quad \frac{N_n^B(\omega)}{n} \xrightarrow{P\text{-p.s.}} 0.} \quad (15.36)$$

Puisque l'on a

$$\sum_{j=1}^n p^{U_{j-1}} = p^A \sum_{i=0}^{n-1} \mathbf{1}_{(U_i=A)} + p^B \sum_{i=0}^{n-1} \mathbf{1}_{(U_i=B)} = p^A N_{n-1}^A + p^B N_{n-1}^B ,$$

il résulte de (15.35) et de (15.36) que

$$\boxed{\frac{G_n}{n} \xrightarrow{P\text{-p.s.}} p^A.}$$

Exercice 15.4. Jeu de pile ou face avec gain dépendant de deux jets consécutifs.

Soit $(Y_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur un espace probabilisé (Ω, \mathcal{A}, P) , indépendantes de même loi $p\delta_1 + q\delta_{-1}$, où $q = 1 - p$. On note S_n la fortune d'un joueur après n parties de pile ou face; on suppose que la règle de gain est telle que

$$S_0 = a \quad \text{et} \quad S_n = a + \sum_{i=1}^n Y_{j-1} Y_j .$$

Les processus considérés seront tous relatifs à la filtration naturelle $(\mathcal{A}_n)_{n \in \mathbb{N}}$ du processus Y .

1. Calculer la probabilité $P(S_n > S_{n-1})$ et vérifier qu'elle est strictement supérieure à $\frac{1}{2}$ si $p \neq q$.
2. Calculer, pour $n \in \mathbb{N}^*$, l'espérance conditionnelle $E^{\mathcal{A}_{n-1}} S_n$. Quelle est la nature du processus S lorsque $p = \frac{1}{2}$?

Étudier la convergence de la suite de terme général ES_n .

3. Soit $s > 0$ quelconque. Calculer, pour $n \in \mathbb{N}^*$, l'espérance conditionnelle $E^{\mathcal{A}_{n-1}}(s^{S_n})$. On pose $u = s + \frac{1}{s}$; démontrer que le processus $(\frac{s^{S_n}}{u^n})_{n \in \mathbb{N}}$ est une surmartingale positive. Étudier les convergences P-p.s. et L^1 de la suite $(\frac{s^{S_n}}{u^n})_{n \in \mathbb{N}}$.

4. Démontrer que S s'écrit de manière unique comme somme d'une martingale de carré intégrable W et d'un processus intégrable prévisible T tel que $T_0 = 0$. Calculer le processus croissant prévisible $\langle W \rangle$ de la martingale W . Étudier la convergence P-p.s. de la suite $(\frac{S_n}{n})_{n \in \mathbb{N}^*}$ et en déduire, dans le cas où $p \neq q$, celle de la suite $(S_n)_{n \in \mathbb{N}}$.

Solution.

1. Puisque $(S_n > S_{n-1}) = (Y_{n-1}Y_n = 1)$ et que les Y_n prennent P-p.s. les valeurs ± 1 , on a

$$P(S_n > S_{n-1}) = P[(Y_{n-1} = 1) \cap (Y_n = 1)] + P[(Y_{n-1} = -1) \cap (Y_n = -1)],$$

soit, par indépendance de Y_{n-1} et Y_n ,

$$P(S_n > S_{n-1}) = p^2 + q^2.$$

Tenant compte de l'égalité $p + q = 1$, on a

$$p^2 + q^2 = 2p^2 - 2p + 1 \equiv h(p);$$

on a alors $h'(p) = 2(2p - 1)$ et $h''(p) = 4$, d'où il résulte que h admet un minimum en $1/2$ et que $h(1/2) = 1/2$, ce qui démontre que, si $p \neq q$, $P(S_n > S_{n-1}) > 1/2$.

2. Pour $n \in \mathbb{N}^*$, on a $S_n = S_{n-1} + Y_{n-1}Y_n$; les variables aléatoires S_{n-1} et Y_{n-1} étant \mathcal{A}_{n-1} -mesurables, on a

$$E^{\mathcal{A}_{n-1}} S_n = S_{n-1} + Y_{n-1} E^{\mathcal{A}_{n-1}} Y_n.$$

Les tribus \mathcal{A}_{n-1} et $\sigma(Y_n)$ étant indépendantes, on a $E^{\mathcal{A}_{n-1}} Y_n = EY_n$, si bien que

$$E^{\mathcal{A}_{n-1}} S_n = S_{n-1} + (p - q)Y_{n-1}. \quad (15.37)$$

En particulier, si $p = q$, S est une martingale.

Il résulte de (15.37) que

$$E(\Delta S_n) = E \left[E^{\mathcal{A}_{n-1}} \Delta S_n \right] = (p - q) EY_{n-1} = (p - q)^2.$$

Il en résulte que

$$E(S_n) = a + n(p - q)^2.$$

Ainsi, si $p \neq q$, on a $\lim_n E(S_n) = +\infty$. La suite est stationnaire si $p = q$.

3. Pour $n \in \mathbb{N}^*$, on a $s^{S_n} = s^{S_{n-1}} s^{Y_{n-1}Y_n}$; la variable aléatoire $s^{S_{n-1}}$ étant \mathcal{A}_{n-1} -mesurable, on a

$$E^{\mathcal{A}_{n-1}} s^{S_n} = s^{S_{n-1}} E^{\mathcal{A}_{n-1}} \left[s^{Y_{n-1}Y_n} \right].$$

Mais, pour tout $(y_0, y_1, \dots, y_{n-1}) \in \{-1, 1\}^n$, on a les relations sur les moyennes conditionnelles

$$m^{(Y_0, Y_1, \dots, Y_{n-1})=(y_0, y_1, \dots, y_{n-1})} ({}_s Y_{n-1} Y_n) = m^{(Y_0, Y_1, \dots, Y_{n-1})=(y_0, y_1, \dots, y_{n-1})} ({}_s y_{n-1} Y_n),$$

ce qui donne, par indépendance de $(Y_0, Y_1, \dots, Y_{n-2})$ et Y_{n-1} ,

$$m^{(Y_0, Y_1, \dots, Y_{n-1})=(y_0, y_1, \dots, y_{n-1})} ({}_s Y_{n-1} Y_n) = E({}_s y_{n-1} Y_n) = p s^{y_{n-1}} + q s^{-y_{n-1}};$$

la tribu \mathcal{A}_{n-1} étant engendrée par $(Y_0, Y_1, \dots, Y_{n-1})$, on a alors $E^{\mathcal{A}_{n-1}} [{}_s Y_{n-1} Y_n] = p s^{Y_{n-1}} + q s^{-Y_{n-1}}$ et donc

$$E^{\mathcal{A}_{n-1}} {}_s S_n = s^{S_{n-1}} [p s^{Y_{n-1}} + q s^{-Y_{n-1}}].$$

Il en résulte que l'on a

$$E^{\mathcal{A}_{n-1}} {}_s S_n = s^{S_{n-1}} \left[\mathbf{1}_{(Y_{n-1}=1)} (ps + \frac{q}{s}) + \mathbf{1}_{(Y_{n-1}=-1)} (\frac{p}{s} + qs) \right],$$

et donc

$$E^{\mathcal{A}_{n-1}} {}_s S_n \leq s^{S_{n-1}} \left(s + \frac{1}{s} \right),$$

soit, en posant $u = s + 1/s$ et en divisant les membres extrêmes par u^n ,

$$E^{\mathcal{A}_{n-1}} \left[\frac{{}_s S_n}{u^n} \right] \leq \frac{s^{S_{n-1}}}{u^{n-1}};$$

autrement dit, **le processus $(s^{S_n}/u^n)_{n \in \mathbb{N}^*}$, est une surmartingale positive.** Ainsi, on sait déjà qu'elle converge P-p.s. En fait, on va voir que cette suite est bornée. En effet, pour tout $n \in \mathbb{N}^*$, on a P-p.s. $a - n \leq S_n \leq a + n$

- si $0 < s < 1$ on a alors $0 \leq s^{S_n} \leq s^{a-n}$ et donc

$$0 \leq \frac{s^{S_n}}{u^n} \leq \frac{s^a}{(us)^n} \leq s^a,$$

puisque $us = 1 + s^2 > 1$;

- si $s \geq 1$, on a de même $0 \leq s^{S_n} \leq s^{a+n}$ et donc

$$0 \leq \frac{s^{S_n}}{u^n} \leq s^a \left(\frac{s}{u} \right)^n \leq s^a,$$

puisque $\frac{u}{s} = 1 + \frac{1}{s^2} > 1$;

Au total, on voit que, pour tout $s > 0$, $\lim_{n \rightarrow \infty} \frac{s^{S_n}}{u^n} = 0$; de plus, on a, pour tout $s > 0$, et tout $n \in \mathbb{N}^*$,

$$0 \leq \frac{s^{S_n}}{u^n} \leq s^a,$$

ce qui implique que la suite est équi-intégrable et qu'elle converge donc aussi vers 0 dans L^1 .

4. Soit W le processus défini par $W_0 = S_0 = a$ et ses accroissements donnés, si $n \geq 1$, par

$$\Delta W_n = \Delta S_n - E^{\mathcal{A}_{n-1}}(\Delta S_n) = Y_{n-1}Y_n - (p-q)Y_{n-1}.$$

soit encore

$$W_0 = a \quad \Delta W_n = Y_{n-1} [Y_n - (p-q)] :$$

par construction, W est une martingale : elle est dans L^2 , puisque les ΔW_n sont P-p.s. bornées. Soit T le processus défini par $T_0 = 0$ et ses accroissements donnés, si $n \geq 1$, par

$$\Delta T_n = E^{\mathcal{A}_{n-1}}(\Delta S_n) = (p-q)Y_{n-1} ;$$

on a bien $S = W + T$ et T est un processus intégrable prévisible. Si $S = W' + T'$ est une autre décomposition du même type, on a, pour tout $n \in \mathbb{N}$, $\Delta(W - W')_n = \Delta(T' - T)_n$. En prenant l'espérance conditionnelle des deux membres et en tenant compte des propriétés de ces processus, il vient

$$0 = E^{\mathcal{A}_{n-1}} [\Delta(W - W')_n] = \Delta(T' - T)_n.$$

ce qui démontre que $T = T'$, puis que $W = W'$. Il y a donc unicité d'une telle décomposition.

Le processus croissant prévisible de W , $\langle W \rangle$, est déterminé par ses accroissements ; puisque $Y_{n-1}^2 = 1$ P-p.s., ils sont donnés, pour tout $n \in \mathbb{N}$, par

$$\Delta \langle W \rangle_n = E^{\mathcal{A}_{n-1}}(\Delta W_n)^2 = E^{\mathcal{A}_{n-1}} [Y_n - (p-q)]^2.$$

soit, par indépendance de \mathcal{A}_{n-1} et $\sigma(Y_n)$,

$$\Delta \langle W \rangle_n = E [Y_n - (p-q)]^2 = \sigma_{Y_n}^2 = EY_n^2 - (EY_n)^2 = 1 - (p-q)^2 = 4pq.$$

Il en résulte que

$$\langle W \rangle_n = 4pq n.$$

La loi forte des grands nombres pour les martingales dans L^2 assure alors que

$\frac{W_n}{\langle W \rangle_n} \xrightarrow{\text{P-p.s.}} 0$. Par ailleurs, on a

$$\frac{S_n}{n} = \frac{W_n}{n} + \frac{T_n}{n} = \frac{W_n}{n} + (p-q) \frac{1}{n} \sum_{j=1}^n Y_{j-1} ;$$

la loi forte des grands nombres pour les suites de variables aléatoires indépendantes, de carré intégrable, et de même loi, donne

$$\frac{1}{n} \sum_{j=1}^n Y_{j-1} \xrightarrow{\text{P-p.s.}} EY_0.$$

Puisque $EY_0 = p - q$, il vient

$$\frac{S_n}{n} \xrightarrow{\text{P-p.s.}} (p-q)^2.$$

Si $p \neq q$, cette limite est strictement positive et il en résulte que la suite de terme général S_n tend vers $+\infty$ P-p.s.

Exercice 15.5. Un modèle de portefeuille d'actions. Le processus $S = (S_n)_{n \in \mathbb{N}}$ d'évolution temporelle d'une action boursière peut être modélisé par la donnée d'un espace probabilisé (Ω, \mathcal{A}, P) sur lequel la suite de variables aléatoires $(S_n)_{n \in \mathbb{N}}$ est définie par

$$S_0 = s_0 > 0 \quad \text{et, si } n \in \mathbb{N}^*, \quad S_n = (1 + \mu)S_{n-1} + \sigma S_{n-1} \varepsilon_n.$$

où $(\varepsilon_n)_{n \in \mathbb{N}^*}$ est un processus de **bruit**, c'est-à-dire ici une suite de variables aléatoires indépendantes de même loi $(\delta_1 + \delta_{-1})/2$, et où les paramètres réels μ et σ (appelés respectivement taux d'actualisation et coefficient de volatilité) satisfont à l'inégalité $|\sigma| < 1 + \mu$. On note λ le réel $\lambda = [(1 + \mu)^2 - \sigma^2]^{\frac{1}{2}}$. On note $(\mathcal{A}_n)_{n \in \mathbb{N}}$ la filtration naturelle du processus S .

1. Soit f la fonction réelle définie sur \mathbb{R}^2 par $f(x, y) = (1 + \mu)x + \sigma xy$. Démontrer qu'elle vérifie les inégalités, pour tout $x \in \mathbb{R}^+$, $f(x, 1) \geq 0$ et $f(x, -1) \geq 0$.

2. Calculer, pour tout $n \in \mathbb{N}^*$, $E^{\mathcal{A}_{n-1}}(\Delta S_n)$. En déduire la nature du processus S , suivant les valeurs des paramètres μ et σ . Dans le cas où $\mu < 0$, démontrer que la suite de terme général S_n converge P-p.s. vers une limite que l'on déterminera.

3. Vérifier que S_n est de carré intégrable et calculer ES_n^2 .

4. On définit le processus Z par, pour tout $n \in \mathbb{N}$, $Z_n = \ln S_n$. Démontrer que Z est, suivant les valeurs de λ , une martingale, une sous ou sur martingale.

Écrire Z_n sous forme d'une somme de variables aléatoires indépendantes et en déduire, suivant les valeurs de λ , la convergence P-p.s. (dans \mathbb{R}) de la suite $(Z_n)_{n \in \mathbb{N}}$ vers une limite à préciser. Comment se traduisent ces résultats pour la suite $(S_n)_{n \in \mathbb{N}}$?

5. Dans le cas particulier où $|\sigma| < \frac{1}{\sqrt{2}}$ et $(1 + \mu)^2 + \sigma^2 < 1$, les résultats précédents prouvent que $-S$ est une sous-martingale (le vérifier). Écrire alors sa décomposition de Doob $-S = M + A$, où M est une martingale intégrable et A un processus croissant prévisible nul en 0.

Vérifier que M est une martingale dans L^2 et calculer son processus croissant prévisible $\langle M \rangle$. En déduire la convergence P-p.s. de la série de terme général S_n^2 .

6. On définit le processus W par, pour tout $n \in \mathbb{N}$, $W_n = \ln\left(\frac{S_n}{\lambda^n}\right)$. Démontrer que W est une martingale dans L^2 et calculer son processus croissant prévisible $\langle W \rangle$ en fonction de $\delta = -\left[\ln\left(\frac{1+\mu+\sigma}{\lambda}\right)\right]\left[\ln\left(\frac{1+\mu-\sigma}{\lambda}\right)\right]$; vérifier que ce nombre est strictement positif.

En déduire la convergence P-p.s. de la suite $(S_n^{1/n})_{n \in \mathbb{N}^*}$ vers une limite à préciser.

7. On définit le processus R par, pour tout $n \in \mathbb{N}$, $R_n = \lambda^{-\sqrt{n}} S_n^{1/\sqrt{n}}$. Démontrer que la suite des lois P_{R_n} converge étroitement vers une loi à densité par rapport à la mesure de Lebesgue : déterminer cette loi.

Solution.

1. On a

$$f(x, 1) = (1 + \mu + \sigma)x \quad \text{et} \quad f(x, -1) = (1 + \mu - \sigma)x;$$

les inégalités

$$-\sigma \leq |\sigma| < 1 + \mu \quad \text{et} \quad \sigma \leq |\sigma| < 1 + \mu$$

montrent alors que, pour tout $x \in \mathbb{R}^+$, $f(x, 1) \geq 0$ et $f(x, -1) \geq 0$. Il en résulte que S est un processus à valeurs P-p.s. dans \mathbb{R}^+ .

2. On a, pour tout $n \in \mathbb{N}^*$, $\Delta S_n = S_{n-1}(\mu + \sigma \varepsilon_n)$; S_{n-1} étant \mathcal{A}_{n-1} -mesurable et les tribus \mathcal{A}_{n-1} et $\sigma(\varepsilon_n)$ étant indépendantes, on a

$$E^{\mathcal{A}_{n-1}}(\Delta S_n) = S_{n-1} E^{\mathcal{A}_{n-1}}[\mu + \sigma \varepsilon_n] = S_{n-1} E[\mu + \sigma \varepsilon_n],$$

soit

$$E^{\mathcal{A}_{n-1}}(\Delta S_n) = \mu S_{n-1}. \quad (15.38)$$

Puisque $S_{n-1} \geq 0$ P-p.s., il en résulte que

$$\begin{cases} S \text{ est une sous-martingale} & \text{si } \mu > 0, \\ S \text{ est une martingale} & \text{si } \mu = 0, \\ S \text{ est une surmartingale} & \text{si } \mu < 0. \end{cases}$$

Dans le cas où $\mu < 0$, S est une surmartingale positive; la suite de terme général S_n converge donc P-p.s. dans $\overline{\mathbb{R}}^+$. Par ailleurs, d'après 15.38, on a

$$ES_n = (1 + \mu) ES_{n-1} = (1 + \mu)^n s_0.$$

Puisque dans ce cas on a $0 < 1 + \mu < 1$, on a $\lim_n ES_n = 0$ et les S_n étant positives, la suite de terme général S_n converge dans L^1 ; on a donc aussi convergence P-p.s. vers 0.

3. Les variables aléatoires S_{n-1} et ε_n étant indépendantes, on a

$$ES_n^2 = ES_{n-1}^2 E(1 + \mu + \sigma \varepsilon_n)^2;$$

mais on a

$$E(1 + \mu + \sigma \varepsilon_n)^2 = \frac{1}{2} [(1 + \mu + \sigma)^2 + (1 + \mu - \sigma)^2] = (1 + \mu)^2 + \sigma^2.$$

Il en résulte que l'on a, pour tout $n \in \mathbb{N}^*$,

$$ES_n^2 = s_0^2 [(1 + \mu)^2 + \sigma^2]^n. \quad (15.39)$$

4. On a $\ln S_0 = \ln s_0$ et, pour tout $n \in \mathbb{N}^*$,

$$Z_n = Z_{n-1} + \ln [1 + \mu + \sigma \varepsilon_n];$$

par indépendance des tribus \mathcal{A}_{n-1} et $\sigma(\varepsilon_n)$, on a

$$E^{\mathcal{A}_{n-1}}(\Delta Z_n) = E \ln [1 + \mu + \sigma \varepsilon_n] = \frac{1}{2} [\ln [1 + \mu + \sigma] + \ln [1 + \mu - \sigma]],$$

soit

$$E^{\mathcal{A}_{n-1}}(\Delta Z_n) = \frac{1}{2} \ln [(1 + \mu)^2 - \sigma^2],$$

c'est-à-dire

$$E^{\mathcal{A}_{n-1}}(\Delta Z_n) = \ln \lambda. \quad (15.40)$$

Ainsi, suivant que λ est strictement supérieur, égal, ou strictement inférieur à 1, Z est une sous-martingale, une martingale, ou une surmartingale.

On a, pour tout $n \in \mathbb{N}^*$,

$$Z_n = Z_0 + \sum_{j=1}^n \ln [1 + \mu + \sigma \varepsilon_j]; \quad (15.41)$$

les variables aléatoires $\ln [1 + \mu + \sigma \varepsilon_n]$ étant indépendantes, de même loi et intégrables, il résulte de la loi forte des grands nombres que

$$\frac{1}{n} \sum_{j=1}^n \ln [1 + \mu + \sigma \varepsilon_j] \xrightarrow{\text{P-p.s.}} E \ln [1 + \mu + \sigma \varepsilon_1],$$

ce qui démontre que $\frac{Z_n}{n} \xrightarrow{\text{P-p.s.}} \ln \lambda$. On a donc

$$\left\{ \begin{array}{l} \text{si } \lambda > 1, \quad Z_n \xrightarrow{\text{P-p.s.}} +\infty, \quad \text{et} \quad S_n \xrightarrow{\text{P-p.s.}} +\infty; \\ \text{si } \lambda = 1, \quad \frac{Z_n}{n} \xrightarrow{\text{P-p.s.}} 0, \quad \text{et} \quad S_n \xrightarrow{\text{P-p.s.}} 1; \\ \text{si } \lambda < 1, \quad Z_n \xrightarrow{\text{P-p.s.}} -\infty, \quad \text{et} \quad S_n \xrightarrow{\text{P-p.s.}} 0. \end{array} \right.$$

5. Puisque l'on a $|\sigma| < 1 + \mu$, on peut réaliser l'inégalité $(1 + \mu)^2 + \sigma^2 < 1$ dès que $|\sigma| < 1/\sqrt{2}$; dans ce cas, on doit avoir $\mu < 0$ et $-S$ est alors une sous-martingale intégrable. Soit $-S = M + A$ sa décomposition de Doob, où M est une martingale intégrable et A un processus croissant prévisible nul en 0. D'après (15.38), on a

$$\Delta A_n = E^{\mathcal{A}_{n-1}} [\Delta(-S)_n] = -\mu S_{n-1},$$

ce qui donne

$$\left\{ \begin{array}{l} A_0 = 0 \text{ et, si } n \in \mathbb{N}^*, \quad A_n = -\mu \sum_{j=0}^{n-1} S_j \\ M_0 = s_0 \text{ et, si } n \in \mathbb{N}^*, \quad M_n = -S_n + \mu \sum_{j=0}^{n-1} S_j. \end{array} \right.$$

Il en résulte que M est une martingale dans L^2 dont le processus croissant prévisible $\langle M \rangle$ est donné par ses accroissements

$$\begin{aligned} \Delta \langle M \rangle_n &= E^{\mathcal{A}_{n-1}} (\Delta M_n)^2 = E^{\mathcal{A}_{n-1}} [\Delta S_n + \Delta A_n]^2 \\ &= E^{\mathcal{A}_{n-1}} [S_{n-1}(\mu + \sigma \varepsilon_n) - \mu S_{n-1}]^2 = \sigma^2 S_{n-1}^2. \end{aligned}$$

Il en résulte que

$$\langle M \rangle_0 = 0 \quad \text{et, si } n \in \mathbb{N}^*, \quad \langle M \rangle_n = \sigma^2 \sum_{j=0}^{n-1} S_j^2.$$

On a alors

$$E \langle M \rangle_n = \sigma^2 \sum_{j=0}^{n-1} E S_j^2,$$

et d'après (15.39).

$$E \langle M \rangle_n = \sigma^2 s_0^2 \frac{1 - [(1 + \mu)^2 + \sigma^2]^n}{1 - s_0^2 [(1 + \mu)^2 + \sigma^2]};$$

puisque, par hypothèse, on a $(1 + \mu)^2 + \sigma^2 < 1$, la suite de terme général $E \langle M \rangle_n$ est alors convergente, ce qui implique que $\sum_{j=0}^{+\infty} E S_j^2 = E[\sum_{j=0}^{+\infty} S_j^2] < +\infty$. Il en résulte que la série de terme général S_n^2 converge P-p.s.

6. On a les égalités $\Delta W_n = \ln \left[\frac{S_n}{\lambda S_{n-1}} \right] = \Delta Z_n - \ln \lambda$. Il résulte alors de (15.40) que, $E^{\mathcal{A}_{n-1}}(\Delta W_n) = 0$, c'est-à-dire que W est une martingale; elle est dans L^2 puisque les variables aléatoires Z_n le sont. Son processus croissant prévisible $\langle W \rangle$ est alors donné par ses accroissements définis, compte tenu de l'indépendance des tribus \mathcal{A}_{n-1} et $\sigma(\varepsilon_n)$, par

$$\Delta \langle W \rangle_n = E^{\mathcal{A}_{n-1}}(\Delta W_n)^2 = E^{\mathcal{A}_{n-1}} \left[\ln \left(\frac{1 + \mu + \sigma \varepsilon_n}{\lambda} \right) \right]^2 = E \left[\ln \left(\frac{1 + \mu + \sigma \varepsilon_n}{\lambda} \right) \right]^2,$$

soit

$$\begin{aligned} \Delta \langle W \rangle_n &= \frac{1}{2} \left[\left(\ln \frac{1 + \mu + \sigma}{\lambda} \right)^2 + \left(\ln \frac{1 + \mu - \sigma}{\lambda} \right)^2 \right] \\ &= \frac{1}{2} \left[\left(\ln \frac{1 + \mu + \sigma}{\lambda} + \ln \frac{1 + \mu - \sigma}{\lambda} \right)^2 - 2 \ln \frac{1 + \mu + \sigma}{\lambda} \ln \frac{1 + \mu - \sigma}{\lambda} \right]; \end{aligned}$$

puisque l'on a

$$\ln \frac{1 + \mu + \sigma}{\lambda} + \ln \frac{1 + \mu - \sigma}{\lambda} = \ln \frac{(1 + \mu)^2 - \sigma^2}{\lambda^2} = 0,$$

il en résulte que $\Delta \langle W \rangle_n = \delta$ (ce qui prouve que $\delta > 0$) et donc que, pour tout $n \in \mathbb{N}$, $\langle W \rangle_n = n\delta$.

La loi forte des grands nombres pour les martingales dans L^2 assure alors que $\frac{W_n}{\langle W \rangle_n} \xrightarrow{\text{P-p.s.}} 0$, soit encore

$$\frac{\ln S_n - n \ln \lambda}{n} \xrightarrow{\text{P-p.s.}} 0,$$

ce qui implique que

$$\frac{1}{S_n^n} \xrightarrow{\text{P-p.s.}} \lambda.$$

7. Puisque $\ln R_n = \frac{1}{\sqrt{n}} Z_n - \sqrt{n} \ln \lambda$, on a, d'après (15.41),

$$\ln R_n = \frac{1}{\sqrt{n}} \left[Z_0 + \sum_{j=1}^n (\ln [1 + \mu + \sigma \varepsilon_j] - \ln \lambda) \right].$$

Par ailleurs, les variables aléatoires $\ln[1 + \mu + \sigma \varepsilon_j]$ sont indépendantes de même loi et admettent un moment d'ordre deux; elles ont pour moyenne

$$E \ln [1 + \mu + \sigma \varepsilon_n] = \frac{1}{2} [\ln(1 + \mu + \sigma) + \ln(1 + \mu - \sigma)] = \ln \lambda,$$

et pour moment d'ordre deux

$$E [\ln(1 + \mu + \sigma \varepsilon_n)]^2 = \frac{1}{2} [(\ln(1 + \mu + \sigma))^2 + (\ln(1 + \mu - \sigma))^2];$$

elles ont donc pour variance

$$\begin{aligned} \sigma_{\ln(1+\mu+\sigma\varepsilon_n)}^2 &= \frac{1}{2} [(\ln(1 + \mu + \sigma))^2 + (\ln(1 + \mu - \sigma))^2] \\ &\quad - \frac{1}{4} [\ln(1 + \mu + \sigma) + \ln(1 + \mu - \sigma)]^2. \end{aligned}$$

soit

$$\begin{aligned} \sigma_{\ln(1+\mu+\sigma\varepsilon_n)}^2 &= \frac{1}{4} [(\ln(1 + \mu + \sigma))^2 + (\ln(1 + \mu - \sigma))^2 \\ &\quad - 2 \ln(1 + \mu + \sigma) \ln(1 + \mu - \sigma)]. \end{aligned}$$

ou encore

$$\sigma_{\ln(1+\mu+\sigma\varepsilon_n)}^2 = \frac{1}{4} \left[\ln \left(\frac{1 + \mu + \sigma}{1 + \mu - \sigma} \right) \right]^2 = \rho^2.$$

Il résulte alors du théorème limite central que

$$\frac{1}{\sqrt{n}|\rho|} \left[\sum_{j=1}^n (\ln [1 + \mu + \sigma \varepsilon_j] - \ln \lambda) \right] \xrightarrow{\mathcal{L}} \mathcal{N}_{\mathbb{R}}(0, 1).$$

et donc que la suite des lois $P_{\ln R_n}$ converge étroitement vers la loi $\mathcal{N}_{\mathbb{R}}(0, \rho^2)$. Il en résulte que, pour tout $f \in \mathcal{C}_b(\mathbb{R})$, on a

$$\lim_n \int_{\mathbb{R}} f(R_n) dP = \int_{\mathbb{R}} f[\exp(\ln R_n)] dP = \int_{\mathbb{R}} f[\exp x] \frac{1}{|\rho|\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\rho^2}\right) dx,$$

puisque $f \circ \exp \in \mathcal{C}_b(\mathbb{R})$. En faisant le changement de variables de \mathbb{R} sur \mathbb{R}^{+*} défini par $y = \exp x$, on obtient

$$\lim_n \int_{\mathbb{R}} f(R_n) dP = \int_{\mathbb{R}^{+*}} f(y) \frac{1}{|\rho|\sqrt{2\pi}} \frac{1}{y} \exp\left(-\frac{(\ln y)^2}{2\rho^2}\right) dy.$$

La suite des lois P_{R_n} converge donc étroitement vers la probabilité de densité par rapport à la mesure de Lebesgue l'application $y \mapsto \mathbf{1}_{\mathbb{R}^{+*}}(y) \frac{1}{|\rho|\sqrt{2\pi}} \frac{1}{y} \exp\left(-\frac{(\ln y)^2}{2\rho^2}\right)$; c'est la **loi Log-normale** de paramètres 0 et $\rho^2 = \frac{1}{4} \left[\ln \left(\frac{1 + \mu + \sigma}{1 + \mu - \sigma} \right) \right]^2$.

Chapitre 16

Chaînes de Markov

Les chaînes de Markov constituent une classe importante de processus stochastiques à temps discret; elles permettent de modéliser des phénomènes aléatoires temporels dont l'évolution probabiliste à tout instant ne dépend que de l'état du système à cet instant et non de toute son évolution antérieure; autrement dit, elles modélisent des phénomènes sans mémoire. On se limite à l'étude des chaînes de Markov à espace d'état dénombrable.

16.1. Introduction

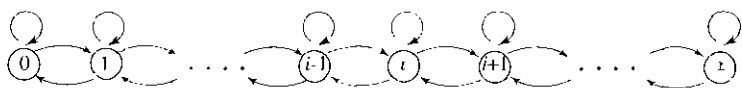
Dans le tome 1 (section 4.3), nous avons déjà montré comment on pouvait traduire qu'un phénomène évolutif, fini en temps et en espace, était **markovien**, c'est-à-dire sans mémoire, autre que la mémoire présente. L'objet de cette introduction est d'illustrer cette **formalisation élémentaire** par l'étude d'un modèle historique, et de mettre ensuite en évidence les difficultés d'**axiomatisation mathématique** qu'elle présente, afin de conduire de manière naturelle à la définition des chaînes de Markov homogènes donnée en 16.12.

Commençons par présenter le **modèle de diffusion de gaz** connu sous le nom de modèle de **Bernoulli-Laplace**.

Exemple 16.1. (Modèle de diffusion de gaz de Bernoulli-Laplace.) Deux urnes numérotées 1 et 2 contiennent chacune m boules; parmi ces $2m$ boules, r ($1 \leq r \leq m$) sont rouges et $2m-r$ sont blanches. L'échelle de temps est choisie discrète, et après réindexation, on la confond avec \mathbb{N} . À chaque instant, on tire au hasard et indépendamment, une boule dans chaque urne, et on remet chacune de ces deux boules dans l'urne d'où elle n'a pas été tirée.

On choisit de représenter l'état du système à l'instant n , $n \geq 1$, par le nombre X_n de boules rouges présentes dans l'urne 1 après le n -ième tirage et la remise des boules tirées dans les urnes; l'état initial est noté X_0 . L'espace d'états possibles est l'intervalle d'entiers $E = [0, 1, \dots, r]$.

On peut s'imaginer l'état X_n comme un **point aléatoire** se déplaçant sur E , ce point ne pouvant, en une étape, que rester en place ou se déplacer vers un de ses voisins les plus proches. Il est d'usage de visualiser ceci par un **graphe** du type :



Il est intuitivement clair que le processus $(X_n)_{n \in \mathbb{N}^*}$ est **markovien**, c'est-à-dire vérifie pour tout $n \geq 0$, et tout $(n+2)$ -uplet d'états x_0, x_1, \dots, x_{n+1} ,

$$P(X_{n+1}=x_{n+1} \mid X_0=x_0, \dots, X_n=x_n) = P(X_{n+1}=x_{n+1} \mid X_n=x_n), \quad (16.1)$$

ces probabilités conditionnelles étant définies au sens élémentaire (les probabilités des événements de conditionnement étant intuitivement non nulles). L'évolution du processus $(X_n)_{n \in \mathbb{N}^*}$ est alors précisée par la détermination de ces probabilités conditionnelles, dites **probabilités de transition**. Si X_n vaut i , avant le $n+1$ -ième tirage, l'urne 1 contient i boules rouges, $m-i$ blanches et l'urne 2 contient $r-i$ boules rouges et $m-(r-i)$ blanches. Pour $j = 1, 2$, notons R_j^{n+1} et B_j^{n+1} , le fait que la couleur de la boule tirée au $n+1$ -ième tirage est respectivement rouge ou blanche. L'indépendance et l'uniformité des tirages permet de dresser les tableaux ci-dessous :

- Si $1 \leq i < r-1$,

configurations du $n+1$ -ième tirage	passage d'un état à un autre	probabilité du passage
$R_1^{n+1} R_2^{n+1}$	$i \rightarrow i$	$\frac{i}{m} \frac{r-i}{m}$
$R_1^{n+1} B_2^{n+1}$	$i \rightarrow i-1$	$\frac{i}{m} \frac{m-(r-i)}{m}$
$B_1^{n+1} R_2^{n+1}$	$i \rightarrow i+1$	$\frac{m-i}{m} \frac{r-i}{m}$
$B_1^{n+1} B_2^{n+1}$	$i \rightarrow i$	$\frac{m-i}{m} \frac{m-(r-i)}{m}$

- Si $i = 0$,

configurations du $n+1$ -ième tirage	passage d'un état à un autre	probabilité du passage
$B_1^{n+1} R_2^{n+1}$	$0 \rightarrow 1$	$\frac{r}{m}$
$B_1^{n+1} B_2^{n+1}$	$0 \rightarrow 0$	$\frac{m-r}{m}$

- Si $i = r$,

configurations du $n+1$ -ième tirage	passage d'un état à un autre	probabilité du passage
$R_1^{n+1} B_2^{n+1}$	$r \rightarrow r-1$	$\frac{r}{m}$
$B_1^{n+1} B_2^{n+1}$	$r \rightarrow r$	$\frac{m-r}{m}$

On s'aperçoit que, dans ces trois cas, les probabilités de passage d'un état à un autre s'écrivent en une formule unique. Les probabilités conditionnelles cherchées, dites **probabilités de transition**, sont alors données, pour tout i tel que $0 \leq i \leq r$, par

$$\begin{cases} P(X_{n+1}=i | X_n=i) &= \frac{i}{m} \frac{r-i}{m} + \frac{m-i}{m} \frac{m-(r-i)}{m}, \\ P(X_{n+1}=i-1 | X_n=i) &= \frac{i}{m} \frac{m-(r-i)}{m}, \\ P(X_{n+1}=i+1 | X_n=i) &= \frac{m-i}{m} \frac{r-i}{m}. \end{cases}$$

Il est d'usage de considérer globalement ces probabilités de transition à l'aide d'une matrice M , dite **matrice de transition**, de terme général

$$M_{ij} = P(X_{n+1} = j | X_n = i).$$

Nous verrons à la proposition 16.14 que la donnée de cette matrice et de la loi de la variable initiale X_0 déterminent entièrement la loi de toute variable aléatoire (X_0, X_1, \dots, X_n) , et détermine donc le comportement probabiliste du processus X . En particulier, il sera possible, après avoir développé la théorie des chaînes de Markov, de décrire rapidement le **comportement asymptotique** qualitatif de ce processus, et de déterminer les **limites des probabilités** d'être dans un état quelconque lorsque le temps croît vers l'infini, problème résolu historiquement par Bernoulli et Laplace.

Citons au passage deux autres modèles célèbres, l'un d'**échange de chaleur entre deux corps isolés**, connu sous le nom de modèle d'**Ehrenfest**, et l'autre, celui de **Pólya**, de **propagation de maladies contagieuses**. Ces phénomènes physiques sont encore représentés par des modèles de tirages de boules dans des urnes. Ils sont abondamment décrits dans le livre de Feller, tome 1, et repris dans de nombreux livres plus récents. Ces modèles serviront d'exemples d'illustration des notions et résultats introduits tout au long de ce chapitre.

Exemple 16.2. (Modèle de diffusion de chaleur entre deux corps isolés de Ehrenfest¹.) Deux récipients, nommés 1 et 2, contiennent au total m particules qui peuvent diffuser d'un récipient à l'autre. On se représente le phénomène de diffusion de ces particules comme, à chaque unité de temps, le choix au hasard d'une particule dans un récipient et le transfert de la particule choisie dans l'autre récipient. Répétant de la même manière ces choix et transferts, on s'intéresse à la **distribution** des particules dans chaque urne après n étapes.

1. Ehrenfest P. und T. Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. *Physikalische Zeitschrift*, vol. 8 (1907), pp. 311-314.

La modélisation sous forme de tirages de boules dans une urne est alors la suivante. On identifie les particules du premier récipient à des boules rouges, et celles du deuxième à des boules blanches (elles sont respectivement, à l'instant initial, au nombre de r et $m - r$). À chaque tirage, la boule tirée est remplacée par une boule de la couleur opposée. L'échelle de temps est choisie discrète, et après réindexation, on la confond avec \mathbb{N} . L'état à l'instant n est le **nombre de boules rouges** contenues dans l'urne.

Exemple 16.3. (Modèle de diffusion de maladies contagieuses de Pólya.) Il s'agit d'un modèle de propagation de maladies contagieuses, qui traduit l'augmentation, ou la diminution, de probabilité de contagion d'un individu à chaque apparition, ou disparition, d'un nouveau cas de maladie parmi une population.

La modélisation sous forme de tirages de boules dans une urne est alors la suivante. On identifie les individus sains de la population à des boules rouges, et les individus malades à des boules blanches (elles sont respectivement, à l'instant initial, au nombre de r et b). Après chaque tirage au hasard, la boule tirée est remise dans l'urne avec c boules de la couleur de la boule tirée. L'état à l'instant n est la **proportion** Y_n de boules blanches contenues dans l'urne **après le n -ième tirage et après avoir rajouté les c boules**.

Si on laisse évoluer indéfiniment ce processus, l'ensemble E des états possibles n'est pas fini, mais est infini dénombrable et, a priori, **contenu** dans l'ensemble des nombres rationnels de l'intervalle $[0, 1]$. Il devient déjà plus difficile de dire, de manière élémentaire, si le processus $(Y_n)_{n \in \mathbb{N}^*}$ est **markovien** ou non, par des formules du type (16.1). En effet, le choix d'un modèle passe d'abord par le choix de l'ensemble E des états possibles ; ceci fait, avant l'étude du modèle, on ne sait pas dire si, pour tout $(n + 1)$ -uple d'états y_0, y_1, \dots, y_n , la probabilité $P(Y_0 = y_0, \dots, Y_n = y_n)$ de l'événement de conditionnement est nulle ou non ; ceci pose un problème de définition de la probabilité conditionnelle au sens élémentaire, et donc aussi, de pertinence d'une définition de la propriété de Markov par les relations du type (16.1). Pour traduire la propriété de Markov, c'est-à-dire que le processus est sans mémoire, autre que la mémoire présente, nous serons conduits à prendre une définition plus adaptée mathématiquement ; au sens de cette définition, le processus de Pólya $(Y_n)_{n \in \mathbb{N}^*}$ sera effectivement **markovien**, ce qui n'est pas tout à fait intuitif.

Un autre exemple, cas particulier des **marches aléatoires** (nous les rencontrerons ci-dessous), met encore en évidence ce même problème de définition. La **propriété de Markov** est toutefois, dans ce cas, intuitivement évidente, si l'idée qu'on en a est de dire que **l'évolution probabiliste du processus après chaque instant n ne dépend que de l'état du processus à cet instant**.

Il montre aussi l'importance de la **loi initiale** dans la modélisation.

Exemple 16.4. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur un espace probabilisé (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{Z} et indépendantes. On suppose que, pour tout $n \in \mathbb{N}^*$, les variables aléatoires X_n ont même loi $p\delta_{-2} + q\delta_2$, avec $p + q = 1$ et $p, q > 0$. On note

$$S_n = \sum_{j=0}^n X_j \quad \text{et} \quad Y_n = \sum_{j=0}^n X_j^2.$$

si bien que $S_{n+1} = S_n + X_{n+1}$ et $Y_{n+1} = Y_n + X_{n+1}^2$. On peut penser à S_n comme à la **position d'une particule** qui saute, à chaque unité de temps, d'un entier à un autre; le processus $(S_n)_{n \in \mathbb{N}}$ est une **marche aléatoire**.

Si $X_0 = 0$, S_n est à valeurs dans $2\mathbb{Z}$, et un calcul de convolution facile assure que tous les entiers **pairs** compris entre $-2n$ et $2n$, et eux seuls, vont être visités par S_n avec une probabilité non nulle. Par contre, si X_0 est de loi $(\delta_0 + \delta_1)/2$, S_n est à valeurs dans \mathbb{Z} ; certaines trajectoires du processus $(S_n)_{n \in \mathbb{N}}$ seront contenues dans $2\mathbb{Z}$, d'autres dans $2\mathbb{Z} + 1$ et, en tout état de cause, pour tout $(n+1)$ -uplet x_0, x_1, \dots, x_n , d'éléments **consécutifs** de l'ensemble d'états possibles \mathbb{Z} , la probabilité $P(X_0 = x_0, \dots, X_n = x_n)$ de l'événement de conditionnement dans la relation du type (16.1) sera nulle, ce qui rend encore caduque, dans ce cas, la définition de la propriété de Markov par de telles relations.

Enfin, l'évolution probabiliste du processus $(Y_n)_{n \in \mathbb{N}}$ en tout instant n ne dépend encore que de l'état de ce processus à cet instant; en ce sens, le processus $(Y_n)_{n \in \mathbb{N}}$ a la propriété de Markov. On peut remarquer, de plus, que l'« **histoire** » probabiliste **au temps** n peut être envisagée de manière plus ou moins exhaustive, suivant que l'on a accès aux valeurs des X_j , ou seulement à celle des X_j^2 , $0 \leq j \leq n$; mais de toute façon, cette histoire au temps n n'influe sur la probabilité que Y_{n+1} prenne une valeur donnée, qu'à travers la connaissance de Y_n . C'est pour tenir compte de ce genre de situation que nous définirons les **chaînes de Markov relativement** à des **filtrations**.

Ceci étant, nous allons définir la propriété de Markov à l'aide de la notion d'**indépendance conditionnelle de tribus**, notion utile dans d'autres contextes, et étudiée dans la section suivante.

16.2. Indépendance conditionnelle

Notations. Dans ce chapitre, on notera $E^{X= \cdot} (Y)$ l'application **moyenne conditionnelle** de la variable aléatoire Y conditionnellement à la variable

aléatoire X . On ne la confondra pas² avec l'**espérance conditionnelle** de Y par rapport à la tribu $\sigma(X)$ engendrée par X , notée $E^{\sigma(X)}(Y)$ ou $E(Y | X)$, suivant le contexte, et qui, elle, est une classe de **variables aléatoires** (suivant l'usage, on note de la même façon un représentant et sa classe).

Commençons par un exemple qui fera bien comprendre la notion d'**indépendance conditionnelle**, notion que nous allons définir et étudier dans sa généralité sitôt après.

Soient X, Y, Z trois variables aléatoires réelles définies sur un espace probabilisé (Ω, \mathcal{A}, P) , indépendantes. Posons

$$U = X + Y \quad \text{et} \quad V = XZ.$$

Soient f et g des fonctions boréliennes bornées quelconques. Puisque les variables aléatoires X et (Y, Z) sont indépendantes, il résulte du théorème de transfert conditionnel (ch. 11, th. 11.9) que la **moyenne conditionnelle** de $f(U)g(V)$ conditionnellement à la variable aléatoire X , vérifie, pour P_X -presque tout x ,

$$E^{X=x} [f(U)g(V)] = E^{X=x} [f(x + Y)g(xZ)] = E [f(x + Y)g(xZ)],$$

soit, en tenant compte de l'indépendance des variables aléatoires Y et Z ,

$$E^{X=x} [f(U)g(V)] = E [f(x + Y)] E [g(xZ)].$$

Il en résulte en particulier, en prenant successivement pour f et g la fonction constante égale à 1, que

$$E^{X=x} [f(U)] = E [f(x + Y)] \quad \text{et} \quad E^{X=x} [g(V)] = E [g(xZ)].$$

Ainsi, on a, **pour P_X -presque tout x ,**

$$E^{X=x} [f(U)g(V)] = E^{X=x} [f(U)] E^{X=x} [g(V)].$$

On en déduit l'égalité des espérances conditionnelles

$$E^{\sigma(X)} [f(U)g(V)] = E^{\sigma(X)} [f(U)] E^{\sigma(X)} [g(V)].$$

Ces deux dernières relations traduisent, de manière équivalente, l'indépendance conditionnelle des **variables aléatoires** U et V par rapport à la variable aléatoire X , ou l'**indépendance** des **tribus** engendrées par les variables aléatoires U et V , **conditionnellement** à la tribu $\sigma(X)$ engendrée par la variable aléatoire X .

Nous donnons à présent la définition générale d'**indépendance conditionnelle de tribus**, notion utilisée de manière essentielle par la suite pour définir la **propriété de Markov**. Nous en étudions ensuite quelques propriétés.

2. La relation entre ces deux notions est établie au lemme 11.31 du chapitre 11.

Les tribus considérées sont des **sous-tribus** définies sur le même espace probabilisé (Ω, \mathcal{A}, P) .

Définition 16.1. Soient trois tribus \mathcal{A}_i , $i = 1, 2, 3$. Les tribus \mathcal{A}_1 et \mathcal{A}_3 sont conditionnellement indépendantes par rapport à \mathcal{A}_2 si, pour $i = 1, 3$ et pour toute variable aléatoire réelle Y_i \mathcal{A}_i -mesurable bornée (ce que l'on notera $Y_i \in b\mathcal{A}_i$), on a

$$E^{\mathcal{A}_2}(Y_1 Y_3) = E^{\mathcal{A}_2}(Y_1) E^{\mathcal{A}_2}(Y_3). \quad (16.2)$$

En particulier, si \mathcal{A}_2 est la tribu $\sigma(X)$ engendrée par une variable aléatoire X , on dit simplement que les tribus \mathcal{A}_1 et \mathcal{A}_3 sont conditionnellement indépendantes par rapport à X . La relation (16.2) s'écrit alors

$$E(Y_1 Y_3 | X) = E^{\mathcal{A}_2}(Y_1 | X) E^{\mathcal{A}_2}(Y_3 | X). \quad (16.3)$$

Remarque. L'indépendance conditionnelle des tribus \mathcal{A}_1 et \mathcal{A}_3 par rapport à \mathcal{A}_2 n'implique pas leur indépendance. Mais si \mathcal{A}_2 est la tribu triviale $\{\Omega, \emptyset\}$, l'indépendance conditionnelle des tribus \mathcal{A}_1 et \mathcal{A}_3 par rapport à \mathcal{A}_2 est alors équivalente à leur indépendance.

On note $P^{\mathcal{B}}(A)$ la probabilité conditionnelle de A sachant la tribu \mathcal{B} définie par $E^{\mathcal{B}}(\mathbf{1}_A)$. Si \mathcal{A}_2 est la tribu $\sigma(X)$ engendrée par une variable aléatoire X , on la note $P(B | X)$.

Lemme 16.2. Les tribus \mathcal{A}_1 et \mathcal{A}_3 sont conditionnellement indépendantes par rapport à \mathcal{A}_2 si et seulement si pour tous événements $A_1 \in \mathcal{A}_1$ et $A_3 \in \mathcal{A}_3$, on a

$$P^{\mathcal{A}_2}(A_1 \cap A_3) = P^{\mathcal{A}_2}(A_1) P^{\mathcal{A}_2}(A_3). \quad (16.4)$$

En particulier, si \mathcal{A}_2 est la tribu $\sigma(X)$ engendrée par une variable aléatoire X , les tribus \mathcal{A}_1 et \mathcal{A}_3 sont conditionnellement indépendantes par rapport à X si et seulement si pour tous événements $A_1 \in \mathcal{A}_1$ et $A_3 \in \mathcal{A}_3$, on a

$$P(A_1 \cap A_3 | X) = P(A_1 | X) P(A_3 | X). \quad (16.5)$$

Démonstration. La condition nécessaire est triviale. Inversement, de (16.4), on déduit que (16.2) est vraie pour toutes les variables aléatoires Y_i \mathcal{A}_i -mesurables étagées, $i = 1, 3$; on passe au cas général par les procédés usuels d'intégration. \square

Si le lemme précédent est anecdotique, le théorème suivant est par contre fondamental pour l'étude que nous ferons des chaînes de Markov.

Théorème 16.3. On note \mathcal{A}_{12} (ou encore $\mathcal{A}_1 \vee \mathcal{A}_2$) la tribu engendrée par \mathcal{A}_1 et \mathcal{A}_2 . Les tribus \mathcal{A}_1 et \mathcal{A}_3 sont conditionnellement indépendantes par rapport à \mathcal{A}_2 si et seulement si pour tout $Y_3 \in \mathcal{L}^1(\Omega, \mathcal{A}_3, P)$, on a

$$E^{\mathcal{A}_{12}}(Y_3) = E^{\mathcal{A}_2}(Y_3). \quad (16.6)$$

Remarque. En fait, par des arguments habituels d'intégration, pour qu'il y ait indépendance conditionnelle, il faut et il suffit que (16.6) soit vraie pour tout $Y_3 \in h\mathcal{A}_3$.

Démonstration. Condition nécessaire. On utilise le principe de prolongement par mesurabilité. On vérifie facilement que la famille d'événements

$$\mathcal{S} = \left\{ A \in \mathcal{A}_{12} \mid \int_A Y_3 dP = \int_A E^{\mathcal{A}_2}(Y_3) dP \quad \forall Y_3 \in \mathcal{L}^1(\Omega, \mathcal{A}_3, P) \right\}$$

est un λ -système. On va démontrer qu'il contient le π -système \mathcal{C} qui engendre \mathcal{A}_{12} défini par

$$\mathcal{C} = \{A_1 \cap A_2 \in \mathcal{A}_{12} \mid A_1 \in \mathcal{A}_1 \text{ et } A_2 \in \mathcal{A}_2\};$$

cela démontrera que \mathcal{S} contient \mathcal{A}_{12} et donc que (16.6) est vraie pour tout $Y_3 \in \mathcal{L}^1(\Omega, \mathcal{A}_3, P)$. Soient donc $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$ quelconques; on a, en utilisant la mesurabilité des variables aléatoires concernées successivement par rapport aux tribus \mathcal{A}_{12} puis \mathcal{A}_2 ,

$$E \left[\mathbf{1}_{A_1} \mathbf{1}_{A_2} E^{\mathcal{A}_{12}}(Y_3) \right] = E \left[\mathbf{1}_{A_1} \mathbf{1}_{A_2} Y_3 \right] = E \left[\mathbf{1}_{A_2} E^{\mathcal{A}_2}(\mathbf{1}_{A_1} Y_3) \right],$$

soit, d'après (16.2),

$$E \left[\mathbf{1}_{A_1} \mathbf{1}_{A_2} E^{\mathcal{A}_{12}}(Y_3) \right] = E \left[\mathbf{1}_{A_2} E^{\mathcal{A}_2}(\mathbf{1}_{A_1}) E^{\mathcal{A}_2}(Y_3) \right],$$

ou encore, puisque $\mathbf{1}_{A_2} E^{\mathcal{A}_2}(Y_3)$ est \mathcal{A}_2 -mesurable,

$$E \left[\mathbf{1}_{A_1} \mathbf{1}_{A_2} E^{\mathcal{A}_{12}}(Y_3) \right] = E \left[\mathbf{1}_{A_1} \mathbf{1}_{A_2} E^{\mathcal{A}_2}(Y_3) \right].$$

La condition nécessaire est établie.

Condition suffisante. Soit, pour $i = 1, 3$, $Y_i \in h\mathcal{A}_i$ quelconques; tenant compte de l'inclusion de tribus $\mathcal{A}_2 \subset \mathcal{A}_{12}$, puis de la \mathcal{A}_{12} -mesurabilité de Y_1 , on a

$$E^{\mathcal{A}_2}(Y_1 Y_3) = E^{\mathcal{A}_2} \left[E^{\mathcal{A}_{12}}(Y_1 Y_3) \right] = E^{\mathcal{A}_2} \left[Y_1 E^{\mathcal{A}_{12}}(Y_3) \right];$$

d'après l'hypothèse (16.6), on a alors

$$E^{\mathcal{A}_2}(Y_1 Y_3) = E^{\mathcal{A}_2} \left[Y_1 E^{\mathcal{A}_2}(Y_3) \right],$$

et par conséquent.

$$E^{\mathcal{A}_2}(Y_1 Y_3) = E^{\mathcal{A}_2}(Y_1) E^{\mathcal{A}_2}(Y_3). \quad \square$$

En utilisant encore le théorème de prolongement par mesurabilité, on peut résoudre l'exercice suivant :

Exercice. Soient quatre tribus \mathcal{A}_i , $i = 1, 2, 3$ et \mathcal{B}_3 . On suppose que $\mathcal{A}_3 = \mathcal{B}_3 \vee \mathcal{A}_2$. Les tribus \mathcal{A}_1 et \mathcal{A}_3 sont conditionnellement indépendantes par rapport à \mathcal{A}_2 si et seulement si les tribus \mathcal{A}_1 et \mathcal{B}_3 sont.

16.3. Chaînes de Markov : propriétés générales

Dans ce chapitre, E est un ensemble **dénombrable** (fini ou infini) muni de la tribu de ses parties \mathcal{E} ; sauf mention du contraire, les processus sont définis sur l'espace probabilisé (Ω, \mathcal{A}, P) .

16.3.1. Propriété de Markov ; matrices de transition

Définition 16.4. Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans $(E, \mathcal{E})^3$. Pour tout $n \in \mathbb{N}$, la tribu $\mathcal{F}_n^l = \sigma(X_j \mid j \geq n)$ est appelée **tribu du futur large** du processus X après le temps n et $\mathcal{F}_n^s = \sigma(X_j \mid j > n)$ est appelée **tribu du futur strict** du processus X après le temps n . La tribu $\mathcal{P}_n = \sigma(X_n)$ est la tribu du **présent** du processus à l'instant n .

Le processus X est une **chaîne de Markov** relativement (ou par rapport) à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$ s'il satisfait aux deux conditions :

- (i) $X = (X_n)_{n \in \mathbb{N}}$ est adapté à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$;
- (ii) X a la **propriété de Markov**, à savoir que, pour tout $n \in \mathbb{N}$, la tribu du **passé** \mathcal{A}_n et la tribu \mathcal{F}_n^l du **futur large** à l'instant n sont conditionnellement indépendantes par rapport à la tribu $\mathcal{P}_n = \sigma(X_n)$ du **présent** à l'instant n .

On dit aussi, dans ce cas, que X est une chaîne de Markov sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P)$.

Remarque. Si X est une chaîne de Markov par rapport à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$, il en est encore une par rapport à sa filtration naturelle $(\mathcal{B}_n)_{n \in \mathbb{N}}$. Dans le cas où la filtration de référence est la filtration naturelle, on parle brièvement de chaîne de Markov.

D'après l'exercice précédent, dans la définition d'une chaîne de Markov, on peut remplacer la tribu du futur large par celle du futur stricte.

Exemple 16.5. Soit $X = (X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes à valeurs dans \mathbb{Z} . Soit pour tout $n \in \mathbb{N}$, $S_n = \sum_{j=0}^n X_j$: les filtrations naturelles des processus X et $S = (S_n)_{n \in \mathbb{N}}$ sont les mêmes et S est une chaîne de Markov, appelée **marche aléatoire sur \mathbb{Z}** .

On donne maintenant une caractérisation des chaînes de Markov.

Théorème 16.5. Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans E adapté à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$. Les propriétés suivantes sont équivalentes :

3. Dans ce contexte, on dira, de façon abrégée, à valeurs dans E .

(i) Le processus X est une chaîne de Markov par rapport à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$.

(ii) Pour tout $n \in \mathbb{N}$ et tout $Y \in b\mathcal{F}_n^1$, on a

$$E^{\mathcal{A}_n}(Y) = E^{\mathcal{P}_n}(Y). \quad (16.7)$$

(iii) Pour tout $n \in \mathbb{N}$ et toute $f \in b\mathcal{E}$ (fonction bornée sur E), on a

$$E^{\mathcal{A}_n}[f(X_{n+1})] = E^{\mathcal{P}_n}[f(X_{n+1})]. \quad (16.8)$$

(iv) Pour tous n et m tels que $n \leq m$, et pour tout $f \in b\mathcal{E}$, on a

$$E^{\mathcal{A}_n}[f(X_m)] = E^{\mathcal{P}_n}[f(X_m)]. \quad (16.9)$$

En particulier, le processus X est une chaîne de Markov **par rapport à sa filtration naturelle** $(\mathcal{B}_n)_{n \in \mathbb{N}}$ si et seulement si pour tout $n \in \mathbb{N}$, pour toute suite croissante finie d'entiers telle que $n_1 \leq n_2 \leq \dots \leq n_k \leq n$ et toute fonction bornée f sur E , on a

$$E^{\sigma(X_{n_1}, \dots, X_{n_k})}[f(X_n)] = E^{\sigma(X_{n_k})}[f(X_n)]. \quad (16.10)$$

Démonstration. L'équivalence de (i) et (ii) résulte du théorème 16.3. Les implications (iv) \Rightarrow (iii), (ii) \Rightarrow (iii) et (ii) \Rightarrow (iv) sont triviales.

Reste à démontrer l'implication (iii) \Rightarrow (ii) : supposons donc que (iii) soit vraie. D'après la remarque précédente, il suffit de démontrer que l'égalité (16.7) est vraie pour tout $Y \in b\mathcal{F}_n^s$. Pour n quelconque fixé, soit

$$\mathcal{H} = \left\{ Y \in b\mathcal{F}_n^s \mid E^{\mathcal{A}_n}(Y) = E^{\mathcal{P}_n}(Y) \right\}.$$

L'ensemble \mathcal{H} est un espace vectoriel contenant les constantes et stable par limite monotone bornée. D'après le théorème de prolongement par mesurabilité, il suffit alors de démontrer que \mathcal{H} contient les indicatrices des éléments du π -système \mathcal{C} engendrant la tribu \mathcal{F}_n^s défini par

$$\mathcal{C} = \left\{ \bigcap_{i=1}^k X_{n+i}^{-1}(E_i) \mid k \in \mathbb{N}^*, E_i \in \mathcal{E} \right\}.$$

Soit donc $Y = \mathbf{1}_{\bigcap_{i=1}^k X_{n+i}^{-1}(E_i)}$. Si $k = 1$, le résultat est trivial. Supposons donc $k \geq 2$. On utilise un procédé classique d'étude des chaînes de Markov, à savoir celui de remonter le temps. On a, puisque $\mathcal{A}_n \subset \mathcal{A}_{n+k-1}$ et que $\prod_{i=1}^{k-1} \mathbf{1}_{E_i} \circ X_{n+i}$ est \mathcal{A}_{n+k-1} -mesurable,

$$E^{\mathcal{A}_n}(Y) = E^{\mathcal{A}_n} \left[\prod_{i=1}^k \mathbf{1}_{E_i} \circ X_{n+i} \right] = E^{\mathcal{A}_n} \left[\left(\prod_{i=1}^{k-1} \mathbf{1}_{E_i} \circ X_{n+i} \right) E^{\mathcal{A}_{n+k-1}}(\mathbf{1}_{E_k} \circ X_{n+k}) \right].$$

et donc, d'après l'hypothèse (iii),

$$E^{\mathcal{A}_n}(Y) = E^{\mathcal{A}_n} \left[\left(\prod_{i=1}^{k-1} \mathbf{1}_{E_i} \circ X_{n+i} \right) E^{\mathcal{P}_{n+k-1}}(\mathbf{1}_{E_k} \circ X_{n+k}) \right];$$

or $E^{\mathcal{P}_{n+k-1}}(\mathbf{1}_{E_k} \circ X_{n+k})$ étant \mathcal{P}_{n+k-1} -mesurable bornée, il existe une fonction réelle (trivialement mesurable) sur E telle que

$$E^{\mathcal{P}_{n+k-1}}(\mathbf{1}_{E_k} \circ X_{n+k}) = g_1(X_{n+k-1}),$$

si bien que l'on a l'égalité

$$E^{\mathcal{A}_n}(Y) = E^{\mathcal{A}_n} \left[\left(\prod_{i=1}^{k-2} \mathbf{1}_{E_i} \circ X_{n+i} \right) (g_1(X_{n+k-1}) \mathbf{1}_{E_{k-1}} \circ X_{n+k-1}) \right].$$

En conditionnant successivement par rapport aux tribus $\mathcal{A}_{n+k-2}, \dots, \mathcal{A}_{n+1}$ et en appliquant l'hypothèse (16.8), puis, en arguant des mêmes arguments de mesurabilité, on obtient, par induction, l'existence de fonctions réelles g_1, g_2, \dots, g_{k-1} bornées sur E telles que l'on ait

$$\begin{aligned} E^{\mathcal{A}_n}(Y) &= E^{\mathcal{A}_n} \left[\left(\prod_{i=1}^{k-2} \mathbf{1}_{E_i} \circ X_{n+i} \right) E^{\mathcal{A}_{n+k-2}}[g_1(X_{n+k-1}) \mathbf{1}_{E_{k-1}} \circ X_{n+k-1}] \right] \\ &= E^{\mathcal{A}_n} \left[\left(\prod_{i=1}^{k-2} \mathbf{1}_{E_i} \circ X_{n+i} \right) E^{\mathcal{P}_{n+k-2}}[g_1(X_{n+k-1}) \mathbf{1}_{E_{k-1}} \circ X_{n+k-1}] \right] \\ &= E^{\mathcal{A}_n} \left[\left(\prod_{i=1}^{k-2} \mathbf{1}_{E_i} \circ X_{n+i} \right) g_2(X_{n+k-2}) \right] \\ &= \dots \\ &= E^{\mathcal{A}_n} [g_{k-1}(X_{n+1})] \\ &= E^{\mathcal{P}_n} [g_{k-1}(X_{n+1})]; \end{aligned}$$

ainsi $E^{\mathcal{A}_n}(Y)$ est \mathcal{P}_n -mesurable, et puisque $\mathcal{P}_n \subset \mathcal{A}_n$, il en résulte que $E^{\mathcal{A}_n}(Y) = E^{\mathcal{P}_n}(Y)$.

En particulier, si le processus X est une chaîne de Markov par rapport à sa **filtration naturelle** $(\mathcal{B}_n)_{n \in \mathbb{N}}$, pour tout $n \in \mathbb{N}$, toute suite croissante finie d'entiers telle que $n_1 \leq n_2 \leq \dots \leq n_k \leq n$ et toute fonction bornée f sur E , on a $\sigma(X_{n_k}) \subset \sigma(X_{n_1}, \dots, X_{n_k}) \subset \mathcal{B}_{n_k}$, et donc

$$E^{\sigma(X_{n_1}, \dots, X_{n_k})}[f(X_n)] = E^{\sigma(X_{n_1}, \dots, X_{n_k})} \left[E^{\mathcal{B}_{n_k}}[f(X_n)] \right] = E^{\sigma(X_{n_k})}[f(X_n)],$$

ce qui démontre l'égalité (16.10). Inversement, si cette dernière propriété est satisfaite, en prenant la suite des entiers consécutifs jusqu'à n , on obtient que l'égalité (16.8) est satisfaite et donc que X est une chaîne de Markov. \square

Voici une situation fréquente dans les applications (elle peut bien sûr se généraliser dans différentes directions).

Exemple 16.6. Soient un ensemble dénombrable E , muni de la tribu de ses parties, et une application mesurable g de $E \times \mathbb{R}$ dans E . On considère une famille de variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{A}, P) et indépendantes; l'une de ces variables aléatoires, X_0 , est à valeurs dans E , les autres forment une suite de variables aléatoires réelles $(U_n)_{n \in \mathbb{N}^*}$ de même loi μ . En particulier, si μ est la loi uniforme sur l'intervalle $[0, 1]$, les U_n peuvent modéliser les **tirages de nombres aléatoires** faits à l'aide d'un générateur de nombres aléatoires lors d'une **simulation**. On construit la suite $(X_n)_{n \in \mathbb{N}}$ en posant

$$X_{n+1} = g(X_n, U_{n+1}).$$

Le processus $X = (X_n)_{n \in \mathbb{N}}$ est dit **auto-régressif**. (On voit bien l'apport du hasard à chaque étape.)

Notons \mathcal{A}_0 la tribu $\sigma(X_0)$ et, si $n \geq 1$, \mathcal{A}_n la tribu $\sigma(X_0, U_1, \dots, U_n)$, tribu qui résume naturellement l'information sur le passé jusqu'au temps n . Le processus X est une chaîne de Markov (homogène⁴) relativement à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$. En effet, par une récurrence facile, on voit que X_n est \mathcal{A}_n -mesurable; autrement dit, X est **adapté** à cette filtration. De plus, pour tout $f \in bE$, on a

$$E^{\mathcal{A}_n} [f(X_{n+1})] = E^{\mathcal{A}_n} [f(g(X_n, U_{n+1}))].$$

Puisque U_{n+1} et \mathcal{A}_n sont indépendantes, il résulte de la proposition 11.22 du chapitre 11 que

$$E^{\mathcal{A}_n} [f(X_{n+1})] = \widehat{f}(X_n) \quad \text{P-p.s.}, \quad (16.11)$$

où la fonction \widehat{f} est définie, pour tout $x \in E$, par

$$\widehat{f}(x) = E[f(x, U_{n+1})] = \int_{\mathbb{R}} f(x, u) d\mu(u). \quad (16.12)$$

L'égalité (16.11) implique l'égalité (16.8), à savoir

$$E^{\mathcal{A}_n} [f(X_{n+1})] = E^{\mathcal{P}^n} [f(X_{n+1})],$$

ce qui démontre que X est une chaîne de Markov par rapport à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$, et donc aussi une chaîne de Markov par rapport à sa filtration naturelle.

Remarque. Puisque E est dénombrable, l'égalité (16.10) est équivalente à l'égalité, pour $P_{(X_{n_1}, \dots, X_{n_k})}$ -presque tous $(x_{n_1}, \dots, x_{n_k}) \in E^k$ et tout $x_n \in E$, des probabilités conditionnelles

$$P^{(X_{n_1}=x_{n_1}, \dots, X_{n_k}=x_{n_k})}(X_n = x_n) = P^{(X_{n_k}=x_{n_k})}(X_n = x_n). \quad (16.13)$$

4. La définition d'une chaîne de Markov **homogène** est donnée ci-après.

Ainsi, X est une chaîne de Markov (relativement à sa filtration naturelle) si et seulement si, pour tous $n, k \in \mathbb{N}$ et pour toute suite croissante finie d'entiers telle que $n_1 \leq n_2 \leq \dots \leq n_k \leq n$, l'égalité (16.13) est satisfaite pour $P_{(X_{n_1}, \dots, X_{n_k})}$ -presque tous $(x_{n_1}, \dots, x_{n_k}) \in E^k$ et tout $x_n \in E$.

Pour des entiers n et m tels que $n \leq m$ et en tous points x, y de l'espace d'états E où cela a un sens, la probabilité conditionnelle $M_{n,m}(x, y) = P^{(X_n=x)}(X_m = y)$ est la **probabilité de transition** de l'état x à l'instant n à l'état y à l'instant m ; on a alors $\sum_{y \in E} P^{(X_n=x)}(X_m = y) = 1$. Pour faciliter le traitement des chaînes de Markov, on est alors amené naturellement aux définitions suivantes :

Définition 16.6. Une famille A de réels positifs ou nuls, bi-indexée sur E est une **matrice de transition**⁵ (ou **matrice stochastique**) si, pour tout $x \in E$, $\sum_{y \in E} A(x, y) = 1$.

Notation. On note bE l'ensemble des fonctions bornées sur E . Si A est une matrice bi-indexée sur E à termes positifs, $A(x, \cdot)$ engendre une mesure μ_x sur E ; si f est positive ou μ_x -intégrable, on note classiquement $A(x, f)$ ou $Af(x)$ son intégrale par rapport à μ_x ; autrement dit

$$A(x, f) = \sum_{y \in E} A(x, y)f(y). \quad (16.14)$$

Si de plus A est telle que les mesures μ_x sont toutes de masse inférieure ou égale à 1 et si f est bornée, la fonction $A(\cdot, f)$ l'est aussi. Ce point de vue fonctionnel se généralise à un espace d'états quelconque. Toutefois ici, E étant dénombrable, il sera souvent utile pour les calculs explicites (tout particulièrement quand E est fini), d'avoir un point de vue vectoriel : on identifiera la fonction f au « vecteur colonne » $f(y)_{y \in E}$, le « vecteur colonne » $A(\cdot, f)$ étant alors le vecteur de composantes $A(x, f)$ données par l'égalité (16.14); ainsi, moyennant ces identifications, on a l'égalité vectorielle $A(\cdot, f) = Af$.

Définition 16.7. Une chaîne de Markov X adaptée à la filtration $(\mathcal{A}_n)_{n \in \mathbb{I}}$ admet une famille de matrices de transition $(M_{n,m})_{n \leq m}$ si, pour tous n et m tels que $n < m$, $M_{n,m}$ est une matrice de transition et si, pour toute fonction f sur E positive ou bornée, on a P-p.s.

$$\mathbb{E}^{\mathcal{A}_n} [f(X_m)] = M_{n,m}(X_n, f). \quad (16.15)$$

Remarque. Pour P_{X_n} -presque tous $x \in E$ et tous $y \in E$, on a alors l'égalité

$$M_{n,m}(x, y) = P^{(X_n=x)}(X_m = y),$$

5. Si E est fini, il s'agit d'une matrice carrée classique, et si E est infini, il s'agit d'une matrice généralisée.

et, si $0 \leq n_1 \leq n_2 \leq \dots \leq n_k = n \leq m$, pour $P_{(X_0, \dots, X_{n_1}, \dots, X_{n_k})}$ -presque tous $(x_0, x_{n_1}, \dots, x_n) \in E^{k+1}$ et tout $x_{n+1} \in E$, on a

$$\begin{aligned} P^{(X_0=X_0, X_{n_1}=x_{n_1}, \dots, X_n=x_n)}(X_{n+1}=x_{n+1}) &= P^{(X_n=x_n)}(X_{n+1}=x_{n+1}) \\ &= M_{n, n+1}(x_n, x_{n+1}). \end{aligned}$$

En effet, puisque $\sigma(X_n) \subset \sigma(X_0, X_{n_1}, \dots, X_n) \subset \mathcal{A}_n$, on a, en prenant pour f la fonction $\mathbf{1}_{\{y\}}$,

$$\begin{aligned} P^{\sigma(X_0, X_{n_1}, \dots, X_n)}(X_m=y) &= E^{\sigma(X_0, X_{n_1}, \dots, X_n)} [\mathbf{1}_{(X_m=y)}] \\ &= E^{\sigma(X_0, X_{n_1}, \dots, X_n)} [E^{\mathcal{A}_n} [\mathbf{1}_{(X_m=y)}]] \\ &= M_{n, m}(X_n, y); \end{aligned}$$

il en résulte que

$$P^{\sigma(X_n)}(X_m=y) = E^{\sigma(X_n)} [E^{\sigma(X_0, X_{n_1}, \dots, X_n)} [\mathbf{1}_{(X_m=y)}]] = M_{n, m}(X_n, y).$$

Remarque. Une matrice de transition A sur E est telle que, pour tout $x \in E$, l'application $A(x, \cdot)$ est un germe de probabilité sur l'ensemble dénombrable E ; en identifiant ce germe et la probabilité engendrée, A peut être vue comme une **probabilité (ou noyau) de transition** sur E , version régulière de la loi conditionnelle de X_m sachant $(X_0, X_{n_1}, \dots, X_n)$. C'est ce point de vue qui permet de traiter les chaînes de Markov à espace d'état général.

Proposition 16.8 (Égalité de Chapman-Kolmogorov). *Soit X une chaîne de Markov adaptée à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$, de famille de matrices de transition $(M_{n, m})_{n \leq m}$; pour tous instants n, r, m tels que $n < r < m$, on a la relation de Chapman-Kolmogorov :*

$$\text{pour tout } y \in E \quad M_{n, m}(X_n, y) = \sum_{z \in E} M_{n, r}(X_n, z) M_{r, m}(z, y).$$

ce qui s'écrit sous forme matricielle

$$\boxed{M_{n, m}(X_n, \cdot) = M_{n, r} M_{r, m}(X_n, \cdot)}. \quad (16.16)$$

Démonstration. Le système des événements $(X_r = z)$, $z \in E$, est un système complet de constituants; en prenant pour f la fonction $\mathbf{1}_{\{y\}}$ dans l'égalité (16.15), on a alors

$$\begin{aligned} M_{n, m}(X_n, y) &= E^{\mathcal{A}_n} [\mathbf{1}_{(X_m=y)}] = E^{\mathcal{A}_n} \left[\sum_{z \in E} \mathbf{1}_{(X_r=z)} \mathbf{1}_{(X_m=y)} \right] \\ &= \sum_{z \in E} E^{\mathcal{A}_n} [\mathbf{1}_{(X_r=z)} \mathbf{1}_{(X_m=y)}] \end{aligned}$$

soit, puisque $(X_r = z) \in \mathcal{A}_r$,

$$\begin{aligned}
M_{n,m}(X_n, y) &= \sum_{z \in E} E^{\mathcal{A}_n} \left[\mathbf{1}_{(X_r=z)} (E^{\mathcal{A}_r} \mathbf{1}_{(X_n=y)}) \right] \\
&= \sum_{z \in E} E^{\mathcal{A}_n} \left[\mathbf{1}_{(X_r=z)} M_{r,m}(X_r, y) \right] \\
&= \sum_{z \in E} E^{\mathcal{A}_n} \left[\mathbf{1}_{(X_r=z)} M_{r,m}(z, y) \right] \\
&= \sum_{z \in E} \left[E^{\mathcal{A}_n} \mathbf{1}_{(X_r=z)} M_{r,m}(z, y) \right] \\
&= \sum_{z \in E} [M_{n,r}(X_n, z) M_{r,m}(z, y)] . \quad \square
\end{aligned}$$

Remarque. Si X est simplement une chaîne de Markov relativement à sa filtration naturelle, on peut donner la **démonstration heuristique**⁶ suivante de l'égalité de Chapman-Kolmogorov ; on a, d'après la formule des probabilités totales, et d'après la propriété de Markov,

$$P(X_m = y \mid X_n = x) = \sum_{z \in E} P(X_m = y \mid X_r = z) P(X_r = z \mid X_n = x).$$

Un système physique dont l'état dépend du temps est **conservatif** si, pour tous t , sa loi de passage d'un état x à l'instant s à l'état y à l'instant $s + t$ est indépendante de s . Si le système est modélisé par une chaîne de Markov, cette notion va se traduire par une homogénéité dans le temps des lois conditionnelles, et donc de la famille des matrices de transition. Ainsi, une chaîne de Markov X de famille de matrices de transition $(M_{n,m})_{n \leq m}$ est (temporellement) **homogène** s'il existe une suite de matrices de transition $(M_{(n)})_{n \in \mathbb{N}}$ indexées sur E telle que l'on ait, pour tous entiers n et m , $M_{n,n+m} = M_{(m)}$. Il résulte alors de (16.16) que pour tous entiers n et m , $M_{n,n+m}(X_n, \cdot) = M_{(1)}^m(X_n, \cdot)$; en particulier, la matrice $M_{(1)}$ est notée M et appelée **matrice de transition** de la chaîne de Markov homogène. On a alors

$$E^{\mathcal{A}_n} [f(X_{n+m})] = M^m(X_n, f),$$

où M^m est la m -ième puissance de la matrice M . En particulier, pour P_{X_0} -presque tous $x \in E$ et tout $y \in E$, on a alors l'égalité

$$M(x, y) = P^{(X_0=x)}(X_1 = y),$$

et pour P_{X_n} -presque tous $x \in E$ et tout $y \in E$, on a l'égalité

$$\boxed{M(x, y) = P^{(X_n=x)}(X_{n+1} = y)}.$$

On est ainsi conduit à la définition suivante :

6. Au sens où on fait abstraction des problèmes de division par 0.

Définition 16.9. Une chaîne de Markov X (relativement à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$), à valeurs dans E , est **homogène de matrice de transition** M si, pour tous entiers n et m tels que $0 \leq n < m$, on a

$$\mathbb{E}^{\mathcal{A}_n} [f(X_m)] = M^{m-n}(X_n, f). \quad (16.17)$$

La proposition suivante permet de **démontrer qu'un processus est une chaîne de Markov homogène de matrice de transition** M .

Proposition 16.10. Le processus X , adapté à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$, est une chaîne de Markov homogène de matrice de transition M si et seulement si pour tous entiers $n \in \mathbb{N}$, on a

$$\mathbb{E}^{\mathcal{A}_n} [f(X_{n+1})] = M(X_n, f). \quad (16.18)$$

Démonstration. La condition nécessaire est triviale. Inversement, supposons que la relation (16.18) soit vraie pour tout entier $n \in \mathbb{N}$. Soient n et m tels que $0 \leq n < m$; on a

$$\begin{aligned} \mathbb{E}^{\mathcal{A}_n} [f(X_m)] &= \mathbb{E}^{\mathcal{A}_n} [\mathbb{E}^{\mathcal{A}_{m-1}} f(X_m)] \\ &= \mathbb{E}^{\mathcal{A}_n} [M(X_{m-1}, f)] \\ &= \mathbb{E}^{\mathcal{A}_n} [\mathbb{E}^{\mathcal{A}_{m-2}} M(X_{m-1}, f)] \\ &= \mathbb{E}^{\mathcal{A}_n} [M(X_{m-2}, M(\cdot, f))] \\ &= \mathbb{E}^{\mathcal{A}_n} [M^2(X_{m-2}, f)]. \end{aligned}$$

Une récurrence facile donne alors la relation (16.17). \square

Ainsi, le processus **auto-régressif** introduit à l'exemple 16.6 est une chaîne de Markov homogène de matrice de transition M déterminée, pour toute fonction bornée f , par

$$M(x, f) = \hat{f}(x) = \int_{\mathbb{R}} f(x, u) d\mu(u).$$

Voici un exemple de chaîne de Markov homogène relativement à une **filtration qui n'est pas sa filtration naturelle**.

Exemple 16.7. (Marche aléatoire conditionnelle.) On considère une famille de variables aléatoires réelles définies sur un même espace probabilisé (Ω, \mathcal{A}, P) et indépendantes; l'une de ces variables aléatoires, Θ , est à valeurs dans l'intervalle $[0, 1]$ et de loi μ , les autres forment une suite de variables aléatoires réelles $(U_n)_{n \in \mathbb{N}}$ de loi uniforme sur l'intervalle $[0, 1]$, modélisant les tirages de nombres aléatoires faits à l'aide d'un générateur de nombres aléatoires. On construit la suite $(X_n)_{n \in \mathbb{N}}$ en posant

$$X_n = \mathbf{1}_{(U_n \leq \Theta)} - \mathbf{1}_{(U_n > \Theta)},$$

autrement dit, conditionnellement à la valeur θ de Θ préalablement tirée, la variable aléatoire X_n suit la loi $\theta\delta_1 + (1 - \theta)\delta_{-1}$. On définit, pour tout $n \in \mathbb{N}$, les variables aléatoires

$$S_n = \sum_{j=0}^n X_j \quad \text{et} \quad Y_n = (\Theta, S_n),$$

et on note $\mathcal{A}_n = \sigma(\Theta, U_0, U_1, \dots, U_n)$ la tribu engendrée par $\Theta, U_0, U_1, \dots, U_n$ (\mathcal{A}_n est la tribu qui résume naturellement l'information sur le passé jusqu'au temps n). Pour rester dans le cadre des chaînes à espace dénombrable, on suppose que la loi μ est à support dans une partie dénombrable E de $[0, 1]$. Pour une valeur θ donnée, la suite $(S_n)_{n \in \mathbb{N}}$ se comporte comme une marche aléatoire sur \mathbb{Z} , et cette valeur de θ doit rester *continuellement* en mémoire pour pouvoir poursuivre la marche. Ainsi, le processus $(S_n)_{n \in \mathbb{N}}$ a une histoire qui dépend constamment de l'instant initial; toutefois, nous allons montrer que le processus $Y = (Y_n)_{n \in \mathbb{N}}$ est une chaîne de Markov homogène à valeurs dans $E \times \mathbb{Z}$, adaptée à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$.

Pour cela, on calcule, pour toute fonction bornée f sur $E \times \mathbb{Z}$,

$$E^{\mathcal{A}_n} [f(Y_{n+1})] = E^{\mathcal{A}_n} [f(\Theta, S_n + X_{n+1})].$$

Définissons la fonction h sur $[0, 1]^2$ par

$$h(\theta, u) = \mathbf{1}_{(u \leq \theta)} - \mathbf{1}_{(u > \theta)}.$$

On a, pour $P_{(\Theta, U_0, U_1, \dots, U_n)}$ -presque tout $(\theta, u_0, u_1, \dots, u_n)$,

$$\begin{aligned} E^{\sigma(\Theta, U_0, U_1, \dots, U_n) = (\theta, u_0, u_1, \dots, u_n)} [f(Y_{n+1})] \\ = E^{\sigma(\Theta, U_0, U_1, \dots, U_n) = (\theta, u_0, u_1, \dots, u_n)} \left[f\left(\theta, \sum_{j=0}^n h(\theta, u_j) + h(\theta, U_{n+1})\right) \right], \end{aligned}$$

soit, par indépendance des variables aléatoires $\Theta, U_0, U_1, \dots, U_n$,

$$E^{\sigma(\Theta, U_0, U_1, \dots, U_n) = (\theta, u_0, u_1, \dots, u_n)} [f(Y_{n+1})] = E \left[f\left(\theta, \sum_{j=0}^n h(\theta, u_j) + h(\theta, U_{n+1})\right) \right].$$

Puisque U_{n+1} est de loi uniforme sur $[0, 1]$, on a

$$\begin{aligned} E^{\sigma(\Theta, U_0, U_1, \dots, U_n) = (\theta, u_0, u_1, \dots, u_n)} [f(Y_{n+1})] \\ = \theta f\left(\theta, \sum_{j=0}^n h(\theta, u_j) + 1\right) + (1 - \theta) f\left(\theta, \sum_{j=0}^n h(\theta, u_j) - 1\right). \end{aligned}$$

On a donc, P-p.s.,

$$E^{\mathcal{A}_n} [f(Y_{n+1})] = \Theta f\left(\Theta, \sum_{j=0}^n h(\Theta, U_j) + 1\right) + (1 - \Theta) f\left(\Theta, \sum_{j=0}^n h(\Theta, U_j) - 1\right).$$

soit

$$E^{\mathcal{A}_n} [f(Y_{n+1})] = \Theta f(\Theta, S_n + 1) + (1 - \Theta) f(\Theta, S_n - 1) \quad \text{P-p.s.}$$

En définissant la **matrice de transition** M sur $E \times \mathbb{Z}$ par

$$M((\theta, s), f) = \theta f(\theta, s + 1) + (1 - \theta) f(\theta, s - 1),$$

on a alors

$$E^{\mathcal{A}_n} [f(Y_{n+1})] = M(Y_n, f) \quad \text{P-p.s.},$$

ce qui démontre que le **processus** Y est une **chaîne de Markov homogène de matrice de transition** M relativement à la filtration $(\mathcal{A}_n)_{n \in \mathbb{N}}$.

Par contre, voici un exemple de chaîne de Markov **non homogène**.

Exemple 16.8. Reprenons l'exemple 16.3 de **modèle de diffusion de maladies contagieuses de Pólya** sous sa forme de tirages de boules dans une urne et montrons que le processus des **proportions** Y_n de boules blanches contenues dans l'urne **après le n -ième tirage et après avoir rajouté les c boules** est une chaîne de Markov **non homogène** et aussi une **martingale**.

On considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires définies sur un espace probabilisé (Ω, \mathcal{A}, P) , à valeurs 0 ou 1 (X_n prend la valeur 0 ou 1 suivant que la n -ième boule tirée est rouge ou blanche). On note $k_n = b + r + nc$; le nombre B_n et la proportion Y_n de boules blanches situées dans l'urne après le n -ième tirage sont respectivement

$$B_n = b + c \sum_{j=1}^n X_j \quad \text{et} \quad Y_n = \frac{B_n}{k_n}.$$

On a

$$P(X_1 = 1) = \frac{b}{b+r} \quad \text{et} \quad P(X_1 = 0) = \frac{r}{b+r}.$$

De plus, les tirages étant tous uniformes, on a, pour tout $n \geq 2$, et pour tout $(x_1, \dots, x_n) \in \{0, 1\}^n$,

$$\begin{cases} P^{(X_1, \dots, X_n) = (x_1, \dots, x_n)}(X_{n+1} = 1) = \frac{b + c \sum_{j=1}^n x_j}{k_n}, \\ P^{(X_1, \dots, X_n) = (x_1, \dots, x_n)}(X_{n+1} = 0) = \frac{r + c(n - \sum_{j=1}^n x_j)}{k_n}. \end{cases} \quad (16.19)$$

Le processus $Y = (Y_n)_{n \in \mathbb{N}^*}$ est à valeurs dans

$$E = \bigcup_{j \in \mathbb{N}^*} \left\{ \frac{j}{k_n} \mid b \leq j \leq k_n \right\},$$

partie infinie dénombrable de l'intervalle $[0, 1]$. Notons d'abord que les tribus $\mathcal{A}_n = \sigma(X_j \mid 1 \leq j \leq n)$ et $\sigma(Y_j \mid 1 \leq j \leq n)$ coïncident. En

effet, il est facile de voir que l'application F_n de \mathbb{R}^n dans lui-même définie par

$$F_n(x_1, \dots, x_n) = (y_1, \dots, y_n) \quad \text{où} \quad y_l = \frac{b + c \sum_{j=1}^l x_j}{k_l}, \quad 1 \leq l \leq n,$$

est une bijection et que $(Y_1, \dots, Y_n) = F_n(X_1, \dots, X_n)$.

Il résulte alors de (16.19) que

$$P^{(Y_1, \dots, Y_n) = (y_1, \dots, y_n)}(X_{n+1} = 1) = P^{(X_1, \dots, X_n) = F_n^{-1}(y_1, \dots, y_n)}(X_{n+1} = 1) = y_n, \quad (16.20)$$

ce qui donne

$$P^{\sigma(Y_1, \dots, Y_n)}(X_{n+1} = 1) = Y_n \quad \text{et} \quad P^{\sigma(Y_1, \dots, Y_n)}(X_{n+1} = 0) = 1 - Y_n. \quad (16.21)$$

Un calcul simple montre que

$$k_{n+1} = k_n + c \quad \text{et} \quad Y_{n+1} = \frac{k_n Y_n + c X_{n+1}}{k_{n+1}}. \quad (16.22)$$

Ainsi, pour toute fonction $f \in bE$, on a

$$E^{\mathcal{A}_n}[f(Y_{n+1})] = E^{\mathcal{A}_n} \left[\mathbf{1}_{(X_{n+1}=1)} f\left(\frac{k_n Y_n + c}{k_{n+1}}\right) \right] + E^{\mathcal{A}_n} \left[\mathbf{1}_{(X_{n+1}=0)} f\left(\frac{k_n Y_n}{k_{n+1}}\right) \right],$$

ce qui donne, en tenant compte de (16.20) :

$$E^{\mathcal{A}_n} [f(Y_{n+1})] = f\left(\frac{k_n Y_n + c}{k_{n+1}}\right) Y_n + f\left(\frac{k_n Y_n}{k_{n+1}}\right) (1 - Y_n).$$

En posant, pour tout $y \in E$ et toute $f \in bE$,

$$M_n(y, f) = f\left(\frac{k_n y + c}{k_{n+1}}\right) y + f\left(\frac{k_n y}{k_{n+1}}\right) (1 - y),$$

on a montré que, pour tout $n \in \mathbb{N}^*$,

$$E^{\mathcal{A}_n} [f(Y_{n+1})] = M_n(Y_n, f). \quad (16.23)$$

La matrice M_n est bien une matrice de transition sur E puisque

$$\sum_{z \in E} M_n(y, z) = \sum_{z \in E} \mathbf{1}_{\left(\frac{k_n y + c}{k_{n+1}} = z\right)} y + \sum_{z \in E} \mathbf{1}_{\left(\frac{k_n y}{k_{n+1}} = z\right)} (1 - y) = y + (1 - y) = 1.$$

Le processus Y est une chaîne de Markov **non homogène** de famille de matrices de transition $(M_n)_{n \in \mathbb{N}^*}$. C'est de plus une **martingale**. En effet, en prenant pour f , dans l'égalité (16.23), l'application identique sur E (qui est bornée), on a

$$E^{\mathcal{A}_n}(Y_{n+1}) = \frac{k_n Y_n + c}{k_{n+1}} Y_n + \frac{k_n Y_n}{k_{n+1}} (1 - Y_n) = Y_n \frac{k_n + c}{k_{n+1}} = Y_n.$$

Il résulte alors de cette propriété de martingale que

$$E(Y_{n+1}) = EY_1 = E \left[\frac{b + cX_1}{b + r + c} \right] = \frac{1}{b + r + c} \left(b + c \frac{b}{b + r} \right),$$

soit

$$E(Y_{n+1}) = \frac{b}{b + r}.$$

La variable aléatoire X_{n+1} ne prend que les valeurs 0 ou 1; sa loi est donc déterminée par sa moyenne, que l'on calcule en utilisant les relations (16.21). On a

$$P(X_{n+1} = 1) = E(X_{n+1}) = E \left[P^{\sigma(Y_1, \dots, Y_n)}(X_{n+1} = 1) \right] = E(Y_n),$$

et donc

$$P(X_{n+1} = 1) = \frac{b}{b + r}.$$

On vient de montrer que **la loi de X_n , pour $n \geq 1$, est indépendante de n et de c** , ce qui n'est, a priori, ni évident, ni intuitif.

La martingale $Y = (Y_n)_{n \in \mathbb{N}^*}$ est bornée; elle converge donc P-p.s. et dans tout L^p vers une variable aléatoire Y_∞ . La loi de Y_∞ est la **loi bêta** $\beta\left(\frac{b}{c}, \frac{r}{c}\right)$ de première espèce sur $[0, 1]$. Une démonstration de ce fait est proposée en exercice en fin de chapitre (ex. 12).

Remarque. Étant donnée une chaîne de Markov homogène de matrice de transition M , pour tous $n \in \mathbb{N}$, pour P_{X_n} -presque tout $x \in E$ et tout $y \in E$, on a l'égalité

$$M(x, y) = P^{(X_n = x)}(X_{n+1} = y).$$

On peut alors reformuler la proposition 16.8 pour les chaînes de Markov **homogènes**.

Proposition 16.11 (Égalité de Chapman-Kolmogorov). *Soit X une chaîne de Markov homogène de matrice de transition M ; pour tous instants $n, r, n + m$ tels que $0 \leq n < r < n + m$, on a la relation de **Chapman-Kolmogorov** :*

$$\text{pour tout } y \in E \quad M^m(X_n, y) = \sum_{z \in E} M^{r-n}(X_n, z) M^{n+m-r}(z, y),$$

ce qui s'écrit sous forme matricielle

$$M^m(X_n, \cdot) = M^{r-n} M^{n+m-r}(X_n, \cdot);$$

en particulier, on a, pour $P_{(X_0, \dots, X_n)}$ -presque tous (x_0, \dots, x_n) et tout $y \in E$,

$$P^{(X_0 = x_0, \dots, X_n = x_n)}(X_{n+m} = y) = P^{(X_n = x_n)}(X_{n+m} = y) = M^m(x, y). \quad (16.24)$$

Proposition 16.12. *Un processus X à valeurs dans E est une chaîne de Markov (relativement à sa filtration naturelle) homogène de matrice de transition M si et seulement si, pour $P_{(X_0, \dots, X_n)}$ -presque tous (x_0, \dots, x_{n-1}, x) et pour tout $y \in E$, on a*

$$P^{(X_0=x_0, \dots, X_n=x)}(X_{n+m} = y) = P^{(X_n=x)}(X_{n+m} = y) = M^m(x, y). \quad (16.25)$$

Démonstration. Il est clair que la condition est nécessaire. Pour la réciproque, on a, si $0 \leq n < m$,

$$\begin{aligned} E^{(X_0=x_0, \dots, X_n=x)} [f(X_m)] &= E^{(X_n=x)} [f(X_m)] \\ &= \sum_{y \in E} [P^{(X_0=x_0, \dots, X_n=x)}(X_m = y) f(y)] = M^{m-n}(x, f), \end{aligned}$$

et, puisque

$$E^{\mathcal{B}_n} [f(X_m)] = E^{(X_0=\cdot, \dots, X_n=\cdot)} [f(X_m)] \circ (X_0, \dots, X_n),$$

on a

$$E^{\mathcal{B}_n} [f(X_m)] = M^{m-n}(X_n, f),$$

ce qui démontre que X est une chaîne de Markov homogène de matrice de transition M . \square

Remarque. Si X est une chaîne de Markov homogène de matrice de transition M et si f est une fonction positive ou bornée telle que $M(\cdot, f) = f$ (une telle fonction est dite **harmonique**), autrement dit si, sous forme vectorielle, f est **vecteur propre** à droite de M associé à la valeur propre 1, on a

$$E^{\mathcal{A}_n} [f(X_{n+1})] = M(X_n, f) = f(X_n),$$

et le processus $(f(X_n))_{n \in \mathbb{N}}$ est une **martingale**.

16.3.2. Propriété de Markov simple ; lois fini-dimensionnelles

On généralise maintenant la formule (16.17) au cas d'une fonctionnelle quelconque du futur⁷ d'une chaîne de Markov **homogène** après l'instant n ; la propriété ainsi mise en évidence est appelée **propriété de Markov simple** (ou **faible**); elle traduit l'homogénéité temporelle de la chaîne X et dit qu'à tout instant n , l'espérance conditionnelle d'une fonctionnelle du futur de la chaîne à partir de cet instant est la valeur moyenne de cette même fonctionnelle évaluée sur toute la vie (à partir de l'instant 0) d'une chaîne de même matrice de transition que X , mais qui, à l'instant 0, vaudrait X_n . Pour

7. On entend par là, une variable aléatoire mesurable par rapport à la tribu du **futur large** après l'instant n ; elle s'écrit sous la forme $f(X_n, X_{n+1}, \dots)$, où f est une fonction mesurable définie sur $E^{\mathbb{N}}$. Le premier temps de passage du processus X en un point après le temps n en est un exemple.

bien formuler cette propriété, on introduit les **opérateurs de translation** θ_n , $n \in \mathbb{N}$, de l'espace $E^{\mathbb{N}}$ des suites à valeurs dans E dans lui-même, définis par, pour tout $y \in E^{\mathbb{N}}$,

$$\theta_n(y) = (y_{n+p})_{p \in \mathbb{N}};$$

la suite $\theta_n(y)$ est la suite y à laquelle on a enlevé les termes y_0, \dots, y_{n-1} . Une **fonctionnelle du futur** d'une chaîne de Markov X **après l'instant** n s'écrit alors $f(\theta_n(X))$, où f est une fonction définie sur $E^{\mathbb{N}}$.

Proposition 16.13 (Propriété de Markov simple). *Soit X une chaîne de Markov homogène de matrice de transition M ; pour tout instant n et toute fonction f sur $E^{\mathbb{N}}$, $\mathcal{E}^{\otimes \mathbb{N}}$ -mesurable positive ou bornée, on a*

$$E^{\mathcal{A}^n} [f(\theta_n(X))] = g(X_n), \quad (16.26)$$

où g est une fonction mesurable sur E définie, pour tout x tel que $P(X_0 = x) > 0$, par

$$g(x) = E^{X=x} [f(X)];$$

elle satisfait en particulier l'égalité

$$g(X_0) = E^{\mathcal{A}^0} [f(X)]. \quad (16.27)$$

Démonstration. On donne deux démonstrations de cette propriété. La première n'est qu'heuristique, puisqu'elle fait abstraction des difficultés apportées par les événements de probabilité nulle; le temps y est utilisé dans son sens naturel. La deuxième est rigoureuse mais plus formelle; le temps y est utilisé dans le sens rétrograde.

• **Démonstration heuristique** : par un argument de prolongement par mesurabilité (utiliser le théorème 8.6, chap. 8), il suffit de démontrer (16.26) pour une fonctionnelle f du type $y \mapsto f_k(y_0, \dots, y_k)$, où $k \in \mathbb{N}$ et f_k est une fonction sur E^{k+1} . Pour une telle fonctionnelle f , l'égalité (16.7) du théorème 16.5 permet d'écrire que

$$E^{\mathcal{A}^n} [f(\theta_n(X))] = E^{\mathcal{P}^n} [f(\theta_n(X))] = E^{\mathcal{P}^n} [f_k(X_n, \dots, X_{n+k})]; \quad (16.28)$$

mais, pour P_{X_n} -presque tout $x \in E$, on a

$$E^{(X_n=x)} [f_k(X_n, \dots, X_{n+k})] = \sum_{(y_1, \dots, y_k) \in E^k} f_k(x, y_1, \dots, y_k) \\ \times P^{(X_n=x)}(X_n = x, X_{n+1} = y_1, \dots, X_{n+k} = y_k).$$

D'après la formule des probabilités conditionnelles en cascade et la propriété de Markov, on a, pour tout $n \in \mathbb{N}$,

$$P^{(X_n=x)}(X_n = x, X_{n+1} = y_1, \dots, X_{n+k} = y_k) \\ = P^{(X_n=x)}(X_{n+1} = y_1) P^{(X_n=x, X_{n+1}=y_1)}(X_{n+2} = y_2) \times \dots \\ \dots \times P^{(X_n=x, \dots, X_{n+k-1}=y_{k-1})}(X_{n+k} = y_k) \\ = P^{(X_n=x)}(X_{n+1} = y_1) P^{(X_{n+1}=y_1)}(X_{n+2} = y_2) \times \dots \\ \dots \times P^{(X_{n+k-1}=y_{k-1})}(X_{n+k} = y_k),$$

et donc

$$P^{(X_n=x)}(X_n=x, X_{n+1}=y_1, \dots, X_{n+k}=y_k) = M(x, y_1)M(y_1, y_2) \dots \times M(y_{k-1}, y_k). \quad (16.29)$$

Il en résulte que, pour tout $n \in \mathbb{N}$,

$$E^{(X_n=x)} [f_k(X_n, \dots, X_{n+k})] = \sum_{(y_1, \dots, y_k) \in E^k} f_k(x, y_1, \dots, y_k) \times M(x, y_1)M(y_1, y_2) \dots M(y_{k-1}, y_k);$$

définissant la fonction g sur E par

$$g(x) = \sum_{(y_1, \dots, y_k) \in E^k} f_k(x, y_1, \dots, y_k) M(x, y_1)M(y_1, y_2) \dots M(y_{k-1}, y_k),$$

on a en particulier, pour $n = 0$,

$$g(x) = E^{(X_0=x)} [f_k(X_0, \dots, X_k)] = E^{(X_0=x)} [f(X)];$$

de plus, il résulte de (16.28) que

$$E^{\mathcal{A}_n} [f(\theta_n(X))] = g(X_n),$$

ce qui démontre le résultat.

• **Démonstration rigoureuse** : par un argument de prolongement par mesurabilité, il suffit de démontrer (16.26) pour une fonctionnelle f du type $y \mapsto \prod_{j=0}^k f_j(y_j)$, où $k \in \mathbb{N}$ et où les f_j sont des fonctions sur E . Le conditionnement par rapport à la tribu \mathcal{A}_{n+k-1} , l'adaptation du processus X et l'utilisation de l'égalité (16.17) permettent d'écrire

$$\begin{aligned} E^{\mathcal{A}_n} [f(\theta_n(X))] &= E^{\mathcal{A}_n} \left[\prod_{j=0}^k f_j(X_{n+j}) \right] \\ &= E^{\mathcal{A}_n} \left[\prod_{j=0}^{k-1} f_j(X_{n+j}) E^{\mathcal{A}_{n+k-1}} f_k(X_{n+k}) \right]. \end{aligned}$$

et donc

$$E^{\mathcal{A}_n} [f(\theta_n(X))] = E^{\mathcal{A}_n} \left[\prod_{j=0}^{k-1} f_j(X_{n+j}) M(X_{n+k-1}, f_k) \right].$$

Le conditionnement par rapport à la tribu \mathcal{A}_{n+k-2} et les mêmes arguments permettent ensuite d'écrire

$$\begin{aligned} E^{\mathcal{A}_n} [f(\theta_n(X))] &= E^{\mathcal{A}_n} \left[\prod_{j=0}^{k-2} f_j(X_{n+j}) E^{\mathcal{A}_{n+k-2}} [M(X_{n+k-1}, f_k) f_{k-1}(X_{n+k-1})] \right] \\ &= E^{\mathcal{A}_n} \left[\prod_{j=0}^{k-2} f_j(X_{n+j}) M(X_{n+k-2}, f_{k-1} M(\cdot, f_k)) \right]. \end{aligned}$$

Mais, pour tout $x \in E$, on a

$$\begin{aligned} M(x, f_{k-1}M(\cdot, f_k)) &= \sum_{y_{k-1} \in E} M(x, y_{k-1}) f_{k-1}(y_{k-1}) M(y_{k-1}, f_k) \\ &= \sum_{y_{k-1} \in E} M(x, y_{k-1}) f_{k-1}(y_{k-1}) \left[\sum_{y_k \in E} M(y_{k-1}, y_k) f_k(y_k) \right] \\ &= \sum_{y_{k-1}, y_k \in E^2} M(x, y_{k-1}) M(y_{k-1}, y_k) f_{k-1}(y_{k-1}) f_k(y_k); \end{aligned}$$

on a alors

$$\begin{aligned} E^{A_n} [f(\theta_n(X))] &= E^{A_n} \left[\prod_{j=0}^{k-2} f_j(X_{n+j}) \right. \\ &\quad \left. \times \sum_{y_{k-1}, y_k \in E^2} M(X_{n+k-2}, y_{k-1}) M(y_{k-1}, y_k) f_{k-1}(y_{k-1}) f_k(y_k) \right]; \end{aligned}$$

Une itération facile de ce procédé conduit à l'égalité

$$\begin{aligned} E^{A_n} [f(\theta_n(X))] &= \sum_{(y_1, \dots, y_k) \in E^k} M(X_n, y_1) M(y_1, y_2) \\ &\quad \dots \times M(y_{k-1}, y_k) f(X_n, y_1, \dots, y_k). \end{aligned}$$

Définissant la fonction g sur E par

$$g(x) = \sum_{(y_1, \dots, y_k) \in E^k} M(x, y_1) M(y_1, y_2) \dots M(y_{k-1}, y_k) f(x, y_1, \dots, y_k).$$

on a alors

$$E^{A_n} [f(\theta_n(X))] = g(X_n);$$

de plus, ce calcul montre clairement que l'on a

$$g(X_0) = E^{A_0} [f(X)],$$

ce qui achève de démontrer le résultat. \square

Les **lois fini-dimensionnelles** (c'est-à-dire les lois de tout vecteur de composantes les états de la chaîne en un nombre fini d'instant) **conditionnelles à l'état initial** X_0 d'une chaîne de Markov **homogène** de matrice de transition M sont entièrement déterminées par la matrice M .

Proposition 16.14. *Soit X une chaîne de Markov homogène de matrice de transition M ; pour tout $x \in E$ tel que $P(X_0 = x) > 0$, on note P_x la probabilité conditionnelle $P^{(X_0=x)}$. Pour toute suite strictement croissante d'instant s_1, s_2, \dots, s_k , et tout x tel que $P(X_0 = x) > 0$, on a, pour tout $(x_1, x_2, \dots, x_k) \in E^k$,*

$$P_x(X_{s_1} = x_1, X_{s_2} = x_2, \dots, X_{s_k} = x_k) = M^{s_1}(x, x_1)M^{s_2-s_1}(x_1, x_2) \dots M^{s_k-s_{k-1}}(x_{k-1}, x_k). \quad (16.30)$$

Il en résulte que

$$P(X_{s_1} = x_1, X_{s_2} = x_2, \dots, X_{s_k} = x_k) = \sum_{x \in E} P(X_0 = x) M^{s_1}(x, x_1) M^{s_2-s_1}(x_1, x_2) \dots M^{s_k-s_{k-1}}(x_{k-1}, x_k). \quad (16.31)$$

Démonstration. Comme pour la proposition 16.13, nous donnons deux démonstrations.

• **Démonstration heuristique** : d'après la formule des probabilités conditionnelles en cascade et la propriété de Markov, on a, pour tout $n \in \mathbb{N}$,

$$\begin{aligned} P_x(X_{s_1}=x_1, X_{s_2}=x_2, \dots, X_{s_k}=x_k) &= P_x(X_{s_1}=x_1)P^{(X_{s_1}=x_1)}(X_{s_2}=x_2) \\ &\dots \times P^{(X_{s_1}=x_1, \dots, X_{s_{k-1}}=x_{k-1})}(X_{s_k}=x_k) \\ &= P_x(X_{s_1}=x_1)P^{(X_{s_1}=x_1)}(X_{s_2}=x_2) \\ &\dots \times P^{(X_{s_{k-1}}=x_{k-1})}(X_{s_k}=x_k), \end{aligned}$$

ce qui démontre l'égalité (16.30), d'après l'égalité de Chapman-Kolmogorov. Il suffit alors d'appliquer la formule des probabilités totales pour obtenir l'égalité (16.31).

• **Démonstration rigoureuse** : le conditionnement par rapport à la tribu $\mathcal{A}_{s_{k-1}}$, l'adaptation du processus X et l'utilisation de l'égalité (16.17) permettent d'écrire

$$\begin{aligned} P_x(X_{s_1} = x_1, X_{s_2} = x_2, \dots, X_{s_k} = x_k) &= E_x \left[\mathbf{1}_{(X_{s_1}=x_1, X_{s_2}=x_2, \dots, X_{s_{k-1}}=x_{k-1})} E_x^{\mathcal{A}_{s_{k-1}}} \mathbf{1}_{(X_{s_k}=x_k)} \right] \\ &= E_x \left[\mathbf{1}_{(X_{s_1}=x_1, X_{s_2}=x_2, \dots, X_{s_{k-1}}=x_{k-1})} M^{s_k-s_{k-1}}(X_{s_{k-1}}, x_k) \right] \\ &= E_x \left[\mathbf{1}_{(X_{s_1}=x_1, X_{s_2}=x_2, \dots, X_{s_{k-1}}=x_{k-1})} \right] M^{s_k-s_{k-1}}(x_{k-1}, x_k); \end{aligned}$$

par itération, on a alors

$$\begin{aligned} P_x(X_{s_1} = x_1, X_{s_2} = x_2, \dots, X_{s_k} = x_k) &= E_x \left[\mathbf{1}_{(X_{s_1}=x_1, X_{s_2}=x_2, \dots, X_{s_{k-1}}=x_{k-1})} \right] M^{s_{k-1}-s_{k-2}}(x_{k-2}, x_{k-1}) M^{s_k-s_{k-1}}(x_{k-1}, x_k) \\ &= \dots \\ &= E_x \left[E^{\mathcal{A}_{01}} \mathbf{1}_{(X_{s_1}=x_1)} \right] M^{s_2-s_1}(x_1, x_2) \dots M^{s_{k-1}-s_{k-2}}(x_{k-2}, x_{k-1}) M^{s_k-s_{k-1}}(x_{k-1}, x_k) \\ &= E_x [M^{s_1}(X_0, x_1)] M^{s_2-s_1}(x_1, x_2) \dots M^{s_{k-1}-s_{k-2}}(x_{k-2}, x_{k-1}) M^{s_k-s_{k-1}}(x_{k-1}, x_k), \end{aligned}$$

ce qui démontre l'égalité (16.30); l'égalité (16.31) s'en déduit immédiatement. \square

On a une caractérisation des chaînes de Markov homogènes relative-
ment à la filtration naturelle.

Proposition 16.15. *Un processus $X = (X_n)_{n \in \mathbb{N}}$ défini sur l'espace probabilisé (Ω, \mathcal{A}, P) est une chaîne de Markov homogène de loi initiale μ (c'est-à-dire telle que, pour tout x , $P(X_0 = x) = \mu(x)$) et de matrice de transition M si et seulement si on a, pour tous $k \in \mathbb{N}^*$ et tous $x_0, x_1, \dots, x_k \in E$,*

$$P(X_0 = x_0, X_1 = x_1, \dots, X_k = x_k) = \mu(x_0)M(x_0, x_1)M(x_1, x_2) \dots M(x_{k-1}, x_k). \quad (16.32)$$

Démonstration. La condition nécessaire est une adaptation simple de la démonstration de la proposition 16.14.

Inversement, supposons que (16.32) soit vraie et soit $f \in bE$ quelconque. On a, pour tous $x_0, x_1, \dots, x_k \in E$,

$$\begin{aligned} & \int_{(X_0=x_0, X_1=x_1, \dots, X_k=x_k)} f(X_{k+1}) dP \\ &= \sum_{x_{k+1}} f(x_{k+1}) P(X_0 = x_0, X_1 = x_1, \dots, X_{k+1} = x_{k+1}) \\ &= \mu(x_0) M(x_0, x_1) M(x_1, x_2) \dots M(x_{k-1}, x_k) \sum_{x_{k+1}} f(x_{k+1}) M(x_k, x_{k+1}) \\ &= \mu(x_0) M(x_0, x_1) M(x_1, x_2) \dots M(x_{k-1}, x_k) M(x_k, f) \\ &= \int_{(X_0=x_0, X_1=x_1, \dots, X_k=x_k)} M(X_k, f) dP. \end{aligned}$$

ce qui démontre, avec les notations antérieures, que

$$E^{\mathcal{F}_k} f(X_{k+1}) = M(X_k, f),$$

et donc que X est une chaîne de Markov homogène de matrice de transition M ; il est évident qu'elle est de loi initiale μ . \square

Remarque. En utilisant le théorème de prolongement de Carathéodory, on démontre que, étant donné une probabilité μ sur E et une matrice stochastique M indexée sur $E \times E$, il existe une unique probabilité P_μ sur l'espace probabilisable $(E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ telle que le processus $X = (X_n)_{n \in \mathbb{N}}$ des coordonnées soit une chaîne de Markov homogène de loi initiale μ et vérifiant, pour tous $k \in \mathbb{N}^*$ et pour tous x_0, x_1, \dots, x_k , l'égalité (16.32). Cette chaîne est appelée chaîne de Markov homogène **canonique** de loi initiale μ et de matrice de transition M . Ce résultat est un cas particulier du théorème de Ionescu-Tulcea.

16.3.3. Loi initiale ; propriété de Markov forte

Dorénavant, on ne s'intéresse plus qu'aux chaînes de Markov homogènes. Elles possèdent la **propriété de Markov forte**, c'est-à-dire l'analogue

de la propriété de Markov simple (ou faible) dans laquelle les temps fixes sont remplacés par des **temps d'arrêt**. C'est à l'aide de cette propriété fondamentale que nous démontrerons les principales propriétés des chaînes de Markov. Pour bien la formuler, il est utile de savoir faire partir une chaîne de Markov homogène selon une **loi initiale donnée**, ce qui justifie la définition suivante :

Définition 16.16. *Un processus X à valeurs dans E est dit chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P)$ de loi initiale ν et de matrice de transition M si*

$$(a) P_{X_0} = \nu.$$

(b) X est une chaîne de Markov homogène de matrice de transition M sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P)$.

Remarque. Avec les notations précédentes, si, pour $x \in E$, X est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P)$ de loi initiale δ_x (on a donc $P(X_0 = x) = 1$) et de matrice de transition M , alors $P_x = P$.

La proposition suivante montre que, si on sait faire partir une chaîne de n'importe quel point x , on sait la faire partir avec une loi initiale ν quelconque.

Proposition 16.17. *Supposons que, pour tout $x \in S \subset E$, X soit une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$ de loi initiale δ_x et de matrice de transition M . Soit ν une probabilité sur E telle que $\nu(S) = 1$. La fonction d'ensembles P_ν définie sur \mathcal{A} par, pour tout $A \in \mathcal{A}$,*

$$P_\nu(A) = \sum_{x \in S} \nu(x) P_x(A)$$

est une probabilité sur (Ω, \mathcal{A}) et X est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_\nu)$ de loi initiale ν et de matrice de transition M .

Démonstration. Bien sûr, $P_\nu(\emptyset) = 0$; de plus, P_ν est σ -additive : si $(A_n)_{n \in \mathbb{N}}$ est une suite d'événements disjoints, on a, les P_x étant des probabilités,

$$P_\nu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{x \in S} \nu(x) \left[\sum_{n \in \mathbb{N}} P_x(A_n) \right],$$

et, puisque les termes sont positifs,

$$P_\nu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \left[\sum_{x \in S} \nu(x) P_x(A_n) \right] = \sum_{n \in \mathbb{N}} P_\nu(A_n).$$

De plus, par les procédés usuels d'intégration, il est facile de montrer que, pour toute variable aléatoire positive ou bornée Y , on a

$$E_\nu(Y) = \sum_{x \in S} \nu(x) E_x(Y),$$

où E_ν (resp. E_x) désigne l'espérance par rapport à P_ν (resp. P_x).

Soient alors des entiers m, n tels que $n < m$ et $f \in bE$; pour tout $A \in \mathcal{A}_n$ on a

$$E_\nu[\mathbf{1}_A f(X_m)] = \sum_{x \in S} \nu(x) E_x[\mathbf{1}_A f(X_m)] = \sum_{x \in S} \nu(x) E_x[\mathbf{1}_A E_x^{\mathcal{A}_n} f(X_m)],$$

et donc, par application de l'égalité (16.17) aux P_x -chaînes de Markov homogènes,

$$E_\nu[\mathbf{1}_A f(X_m)] = \sum_{x \in S} \nu(x) E_x[\mathbf{1}_A M^{m-n}(X_n, f)] = E_\nu[\mathbf{1}_A M^{m-n}(X_n, f)];$$

$M^{m-n}(X_n, f)$ étant \mathcal{A}_n -mesurable, il en résulte que

$$E_\nu^{\mathcal{A}_n} [f(X_m)] = M^{m-n}(X_n, f), \quad (16.33)$$

ce qui démontre que X est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_\nu)$. De plus, pour tout $B \in \mathcal{E}$, on a, par hypothèse,

$$P_\nu(X_0 \in B) = \sum_{x \in S} \nu(x) P_x(X_0 \in B) = \sum_{x \in S} \nu(x) \delta_x(B) = \nu(S \cap B) = \nu(B),$$

ce qui démontre que X est de loi initiale ν pour P_ν . □

Pour compléter cette proposition, on montre comment, disposant d'une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P)$, on peut faire partir cette chaîne de P_{X_0} -presque tout point de E .

Proposition 16.18. *Soit X une chaîne de Markov homogène de matrice de transition M sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P)$. Pour tout $x \in E$ tel que $P(X_0 = x) > 0$, on définit la probabilité $P_x = P(\cdot \mid X_0 = x)$. Alors, X est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$ de loi initiale δ_x et de matrice de transition M .*

Démonstration. On a bien, par définition,

$$P_x(X_0 = x) = P^{(X_0=x)}(X_0 = x) = 1$$

et donc $X_0(P_x) = \delta_x$. Par ailleurs, pour toute variable aléatoire positive ou bornée Y , on a

$$E_x(Y) = \frac{1}{P(X_0 = x)} E(\mathbf{1}_{(X_0=x)} Y).$$

Soient alors des entiers m, n tels que $n < m$ et $f \in bE$; pour tout $A \in \mathcal{A}_n$ on a, puisque $(X_0 = x) \cap A \in \mathcal{A}_n$,

$$\begin{aligned} E_x [\mathbf{1}_A f(X_m)] &= \frac{1}{P(X_0 = x)} E [\mathbf{1}_{(X_0=x)} \mathbf{1}_A f(X_m)] \\ &= \frac{1}{P(X_0 = x)} E [\mathbf{1}_{(X_0=x)} \mathbf{1}_A E^{\mathcal{A}_n} f(X_m)] \\ &= \frac{1}{P(X_0 = x)} E [\mathbf{1}_{(X_0=x)} \mathbf{1}_A M^{m-n}(X_n, f)] \\ &= E_x [\mathbf{1}_A M^{m-n}(X_n, f)]; \end{aligned}$$

$M^{m-n}(X_n, f)$ étant \mathcal{A}_n -mesurable, il en résulte que

$$E_x^{\mathcal{A}_n} [f(X_m)] = M^{m-n}(X_n, f),$$

ce qui démontre que X est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$. \square

Remarque. Les deux dernières propositions permettent donc, partant d'une chaîne de Markov homogène X de matrice de transition M sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P)$ de construire, pour toute probabilité ν sur E de même support que P_{X_0} , une probabilité P_ν sur (Ω, \mathcal{A}) telle que X soit une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_\nu)$ de loi initiale ν et de matrice de transition M .

Avec les notations précédentes, on peut reformuler ainsi la propriété de Markov simple énoncée à la proposition 16.13 :

Proposition 16.19 (Propriété de Markov simple). *Soit X un processus qui, pour tout $x \in E$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$ de loi initiale δ_x et de matrice de transition M . Pour tout fonction f sur $E^{\mathbb{N}}$, $\mathcal{E}^{\otimes \mathbb{N}}$ -mesurable positive ou bornée, on a*

$$\boxed{E_x^{\mathcal{A}_n} [f(\theta_n(X))] = E_{X_n} [f(X)]} \quad (16.34)$$

Remarque. Il faut bien comprendre que $E_{X_n} [f(X)]$ désigne la valeur en X_n de l'application $x \mapsto E_x [f(X)]$.

Proposition 16.20 (Propriété de Markov forte). *Soit X un processus qui, pour tout $x \in E$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$ de loi initiale δ_x et de matrice de transition M . Alors X a la propriété de Markov forte, c'est-à-dire que, pour tout fonction f sur $E^{\mathbb{N}}$, $\mathcal{E}^{\otimes \mathbb{N}}$ -mesurable positive ou bornée, pour tout temps d'arrêt T et tout $x \in E$, on a*

$$\boxed{E_x^{\mathcal{A}_T} [\mathbf{1}_{(T < +\infty)} f(\theta_T(X))] = \mathbf{1}_{(T < +\infty)} E_{X_T} [f(X)]} \quad (16.35)$$

Démonstration. Pour tout $n \in \overline{\mathbb{N}}$, on a

$$\mathbf{1}_{(T=n)} E_x^{\mathcal{A}_T} [\mathbf{1}_{(T<+\infty)} f(\theta_T(X))] = \mathbf{1}_{(T=n)} E_x^{\mathcal{A}_n} [\mathbf{1}_{(T<+\infty)} f(\theta_T(X))],$$

et donc, puisque $(T = n) \in \mathcal{A}_n$,

$$\mathbf{1}_{(T=n)} E_x^{\mathcal{A}_T} [\mathbf{1}_{(T<+\infty)} f(\theta_T(X))] = E_x^{\mathcal{A}_n} [\mathbf{1}_{(T=n)} \mathbf{1}_{(T<+\infty)} f(\theta_T(X))].$$

Il en résulte que

$$\mathbf{1}_{(T=n)} E_x^{\mathcal{A}_T} [\mathbf{1}_{(T<+\infty)} f(\theta_T(X))] = \begin{cases} E_x^{\mathcal{A}_n} [\mathbf{1}_{(T=n)} f(\theta_n(X))] & \text{si } n \in \mathbb{N}. \\ 0 & \text{si } n = +\infty. \end{cases}$$

Ainsi, d'après la propriété de Markov simple, on a, si $n \in \mathbb{N}$,

$$\begin{aligned} \mathbf{1}_{(T=n)} E_x^{\mathcal{A}_T} [\mathbf{1}_{(T<+\infty)} f(\theta_T(X))] &= \mathbf{1}_{(T=n)} E_x^{\mathcal{A}_n} [f(\theta_n(X))] \\ &= \mathbf{1}_{(T=n)} E_{X_n} [f(X)] \\ &= \mathbf{1}_{(T=n)} E_{X_T} [f(X)]. \end{aligned}$$

On a alors, puisque le terme correspondant à $n = +\infty$ est nul :

$$E_x^{\mathcal{A}_T} [\mathbf{1}_{(T<+\infty)} f(\theta_T(X))] = \sum_{n \in \mathbb{N}} \mathbf{1}_{(T=n)} E_{X_T} [f(X)] = \mathbf{1}_{(T<+\infty)} E_{X_T} [f(X)]. \quad \square$$

Corollaire 16.21. *Avec les mêmes hypothèses qu'à la proposition 16.20, soit T un temps d'arrêt fini; on définit le processus Y et la filtration $(\mathcal{B}_n)_{n \in \mathbb{N}}$ par $Y_n = X_{T+n}$ et $\mathcal{B}_n = \mathcal{A}_{T+n}$, si $n \in \mathbb{N}$. Pour tout $x \in E$, le processus Y est une chaîne de Markov homogène de matrice de transition M sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{B}_n)_{n \in \mathbb{N}}, P_x)$.*

Démonstration. Pour tous $f \in bE$ et tous entiers m et n tels que $m < n$, on a

$$E_x^{\mathcal{B}_n} [f(Y_n)] = E_x^{\mathcal{A}_{T+m}} [f(\{(\theta_{T+m}(X))_{n-m}\})],$$

soit, d'après la propriété de Markov forte,

$$E_x^{\mathcal{B}_n} [f(Y_n)] = E_{X_{T+m}} [f(X_{n-m})],$$

et donc, d'après (16.17),

$$E_x^{\mathcal{B}_n} [f(Y_n)] = M^{n-m}(X_{T+m}, f) = M^{n-m}(Y_m, f).$$

ce qui démontre le résultat. □

16.4. Visites à un état fixe

Dans cette section, on se donne un processus X qui, pour tout $x \in E$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$ de loi initiale δ_x et de matrice de transition M . On étudie les temps d'entrée

des trajectoires de X dans une partie B et plus spécialement, lorsque B est un singleton, les temps de passage en des points de E . Cette étude conduira à une classification des points de E suivant le comportement qu'a la chaîne vis à vis d'eux.

Notations. Si B est une partie de E , on note, avec la convention $\inf \emptyset = +\infty$,

$$T_B = \inf(n \in \mathbb{N}^* \mid X_n \in B) \quad \text{et} \quad N_B = \sum_{j \in \mathbb{N}} \mathbf{1}_{(X_j \in B)};$$

ce sont respectivement le **premier temps d'entrée** dans B après l'instant 1 et le **temps passé** dans B par la chaîne durant toute la vie de ce processus; en particulier, si $B = \{y\}$, où $y \in E$, ces quantités sont notées simplement T_y et N_y . On définit de manière analogue des fonctionnelles τ_B, n_B, τ_y et n_y sur $E^{\mathbb{N}}$ par, pour tout $u \in E^{\mathbb{N}}$,

$$\tau_B(u) = \inf(n \in \mathbb{N}^* \mid u_n \in B) \quad \text{et} \quad n_B(u) = \sum_{j \in \mathbb{N}} \mathbf{1}_{(u_j \in B)};$$

on note $\tau_y = \tau_{\{y\}}$ et $n_y = n_{\{y\}}$. Le lemme suivant sera d'application constante par la suite et permettra de se mettre en situation d'appliquer la propriété de Markov forte.

Lemme 16.22. *Avec les notations précédentes, on a $T_B = \tau_B(X)$ et, pour tout $p \in \mathbb{N}^*$,*

$$\text{sur } (T_B > p) \quad T_B = p + \tau_B[\theta_p(X)]; \quad (16.36)$$

en particulier

$$\boxed{\text{sur } (T_y > p) \quad T_y = p + \tau_y[\theta_p(X)]}. \quad (16.37)$$

On en déduit que, pour tout temps d'arrêt T , on a

$$\text{sur } (T_B > T) \cap (T < +\infty) \quad T_B = T + \tau_B[\theta_T(X)]; \quad (16.38)$$

en particulier

$$\boxed{\text{sur } (T_y > T) \cap (T < +\infty) \quad T_y = T + \tau_y[\theta_T(X)]}. \quad (16.39)$$

Démonstration. Il suffit de constater que, sur $(T_B > p)$,

$$\begin{aligned} \tau_B[\theta_p(X)] &= \inf(n \in \mathbb{N}^* \mid X_{n+p} \in B) \\ &= \inf(n \geq p+1 \mid X_n \in B) - p = T_B - p. \end{aligned}$$

Pour obtenir (16.38), il suffit alors d'appliquer (16.36) sur les ensembles $(T = p) \cap (T_y > p)$, $p \in \mathbb{N}^*$. \square

Notations. On définit de même par récurrence la suite des temps d'entrée dans B par

$$T_B^0 = 0, \quad T_B^1 = T_B, \quad T_B^{p+1} = \inf(n > T_B^p \mid X_n \in B);$$

en particulier, si $B = \{y\}$, ces temps sont notés simplement T_y^p (T_y^p est la date du p -ième passage en y). On démontre alors de même la relation

$$\text{sur } (T_B < +\infty) \quad T_B^{p+1} = T_B^p + \tau_B \left[\theta_{T_B^p}(X) \right]; \quad (16.40)$$

en particulier,

$$\boxed{\text{sur } (T_y^p < +\infty) \quad T_y^{p+1} = T_y^p + \tau_y \left[\theta_{T_y^p}(X) \right].} \quad (16.41)$$

16.4.1. Étude de la suite des temps de passage en un point

Proposition 16.23. Avec les notations précédentes, pour tout $p \in \mathbb{N}^*$, T_B^p est un temps d'arrêt. Pour tout $x, y \in E$, la suite $(T_y^p)_{p \in \mathbb{N}}$ est une chaîne de Markov homogène à valeurs dans $\overline{\mathbb{N}^*}$ sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_{T_y^p})_{p \in \mathbb{N}^*}, P_x)$.

Si $P_y(T_y^1 < +\infty) = 1$ (c'est-à-dire si la chaîne partant de y retourne en y en un temps fini P_y -p.s.), on a, pour tout $p \in \mathbb{N}^*$, $P_y(T_y^p < +\infty) = 1$, et par conséquent aussi $P_y(N_y = +\infty) = 1$; de plus, la suite $(T_y^{p+1} - T_y^p)_{p \in \mathbb{N}^*}$ des intervalles de temps entre deux passages en y est une suite de variables aléatoires (définies et finies P_y -p.s.) P_y -indépendantes, de même loi (sous P_y) que celle de T_y^1 .

Démonstration. • On sait déjà que T_B^1 est un temps d'arrêt; si $p \geq 2$, X étant adapté, on a, pour tout $n \in \mathbb{N}^*$,

$$(T_B^p \leq n) = \left(\sum_{j=1}^n \mathbf{1}_B(X_j) \geq p \right) \in \mathcal{A}_n,$$

ce qui démontre que T_B^p est un temps d'arrêt.

• Pour toute fonction bornée f sur $\overline{\mathbb{N}^*}$ et tout x de E , on a, d'après l'égalité (16.38),

$$\begin{aligned} E_x^{\mathcal{A}_{T_y^p}} [f(T_y^{p+1})] &= E_x^{\mathcal{A}_{T_y^p}} \left[\mathbf{1}_{(T_y^p < +\infty)} f(T_y^p + \tau_y [\theta_{T_y^p}(X)]) \right] \\ &\quad + E_x^{\mathcal{A}_{T_y^p}} \left[\mathbf{1}_{(T_y^p = +\infty)} f(+\infty) \right] \end{aligned} \quad (16.42)$$

soit, puisque $(T_y^p = +\infty) \in \mathcal{A}_{T_y^p}$ et que $(T_y^p = i) \in \mathcal{A}_i$,

$$E_x^{\mathcal{A}_{T_y^p}} [f(T_y^{p+1})] = \sum_{i \in \mathbb{N}^*} \mathbf{1}_{(T_y^p = i)} E_x^{\mathcal{A}_i} [f(i + \tau_y [\theta_i(X)])] + \mathbf{1}_{(T_y^p = +\infty)} f(+\infty).$$

Par application de la propriété de Markov simple (la propriété de Markov forte ne s'applique pas ici), on a

$$E_x^{A_{T_y^p}} [f(T_y^{p+1})] = \sum_{i \in \mathbb{N}^*} \mathbf{1}_{(T_y^p=i)} E_{X_i} [f(i + \tau_y(\mathbf{X}))] + \mathbf{1}_{(T_y^p=+\infty)} f(+\infty).$$

Puisque $\tau_y(\mathbf{X}) = T_y^1$ et que, par définition de T_y^p , sur $(T_y^p = i)$ on a $X_i = X_{T_y^p} = y$, il vient

$$E_x^{A_{T_y^p}} [f(T_y^{p+1})] = \sum_{i \in \mathbb{N}^*} \mathbf{1}_{(T_y^p=i)} E_y [f(i + T_y^1)] + \mathbf{1}_{(T_y^p=+\infty)} f(+\infty).$$

Définissant la probabilité de transition N sur $\overline{\mathbb{N}^*}$ par

$$N(i, f) = \begin{cases} E_y [f(i + T_y^1)] & \text{si } i \in \mathbb{N}^*, \\ f(+\infty) & \text{si } i = +\infty, \end{cases} \quad (16.43)$$

on a alors

$$E_x^{A_{T_y^p}} [f(T_y^{p+1})] = \sum_{i \in \mathbb{N}^*} \mathbf{1}_{(T_y^p=i)} N(i, f) + \mathbf{1}_{(T_y^p=+\infty)} N(+\infty, f).$$

soit encore

$$E_x^{A_{T_y^p}} [f(T_y^{p+1})] = N(T_y^p, f); \quad (16.44)$$

ceci démontre que le processus $(T_y^p)_{p \in \mathbb{N}}$ est une chaîne de Markov homogène à valeurs dans $\overline{\mathbb{N}^*}$ de matrice de transition N donnée, pour $i, j \in \overline{\mathbb{N}^*}$, par (prendre $f = \mathbf{1}_{\{j\}}$)

$$N(i, j) = \begin{cases} P_y(T_y^1 = j - i) & \text{si } i, j \in \mathbb{N}^* \text{ et } j - i \geq 1, \\ 0 & \text{si } j \leq i \\ P_y(T_y^1 = +\infty) & \text{si } i \in \mathbb{N}^* \text{ et } j = +\infty. \\ 1 & \text{si } i = j = +\infty. \end{cases}$$

• Prenant $f = \mathbf{1}_{\mathbb{N}^*}$ dans l'égalité (16.43), il vient

$$N(i, \mathbf{1}_{\mathbb{N}^*}) = \begin{cases} P_y(T_y^1 < +\infty) & \text{si } i \in \mathbb{N}^*, \\ 0 & \text{si } i = +\infty. \end{cases}$$

ce qui conduit, en reportant dans (16.44), à l'égalité

$$E_x^{A_{T_y^p}} [\mathbf{1}_{(T_y^{p+1} < +\infty)}] = \mathbf{1}_{(T_y^p < +\infty)} P_y(T_y^1 < +\infty).$$

Il en résulte, en prenant la E_x -moyenne de chacun des membres de l'égalité précédente, que

$$P_x(T_y^{p+1} < +\infty) = P_x(T_y^p < +\infty) P_y(T_y^1 < +\infty).$$

et qu'en particulier,

$$P_y(T_y^{p+1} < +\infty) = P_y(T_y^p < +\infty) P_y(T_y^1 < +\infty).$$

Si $P_y(T_y^1 < +\infty) = 1$, on a donc, pour tout $p \in \mathbb{N}^*$, $P_y(T_y^p < +\infty) = 1$; de plus, puisque la suite d'événements $(T_y^p < +\infty)$ est décroissante et que $(N_y = +\infty) = \bigcap_{p \in \mathbb{N}^*} (T_y^p < +\infty)$, on a

$$P_y(N_y = +\infty) = \lim_p P_y(T_y^p < +\infty) = 1.$$

Enfin, pour toute partie D de $\overline{\mathbb{N}^*}$, sous cette hypothèse, on a, d'après l'égalité (16.41),

$$E_y^{\mathcal{A}_{T_y^p}} [\mathbf{1}_D (T_y^{p+1} - T_y^p)] = E_y^{\mathcal{A}_{T_y^p}} \left[\mathbf{1}_{(T_y^p < +\infty)} \mathbf{1}_D(\tau_y [\theta_{T_y^p}(X)]) \right].$$

soit, d'après la propriété de Markov forte,

$$\begin{aligned} E_y^{\mathcal{A}_{T_y^p}} [\mathbf{1}_D (T_y^{p+1} - T_y^p)] &= \mathbf{1}_{(T_y^p < +\infty)} E_{X_{T_y^p}} [\mathbf{1}_D(\tau_y(X))] \\ &= \mathbf{1}_{(T_y^p < +\infty)} E_y [\mathbf{1}_D(T_y^1)], \end{aligned}$$

et donc

$$E_y^{\mathcal{A}_{T_y^p}} [\mathbf{1}_D (T_y^{p+1} - T_y^p)] = P_y [T_y^1 \in D].$$

Il en résulte tout d'abord que les tribus $\mathcal{A}_{T_y^p}$ et $\sigma(T_y^{p+1} - T_y^p)$ sont P_y -indépendantes: puisque ceci est vrai pour tout p , on en déduit facilement la P_y -indépendance des variables aléatoires $T_y^{p+1} - T_y^p$. De plus, en prenant la E_y -moyenne de chacun des membres de l'égalité précédente, on a

$$P_y [(T_y^{p+1} - T_y^p) \in D] = P_y [T_y^1 \in D],$$

ce qui, vu l'arbitraire de D montre que $T_y^{p+1} - T_y^p$ et T_y^1 ont même loi sous P_y . \square

16.4.2. Loïs du nombre de visites d'un point et du premier temps de passage en ce point

Comme nous allons le voir, nombre de visites d'un point y et premier temps de passage en ce point sont intimement liés.

Proposition 16.24. *La loi du nombre N_y de passages en y pendant toute la vie du processus est donnée (avec la convention d'écriture $0^0 = 1$), par*

- si $x \neq y$,

$$P_x(N_y = m) = P_x(T_y^1 < +\infty) P_y(T_y^1 = +\infty) [P_y(T_y^1 < +\infty)]^{m-1}$$

si $m \in \mathbb{N}^*$, et

$$P_x(N_y = 0) = P_x(T_y^1 = +\infty);$$

• si $x = y$,

$$P_y(N_y = m) = P_y(T_y^1 = +\infty) [P_y(T_y^1 < +\infty)]^{m-1} \quad \text{si } m \in \mathbb{N}^*.$$

Autrement dit, si $0 < P_y(T_y^1 < +\infty) < 1$, la loi de N_y sous P_y est la loi géométrique sur \mathbb{N}^* de paramètre $P_y(T_y^1 = +\infty)$.

Démonstration. L'événement $(N_y = m) = (\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j=y)} = m)$ est l'ensemble des trajectoires qui passent exactement m fois par y à partir du temps 0; or, pour tout $m \in \mathbb{N}^*$, on a

$$\left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j=y)} = m \right) = (T_y^1 < +\infty) \cap (T_y^2 < +\infty) \cap \dots \\ \dots \cap (T_y^m < +\infty) \cap (T_y^{m+1} - T_y^m = +\infty),$$

ce qui, en vertu de l'égalité (16.41), s'écrit

$$\left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j=y)} = m \right) = (T_y^1 < +\infty) \cap (T_y^2 < +\infty) \cap \\ \dots \cap (T_y^m < +\infty) \cap (\tau_y [\theta_{T_y^m}(X)] = +\infty);$$

En intégrant par rapport à P_x et en remarquant que

$$(T_y^1 < +\infty) \cap (T_y^2 < +\infty) \cap \dots \cap (T_y^m < +\infty) \in \mathcal{A}_{T_y^m},$$

on obtient, par conditionnement par rapport à la tribu $\mathcal{A}_{T_y^m}$,

$$P_x \left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j=y)} = m \right) = E_x \left[\mathbf{1}_{(T_y^1 < +\infty) \cap (T_y^2 < +\infty) \cap \dots \cap (T_y^m < +\infty)} \right. \\ \left. \times E_x^{\mathcal{A}_{T_y^m}} \left(\mathbf{1}_{(T_y^m < +\infty)} \mathbf{1}_{(\tau_y [\theta_{T_y^m}(X)] = +\infty)} \right) \right];$$

or, par application de la propriété de Markov forte (calcul déjà rencontré), on a

$$E_x^{\mathcal{A}_{T_y^m}} \left(\mathbf{1}_{(T_y^m < +\infty)} \mathbf{1}_{(\tau_y [\theta_{T_y^m}(X)] = +\infty)} \right) = \mathbf{1}_{(T_y^m < +\infty)} E_{X_{T_y^m}} \left(\mathbf{1}_{(\tau_y(X) = +\infty)} \right) \\ = \mathbf{1}_{(T_y^m < +\infty)} P_y(T_y^1 = +\infty),$$

ce qui, en portant dans l'égalité précédente, donne

$$P_x \left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j=y)} = m \right) = E_x \left[\mathbf{1}_{(T_y^1 < +\infty) \cap (T_y^2 < +\infty) \cap \dots \cap (T_y^m < +\infty)} \right] \\ \times P_y(T_y^1 = +\infty). \quad (16.45)$$

En tenant compte de l'égalité

$$\begin{aligned} (\Gamma_y^{m-1} < +\infty) \cap (\Gamma_y^m < +\infty) &= (\Gamma_y^{m-1} < +\infty) \cap (\Gamma_y^m - \Gamma_y^{m-1} < +\infty) \\ &= (\Gamma_y^{m-1} < +\infty) \cap \left(\tau_y \left[\theta_{\Gamma_y^{m-1}}(X) \right] < +\infty \right), \end{aligned}$$

on obtient, par le même procédé de conditionnement par rapport à la tribu $\mathcal{A}_{\Gamma_y^{m-1}}$, puis application de la propriété de Markov forte,

$$\begin{aligned} &E_x \left[\mathbf{1}_{(\Gamma_y^1 < +\infty) \cap (\Gamma_y^2 < +\infty) \cap \dots \cap (\Gamma_y^m < +\infty)} \right] \\ &= E_x \left[\mathbf{1}_{(\Gamma_y^1 < +\infty) \cap (\Gamma_y^2 < +\infty) \cap \dots \cap (\Gamma_y^{m-1} < +\infty)} E_x^{\mathcal{A}_{\Gamma_y^{m-1}}} \left(\mathbf{1}_{(\tau_y[\theta_{\Gamma_y^{m-1}}(X)] < +\infty)} \right) \right] \\ &= E_x \left[\mathbf{1}_{(\Gamma_y^1 < +\infty) \cap (\Gamma_y^2 < +\infty) \cap \dots \cap (\Gamma_y^{m-1} < +\infty)} E_{X_{\Gamma_y^{m-1}}} \left(\mathbf{1}_{(\tau_y(X) < +\infty)} \right) \right] \\ &= E_x \left[\mathbf{1}_{(\Gamma_y^1 < +\infty) \cap (\Gamma_y^2 < +\infty) \cap \dots \cap (\Gamma_y^{m-1} < +\infty)} \right] P_y(\Gamma_y^1 < +\infty), \end{aligned}$$

ce qui, en reportant dans l'égalité (16.45), donne

$$\begin{aligned} P_x \left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j = y)} = m \right) &= E_x \left[\mathbf{1}_{(\Gamma_y^1 < +\infty) \cap (\Gamma_y^2 < +\infty) \cap \dots \cap (\Gamma_y^{m-1} < +\infty)} \right] \\ &\quad \times P_y(\Gamma_y^1 < +\infty) P_y(\Gamma_y^1 = +\infty). \end{aligned}$$

Par itération rétrograde et par le même procédé, on obtient alors l'égalité

$$P_x \left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j = y)} = m \right) = P_x(\Gamma_y^1 < +\infty) P_y(\Gamma_y^1 = +\infty) [P_y(\Gamma_y^1 < +\infty)]^{m-1}. \quad (16.46)$$

– Si $x \neq y$ et $m \in \mathbb{N}^*$, on a

$$P_x(N_x = m) = P_x \left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j = y)} = m \right) = P_x \left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j = y)} = m \right),$$

et l'égalité (16.46) donne le résultat annoncé.

– Si $x \neq y$ et $m = 0$, on a

$$P_x(N_x = 0) = P_x(\Gamma_y^1 = +\infty).$$

– Enfin, si $x = y$, on a $P_y(N_y = 0) = 0$ et, si $m \in \mathbb{N}^*$,

$$P_y(N_y = m) = P_y \left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j = y)} = m \right) = P_x \left(\sum_{j \in \mathbb{N}^*} \mathbf{1}_{(X_j = y)} = m - 1 \right),$$

et l'égalité (16.46) donne encore le résultat annoncé. \square

On étudie maintenant le temps T_y^1 .

Notations. Pour $k \in \mathbb{N}^*$ et tous $x, y \in E$, on note

$$F_k(x, y) = P_x(\Gamma_y^1 = k) \text{ et } F(x, y) = P_x(\Gamma_y^1 < +\infty).$$

Proposition 16.25. *La suite des matrices F_k est solution du système itératif suivant : pour tous $x, y \in E$,*

$$\begin{cases} F_1(x, y) = M(x, y). \\ F_k(x, y) = \sum_{z \in E \setminus \{y\}} M(x, z) F_{k-1}(z, y) & \text{si } k \geq 2. \end{cases} \quad (16.47)$$

En conséquence, la matrice F est solution de l'équation matricielle déterminée par : pour tous $x, y \in E$,

$$F(x, y) = M(x, y) + \sum_{z \in E \setminus \{y\}} M(x, z) F(z, y). \quad (16.48)$$

Démonstration. • On a $F_1(x, y) = P_x(X_1 = y) = M(x, y)$.

• Si $k \geq 2$, sur $(T_y^1 > 1)$, on a $T_y^1 = 1 + \tau_y[\theta_1(X)]$ et donc, par conditionnement par rapport à la tribu \mathcal{A}_1 et application de la propriété de Markov simple (après avoir noté que X_1 est \mathcal{A}_1 -mesurable)

$$\begin{aligned} F_k(x, y) &= E_x \left[\mathbf{1}_{(X_1 \neq y)} E_x^{\mathcal{A}_1} (\mathbf{1}_{(1+\tau_y[\theta_1(X)] = k)}) \right] \\ &= E_x \left[\mathbf{1}_{(X_1 \neq y)} E_{X_1} (\mathbf{1}_{(1+\tau_y(X) = k)}) \right]. \end{aligned}$$

soit encore

$$F_k(x, y) = E_x \left[\mathbf{1}_{(X_1 \neq y)} P_{X_1}(T_y^1 = k - 1) \right] = \sum_{z \in E \setminus \{y\}} P_x(X_1 = z) P_z(T_y^1 = k - 1),$$

ce qui démontre (16.47).

• On a

$$F(x, y) = P_x(T_y^1 < +\infty) = \sum_{k \in \mathbb{N}^*} P_x(T_y^1 = k) = \sum_{k \in \mathbb{N}^*} F_k(x, y),$$

et donc, d'après (16.47),

$$F(x, y) = M(x, y) + \sum_{k \geq 2} \left[\sum_{z \in E \setminus \{y\}} M(x, z) F_{k-1}(z, y) \right].$$

ce qui démontre (16.48), après permutation des sommes (à termes positifs). \square

On reformule les résultats obtenus à la proposition 16.24 à l'aide de la matrice F et en donne, sans démonstration, des conséquences immédiates :

Proposition 16.26. *Avec les notations ci-dessus, on a*

(a) *si $x \neq y$,*

$$P_x(N_y = m) = \begin{cases} 1 - F(x, y) & \text{si } m = 0, \\ F(x, y) [1 - F(y, y)] [F(y, y)]^{m-1} & \text{si } m \in \mathbb{N}^*; \end{cases}$$

(b) si $x = y$, (avec la convention d'écriture $0^0 = 1$),

$$P_y(N_y = m) = [1 - F(y, y)] [F(y, y)]^{m-1} \quad \text{si } m \in \mathbb{N}^*.$$

(c) On a l'alternative suivante

$$P_y(N_y < +\infty) = \begin{cases} 1 & \text{si } F(y, y) < 1, \\ 0 & \text{si } F(y, y) = 1. \end{cases}$$

- Si $F(y, y) = 1$, on a $P_y(N_y = +\infty) = 1$ et donc $E_y(N_y) = +\infty$,

- si $0 < F(y, y) < 1$, la loi de N_y sous P_y est la loi géométrique sur \mathbb{N}^* de paramètre $1 - F(y, y)$,

- si $F(y, y) = 0$, on a P_y -p.s. $N_y = 1$.

En particulier, le nombre moyen $E_y(N_y)$ de passages en y par la chaîne partant de y à l'instant 0 est

$$E_y(N_y) = \begin{cases} \frac{1}{1 - F(y, y)} & \text{si } F(y, y) < 1, \\ +\infty & \text{si } F(y, y) = 1. \end{cases}$$

Définition 16.27. La matrice R (à termes dans $\overline{\mathbb{N}}$) définie, pour tous $x, y \in E$, par $R(x, y) = E_x(N_y)$, nombre moyen de passages en y par la chaîne partant de x à l'instant 0, est appelée **matrice potentiel** de la chaîne.

De la proposition 16.26, on déduit le corollaire :

Corollaire 16.28. Avec les conventions $\frac{1}{0} = +\infty$ et $0 \cdot \infty = 0$, on a

$$R(x, y) = \begin{cases} \frac{1}{1 - F(y, y)} & \text{si } x = y, \\ F(x, y)R(y, y) & \text{si } x \neq y. \end{cases} \quad (16.49)$$

Remarque. En pratique, il est en général plus facile de calculer d'abord R (on en verra une méthode de calcul ultérieurement) et d'en déduire F . La proposition suivante montre que R est solution d'une équation matricielle ; ceci permet, en particulier dans le cas où E est fini, de calculer R , après avoir identifié ses éléments infinis.

Proposition 16.29. La matrice potentiel R vérifie l'égalité

$$R = \sum_{n=0}^{+\infty} M^n,$$

au sens où, pour tous $x, y \in E$, on a l'égalité, dans $\overline{\mathbb{R}}^+$, $R(x, y) = \sum_{n=0}^{+\infty} M^n(x, y)$; de plus, R est solution de l'équation matricielle

$$R(I - M) = (I - M)R = I, \quad (16.50)$$

où I est la matrice identité ($I(x, y) = 1$ si $x = y$, 0 sinon).

En particulier, si E est fini et si $\mathbf{R}n$ a que des termes finis, $I - M$ admet un inverse et $\mathbf{R} = (I - M)^{-1}$.

Démonstration. Par convergence monotone, on a

$$\mathbf{R}(x, y) = \sum_{n=0}^{+\infty} E_x(\mathbf{1}_{\{X_n=y\}}) = \sum_{n=0}^{+\infty} M^n(x, y);$$

il en résulte que

$$\mathbf{R}M = M\mathbf{R} = \sum_{n=1}^{+\infty} M^n = \mathbf{R} - I,$$

ce qui donne l'égalité (16.50). □

16.5. Classification des états

Dans cette section, on se donne un processus X qui, pour tout $x \in E$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$ de loi initiale δ_x et de matrice de transition M . On classe les points de E suivant leur fréquentation par les trajectoires de X ,

16.5.1. Communication ; périodicité

Définition 16.30. Si B est une partie de E , on dit que le point $x \in E$ **conduit à B** si $P_x(T_B < +\infty) > 0$. On note cette relation $x \rightarrow B$; en particulier, si $B = \{y\}$, où $y \in E$, on dit que x **conduit à y** et on note cette relation $x \rightarrow y$.

Proposition 16.31. La relation de conduction $x \rightarrow y$ est transitive. De plus, x conduit à y si et seulement si il existe $n \in \mathbb{N}^*$ tel que $M^n(x, y) > 0$.

Démonstration. • Supposons que $x \rightarrow y$ et $y \rightarrow z$. L'ensemble des trajectoires passant en z après être passées par y est contenu dans celui des trajectoires passant en z , ce qui donne l'inclusion des événements

$$(T_y < +\infty) \cap (\tau_z [\theta_{T_y}(X) < +\infty]) \subset (T_z < +\infty);$$

ainsi, on a

$$P_x [(T_y < +\infty) \cap (\tau_z [\theta_{T_y}(X) < +\infty])] \leq P_x(T_z < +\infty).$$

Conditionnement par rapport à \mathcal{A}_{T_y} et propriété de Markov forte (calcul maintenant classique) conduisent à la suite d'égalités

$$\begin{aligned}
& P_x[(T_y < +\infty) \cap (\tau_z[\theta_{T_y}(X)] < +\infty)] \\
&= E_x \left[E_x^{\theta_{T_y}} (\mathbf{1}_{(T_y < +\infty)} \mathbf{1}_{(\tau_z[\theta_{T_y}(X)] < +\infty)}) \right] \\
&= E_x \left[\mathbf{1}_{(T_y < +\infty)} E_{X_{T_y}} (\mathbf{1}_{(T_z < +\infty)}) \right] \\
&= P_x(T_y < +\infty) P_y(T_z < +\infty).
\end{aligned}$$

ce qui démontre que

$$0 < P_x(T_y < +\infty) P_y(T_z < +\infty) \leq P_x(T_z < +\infty),$$

et donc que $x \rightarrow z$.

• Si $x \rightarrow y$; puisque $(T_y < +\infty) = \bigcup_{n \in \mathbb{N}^*} (X_n = y)$, on a

$$0 < P_x(T_y < +\infty) \leq \sum_{n \in \mathbb{N}^*} P_x(X_n = y) = \sum_{n \in \mathbb{N}^*} M^n(x, y),$$

ce qui démontre qu'il existe $n \in \mathbb{N}^*$ tel que $M^n(x, y) > 0$.

Inversement, soit un tel n ; on a

$$0 < M^n(x, y) = P_x(X_n = y) \leq P_x(T_y < +\infty),$$

et donc x conduit à y . □

On déduit de cette relation transitive, une relation d'équivalence en la symétrisant :

Définition 16.32. On dit que x **communique avec** y si x conduit à y et y conduit à x ou si x et y coïncident; on note cette relation $x \longleftrightarrow y$.

La relation de communication est une relation d'équivalence et ses classes d'équivalence sont appelées **classes de communication** ou **classes irréductibles**; en particulier, s'il n'existe qu'une classe de communication (c'est-à-dire si tous les points communiquent), on dit que la **chaîne est irréductible**.

Il est habituel d'associer à une chaîne de Markov de matrice de transition M un **graphe** dont les sommets sont les points de E , ces sommets étant reliés, s'ils communiquent, par des flèches indiquant le sens de communication; par exemple, ce graphe



résume le fait que x conduit à lui-même, x conduit à y , y communique avec z , z conduit à t et t conduit à lui-même. Les classes de communication sont $\{x\}$, $\{y, z\}$ et $\{t\}$. Sur ce graphe, on peut aussi mentionner les probabilités de passage d'un point x à un point y du temps 0 au temps 1, autrement dit les probabilités $M(x, y)$, mais cela a moins d'intérêt.

Exemple 16.9. 1. Si $E = \{1, 2, 3, 4, 5\}$ et si M est la matrice

$$\begin{pmatrix} \frac{2}{5} & \frac{3}{5} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{ou, si on veut visualiser les états,} \quad \begin{matrix} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\ \mathbf{1} & \begin{pmatrix} \frac{2}{5} & \frac{3}{5} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

Le graphe associé est



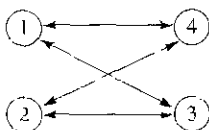
il y a donc exactement deux classes de communication $E_1 = \{1, 2\}$ et $E_2 = \{3, 4, 5\}$; ici apparaissent deux chaînes de Markov homogènes sous-jacentes à valeurs respectivement dans E_1 et E_2 de matrices de transition les sous-matrices de M , M_1 et M_2 , données par

$$M_1 = \begin{pmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} \quad \text{et} \quad M_2 = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 1 & 0 \end{pmatrix}.$$

2. Si $E = \{1, 2, 3, 4\}$ et si M est la matrice

$$\begin{pmatrix} 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{pmatrix} \quad \text{ou, si on veut visualiser les états,} \quad \begin{matrix} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{1} & \begin{pmatrix} 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{pmatrix} \end{matrix}$$

Le graphe associé est



On voit que tous les états communiquent; la chaîne est **irréductible**. Cependant apparaissent deux sous-classes $C_1 = \{1, 2\}$ et $C_2 = \{3, 4\}$ telles que si $X_n \in C_0$ alors $X_{n+1} \in C_1$ et si $X_n \in C_1$ alors $X_{n+1} \in C_0$; ce sont ce qu'on appelle des **classes cycliques**. On est ainsi naturellement amené à définir la notion de période d'un point.

Définition 16.33. Soit $x \in E$; le plus grand entier d tel que l'on ait l'inclusion

$$\{n \in \mathbb{N}^* \mid M^n(x, x) > 0\} \subset d \mathbb{N}^*$$

est appelé **période** de x et noté $d(x)$; c'est le pgcd de l'ensemble $\{n \in \mathbb{N}^* \mid M^n(x, x) > 0\}$; si cet ensemble est vide, on pose $d(x) = 0$. Si $d(x) = 1$, on dit que x est **apériodique**.

Proposition 16.34 (et définition). Soit C une classe de communication. Tous les éléments de C ont même période, notée $d(C)$, et appelée **période de la classe** C . Si $d(C) = 1$, on dit que la **classe** est **apériodique**. Une **chaîne** de Markov homogène irréductible et dont un point est apériodique est alors dite **apériodique**.

Démonstration. Soient $x, y \in C$. Puisque x et y communiquent, il existe k et $l \in \mathbb{N}^*$ tels que $M^k(x, y) > 0$ et $M^l(y, x) > 0$; il en résulte que

$$M^{k+l}(x, x) \geq M^k(x, y)M^l(y, x) > 0,$$

et donc que, d'une part, $d(x) \geq 1$, et d'autre part que $k+l \equiv 0 \pmod{d(x)}$. Remarquons que pour tout n non multiple de $d(x)$, il en est de même pour $n+k+l$ et qu'ainsi $M^{n+k+l}(x, x) = 0$; on en déduit que

$$0 = M^{n+k+l}(x, x) \geq M^k(x, y)M^n(y, y)M^l(y, x) \geq 0,$$

et donc que $M^n(y, y) = 0$. Par contraposition, on vient de montrer que si $M^n(y, y) > 0$, alors n est multiple de $d(x)$. Il en résulte en particulier que, puisque $M^{d(y)}(y, y) > 0$, $d(y)$ est multiple de $d(x)$, et donc que $d(y) \geq d(x)$. Par symétrie, on a aussi $d(x) \geq d(y)$, ce qui démontre l'égalité $d(x) = d(y)$. \square

Soit C une classe de communication de période $d > 1$ et soit $x_0 \in C$. Tout point $x \in C$ communique avec x_0 ; soit $k \in \mathbb{N}^*$ le plus petit entier tel que $M^k(x, x_0) > 0$. On a, pour tout $n \in \mathbb{N}^*$,

$$M^{n+k}(x_0, x_0) \geq M^n(x_0, x)M^k(x, x_0),$$

ce qui montre que, pour tout $n \in \mathbb{N}^*$ tel que $M^n(x_0, x) > 0$, on a $M^{n+k}(x_0, x_0) > 0$, et donc, puisque x_0 est de période d , que $n+k \equiv 0 \pmod{d}$. Ainsi, il existe un unique entier $j \in \{0, 1, \dots, d-1\}$ (j est le reste de la division euclidienne de $-k$ par d) tel que l'on ait l'implication

$$M^n(x_0, x) > 0 \implies n \equiv j \pmod{d}.$$

On définit alors les **classes cycliques** C_j , $j = 0, 1, \dots, d-1$, de C de la manière suivante :

$$C_j = \{y \in C \mid M^n(x_0, y) > 0 \implies n \equiv j \pmod{d}\}.$$

autrement dit $y \in C_j$, si et seulement si

$$\{n \in \mathbb{N}^* \mid M^n(x_0, y) > 0\} \subset j + d\mathbb{N}^*.$$

Les C_j , $j = 0, 1, \dots, d-1$, forment une partition de C . De plus, si $x \in C_j$ et si y est tel que $M(x, y) > 0$, alors $y \in C_{j+1} \pmod{d}$. En effet, soit n tel que $M^n(x_0, x) > 0$; il est alors congru à j modulo d et on a donc $n+1 \equiv j+1 \pmod{d}$ et, d'après ce qui précède, $y \in C_{j+1}$.

Pour tout $n \in \mathbb{N}^*$, $X_{j+nd} \in C_j$ P. _{x_0} -p.s. et la sous-chaîne $(X_{j+nd})_{n \in \mathbb{N}^*}$, partant de $x_0 \in C_0$ à l'instant 0, est une chaîne de Markov homogène à valeurs dans C_j , de matrice de transition $(M^d(x, y))_{x, y \in C_j}$, irréductible et apériodique.

Pour illustrer ces premières notions, on reprend le modèle de diffusion d'Ehrenfest.

Exemple 16.10. (Modèle de diffusion de chaleur de Ehrenfest; suite.) On considère le modèle de Ehrenfest, décrit sous sa forme de tirages de boules dans une urne (voir l'exemple 16.2, dont on reprend les notations); on note X_n le nombre de boules rouges contenues dans l'urne à l'instant n . Puisque les tirages successifs sont uniformes, le processus $X = (X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov homogène à valeurs dans l'intervalle d'entiers $[0, 1, \dots, m]$, de matrice de transition M donnée par

$$\begin{aligned} & - \text{si } 1 \leq k \leq m-1, \quad M(k, k+1) = 1 - \frac{k}{m}, \quad M(k, k-1) = \frac{k}{m}, \text{ et} \\ & \quad M(k, l) = 0 \text{ si } l \neq k-1 \text{ ou } k+1, \\ & - \text{et (conditions frontières)} \end{aligned}$$

$$M(0, 1) = 1, \quad M(m, m-1) = 1. \quad (16.51)$$

Autrement dit, la matrice de transition M est donnée par, pour tout $k \in [0, 1, \dots, m]$,

$$M(k, k+1) = p_k, \quad M(k, k-1) = q_k.$$

où

$$p_k = 1 - \frac{k}{m}, \quad q_k = \frac{k}{m}.$$

Il est clair que tous les points de E communiquent; la chaîne est donc irréductible. De plus, 0 est apériodique; il en est donc de même de tous les points de E . Ainsi, **la chaîne de Ehrenfest est irréductible apériodique.**

Remarque. Ce modèle est un cas particulier des **processus de naissance et de mort** (voir ch. 16, exercice 4); ici les **barrières** 0 et m sont **réfléchissantes**, c'est-à-dire vérifient les conditions (16.51).

16.5.2. Récurrence

Définition 16.35. Un point x est

- **récurrent** si $P_x(T_x^1 < +\infty) = 1$,
- **récurrent nul** s'il est récurrent et si $E_x(T_x^1) = +\infty$,
- **récurrent positif** s'il est récurrent et si $E_x(T_x^1) < +\infty$,
- **transitoire** s'il n'est pas récurrent, autrement dit si

$$P_x(T_x^1 < +\infty) < 1.$$

Remarque. L'état x est récurrent nul si la chaîne, partant de x , retourne presque sûrement en x en un temps fini, mais « lentement ». Nous verrons ultérieurement la raison de cette terminologie en étudiant le problème de l'existence d'une probabilité invariante (cf. théorème 16.54).

Le lemme suivant est préliminaire au théorème de classification des états.

Lemme 16.36. Pour tout $x \in E$, on a

(a) pour tout $p \in \mathbb{N}^*$, $P_x(T_x^p < +\infty) = [P_x(T_x^1 < +\infty)]^p$;

(b) L'ensemble R_x des trajectoires qui passent une infinité de fois en x , défini par

$$R_x = \limsup_n (X_n = x) = (N_x = +\infty),$$

est égal à $\bigcap_{p \in \mathbb{N}^*} (T_x^p < +\infty)$ et vérifie

$$P_x(R_x) = \lim_p \searrow [P_x(T_x^1 < +\infty)]^p :$$

(c) Le potentiel $R(x, x)$, c'est-à-dire le nombre moyen de passages en x lorsque la chaîne part de x à l'instant 0, est donné par

$$R(x, x) = \sum_{p=0}^{+\infty} [P_x(T_x^1 < +\infty)]^p.$$

Démonstration. (a) C'est un corollaire de la proposition 16.23; nous en donnons toutefois une démonstration directe. Puisque

$$(T_x^{p+1} < +\infty) \subset (T_x^p < +\infty),$$

la relation (16.41) permet d'écrire (en conditionnant par rapport à $\mathcal{A}_{T_x^p}$, et en appliquant la propriété de Markov forte) la suite d'égalités :

$$\begin{aligned} P_x(T_x^{p+1} < +\infty) &= E_x \left[E_x^{\mathcal{A}_{T_x^p}} \left(\mathbf{1}_{(T_x^p < +\infty)} \mathbf{1}_{(\tau_x^1(\theta_{T_x^p}(X)) < +\infty)} \right) \right] \\ &= E_x \left[\mathbf{1}_{(T_x^p < +\infty)} E_{X_{T_x^p}} \left(\mathbf{1}_{(\tau_x^1(X) < +\infty)} \right) \right], \end{aligned}$$

soit, puisque $\tau_x^1(X) = T_x^1$ et que $X_{T_x^p} = x$,

$$P_x(T_x^{p+1} < +\infty) = P_x(T_x^p < +\infty) P_x(T_x^1 < +\infty),$$

ce qui donne le résultat par itération.

(b) Par définition de R_x et des temps de passage en x , on a l'égalité $R_x = \bigcap_{p \in \mathbb{N}^*} (T_x^p < +\infty)$; il suffit alors de remarquer que la suite des ensembles $(T_x^p < +\infty)$ est décroissante pour obtenir le résultat.

(c) Par définition, on a

$$R(x, x) = \sum_{n \in \mathbb{N}} E_x(\mathbf{1}_{(X_n = x)}).$$

On note $T_x^0 = 0$. La suite des temps T_x^p est strictement croissante : de plus, pour tout p , on a $p \leq T_x^p$, ce qui implique que $\lim_p T_x^p = +\infty$. On peut donc partitionner \mathbb{N} à l'aide des intervalles aléatoires $[T_x^p, T_x^{p+1}[$, $p \in \mathbb{N}$, et écrire

$$R(x, x) = \sum_{p \in \mathbb{N}} E_x \left[\sum_{n \in [T_x^p, T_x^{p+1}[} \mathbf{1}_{(X_n = x)} \right];$$

en remarquant que l'intervalle $[T_x^p, T_x^{p+1}[$ est vide dès que $T_x^p = +\infty$, que, par définition des temps T_x^p , on a $\mathbf{1}_{(X_n = x)} = 0$ pour tout $n \in]T_x^p, T_x^{p+1}[$, et que $\mathbf{1}_{(X_{T_x^p} = x)} = 1$ sur $(T_x^p < +\infty)$, on obtient

$$\begin{aligned} R(x, x) &= \sum_{p \in \mathbb{N}} E_x \left[\mathbf{1}_{(T_x^p < +\infty)} \right] = \sum_{p \in \mathbb{N}} P_x(T_x^p < +\infty) \\ &= \sum_{p \in \mathbb{N}} [P_x(T_x^1 < +\infty)]^p. \end{aligned} \quad \square$$

Théorème 16.37 (Classification des états). *On a l'alternative*

1. x est récurrent : dans ce cas, $P_x(R_x) = 1$ et $R(x, x) = +\infty$;
2. x est transitoire ; dans ce cas, $P_x(R_x) = 0$ et $R(x, x) < +\infty$.

De plus, si x est récurrent et si x conduit à y alors y conduit à x , y est récurrent et $P_y(T_y^1 < +\infty) = 1$.

Démonstration. L'alternative résulte immédiatement du lemme 16.36. Démontrons la dernière assertion. Supposons que x soit récurrent et que x conduise à y . Dire que la chaîne passe par y après être passée par x puis ne repasse plus par x implique qu'elle ne passe qu'un nombre fini de fois par x ; on a donc l'inclusion des ensembles

$$(T_x^1 < +\infty) \cap (\tau_y^1[\theta_{T_x^1}^1(X)] < +\infty) \cap (\tau_x^1[\theta_{\tau_y^1[\theta_{T_x^1}^1(X)]}^1(X)] < +\infty) \subset R_x^c. \quad (16.52)$$

Les deux derniers ensembles du premier membre font intervenir des fonctionnelles du futur du processus après le temps T_x^1 . Ainsi, en conditionnant par rapport à $\mathcal{A}_{T_x^1}$, en appliquant la propriété de Markov forte et en tenant compte de ce que $X_{T_x^1} = x$, on a :

$$\begin{aligned} P_x \left[(T_x^1 < +\infty) \cap (\tau_y^1[\theta_{T_x^1}^1(X)] < +\infty) \cap (\tau_x^1[\theta_{\tau_y^1[\theta_{T_x^1}^1(X)]}^1(X)] = +\infty) \right] \\ = E_x \left[\mathbf{1}_{(T_x^1 < +\infty)} E_x(\mathbf{1}_{(T_x^1 < +\infty)} \mathbf{1}_{(\tau_y^1[\theta_{T_x^1}^1(X)] = +\infty)}) \right] \\ = P_x(T_x^1 < +\infty) E_x \left[\mathbf{1}_{(T_x^1 < +\infty)} \mathbf{1}_{(\tau_x^1[\theta_{T_x^1}^1(X)] = +\infty)} \right]. \end{aligned}$$

En conditionnant maintenant par rapport à $\mathcal{A}_{T_y^1}$, en appliquant la propriété de Markov forte et en tenant compte de ce que $X_{T_y^1} = y$, on a alors

$$\begin{aligned} P_x \left[(T_x^1 < +\infty) \cap (\tau_y^1[\theta_{T_x^1}^1(X)] < +\infty) \cap (\tau_x^1[\theta_{\tau_y^1[\theta_{T_x^1}^1(X)]}^1(X)] = +\infty) \right] \\ = P_x(T_x^1 < +\infty) E_x \left[\mathbf{1}_{(T_y^1 < +\infty)} E_y(\mathbf{1}_{(\tau_x^1(X) = +\infty)}) \right] \\ = P_x(T_x^1 < +\infty) P_x(T_y^1 < +\infty) P_y(T_x^1 = +\infty). \end{aligned}$$

Il résulte alors de l'inclusion (16.52) que :

$$P_x(T_x^1 < +\infty) P_x(T_y^1 < +\infty) P_y(T_x^1 = +\infty) \leq P_x(R_x^c). \quad (16.53)$$

Le point x étant récurrent, on a, comme on vient de le voir, $P_x(R_x^c) = 0$. De plus, puisque x conduit à y , on a $P_x(T_y^1 < +\infty) > 0$. Il résulte alors de (16.53) que $P_y(T_x^1 = +\infty) = 0$, ou encore que $P_y(T_x^1 < +\infty) = 1$; en particulier y conduit à x .

Enfin, y est récurrent; en effet, puisque x et y communiquent, il existe i et $j \in \mathbb{N}^*$ tels que $M^i(x, y) > 0$ et $M^j(y, x) > 0$. De plus, pour tout $n \in \mathbb{N}^*$, on a

$$M^{n+i+j}(y, y) \geq M^j(y, x) M^n(x, x) M^i(x, y),$$

et, x étant récurrent, on a aussi $R(x, x) = \sum_{n \in \mathbb{N}} M^n(x, x) = +\infty$. Il en résulte que $\sum_{n \in \mathbb{N}} M^{n+i+j}(y, y) = +\infty$, et donc que $R(y, y) = +\infty$, ce qui démontre que y est récurrent. \square

16.5.3. Comportement asymptotique et classification

La loi de X_n est donnée par $P_x(X_n = y) = M^n(x, y)$; il est souvent impossible de la calculer effectivement. dès que la matrice M est trop grande ou pas assez creuse. Il est donc important d'obtenir des résultats asymptotiques.

Proposition 16.38. *Si y est transitoire, on a, pour tout $x \in E$, $R(x, y) < +\infty$ et $\lim_n M^n(x, y) = 0$.*

Démonstration. On rappelle que (corollaire 16.28)

$$R(x, y) = \begin{cases} \frac{1}{1 - F(y, y)} & \text{si } x = y, \\ F(x, y)R(y, y) & \text{si } x \neq y. \end{cases}$$

Le point y étant transitoire, on a $R(y, y) < +\infty$, et donc aussi $R(x, y) < +\infty$. Puisque $R(x, y) = \sum_{n=0}^{+\infty} M^n(x, y)$, le terme général de cette série convergente tend vers zéro. \square

Pour démontrer la proposition suivante, on s'appuie sur un lemme d'analyse que nous donnons ci-dessous, sans démonstration. Il faut savoir que ce n'est pas un résultat immédiat ; on peut en trouver une démonstration dans le livre de W. Feller (*An introduction to probability theory and its applications*, vol. 1, 1964, John Wiley and Sons Publishers, section XIII.10 p. 306).

Lemme 16.39. *Soit $(f_j)_{j \in \mathbb{N}^*}$ une suite de réels positifs telle que $\sum_{j \in \mathbb{N}^*} f_j = 1$ et $\text{pgcd}\{j \mid f_j > 0\} = 1$. Soit $(u_n)_{n \in \mathbb{N}}$ une suite de réels vérifiant*

$$u_0 = 1, \quad \text{pour tout } n \in \mathbb{N}^*, \quad u_n = \sum_{j=1}^n f_j u_{n-j}.$$

La suite $(u_n)_{n \in \mathbb{N}}$ est convergente et

$$\lim_n u_n = \frac{1}{\sum_{j=1}^{+\infty} j f_j}.$$

Proposition 16.40. *Si y est récurrent apériodique, pour tout $x \in E$, la suite de terme général $M^n(x, y)$ est convergente et*

$$\lim_n M^n(x, y) = \frac{F(x, y)}{E_y(T_y^1)}, \quad (16.54)$$

avec la convention $\frac{1}{\infty} = 0$.

Démonstration. Puisque $(X_n = y) \subset (T_y^1 \leq n)$, on a

$$\begin{aligned} M^n(x, y) &= E_x \left[\mathbf{1}_{(X_n=y)} \mathbf{1}_{(T_y^1 \leq n)} \right] = \sum_{j=1}^n E_x \left[\mathbf{1}_{(X_n=y)} \mathbf{1}_{(T_y^1=j)} \right] \\ &= \sum_{j=1}^n E_x \left[\mathbf{1}_{(T_y^1=j)} \mathbf{1}_{(X_{T_y^1+n-j}=y)} \right], \end{aligned}$$

ce qui donne, par application de la propriété de Markov forte (après conditionnement par rapport à $\mathcal{A}_{T_y^1}^y$),

$$\begin{aligned} M^n(x, y) &= \sum_{j=1}^n E_x[\mathbf{1}_{(T_y^1=j)} E_{X_{T_y^1}} \mathbf{1}_{(X_{n-j}=y)}] \\ &= \sum_{j=1}^n E_x[\mathbf{1}_{(T_y^1=j)}] E_y[\mathbf{1}_{(X_{n-j}=y)}], \end{aligned}$$

c'est-à-dire

$$M^n(x, y) = \sum_{j=1}^n F_j(x, y) M^{n-j}(y, y). \quad (16.55)$$

Supposons d'abord que $x = y$. On applique le lemme 16.39 avec $f_j = F_j(y, y)$ et $u_n = M^n(y, y)$. La relation (16.55) s'écrit alors, pour tout $n \in \mathbb{N}^*$, $u_n = \sum_{j=1}^n f_j u_{n-j}$. Mais, si $d = \text{pgcd}\{j \mid f_j > 0\}$, on montre, par récurrence à partir de cette dernière relation, que $\{n \mid u_n > 0\} \subset d\mathbb{N}^*$; puisque y est apériodique, on a $d = 1$. Par ailleurs, y étant récurrent, on a $\sum_{j \in \mathbb{N}^*} F_j(y, y) = P_y(T_y^1 < +\infty) = 1$. Tenant compte des égalités

$$F(y, y) = 1 \quad \text{et} \quad \sum_{j=1}^{+\infty} j f_j = \sum_{j=1}^{+\infty} j P_y(T_y^1 = j) = E_y(T_y^1),$$

le lemme 16.39 établit l'égalité (16.54).

Si $x \neq y$, l'égalité (16.55) peut s'écrire

$$M^n(x, y) = \sum_{j=1}^{+\infty} [\mathbf{1}_{(j \leq n)} M^{n-j}(y, y)] F_j(x, y). \quad (16.56)$$

Interprétons cette somme comme l'intégrale de l'application

$$j \mapsto \mathbf{1}_{(j \leq n)} M^{n-j}(y, y)$$

par rapport à la mesure $\sum_{j=1}^{+\infty} F_j(x, y) \delta_j$ de masse finie égale à

$$\sum_{j=1}^{+\infty} F_j(x, y) = P_x(T_y^1 < +\infty) \leq 1;$$

remarquant que, pour tout $n \in \mathbb{N}^*$, on a $0 \leq \mathbf{1}_{(j \leq n)} M^{n-j}(y, y) \leq 1$, le théorème de convergence dominée assure la convergence de la suite de terme général $M^n(x, y)$ et démontre, compte tenu de la première partie, que

$$\begin{aligned} \lim_n M^n(x, y) &= \sum_{j=1}^{+\infty} \left[\lim_n \mathbf{1}_{(j \leq n)} M^{n-j}(y, y) \right] F_j(x, y) \\ &= \sum_{j=1}^{+\infty} \left[\frac{1}{E_y(T_y^1)} \right] F_j(x, y) = \frac{F(x, y)}{E_y(T_y^1)}. \quad \square \end{aligned}$$

La proposition suivante donne le **comportement asymptotique** de la chaîne relativement à une classe **récurrente périodique**.

Proposition 16.41. *Soit y un état récurrent périodique de période $d > 1$.*

(a) *Si x communique avec y et si $x \in C_r$ et $y \in C_{r+a}$, où C_j , $j = 0, 1, \dots, d-1$, sont les classes cycliques de C , la suite de terme général $M^{nd+a}(x, y)$ est convergente et*

$$\lim_n M^{nd+a}(x, y) = \frac{d}{E_y(T_y^1)}. \quad (16.57)$$

(b) *Si x est quelconque, on a, pour tout $a = 0, 1, \dots, d-1$,*

$$\lim_n M^{nd+a}(x, y) = \left[\sum_{j=1}^{+\infty} F_{jd+a}(x, y) \right] \frac{d}{E_y(T_y^1)}. \quad (16.58)$$

Démonstration. (a) Si $a = 0$, y est récurrent apériodique pour la chaîne de Markov homogène $(X_{nd})_{n \in \mathbb{N}}$ de matrice de transition M^d ; on a alors, d'après la proposition 16.40,

$$\lim_n M^{nd}(x, y) = \frac{1}{E_y(S_y^1)},$$

où $S_y^1 = \inf\{n \in \mathbb{N}^* \mid X_{nd} = y\}$. Puisque $P_y(S_y^1 = k) = P_y(T_y^1 = kd)$, on a $E_y(S_y^1) = \frac{1}{d} E_y(T_y^1)$, ce qui démontre (16.57) dans ce cas.

Supposons alors le résultat vrai jusqu'à l'ordre $a < d-1$ et démontrons le à l'ordre $a+1$. On a

$$M^{nd+a+1}(x, y) = \sum_{z \in E} M(x, z) M^{nd+a}(z, y);$$

l'hypothèse de récurrence et le théorème de convergence dominée donnent

$$\lim_n M^{nd+a+1}(x, y) = \sum_{z \in E} \left[M(x, z) \frac{d}{E_y(T_y^1)} \right] = \frac{d}{E_y(T_y^1)}.$$

(b) D'après l'égalité (16.55), on a

$$M^{nd+a}(x, y) = \sum_{j=1}^{nd+a} F_j(x, y) M^{nd+a-j}(y, y).$$

Le point y ayant pour période d , $M^{nd+a-j}(y, y) = 0$, sauf si $a-j \in d\mathbb{Z}$; on a donc

$$\begin{aligned} M^{nd+a}(x, y) &= \sum_{k=0}^n F_{kd+a}(x, y) M^{(n-k)d}(y, y) \\ &= \sum_{k=0}^{+\infty} [\mathbf{1}_{(k \leq n)} M^{(n-k)d}(y, y)] F_{kd+a}(x, y). \end{aligned}$$

Puisque

$$\sum_{k=0}^{+\infty} F_{kd+a}(x, y) = \sum_{k=0}^{+\infty} P_x(T_y^1 = kd + a) \leq P_x(T_y^1 < +\infty) \leq 1,$$

on obtient (16.58), par application de (16.57) et du théorème de convergence dominée. \square

On déduit de cette proposition deux corollaires qui permettent de trouver la **nature des classes de communication**.

Corollaire 16.42. *Tous les états d'une classe de communication C sont de même nature, à savoir, transitoires, récurrents positifs, récurrents nuls, apériodiques ou de même période. La nature de la classe C est alors, par définition, celle de l'un quelconque de ses points.*

Démonstration. Soient $x, y \in C$;

- Puisque x conduit à y , si y est transitoire, x l'est aussi (sinon, d'après le théorème 16.37, y serait récurrent).
- Si y est récurrent apériodique, il en est de même de x , puisque x et y communiquent (cf le théorème 16.37 et la proposition 16.34). De plus, si y est récurrent nul, il résulte des propositions 16.40 et 16.41 que $\lim_n M^n(y, y) = 0$. Puisque x et y communiquent, il existe k et $l \in \mathbb{N}^*$ tels que $M^k(x, y) > 0$ et $M^l(y, x) > 0$; comme on a, pour tout $n \in \mathbb{N}^*$,

$$M^{n+k+l}(y, y) \geq M^l(y, x) M^n(x, x) M^k(x, y),$$

il vient $\lim_n M^n(x, x) = 0$, ce qui, toujours d'après la proposition 16.40, démontre que x est récurrent nul.

- Si y est récurrent apériodique **positif**, il en est de même de x , sinon, d'après l'assertion précédente, y serait récurrent apériodique nul.
- Le reste de l'énoncé est la proposition 16.34. \square

Définition 16.43. *Une classe de communication C est fermée (ou absorbante) si, pour tout $x \in C$, $P_x(T_C < +\infty) = 0$. (Arrivant dans une telle classe, on ne la quitte plus.) En particulier, si la classe fermée C est réduite à un point, on dit que ce point est absorbant. (Arrivant en un tel point, on ne le quitte plus.)*

Corollaire 16.44. Soit C une classe de communication fermée. Si C est de cardinal fini, elle ne contient ni état transitoire, ni état récurrent nul.

En particulier, une chaîne de Markov homogène irréductible et finie ne contient que des points récurrents positifs.

Démonstration. Si les états de C étaient soit transitoires, soit récurrents nuls, il résulterait des propositions 16.40 et 16.41, quitte à prendre une sous-suite, que $\lim_n M^n(x, y) = 0$ pour tous $x, y \in C$; puisque C est fini, on aurait

$$\lim_n \sum_{y \in C} M^n(x, y) = 0.$$

Il y aurait contradiction avec le fait que C est fermée, puisque l'on aurait, pour tout $n \in \mathbb{N}^*$,

$$0 = P_x(T_C < +\infty) = P_x\left(\bigcup_{k \in \mathbb{N}^*} (X_k \notin C)\right) \geq P_x(X_n \notin C) = 1 - \sum_{y \in C} M^n(x, y),$$

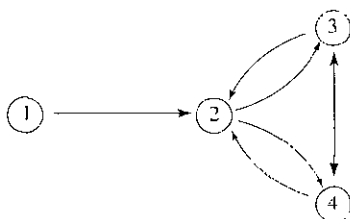
et, en passant à la limite, $0 \geq 1$. □

Définition 16.45. Un état x est **inessentiel** s'il existe $m \in \mathbb{N}^*$ et $y \neq x$ tels que $M^m(x, y) > 0$ et, pour tout $n > m$ et tout $z \in E$, $M^n(z, x) = 0$. Autrement dit, un état inessentiel est un état qu'avec une probabilité positive la chaîne quitte après un nombre fini (non aléatoire) d'étapes sans jamais y revenir.

Par exemple, pour la chaîne de Markov homogène X , à valeurs dans $E = \{1, 2, 3, 4\}$, et de matrice de transition M

$$M = \begin{array}{c} \begin{array}{cccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{1} & \left(\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{array} \right) \\ \mathbf{2} & \\ \mathbf{3} & \\ \mathbf{4} & \end{array} \end{array}$$

le graphe associé est



et on voit que 1 est un état **inessentiel**. Les autres états communiquent tous; la chaîne restreinte à l'espace d'états $\{2, 3, 4\}$ est alors irréductible, récurrente positive de période 2.

Proposition 16.46. *Un état inessentiel est transitoire.*

Démonstration. Soit x un état inessentiel; soient $m \in \mathbb{N}^*$ et $y \neq x$ tels que $M^m(x, y) > 0$ et, pour tout $n > m$ et tout $z \in E$, $M^n(z, x) = 0$. On a alors, pour tout $n > m$,

$$M^{n+m}(x, x) = \sum_{z \in E} M^m(x, z) M^n(z, x) = 0.$$

Il en résulte que $R(x, x) < +\infty$, et donc que x est transitoire. \square

En résumé, pour étudier le comportement d'une chaîne de Markov homogène, on cherche les états inessentiels (ils sont alors transitoires), puis les classes de communication des états essentiels (ceux qui ne sont pas inessentiels). On cherche ensuite à préciser la nature de chaque classe à l'aide des caractérisations étudiées précédemment.

L'exemple suivant met en évidence un des nombreux liens entre martingales et chaînes de Markov.

Exemple 16.11. (Chaîne de Markov et martingale; un exemple de modèle génétique.) Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène à valeurs dans l'ensemble des entiers $E = \{0, 1, \dots, N\}$, de matrice de transition M . Si le processus X est aussi une **martingale** par rapport à sa filtration naturelle $(\mathcal{A}_n)_{n \in \mathbb{N}}$, pour toute probabilité P_x qui fait partir la chaîne de x à l'instant 0, **les points frontière 0 et N sont absorbants**. De plus, si on définit les premiers temps de visite en x par

$$\tau_x = \inf\{n \in \mathbb{N}^* \mid X_n = x\} \quad \text{avec} \quad \inf \emptyset = +\infty,$$

on a, pour tout $x \in E$,

$$\boxed{P_x(\tau_N < \tau_0) = \frac{x}{N}}. \quad (16.59)$$

En effet, pour tout $x \in E$ et tout $f \in bE$, on a

$$E_x^{\mathcal{A}_n} \{f(X_{n+1})\} = M(X_n, f) = \sum_{y \in E} f(y) M(X_n, y):$$

en particulier, en prenant pour f l'application identique sur E (elle est bornée), on obtient

$$E_x^{\mathcal{A}_0}(X_1) = \sum_{y \in E} y M(X_0, y).$$

Puisque X est une martingale, en prenant la moyenne des deux membres de l'inégalité précédente par rapport à P_x , on a alors

$$E_x(X_0) = E_x(X_1) = \sum_{y \in E} y E_x[M(X_0, y)] = \sum_{y \in E} y M(x, y). \quad (16.60)$$

En prenant $x = 0$ dans (16.60), on a

$$0 = E_0(X_0) = \sum_{y \in E} y M(0, y) = \sum_{y=1}^N y M(0, y),$$

ce qui démontre, puisque tous les termes sont positifs ou nuls, que $M(0, y) = 0$ pour tout $y = 1, \dots, N$, et donc que $M(0, 0) = 1$, c'est-à-dire que le point 0 est **absorbant**.

De même, en prenant $x = N$ dans (16.60), on a

$$N = E_N(X_0) = \sum_{y \in E} y M(N, y) = \sum_{y=1}^{N-1} y M(N, y) + N M(N, N);$$

puisque tous les termes sont positifs ou nuls et que $\sum_{y=0}^N M(N, y) = 1$, on a $M(N, y) = 0$ pour tout $y = 0, \dots, N-1$, et donc $M(N, N) = 1$. Ainsi, le point N est **absorbant**.

Puisque E est borné, la martingale X est équi-intégrable; le **deuxième théorème d'arrêt** appliqué au temps d'arrêt $\tau_0 \wedge \tau_N$ assure donc que

$$E_x(X_0) = E_x(X_{\tau_0 \wedge \tau_N}).$$

En particulier, on a

$$x = E_x(X_0) = E_x[\mathbf{1}_{\{\tau_N < \tau_0\}} \cdot N + \mathbf{1}_{\{\tau_N > \tau_0\}} \cdot 0] = N P_x(\tau_N < \tau_0),$$

ce qui démontre (16.59).

Application à un modèle génétique⁸.

On considère une population qui se reproduit en gardant la même taille N . Un individu d'une génération possède deux allèles⁹ de type G ou g , si bien qu'à une génération donnée, le nombre d'allèles est de $2N$. On suppose que les mariages dans une génération donnée sont indépendants et uniformes parmi les individus de la population (en anglais, *random mating*), si bien que, si X_n est le nombre d'allèles de type G existant parmi la population de la n -ième génération, le processus $X = (X_n)_{n \in \mathbb{I}}$ est une chaîne de Markov homogène de matrice de transition M engendrant les probabilités (on identifie le germe et la probabilité)

$$M(i, \cdot) = \begin{cases} \mathcal{B}(2N, \frac{i}{2N}) & \text{si } 1 \leq i \leq 2N-1, \\ \delta_0 & \text{si } i = 0, \\ \delta_N & \text{si } i = N. \end{cases}$$

8. Problème étudié par R.A. Fisher et S. Wright et formulé en termes de chaîne de Markov par G. Malécot (C.R.A.S. 1944, pp. 379-381).

9. Pour quelques explications d'ordre génétique, voir l'exemple 16.13 ci-après.

Ainsi, on a

$$E^{X_n=i}(X_{n+1}) = \begin{cases} (2N) \cdot \frac{i}{2N} = i & \text{si } 1 \leq i \leq 2N-1, \\ 0 & \text{si } i = 0, \\ N & \text{si } i = N. \end{cases}$$

En appelant $E = \{0, 1, \dots, 2N\}$ l'ensemble d'états, cela démontre que $E_x^{\sigma(X_n)}(X_{n+1}) = X_n$ pour tout $x \in E$. Avec les notations ci-dessus, puisque X est une chaîne de Markov et que E est fini, on a alors, pour tout $x \in E$,

$$E_x^{A_n}[X_{n+1}] = X_n.$$

Il en résulte que les points 0 et $2N$ sont absorbants; **autrement dit, à long terme, il ne restera plus que des allèles d'un seul type.**

16.5.4. Critère analytique de récurrence

On donne d'abord un moyen de calcul de la probabilité que la chaîne reste toujours dans une partie A de E . Pour cela, on note Q la restriction de M à A , c'est-à-dire la matrice indexée sur $A \times A$ définie par, pour tous $x, y \in A$, $Q(x, y) = M(x, y)$. On a, pour tout $n \geq 2$,

$$\begin{aligned} Q^n(x, y) &= \sum_{x_1 \in A} \sum_{x_2 \in A} \dots \sum_{x_{n-1} \in A} Q(x, x_1)Q(x_1, x_2) \dots Q(x_{n-1}, y) \\ &= P_x(X_1 \in A, \dots, X_{n-1} \in A, X_n = y). \end{aligned}$$

si bien que

$$P_x(X_1 \in A, \dots, X_{n-1} \in A, X_n \in A) = \sum_{y \in A} Q^n(x, y) \equiv Q^n(x, A).$$

On note, si $x \in A$,

$$f_n(x) = P_x \left[\bigcap_{j=1}^n (X_j \in A) \right] = \sum_{y \in A} Q^n(x, y);$$

la suite de terme général $f_n(x)$ est alors décroissante et converge vers $f(x) = P_x \left[\bigcap_{j \in \mathbb{N}^*} (X_j \in A) \right]$.

Proposition 16.47. *La fonction f définie sur A est solution maximale du système*

$$h = Qh \quad 0 \leq h \leq 1, \quad (16.61)$$

où, comme, précédemment¹⁰, on note $h(x) = \sum_{y \in A} Q(x, y)h(y)$. De plus, soit $f = 0$, soit $\sup_{x \in A} f(x) = 1$.

¹⁰ Pour l'interprétation vectorielle des fonctions, voir la notation suivant la définition 16.6. En particulier, on emploiera ici la notation Qf , bien adaptée au calcul vectoriel, en place de $Q(\cdot, f)$.

Démonstration. Puisque, pour tout $x \in A$, on a, par associativité et commutativité des sommes (les termes étant positifs)

$$\begin{aligned} f_{n+1}(x) &= \sum_{y \in A} \left[\sum_{z \in A} Q(x, z) Q^n(z, y) \right] = \sum_{z \in A} Q(x, z) \left[\sum_{y \in A} Q^n(z, y) \right] \\ &= Q f_n(x), \end{aligned}$$

on a $f_{n+1} = Q f_n$. Il en résulte que, par application du théorème de convergence dominée, on a $f = Q f$; de plus, on a bien sûr $0 \leq f \leq 1$. Ainsi, f est solution du système (16.61); montrons qu'elle est maximale. Soit h une autre solution du système. En notant 1 la fonction constante, on a alors, pour tout $n \in \mathbb{N}^*$,

$$h = Q^n h \leq Q^n 1 = f_n,$$

ce qui implique, par passage à la limite, $h \leq f$.

Enfin, si f est non nulle, notons $c = \sup_{x \in A} f(x)$. Par le même calcul, on a, pour tout $n \in \mathbb{N}^*$,

$$f = Q^n f \leq Q^n c = c f_n,$$

et donc, par passage à la limite, $f \leq c f$, ce qui implique que $c \geq 1$; puisque de plus $c \leq 1$ (car on a $0 \leq f \leq 1$), on a $c = 1$. \square

Corollaire 16.48. *Soit X une chaîne de Markov homogène irréductible de matrice de transition M ; soit x_0 un point quelconque de E et Q la matrice restriction de M à $E \setminus \{x_0\}$. Pour que X soit récurrente il faut et il suffit que le système*

$$h = Q h \quad 0 \leq h \leq 1, \quad (16.62)$$

ait pour unique solution $h = 0$.

Démonstration. Supposons que 0 soit l'unique solution du système (16.62). La chaîne étant irréductible, tous les états sont de même nature. De plus, en notant $A = E \setminus \{x_0\}$, il existe $\gamma \in A$ tel que x_0 conduise à γ . D'après la proposition 16.47, la fonction f , définie relativement à cette partie A , est solution maximale du système (16.62), ce qui implique, par hypothèse, que $P_\gamma \left[\bigcap_{j \in \mathbb{N}^*} (X_j \in A) \right] = 0$, soit encore que $P_\gamma(T_{x_0} < +\infty) = 1$. Montrons que cela entraîne que x_0 est récurrent.

Par la propriété de Markov simple, on a, pour tout $x \in A$ et tous $n, p \in \mathbb{N}^*$,

$$P_x \left[\bigcap_{j=p}^{n+p} (X_j \in A) \right] = E_x \left[\mathbf{1}_{(X_p \in A)} Q^n(X_p, A) \right] = \sum_{z \in A} M^p(x, z) Q^n(z, A);$$

en passant à la limite en n (le membre de gauche est décroissant en n et le membre de droite est susceptible d'application du théorème de convergence

dominée), on a, puisque $f = 0$,

$$P_x \left[\bigcap_{j=p}^{+\infty} (X_j \in A) \right] = \sum_{z \in A} M^p(x, z) \left[\lim_n Q^n(z, A) \right] = \sum_{z \in A} M^p(x, z) f(z) = 0.$$

Il en résulte que $P_x(\liminf_n (X_n \in A)) = 0$, et donc que, pour tout $x \in A$,

$$P_x(N_{x_0} = +\infty) \geq P_x(\limsup_n (X_n = x_0)) = 1. \quad (16.63)$$

Mais, dire que la chaîne atteint y en un temps fini et passe en x_0 une infinité de fois est équivalent à dire qu'elle atteint y en un temps fini et passe en x_0 une infinité de fois après avoir atteint y une première fois, ce qui s'écrit

$$(N_{x_0} = +\infty) \cap (T_y^1 < +\infty) = (T_y^1 < +\infty) \cap (n_{x_0}[\theta_{T_y^1}(X)] = +\infty):$$

en conditionnant par rapport à $\mathcal{A}_{T_y^1}$ et en appliquant la propriété de Markov forte, on a alors

$$\begin{aligned} P_{x_0}[(N_{x_0} = +\infty) \cap (T_y^1 < +\infty)] \\ &= E_{x_0} \left[\mathbf{1}_{(T_y^1 < +\infty)} E_{x_0}^{\mathcal{A}_{T_y^1}} (\mathbf{1}_{(n_{x_0}[\theta_{T_y^1}(X)] = +\infty)}) \right] \\ &= E_{x_0} \left[\mathbf{1}_{(T_y^1 < +\infty)} E_y (\mathbf{1}_{(n_{x_0}(X) = +\infty)}) \right]. \end{aligned}$$

et donc

$$P_{x_0}[(N_{x_0} = +\infty) \cap (T_y^1 < +\infty)] = P_{x_0}(T_y^1 < +\infty) P_y(N_{x_0} = +\infty).$$

Puisque, de plus, x_0 conduit à y , il résulte de cette dernière égalité et de (16.63) que

$$P_{x_0}(N_{x_0} = +\infty) \geq P_{x_0}(T_y^1 < +\infty) P_y(N_{x_0} = +\infty) = P_{x_0}(T_y^1 < +\infty) > 0.$$

Mais $P_{x_0}(N_{x_0} = +\infty)$ ne prend que les valeurs 0 ou 1; on a donc $P_{x_0}(N_{x_0} = +\infty) = 1$, ce qui démontre que x_0 est récurrent.

Inversement, si x_0 est récurrent, pour tout $z \in A$, on a $P_z(T_{x_0}^1 < +\infty) = 1$ et donc $f(z) = P_z(\bigcap_{j \in \mathbb{N}^*} (X_j \in A)) = 0$; la proposition 16.47 affirmant que f est solution maximale du système (16.62), f est l'unique solution de ce système. \square

Remarque. Le corollaire 16.48, d'apparence restrictive, puisqu'énoncé pour une chaîne irréductible, est en fait d'usage général pour déterminer si une classe C de communication est récurrente: il suffit d'appliquer ce corollaire à la chaîne restreinte à la classe C qui, elle, est bien une chaîne irréductible. Évidemment, tout ceci n'a d'intérêt que si E est infini.

16.6. Calcul de la matrice potentiel et de $P_x(T_y^1 < +\infty)$

16.6.1. Calcul de la matrice potentiel

Soient x et $y \in E$.

Si y est récurrent, il résulte de l'égalité (16.49) du corollaire 16.28 que

$$R(x, y) = \begin{cases} 0 & \text{si } F(x, y) = 0, \\ +\infty & \text{si } F(x, y) > 0. \end{cases}$$

Si y est transitoire,

– **si x est récurrent**, x ne conduit pas à y ; par conséquent, $F(x, y) = 0$ et $R(x, y) = 0$;

– **si x est transitoire**; notons D l'ensemble des points transitoires, Q et S , respectivement, les restrictions à $D \times D$ des matrices M et R . Pour résoudre ce cas, on s'intéresse globalement au calcul de S . Après une éventuelle réindexation des points de E , en indexant en premier les points récurrents, la matrice M a la structure de blocs

$$M = \begin{pmatrix} K & 0 \\ L & Q \end{pmatrix};$$

on a donc

$$M^n = \begin{pmatrix} K^n & 0 \\ L_n & Q^n \end{pmatrix}$$

(prendre garde que L_n n'est pas une puissance de L) et, par conséquent,

$$R = \sum_{n=0}^{+\infty} M^n = \begin{pmatrix} \sum_{n=0}^{+\infty} K^n & 0 \\ \sum_{n=0}^{+\infty} L_n & \sum_{n=0}^{+\infty} Q^n \end{pmatrix}.$$

Ainsi, on a

$$S = \sum_{n=0}^{+\infty} Q^n.$$

En notant I la matrice identité sur D , on a alors

$$SQ = QS = S - I,$$

soit

$$(I - Q)S = S(I - Q) = I;$$

en particulier, **si l'ensemble D des points transitoires est fini**, on a

$$S = (I - Q)^{-1}.$$

16.6.2. Calcul de $F(x, y) \equiv P_x(T_y^1 < +\infty)$

Si x et y sont récurrents,

$$F(x, y) = \begin{cases} 1 & \text{s'ils appartiennent à la même classe de communication,} \\ 0 & \text{sinon.} \end{cases}$$

Si x est récurrent et y est transitoire, $F(x, y) = 0$.

Si x et y sont transitoires, il résulte de l'égalité (16.49) du corollaire 16.28 que

$$F(x, y) = \begin{cases} \frac{R(x, y)}{R(y, y)} & \text{si } x \neq y, \\ 1 - \frac{1}{R(y, y)} & \text{sinon.} \end{cases}$$

Si x est transitoire et y est récurrent, la réponse est donnée par les propositions suivantes.

Proposition 16.49. *Soit C une classe de communication récurrente. Pour tout point x transitoire, on a, pour tout $y \in C$,*

$$F(x, y) = P_x(T_C < +\infty).$$

Démonstration. On a évidemment, puisque $y \in C$,

$$P_x(T_y^1 < +\infty) \leq P_x(T_C < +\infty).$$

Inversement,

$$P_x(T_C < +\infty) = E_x \left[\mathbf{1}_{(T_y^1 < +\infty)} \mathbf{1}_{(T_C < +\infty)} \right] + E_x \left[\mathbf{1}_{(T_y^1 = +\infty)} \mathbf{1}_{(T_C < +\infty)} \right],$$

et par conséquent, en remarquant que

$$(T_y^1 = +\infty) \cap (T_C < +\infty) \subset (T_C < +\infty) \cap \left(\tau_y \{ \theta_{T_C}(X) \} = +\infty \right),$$

en conditionnant par rapport à \mathcal{A}_{T_C} , et en appliquant la propriété de Markov forte, on a :

$$P_x(T_C < +\infty) \leq E_x \left[\mathbf{1}_{(T_y^1 < +\infty)} \right] + E_x \left[\mathbf{1}_{(T_C < +\infty)} E_{X_{T_C}}(\mathbf{1}_{(T_y^1 = +\infty)}) \right];$$

puisque C est récurrente et que $y \in C$, on a $E_{X_{T_C}}(\mathbf{1}_{(T_y^1 = +\infty)}) = 0$, ce qui donne ainsi l'autre inégalité

$$P_x(T_C < +\infty) \leq P_x(T_y^1 < +\infty). \quad \square$$

On note D l'ensemble des points transitoires et $(C_j)_{j \in J}$ la famille des classes de communication récurrentes. On va donner un **procédé de calcul** de $P_x(T_{C_j} < +\infty)$, pour tout $x \in D$ et tout $j \in J$. Ce calcul est important,

puisque, la chaîne partant d'un point $x \in D$ est « condamnée » à aller passer la fin de sa vie dans une unique classe C_j .

Pour cela, on définit le processus $Y = (Y_n)_{n \in \mathbb{N}}$ (sur les mêmes bases de processus que X), à valeurs dans¹¹ $D \cup J$ par

$$Y_n = \begin{cases} X_n & \text{si } X_n \in D, \\ j & \text{si } X_n \in C_j, \quad j \in J. \end{cases}$$

Proposition 16.50. *Le processus Y est une chaîne de Markov homogène de matrice de transition \widehat{M} donnée par*

$$\begin{cases} \widehat{M}(x, y) = M(x, y) & \text{si } x, y \in D, \\ \widehat{M}(x, j) = \sum_{z \in C_j} M(x, z) \equiv b_j(x) & \text{si } x \in D, \quad j \in J, \\ \widehat{M}(i, j) = \delta_{i,j} & \text{si } i, j \in J, \\ \widehat{M}(i, x) = 0 & \text{si } i \in J, \quad x \in D. \end{cases}$$

On utilise les mêmes notations qu'à la section précédente, à savoir Q est la restriction à $D \times D$ de la matrice M et $S = \sum_{n=0}^{+\infty} Q^n$. On définit la matrice B indexée sur $D \times J$ par

$$\forall (x, j) \in D \times J \quad B(x, j) = b_j(x);$$

alors, pour tout $(x, j) \in D \times J$,

$$\boxed{P_x(T_{C_j} < +\infty) = (SB)(x, j).} \quad (16.64)$$

On rappelle, qu'en particulier, si D est fini, $S = (I - Q)^{-1}$.

Démonstration. Si g est l'application de E dans $D \cup J$ définie par

$$g(x) = \begin{cases} x & \text{si } x \in D, \\ j & \text{si } x \in C_j, \quad j \in J, \end{cases}$$

on a $Y_n = g(X_n)$. Pour toute fonction f définie sur $D \cup J$ et bornée, on a

$$E_x^{A_n} [f(Y_{n+1})] = E_x^{A_n} [f \circ g(X_{n+1})] = M(X_n, f \circ g). \quad (16.65)$$

Puisque

$$f \circ g = f \mathbf{1}_D + \sum_{j \in J} f(j) \mathbf{1}_{C_j},$$

on a, pour tout $x \in E$, par linéarité de $M(x, \cdot)$,

11. Par abus de notation, on note $D \cup J$ l'ensemble de tous les éléments de D auquel on adjoint les éléments de J .

$$\begin{aligned}
M(x, f \circ g) &= M(x, f \mathbf{1}_D) + \sum_{j \in J} f(j) M(x, \mathbf{1}_{C_j}) \\
&= \left[\mathbf{1}_D(x) M(x, f \mathbf{1}_D) + \sum_{j \in J} \mathbf{1}_{C_j}(x) M(x, f \mathbf{1}_D) \right] \\
&+ \left[\sum_{j \in J} f(j) \mathbf{1}_D(x) M(x, \mathbf{1}_{C_j}) + \sum_{i \in J} \mathbf{1}_{C_i}(x) \left[\sum_{j \in J} f(j) M(x, \mathbf{1}_{C_j}) \right] \right];
\end{aligned}$$

en remarquant que, si $x \in D$, $M(x, \mathbf{1}_{C_j}) = b_j(x) = \widehat{M}(x, j)$ et que, pour tout $x \in E$, $\mathbf{1}_{C_j}(x) M(x, f \mathbf{1}_D) = 0$, et $\mathbf{1}_{C_i}(x) M(x, \mathbf{1}_{C_j}) = \delta_{ij} \mathbf{1}_{C_i}(x) = \widehat{M}(i, j) \mathbf{1}_{C_i}(x)$, il vient :

$$\begin{aligned}
M(x, f \circ g) &= \mathbf{1}_D(x) \left[\widehat{M}(x, f \mathbf{1}_D) + \sum_{j \in J} f(j) \widehat{M}(x, j) \right] \\
&+ \sum_{i \in J} \mathbf{1}_{C_i}(x) \left[\sum_{j \in J} f(j) \widehat{M}(i, j) \right] \\
&= \mathbf{1}_D(x) \widehat{M}(x, f) + \sum_{i \in J} \mathbf{1}_{C_i}(x) \widehat{M}(i, f).
\end{aligned}$$

On a alors

$$M(X_n, f \circ g) = \mathbf{1}_D(Y_n) \widehat{M}(Y_n, f) + \sum_{i \in J} \mathbf{1}_i(Y_n) \widehat{M}(i, f) = \widehat{M}(Y_n, f),$$

ce qui, en reportant dans l'égalité (16.65), donne

$$E_x^{\mathcal{A}^n} [f(Y_{n+1})] = \widehat{M}(Y_n, f),$$

et démontre que Y est une chaîne de Markov homogène de matrice de transition \widehat{M} .

Puisque

$$T_{C_j} = \inf(n \in \mathbb{N}^* \mid X_n \in C_j) = \inf(n \in \mathbb{N}^* \mid Y_n = j) \equiv \widehat{T}_j,$$

on a, pour tout $x \in D$,

$$P_x(T_{C_j} < +\infty) = P_x(\widehat{T}_j < +\infty) = P_x \left[\bigcup_{n \in \mathbb{N}^*} (Y_n = j) \right];$$

mais, la classe C_j étant récurrente, on a, P_x -p.s.,

$$(Y_n = j) \subset (Y_{n+1} = j)$$

et donc

$$P_x(T_{C_j} < +\infty) = \lim_n \nearrow P_x(Y_n = j) = \lim_n \widehat{M}^n(x, j).$$

Or, la matrice \widehat{M} ayant la structure de blocs

$$\widehat{M} = \begin{pmatrix} Q & B \\ 0 & I \end{pmatrix},$$

on a, pour tout $n \in \mathbb{N}^*$,

$$\widehat{M}^n = \begin{pmatrix} Q^n & B_n \\ 0 & I \end{pmatrix},$$

avec

$$B_n = (I + Q + Q^2 + \cdots + Q^n)B.$$

Il en résulte que

$$\lim_n B_n = \left(\sum_{n=0}^{+\infty} Q^n \right) B = SB,$$

ce qui donne le résultat annoncé. \square

16.7. Mesures invariantes

La notion de mesure invariante pour une chaîne de Markov homogène de matrice de transition M est essentiellement liée à son comportement asymptotique.

Par la suite, E étant dénombrable, on identifie une mesure ν sur E et son germe. De plus, tenant compte de la dualité entre fonctions et mesures, et conservant le point de vue vectoriel précédemment adopté, on identifie la mesure ν au « vecteur ligne » $(\nu(x))_{x \in E}$.

Définition 16.51. Soit M une matrice de transition sur E . À toute mesure ν sur E , on associe la mesure νM définie par, pour tout $y \in E$,

$$\nu M(y) = \sum_{x \in E} \nu(x) M(x, y). \quad (16.66)$$

On dit alors que ν est une **mesure invariante** (relativement à la matrice de transition M) si $\nu M = \nu$, autrement dit, avec les identifications ci-dessus, si ν est vecteur propre à gauche de M associé à la valeur propre 1. Une mesure invariante qui est une probabilité est dite **probabilité invariante**.

Soient X une chaîne de Markov homogène de matrice de transition M et ν une mesure invariante (relativement à M); ν est appelée **mesure invariante de la chaîne**.

Remarque. Si ν est une mesure invariante, pour tout $a \geq 0$, il en est de même de la mesure $a\nu$. De plus, si ν_1 et ν_2 sont des probabilités invariantes, toute combinaison convexe de ν_1 et ν_2 est encore une probabilité invariante; ainsi,

l'existence de deux probabilités invariantes distinctes implique l'existence d'une infinité de probabilités invariantes.

Proposition 16.52. *Soit X une chaîne de Markov homogène de matrice de transition M admettant une probabilité invariante ν . Si X est de loi initiale ν , pour tout $n \in \mathbb{N}^*$, X_n est de loi ν .*

Démonstration. Pour tout $y \in E$, on a, en notant $E' = \{x \in E \mid \nu(x) \neq 0\}$,

$$P(X_n = y) = \sum_{x \in E'} P(X_0 = x) P^{(X_0=x)}(X_n = y) = \sum_{x \in E} \nu(x) M^n(x, y);$$

or

$$\begin{aligned} \sum_{x \in E} \nu(x) M^n(x, y) &= \sum_{x \in E} \nu(x) \left[\sum_{z \in E} M(x, z) M^{n-1}(z, y) \right] \\ &= \sum_{z \in E} \left[\sum_{x \in E} \nu(x) M(x, z) \right] M^{n-1}(z, y), \end{aligned}$$

soit, puisque ν est invariante,

$$\sum_{x \in E} \nu(x) M^n(x, y) = \sum_{z \in E} \nu(z) M^{n-1}(z, y);$$

autrement dit, pour tout $n \in \mathbb{N}^*$, on a $\nu M^n = \nu M^{n-1}$, et donc $\nu M^n = \nu$. Il en résulte que, pour tout $y \in E$, on a $P(X_n = y) = \nu(y)$. \square

On étudie le problème d'existence et d'unicité d'une probabilité invariante, puis son lien avec l'existence d'une probabilité limite. On en déduit un critère de récurrence positive pour une chaîne de Markov homogène en terme de probabilité invariante.

Proposition 16.53. *Soit M une matrice de transition telle que, pour tous $x, y \in E$, la suite de terme général $M^n(x, y)$ soit convergente et de limite $\pi(y)$ indépendante de x . Alors*

(a) *la mesure π est invariante et de masse inférieure ou égale à 1, c'est-à-dire que l'on a*

$$\pi M = \pi \quad \text{et} \quad \sum_{y \in E} \pi(y) \leq 1;$$

(b) *soit $\pi = 0$ (c'est-à-dire que, pour tout $y \in E$, on a $\pi(y) = 0$), soit π est une probabilité invariante.*

(c) *Si $\pi = 0$, il n'existe pas de probabilité invariante pour M . Si π est une probabilité invariante, π est l'unique probabilité invariante pour M .*

Démonstration. (a) D'après le lemme de Fatou, on a

$$\sum_{y \in E} \pi(y) = \sum_{y \in E} \lim_n M^n(x, y) \leq \liminf_n \sum_{y \in E} M^n(x, y) = 1.$$

De plus, pour tout $y \in E$, on a, pour tout $x \in E$, $\pi(x) = \lim_n M^n(y, x)$, et donc

$$\begin{aligned} \pi M(y) &= \sum_{x \in E} \left[\lim_n M^n(y, x) \right] M(x, y) \leq \lim_n \inf \sum_{x \in E} M^n(y, x) M(x, y) \\ &= \lim_n \inf M^{n+1}(y, y) = \pi(y) . \end{aligned}$$

Autrement dit, $\pi M \leq \pi$. Supposons qu'il n'y ait pas égalité; il existerait alors y_0 tel que $\pi M(y_0) < \pi(y_0)$ et on aurait (puisque $\pi M \leq \pi$)

$$\sum_{y \in E} \pi(y) > \sum_{y \in E} \left[\sum_{x \in E} \pi(x) M(x, y) \right] = \sum_{x \in E} \pi(x) \left[\sum_{y \in E} M(x, y) \right] = \sum_{x \in E} \pi(x),$$

ce qui est absurde. Par conséquent, on a $\pi M = \pi$; ainsi, π est une mesure invariante de masse inférieure ou égale à 1.

(b) Remarquons que si ν est une mesure invariante de masse inférieure ou égale à 1, on a, pour tout $n \in \mathbb{N}^*$, $\nu M^n = \nu$. Pour tout $y \in E$, on a alors

$$\nu(y) = \lim_n \nu M^n(y) = \lim_n \sum_{x \in E} \nu(x) M^n(x, y),$$

soit, par le théorème de convergence dominée et par définition de π ,

$$\nu(y) = \sum_{x \in E} \nu(x) \left[\lim_n M^n(x, y) \right] = \sum_{x \in E} \nu(x) [\pi(y)] \quad (16.67)$$

et donc

$$\nu(y) = \pi(y) \left[\sum_{x \in E} \nu(x) \right]. \quad (16.68)$$

En particulier, si $\nu = \pi$, on a

$$\pi(y) = \pi(y) \left[\sum_{x \in E} \pi(x) \right],$$

ce qui implique le résultat annoncé.

(c) Si $\pi = 0$, il résulte de l'égalité (16.68) que $\nu = 0$ et donc que, dans ce cas, il n'existe pas de probabilité invariante pour M . Toujours d'après cette même égalité, si ν est une probabilité invariante, elle coïncide avec π . \square

Le théorème suivant permet, dans le cas où E est infini, de savoir si une classe de communication est récurrente positive et d'évaluer le temps moyen de retour en un point.

Théorème 16.54 (Critère de récurrence positive). *Une chaîne de Markov homogène de matrice de transition M possède une probabilité invariante π unique si et seulement si elle a exactement une classe de communication récurrente positive C . Dans ce cas, on a*

$$\pi(x) = \begin{cases} \frac{1}{E_x(T_x^1)} & \text{si } x \in C, \\ 0 & \text{sinon.} \end{cases} \quad (16.69)$$

Démonstration. 1. Si X n'admet pas de classe récurrente positive, X n'a que des états transitoires ou récurrents nuls, et, dans ces deux cas, il résulte des propositions 16.40, 16.38 et 16.41 que $\lim_n M^n(x, y) = 0$, pour tous $x, y \in E$. La proposition 16.53 montre alors qu'il n'existe pas de probabilité invariante.

2. Si X admet une seule classe récurrente positive C , deux cas se présentent.

• Si C est apériodique, alors (proposition 16.40), pour tous $x, y \in C$, on a

$$\lim_n M^n(x, y) = \frac{1}{E_y(T_y^1)} = \pi(y) > 0,$$

limite indépendante de $x \in C$; la restriction de M à $C \times C$ étant dans ce cas une matrice de transition, il résulte de la proposition 16.53 appliquée à cette matrice que $\pi|_C$ est l'unique probabilité invariante relativement à la matrice $M|_{C \times C}$. Il en résulte en particulier que la mesure π est une probabilité sur E , invariante pour M . En effet,

– pour tout $y \in C$, on a alors, puisque, pour tout $x \notin C$, $\pi(x) = 0$,

$$\begin{aligned} \pi(y) &= \pi|_C(y) = \sum_{x \in C} \pi|_C(x) M|_{C \times C}(x, y) \\ &= \sum_{x \in C} \pi(x) M(x, y) = \sum_{x \in E} \pi(x) M(x, y) = \pi M(y); \end{aligned}$$

– pour tout $y \notin C$, remarquant que C est une classe fermée, on a $M(x, y) = 0$ pour tout $x \in C$, et donc

$$\sum_{x \in E} \pi(x) M(x, y) = \sum_{x \in C^c} \pi(x) M(x, y) = 0 = \pi(y).$$

L'existence d'une probabilité invariante pour M est ainsi démontrée.

Démontrons l'unicité. Si ν est une probabilité invariante sur E , il résulte de (16.67) que, pour tout $y \in C$, on a

$$\nu(y) = \pi(y) \sum_{x \in C} \nu(x) + r(y),$$

où

$$r(y) = \sum_{x \in C^c} \nu(x) \left[\lim_n M^n(x, y) \right];$$

autrement dit, pour tout $y \in C$, on a

$$\nu(y) = \pi(y) \nu(C) + r(y).$$

Il en résulte, en sommant sur les y de C , que

$$v(C) = \pi(C)v(C) + r(C);$$

puisque π est une probabilité sur C , on a alors $r(C) = 0$, et donc, pour tout $y \in C$, on a

$$v(y) = \pi(y)v(C). \quad (16.70)$$

Par ailleurs, si $y \notin C$, par hypothèse, y est transitoire ou récurrent nul; de toute façon, pour tout $x \in E$, on a $\lim_n M^n(x, y) = 0$, et, d'après (16.67), $v(y) = \pi(y) = 0$.

Ainsi, $v(C) = 1$, et il résulte de (16.70) que pour tout $y \in C$, on a $v(y) = \pi(y)$. La probabilité π est donc l'unique probabilité invariante.

• Si C est périodique de période d , on note C_k , $k = 0, 1, \dots, d-1$, les classes cycliques de C , indexées comme à la proposition 16.41. On rappelle que, pour tous $k = 0, 1, \dots, d-1$ et tous $x, y \in C_k$, on a

$$\lim_n M^{nd}(x, y) = \frac{d}{m(y)},$$

où $m(y) = E_y(T_y^1)$. Démontrons que la mesure π définie par (16.69), soit encore

$$\pi(x) = \begin{cases} \frac{1}{m(x)} & \text{si } x \in C, \\ 0 & \text{sinon,} \end{cases}$$

est une probabilité invariante.

– Si $x \in C$, on a

$$M^{nd}(x, x) = \sum_{y \in C} M^{nd-1}(x, y)M(y, x),$$

et, d'après le lemme de Fatou,

$$\begin{aligned} \frac{d}{m(x)} &= \lim_n M^{nd}(x, x) \geq \sum_{y \in C} \lim_n \inf M^{nd-1}(x, y)M(y, x) \\ &= \sum_{k=0}^{d-1} \left[\sum_{y \in C_k} \lim_n \inf M^{nd-1}(x, y)M(y, x) \right], \end{aligned}$$

soit, si $x \in C_{k_0}$, et si $k_1 = k_0 - 1 \pmod{d}$ avec $0 \leq k_1 \leq d-1$,

$$\frac{d}{m(x)} \geq \sum_{y \in C_{k_1}} \lim_n \inf M^{nd-1}(x, y)M(y, x) = \sum_{y \in C_{k_1}} \frac{d}{m(y)} M(y, x),$$

ce qui donne, puisque, $M(y, x) = 0$ si $y \notin C_{k_1}$,

$$\frac{d}{m(x)} \geq \sum_{k=0}^{d-1} \left[\sum_{y \in C_k} \frac{d}{m(y)} M(y, x) \right] = \sum_{y \in C} \frac{d}{m(y)} M(y, x);$$

ainsi, on a l'inégalité

$$\frac{1}{m(x)} \geq \sum_{y \in C} \frac{1}{m(y)} M(y, x). \quad (16.71)$$

Démontrons, qu'en fait, pour tout $x \in C$, il y a égalité dans l'inégalité (16.71). Supposons qu'il n'y ait pas égalité ; il existerait alors x_0 tel que

$$\frac{1}{m(x_0)} > \sum_{y \in C} \frac{1}{m(y)} M(y, x_0).$$

et on aurait

$$\begin{aligned} \sum_{x \in C} \frac{1}{m(x)} &> \sum_{x \in C} \left[\sum_{y \in C} \frac{1}{m(y)} M(y, x) \right] = \sum_{y \in C} \left[\frac{1}{m(y)} \sum_{x \in C} M(y, x) \right] \\ &= \sum_{y \in C} \frac{1}{m(y)}, \end{aligned}$$

ce qui est absurde.

Puisque π est à support dans C , on a alors

$$\frac{1}{m(x)} = \sum_{y \in C} \frac{1}{m(y)} M(y, x) = \sum_{y \in E} \pi(y) M(y, x) = \pi M(x).$$

– Si $x \notin C$. Pour tout $y \notin C$, on a $\pi(y) = 0$, et, remarquant que C est une classe fermée, pour tout $y \in C$, $M(y, x) = 0$; ainsi $\pi(y) M(y, x) = 0$ pour tout $y \in E$, ce qui conduit à l'égalité

$$\pi M(x) = \pi(x) = 0$$

et achève de démontrer que π est une mesure invariante.

Reste à démontrer que c'est une probabilité ; or, la chaîne restreinte à C_k de matrice de transition $M_{|_{C_k \times C_k}}^d$ étant apériodique, le point précédemment démontré assure que $d\pi_{|_{C_k}}$ est l'unique probabilité invariante pour cette chaîne. Il en résulte que $\pi(C) = 1$.

• Démontrons l'**unicité**. Si ν est une probabilité invariante sur E , on a, pour tout $n \in \mathbb{N}^*$ et tout $x \in C$,

$$\nu(x) = \sum_{k=0}^{d-1} \sum_{y \in C_k} \nu(y) M^{nd}(y, x) + r_n(x),$$

où

$$r_n(x) = \sum_{y \notin C} \nu(y) M^{nd}(y, x).$$

Si $x \in C_{k_0}$, on a

$$\lim_n M^{nd}(y, x) = \begin{cases} \frac{d}{m(x)} & \text{si } y \in C_{k_0} \\ 0 & \text{sinon,} \end{cases}$$

si bien que, par le théorème de convergence dominée appliqué à chaque somme sur C_k , on obtient la convergence de la suite de terme général $r_n(x)$ et l'égalité

$$v(x) = \sum_{y \in C_{k_0}} \left[v(y) \frac{d}{m(x)} \right] + \lim_n r_n(x),$$

soit

$$v(x) = d \pi_{|C_{k_0}}(x) v(C_{k_0}) + \lim_n r_n(x). \quad (16.72)$$

En sommant sur les x de C_{k_0} , on obtient l'égalité

$$v(C_{k_0}) = d \pi_{|C_{k_0}}(C_{k_0}) v(C_{k_0}) + \sum_{x \in C_{k_0}} \lim_n r_n(x);$$

puisque $d \pi_{|C_{k_0}}(C_{k_0}) = 1$, on a ainsi $\sum_{x \in C_{k_0}} \lim_n r_n(x) = 0$ et donc, pour tout $x \in C_{k_0}$, $\lim_n r_n(x) = 0$. Il résulte alors de (16.72) que, pour tout k_0 tel que $0 \leq k_0 \leq d-1$ et tout $x \in C_{k_0}$, on a

$$v(x) = d \pi_{|C_{k_0}}(x) v(C_{k_0}). \quad (16.73)$$

Par ailleurs, si $y \notin C$, y n'est pas récurrent positif et, toujours par le même raisonnement, on a

$$v(y) = \sum_{x \in E} v(x) \left[\lim_n M^n(x, y) \right] = 0.$$

ce qui montre que v est portée par C .

Enfin, v étant invariante, si $x \in C_{k_0}$, et si $k_1 = k_0 - 1 \pmod{d}$ avec $0 \leq k_1 \leq d-1$, il résulte de (16.73) que l'on a

$$v(x) = d \pi_{|C_{k_0}}(x) v(C_{k_0}) = \sum_{k=0}^{d-1} \sum_{y \in C_k} v(y) M(y, x) = \sum_{y \in C_{k_1}} v(y) M(y, x)$$

soit, toujours d'après (16.73),

$$v(x) = \sum_{y \in C_{k_1}} \left[d \pi_{|C_{k_1}}(y) v(C_{k_1}) \right] M(y, x) = d v(C_{k_1}) \sum_{y \in C_{k_1}} \frac{1}{m(y)} M(y, x).$$

Puisque $x \in C_{k_0}$ et que π est à support dans C , cela s'écrit encore

$$v(x) = d v(C_{k_1}) \sum_{y \in C} \pi(y) M(y, x) = d v(C_{k_1}) \pi M(x)$$

soit, compte tenu de l'invariance de π ,

$$v(x) = d v(C_{k_1}) \pi(x).$$

Il en résulte, en sommant sur les x de C_{k_0} , que

$$v(C_{k_0}) = dv(C_{k_1})\pi(C_{k_0});$$

puisque $d\pi|_{C_{k_0}}$ est une probabilité, on a ainsi montré que $v(C_{k_0}) = v(C_{k_1})$, pour tout k_0 tel que $0 \leq k_0 \leq d-1$. Il en résulte que $v = \pi$.

3. Si X admet N ($N \geq 2$) classes récurrentes positives C^1, C^2, \dots, C^N , définissons, pour tous réels positifs a_j , tels que $\sum_{j=1}^N a_j = 1$, la mesure μ par

$$\mu(x) = \begin{cases} \frac{a_j}{m(x)} & \text{si } x \in C^j \quad j = 1, \dots, N \\ 0 & \text{si } x \notin \bigcup_{j=1}^N C^j. \end{cases}$$

Pour $j_0 = 1, \dots, N$ quelconque et tout $x \in C^{j_0}$, les classes C^1, C^2, \dots, C^N étant fermées, on a

$$\begin{aligned} \mu M(x) &= \sum_{y \in E} \mu(y)M(y, x) = \sum_{j=1}^N a_j \left[\sum_{y \in C^j} \frac{1}{m(y)} M(y, x) \right] \\ &= a_{j_0} \sum_{y \in C^{j_0}} \frac{1}{m(y)} M(y, x); \end{aligned} \quad (16.74)$$

la chaîne restreinte à la classe fermée C^{j_0} étant, par essence, irréductible, le résultat démontré au point 2 assure que la mesure définie sur C^{j_0} par $\mu_{j_0}(y) = \frac{1}{m(y)}$ pour tout $y \in C^{j_0}$ est l'unique probabilité invariante de cette chaîne. Reportant dans (16.74), on en déduit l'égalité

$$\mu M(x) = a_{j_0} \mu_{j_0}(x) = a_{j_0} \frac{1}{m(x)},$$

ce qui démontre que, pour tout $x \in C^{j_0}$, $\mu M(x) = \mu(x)$.

De plus, si $x \notin \bigcup_{j=1}^N C^j$, on a $M(y, x) = 0$ pour tout $j = 1, \dots, N$ et tout $y \in C^j$, ce qui implique que $\mu M(x) = 0$; comme on a aussi $\mu(x) = 0$, on a encore $\mu M(x) = \mu(x)$ et μ est une mesure invariante. C'est en fait une probabilité; en effet, μ_j étant pour tout j une probabilité, on a $\sum_{x \in C^j} \frac{1}{m(x)} = 1$, et donc

$$\mu(E) = \sum_{j=1}^N a_j \left[\sum_{x \in C^j} \frac{1}{m(x)} \right] = \sum_{j=1}^N a_j = 1.$$

On vient de démontrer que, dans ce cas, il existe une infinité non dénombrable de probabilités invariantes. \square

On reprend le modèle de diffusion d'**Ehrenfest** et on détermine la **probabilité invariante** de la chaîne associée.

Exemple 16.12. (Modèle de diffusion de chaleur de Ehrenfest; suite.) On considère à nouveau le modèle de Ehrenfest, décrit sous sa forme de tirages de boules dans une urne (voir les exemples 16.2 et 16.10, dont on reprend les notations). On rappelle que X_n représente le nombre de boules rouges contenues dans l'urne à l'instant n , et que le processus $X = (X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov homogène, à valeurs dans l'intervalle d'entiers $E = \{0, 1, \dots, m\}$, dont la matrice de transition M est donnée, pour tout $k \in \{0, 1, \dots, m\}$, par

$$M(k, k+1) = p_k, \quad M(k, k-1) = q_k,$$

où

$$p_k = 1 - \frac{k}{m}, \quad q_k = \frac{k}{m}.$$

Cette chaîne est **irréductible, apériodique, récurrente positive** (ce dernier point est une conséquence du corollaire 16.44). Il résulte donc du théorème 16.54 que X admet une **probabilité invariante unique**. On se propose de la déterminer.

On cherche d'abord une **mesure** invariante μ ; une telle mesure est solution du système d'équations

$$\mu(y) = \sum_{x \in E} \mu(x) M(x, y),$$

système qui s'écrit, en notant μ_k pour $\mu(k)$,

$$\begin{cases} \mu_k = \mu_{k-1} p_{k-1} + \mu_{k+1} q_{k+1} & \text{si } 1 \leq k \leq m-1, \\ \mu_0 = \mu_1 q_1, \\ \mu_m = \mu_{m-1} p_{m-1}. \end{cases} \quad (16.75)$$

Puisque $p_{k+1} + q_{k+1} = 1$, on a alors, si $1 \leq k \leq m-1$,

$$\mu_{k+1} - \mu_k = \mu_{k+1} p_{k+1} - \mu_{k-1} p_{k-1}, \quad (16.76)$$

et donc

$$\sum_{j=k}^{m-1} (\mu_{j+1} - \mu_j) = \sum_{j=k}^{m-1} (\mu_{j+1} p_{j+1} - \mu_{j-1} p_{j-1}),$$

ou encore, en faisant des changements d'indices,

$$\mu_m - \mu_k = \sum_{j=k+1}^m (\mu_j p_j) - \sum_{j=k-1}^{m-2} (\mu_j p_j),$$

soit

$$\mu_m - \mu_k = \mu_m p_m + \mu_{m-1} p_{m-1} - \mu_{k-1} p_{k-1} - \mu_k p_k. \quad (16.77)$$

Ainsi, puisque

$$p_k + q_k = 1, \quad p_m = 0, \quad \text{et} \quad p_{m-1} = \frac{1}{m},$$

on a

$$\mu_m - \frac{1}{m} \mu_{m-1} = \mu_k q_k - \mu_{k-1} p_{k-1}.$$

En remarquant que, d'après (16.75),

$$\mu_m - \frac{1}{m} \mu_{m-1} = 0,$$

on a, dès que $1 \leq k \leq m-1$,

$$\mu_k = \frac{p_{k-1}}{q_k} \mu_{k-1}, \quad (16.78)$$

soit, par itération,

$$\mu_k = \frac{p_{k-1} p_{k-2} \cdots p_1 p_0}{q_k q_{k-1} \cdots q_2 q_1} \mu_0. \quad (16.79)$$

Puisque

$$\begin{aligned} \frac{p_{k-1} p_{k-2} \cdots p_1 p_0}{q_k q_{k-1} \cdots q_2 q_1} &= \frac{\frac{m-k+1}{m} \cdot \frac{m-k+2}{m} \cdots \frac{m-1}{m} \cdot \frac{m}{m}}{\frac{k}{m} \cdot \frac{k-1}{m} \cdots \frac{2}{m} \cdot \frac{1}{m}} \\ &= \frac{m(m-1) \cdots (m-k+1)}{k!} = \binom{m}{k}, \end{aligned}$$

on a

$$\mu_k = \binom{m}{k} \mu_0 \quad \text{si } 1 \leq k \leq m-1. \quad (16.80)$$

Par ailleurs, d'après (16.75) et (16.80), on a

$$\mu_m = \frac{1}{m} \mu_{m-1} = \frac{1}{m} m \mu_0 = \mu_0,$$

soit encore

$$\mu_m = \binom{m}{m} \mu_0.$$

Ainsi, toute **mesure** invariante μ est déterminée par

$$\mu_k = \binom{m}{k} \mu_0 \quad \text{si } 1 \leq k \leq m. \quad (16.81)$$

Il existe donc une unique **probabilité** invariante μ déterminée par les égalités équivalentes

$$\mu_0 + \sum_{k=1}^m \binom{m}{k} \mu_0 = 1 \iff \left[\sum_{k=0}^m \binom{m}{k} \right] \mu_0 = 1,$$

ce qui donne

$$\mu_0 = \frac{1}{2^m}.$$

La **probabilité invariante** μ est donc donnée par

$$\mu_k = \frac{\binom{m}{k}}{2^m} \quad \text{si } 0 \leq k \leq m;$$

autrement dit, la **probabilité invariante** μ est la loi binomiale $\mathcal{B}(m, \frac{1}{2})$.

Ainsi, la chaîne d'Ehrenfest est **irréductible, récurrente positive apériodique**, et admet, d'après le théorème 16.55 ci-dessous, **une probabilité limite** μ qui est la **loi binomiale** $\mathcal{B}(m, 1/2)$. Autrement dit, le **régime stationnaire** est comme si on avait autant de boules de chaque couleur, et qu'on les tirait au hasard.

De plus, d'après le théorème 16.54, le **temps moyen de retour** en k , partant de k , est $2^m / \binom{m}{k}$.

On donne maintenant une condition nécessaire et suffisante pour l'existence d'une **probabilité limite**, c'est-à-dire d'une probabilité μ qui soit telle que, pour tous $x, y \in E$, la suite de terme général $M^n(x, y)$ soit convergente de limite $\mu(y)$, indépendante de x . Pour une telle probabilité μ , on a alors, pour tout $y \in E$,

$$\lim_n P_x(X_n = y) = \mu(y),$$

et ceci, **indépendamment** du point de départ x de la chaîne à l'instant initial.

Théorème 16.55. *Une chaîne de Markov homogène de matrice de transition M possède une **probabilité limite** si et seulement si elle admet une unique classe récurrente positive apériodique C telle que $P_x(T_y^1 < +\infty) = 1$, pour tous $x \in E$ et $y \in C$.*

Démonstration. S'il existe une probabilité limite μ , il résulte de la proposition 16.53 que c'est l'unique probabilité invariante; le théorème 16.54 assure alors l'existence d'une *unique classe récurrente positive* C . Elle est *apériodique*. Supposons en effet qu'elle soit *périodique* et notons C_k , $k = 0, 1, \dots, d-1$, les classes cycliques de C , indexées comme à la proposition 16.41. Dans ce cas, pour tous $x \in C_0$ et $y \in C_1$, on aurait

$$\lim_n M^{nd+1}(x, y) = \frac{d}{E_y(T_y^1)} > 0 \quad \text{et, pour tout } n \in \mathbb{N}^*, M^{nd}(x, y) = 0,$$

ce qui est en contradiction avec l'existence d'une probabilité limite.

Enfin, pour tous $x \in E$ et $y \in C$, on a

$$\lim_n M^n(x, y) = \frac{P_v(T_y^1 < +\infty)}{E_y(T_y^1)} = \mu(y) > 0,$$

ce qui démontre que l'application $x \mapsto P_x(T_y^1 < +\infty)$ est constante; mais, y étant récurrent, on a $P_y(T_y^1 < +\infty) = 1$. Il en résulte que $P_x(T_y^1 < +\infty) = 1$, pour tous $x \in E$ et $y \in C$.

Inversement, supposons qu'il existe une unique classe récurrente positive apériodique C telle que $P_x(T_y^1 < +\infty) = 1$, pour tous $x \in E$ et $y \in C$. Pour de tels points, on a alors

$$\lim_n M^n(x, y) = \frac{1}{E_y(T_y^1)} > 0.$$

Par ailleurs, si $x \in E$ et $y \notin C$, y est récurrent nul ou transitoire et on a $\lim_n M^n(x, y) = 0$. En définitive, pour tous $x, y \in E$, la suite de terme général $M^n(x, y)$ est convergente de limite $\pi(y)$ indépendante de x . Puisque la mesure π ainsi définie est non nulle, il résulte de la proposition 16.53 que c'est l'unique probabilité invariante. \square

Nous terminons ce paragraphe en faisant l'étude d'un modèle génétique.

Exemple 16.13. (Modèle génétique.) Un caractère héréditaire chez un individu dépend, en général, de la possession dans son patrimoine génétique de la présence de gènes de deux types G et g , appelés **allèles**¹², qui apparaissent par couple GG , gg , Gg et gG , ces deux derniers étant génétiquement les mêmes; ainsi, pour la mise en évidence du caractère, seuls comptent les couples non ordonnés GG , gg , Gg , appelés **génotypes**. Le gène G est souvent **prédominant**, et les génotypes Gg et GG donnent le même caractère héréditaire, appelé **phénotype**¹³. Suivant qu'un individu possède le génotype GG , gg , ou Gg , il est dit **dominant**, **récessif**, ou **hybride**.

Un individu reçoit indépendamment de chaque parent un gène de manière aléatoire.

- Si chacun des parents est dominant (respectivement, récessif), il est lui-même dominant (respectivement, récessif).
- Si l'un des parents est dominant et l'autre récessif, il est lui-même hybride.

12. Du grec, allélôn, qui signifie *les uns les autres*. Un allèle, ou allélomorphe, désigne un caractère héréditaire qui s'oppose à un autre (graines de pois lisses ou ridées) ou le gène porteur de ce caractère.

13. **Phénotype** : aspect extérieur de l'être vivant, par opposition au génotype (patrimoine héréditaire). Plusieurs génotypes différents peuvent conférer le même aspect à un individu (mais les différences reparaîtront aux générations suivantes).

- Si l'un des parents est dominant et l'autre hybride, il reçoit le gène G du parent dominant et reçoit le gène G ou g de l'autre parent avec la même probabilité; ainsi, il a même probabilité d'être dominant ou hybride.

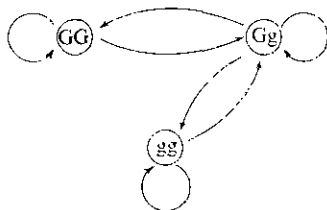
De même, si l'un des parents est récessif et l'autre hybride, il a même probabilité d'être récessif ou hybride.

- Si les deux parents sont hybrides, il a la même probabilité de recevoir de chaque parent le gène G ou g . Il sera donc dominant avec la probabilité $\frac{1}{4}$, récessif avec la probabilité $\frac{1}{4}$ et hybride avec la probabilité $\frac{1}{2}$.

Considérons le processus suivant : un individu de caractère donné se marie avec un hybride et donne naissance à des enfants. On choisit au hasard un des enfants et le marie à nouveau avec un hybride, et ainsi de suite. Désignons par X_n le type génétique du n -ième descendant ainsi observé. Le processus $(X_n)_{n \in \mathbb{N}^*}$ est une chaîne de Markov de matrice de transition

$$M_h = \begin{array}{c} \begin{array}{ccc} & \mathbf{GG} & \mathbf{Gg} & \mathbf{gg} \\ \mathbf{GG} & \left(\begin{array}{ccc} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{array} \right) \\ \mathbf{Gg} & & & \\ \mathbf{gg} & & & \end{array} \end{array}$$

Le graphe associé à cette chaîne de Markov est



La chaîne est **irréductible aperiodique et récurrente positive**. Elle admet donc une probabilité invariante unique. Calculons la. Déterminons d'abord les mesures invariantes $v = (a, b, c)$; v est valeur propre à gauche de M_h , associée à la valeur propre 1 et est donc solution du système

$$\begin{cases} \frac{1}{2}a + \frac{1}{4}b = a \\ \frac{1}{2}a + \frac{1}{2}b + \frac{1}{2}c = b \\ \frac{1}{4}b + \frac{1}{2}c = c, \end{cases}$$

qui a pour solution $(a, 2a, a)$. Les mesures invariantes sont alors données par $(a, 2a, a)$, avec $a \geq 0$ quelconque. Il existe donc une seule probabilité invariante v_0 ; elle est déterminée par la condition $a + 2a + a = 1$, ce qui donne $v_0 = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

Si on note $E = \{GG, gg, Gg\}$ l'espace d'états, il en résulte, puisque la chaîne est irréductible apériodique, que, pour tout $x, y \in E$,

$$\lim_n M_n^h(x, y) = \frac{1}{E_y(T_y^1)} = v_0(y).$$

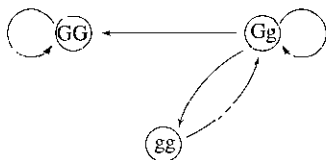
En particulier, on obtient les temps moyens de retour en un point :

$$E_{GG}(T_{GG}^1) = E_{gg}(T_{gg}^1) = 4 \quad \text{et} \quad E_{Gg}(T_{Gg}^1) = 2.$$

Si, au lieu de marier l'un des enfants choisi au hasard avec un hybride, on le marie avec un dominant, le processus $(X_n)_{n \in \mathbb{N}^*}$ est une chaîne de Markov de matrice de transition

$$M = \begin{array}{c} \begin{array}{ccc} & \mathbf{GG} & \mathbf{Gg} & \mathbf{gg} \\ \mathbf{GG} & \left(\begin{array}{ccc} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 2 & 0 \\ 0 & 1 & 0 \end{array} \right) \\ \mathbf{Gg} & & & \\ \mathbf{gg} & & & \end{array} \end{array}$$

Le graphe associé à cette chaîne de Markov est



Chaque état forme une classe de communication. L'état gg est **inessentiel**, les états gg et Gg sont **transitoires** et GG est un état **absorbant**.

16.8. Loi forte des grands nombres

On donne un théorème de loi forte des grands nombres pour une chaîne de Markov homogène que l'on applique ensuite, dans le cas fini, à l'estimation de sa matrice de transition.

16.8.1. Théorème de loi forte

Théorème 16.56 (Théorème de Chacon-Orstein). Soit X un processus qui, pour tout $x \in E$, est une chaîne de Markov homogène sur la base de

processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$ de loi initiale δ_x et de matrice de transition M . On suppose que X admet une unique classe récurrente positive C (on rappelle qu'il existe alors une unique probabilité invariante π). On suppose de plus¹⁴ qu'il existe un $y \in C$ tel que, pour tout $x \in E$, $P_x(T_y^1 < +\infty) = 1$.

Soient f et g des fonctions définies sur E , π -intégrables; on suppose que g ne s'annule pas. Alors, pour tout $x \in E$, la suite de terme général $\frac{\sum_{j=1}^n f(X_j)}{\sum_{j=1}^n g(X_j)}$ est P_x -p.s. convergente et on a

$$\lim_n \frac{\sum_{j=1}^n f(X_j)}{\sum_{j=1}^n g(X_j)} = \frac{\sum_{x \in E} f(x)\pi(x)}{\sum_{x \in E} g(x)\pi(x)} \quad P_x\text{-p.s.}$$

Démonstration. • Pour tout $x \in E$, on a $P_x(R_y) = 1$. En effet, une modification triviale de la démonstration du lemme 16.36 permet d'établir que, pour tout $p \in \mathbb{N}^*$, on a

$$P_x(T_y^{p+1} < +\infty) = P_x(T_y^1 < +\infty) [P_y(T_y^1 < +\infty)]^p;$$

il résulte alors des hypothèses que

$$P_x(R_y) = \lim_n \searrow P_x(T_y^1 < +\infty) [P_y(T_y^1 < +\infty)]^p = 1.$$

• Soit μ la mesure sur E définie par, pour tout $x \in E$,

$$\mu(x) = E_y \left[\sum_{n=1}^{T_y^1} \mathbf{1}_{(X_n=x)} \right],$$

nombre moyen de passages en x , avant le premier retour en y , par la chaîne qui part de y à l'instant initial. C'est une mesure invariante. En effet, on a

$$\begin{aligned} \mu M(x) &= \sum_{z \in E} \mu(z) M(z, x) = \sum_{z \in E} \left[E_y \left(\sum_{n=1}^{T_y^1} \mathbf{1}_{(X_n=z)} \right) \right] M(z, x) \\ &= E_y \left[\sum_{n=1}^{T_y^1} \left(\sum_{z \in E} \mathbf{1}_{(X_n=z)} M(z, x) \right) \right] = E_y \left[\sum_{n=1}^{T_y^1} M(X_n, x) \right], \end{aligned}$$

soit, en partitionnant,

$$\mu M(x) = E_y \left[\mathbf{1}_{(T_y^1=1)} M(X_1, x) \right] + E_y \left[\mathbf{1}_{(T_y^1 \geq 2)} \left(\sum_{n=1}^{T_y^1-1} M(X_n, x) + M(X_{T_y^1}, x) \right) \right];$$

14. Si E est fini, cette hypothèse est automatiquement satisfaite.

en remarquant que $X_{T_y^1} = X_0 = y$ P_y-p.s., on a alors

$$\mu M(x) = E_y \left[\mathbf{1}_{(T_y^1=1)} M(X_0, x) \right] + E_y \left[\mathbf{1}_{(T_y^1 \geq 2)} \sum_{n=0}^{T_y^1-1} M(X_n, x) \right],$$

soit encore

$$\mu M(x) = E_y \left[\sum_{n=0}^{T_y^1-1} M(X_n, x) \right].$$

Cette égalité peut s'écrire, par la propriété de Markov,

$$\mu M(x) = \sum_{n=0}^{+\infty} E_y \left[\mathbf{1}_{(n < T_y^1)} P_{X_n}(X_{n+1}=x) \right] = \sum_{n=0}^{+\infty} E_y \left[\mathbf{1}_{(n < T_y^1)} E_y^{\mathcal{A}_n} \mathbf{1}_{(X_{n+1}=x)} \right],$$

soit, puisque $(n < T_y^1) \in \mathcal{A}_n$,

$$\mu M(x) = \sum_{n=0}^{+\infty} E_y \left[\mathbf{1}_{(n < T_y^1)} \mathbf{1}_{(X_{n+1}=x)} \right] = E_y \left[\sum_{n=1}^{T_y^1} \mathbf{1}_{(X_n=x)} \right] = \mu(x),$$

ce qui démontre que μ est une mesure invariante. C'est une mesure bornée, puisque, E étant dénombrable et y étant récurrent positif, on a

$$\mu(E) = \sum_{x \in E} E_y \left[\sum_{n=1}^{T_y^1} \mathbf{1}_{(X_n=x)} \right] = E_y \left[\sum_{n=1}^{T_y^1} \left(\sum_{x \in E} \mathbf{1}_{(X_n=x)} \right) \right] = E_y [T_y^1] < +\infty.$$

Il en résulte que $\frac{\mu}{E_y[T_y^1]}$ est une probabilité invariante; par unicité de la probabilité invariante, on a donc

$$\pi = \frac{\mu}{E_y [T_y^1]},$$

ce qui donne, dans ce cas, une interprétation intuitive de la probabilité invariante.

Si f est une fonction positive sur E, on a alors, par un calcul simple,

$$\int f d\mu = \sum_{x \in E} f(x) \mu(x) = E_y \left[\sum_{n=1}^{T_y^1} f(X_n) \right] = E_y \left[\sum_{n=0}^{T_y^1-1} f(X_n) \right]; \quad (16.82)$$

si f est de signe quelconque, on en déduit classiquement un critère d'intégrabilité; les formules (16.82) restant vraies pour les fonctions μ -intégrables.

• Le principe de la démonstration est maintenant de découper les sommes étudiées selon les différents temps de passage en y , les morceaux étant **indépendants de même loi**; on utilise alors la loi des grands nombres pour les variables aléatoires indépendantes.

Soient donc f une fonction μ -intégrable et, pour tout $p \in \mathbb{N}$, Z_p la variable aléatoire définie par

$$Z_p = \begin{cases} \sum_{n=T_y^p}^{T_y^{p+1}-1} f(X_n) & \text{sur } (T_y^p < +\infty) \\ 0 & \text{sur } (T_y^p = +\infty), \end{cases}$$

(on rappelle que $T_y^0 = 0$). Puisque, pour tout $x \in E$ et tout $p \in \mathbb{N}^*$, on a $P_x(T_y^p < +\infty) = 1$, **les variables aléatoires Z_p sont P_x -p.s.-finies**. Démontrons que, pour tout $x \in E$, elles sont **P_x -indépendantes et de même loi**¹⁵. En effet, soit, pour tout $p \in \mathbb{N}^*$, un borélien quelconque A_p de \mathbb{R} ; puisque, P_x -p.s., on a

$$T_y^{p+1} = T_y^p + \tau_y^1 \left[\theta_{T_y^p}(X) \right],$$

il résulte de la propriété de Markov forte que

$$E_x^{\mathcal{A}_{T_y^p}} \left[\mathbf{1}_{(Z_p \in A_p)} \right] = E_{X_{T_y^p}} \left[\mathbf{1}_{\left(\sum_{n=0}^{T_y^p-1} f(X_n) \in A_p \right)} \right] = P_y(Z_0 \in A_p).$$

Il en résulte que, pour tout N , on a, par un conditionnement classique,

$$\begin{aligned} E_x \left[\prod_{p=1}^N \mathbf{1}_{(Z_p \in A_p)} \right] &= E_x \left[\left(\prod_{p=1}^{N-1} \mathbf{1}_{(Z_p \in A_p)} \right) E_x^{\mathcal{A}_{T_y^{N-1}}} \left[\mathbf{1}_{(Z_N \in A_N)} \right] \right] \\ &= E_x \left[\prod_{p=1}^{N-1} \mathbf{1}_{(Z_p \in A_p)} \right] P_y(Z_0 \in A_N). \end{aligned}$$

Une itération rétrograde conduit alors à l'égalité

$$E_x \left[\prod_{p=1}^N \mathbf{1}_{(Z_p \in A_p)} \right] = \prod_{p=1}^N P_y(Z_0 \in A_p),$$

ce qui prouve que les Z_p , $p \in \mathbb{N}^*$, ont même loi sous P_x que Z_0 sous P_y et que les Z_p sont P_x -indépendantes.

• Montrons que Z_1 est P_x -intégrable (il en est alors de même de tous les Z_p). On a

$$E_x(|Z_1|) = E_x \left[\left| \sum_{n=T_y^1}^{T_y^2-1} f(X_n) \right| \right] = E_x \left[E_x^{\mathcal{A}_{T_y^1}} \left[\left| \sum_{n=T_y^1}^{T_y^2-1} f(X_n) \right| \right] \right],$$

15. En utilisant cette remarque, il est clair que l'on peut déduire un **théorème limite central** pour les chaînes de Markov homogènes satisfaisant aux hypothèses de ce présent théorème, à partir d'un théorème limite central pour les suites de variables aléatoires indépendantes.

soit, d'après la propriété de Markov forte,

$$E_x(|Z_1|) = E_x \left[E_{X_{T_y^1}} \left| \sum_{n=0}^{T_y^1-1} f(X_n) \right| \right] = E_y \left[\left| \sum_{n=0}^{T_y^1-1} f(X_n) \right| \right].$$

Il en résulte que

$$E_x(|Z_1|) \leq E_y \left[\sum_{n=0}^{T_y^1-1} |f(X_n)| \right] = \int |f| d\mu < +\infty.$$

Un calcul similaire montre alors que $E_x(Z_1) = \int f d\mu$.

• Il résulte alors de la deuxième loi forte des grands nombres pour les variables aléatoires indépendantes que

$$\frac{1}{n} \sum_{p=1}^{n-1} Z_p = \frac{1}{n} \sum_{k=T_y^1}^{T_y^n-1} f(X_k) \xrightarrow{P_x\text{-p.s.}} \int f d\mu. \quad (16.83)$$

Soit alors la suite croissante des entiers aléatoires $v(n) = \sum_{j=1}^n \mathbf{1}_{(X_j=y)}$, nombre de passages en y jusqu'à l'instant n . On a, par hypothèse,

$$P_x(\lim_n v(n) = +\infty) = P_x(R_y) = 1$$

et, par définition de $v(n)$,

$$T_y^{v(n)} \leq n < T_y^{v(n)+1}.$$

Si de plus f est positive, on en déduit les inégalités

$$\frac{\sum_{k=0}^{T_y^{v(n)}} f(X_k)}{v(n)} \leq \frac{\sum_{k=0}^n f(X_k)}{n} \leq \frac{\sum_{k=0}^{T_y^{v(n)+1}} f(X_k)}{v(n)};$$

D'après (16.83), les termes extrêmes convergent P_x -p.s.; il en est de même du terme médian. Dans ce cas, le théorème en résulte de suite, en se rappelant que π est proportionnelle à μ . On en déduit le théorème dans le cas général où f est de signe quelconque, en décomposant f en ses parties positive et négative. \square

En particulier, on obtient la formulation traditionnelle de l'énoncé de la loi forte des grands nombres pour les chaînes de Markov homogènes.

Corollaire 16.57 (Loi forte des grands nombres). *Sous les hypothèses du théorème de Chacon-Orstein 16.56, on a, pour toute fonction f π -intégrable,*

$$\boxed{\frac{1}{n} \sum_{j=1}^n f(X_j) \xrightarrow{P_x\text{-p.s.}} \int f d\pi.}$$

Démonstration. Il suffit d'appliquer le théorème de Chacon-Orstein en prenant pour g la fonction constante égale à 1. \square

Remarque. Sous les mêmes hypothèses, en prenant pour f l'indicatrice d'un singleton, on obtient que, pour tous $x, y \in E$,

$$\boxed{\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(X_j=y)} \xrightarrow{P_x\text{-p.s.}} \pi(y);}$$

le quotient $\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(X_j=y)}$ représente le **temps moyen** passé par une trajectoire dans l'état y entre les instants 1 et n . Ce résultat donne un procédé d'**estimation de la probabilité invariante**.

16.8.2. Estimation de la matrice de transition

On suppose ici que $E = \{x_1, \dots, x_L\}$ est fini et que X est une chaîne de Markov homogène irréductible de matrice de transition M ; elle est alors récurrente positive et il existe une unique probabilité invariante π . On définit, pour $i, j \in \{1, 2, \dots, L\}$ et $n \in \mathbb{N}^*$, les variables aléatoires N_i^n et $N_{i,j}^n$ par

$$N_i^n = \sum_{l=0}^{n-1} \mathbf{1}_{(X_l=x_i)} \quad \text{et} \quad N_{i,j}^n = \sum_{l=0}^{n-1} \mathbf{1}_{(X_l=x_i)} \mathbf{1}_{(X_{l+1}=x_j)},$$

qui représentent respectivement le nombre de visites en x_i et le nombre de passages de x_i en x_j jusqu'au temps n . On a

$$\sum_{j=1}^L N_{i,j}^n = \sum_{l=0}^{n-1} \left[\sum_{j=1}^L \mathbf{1}_{(X_l=x_i)} \mathbf{1}_{(X_{l+1}=x_j)} \right] = \sum_{l=0}^{n-1} \mathbf{1}_{(X_l=x_i)},$$

c'est-à-dire que $N_i^n = \sum_{j=1}^L N_{i,j}^n$. On note $\widehat{M}_{i,j}^n = \frac{N_{i,j}^n}{N_i^n}$; on étudie, pour tout $x \in E$, la convergence P_x -p.s. de la suite de terme général $\widehat{M}_{i,j}^n$.

Proposition 16.58. *Avec les notations et hypothèses précédentes, pour $i, j \in \{1, 2, \dots, L\}$ et pour tout $x \in E$,*

$$\widehat{M}_{i,j}^n \xrightarrow{P_x\text{-p.s.}} M(x_i, x_j).$$

Démonstration. D'après la remarque précédente, on a déjà, pour $i \in \{1, 2, \dots, L\}$,

$$\frac{N_i^n}{n} \xrightarrow{P_X - p.s.} \pi(x_i). \quad (16.84)$$

Soit $\pi \otimes M$ la probabilité sur $E \times E$ définie, pour toute partie A de $E \times E$ par

$$\begin{aligned} (\pi \otimes M)(A) &= \int \left[\sum_{i=1}^L \mathbf{1}_A(x, x_i) M(x, x_i) \right] d\pi(x) \\ &= \sum_{x \in E} \left[\sum_{i=1}^L \mathbf{1}_A(x, x_i) M(x, x_i) \right] \pi(x) : \end{aligned}$$

on a, bien sûr,

$$\int f d(\pi \otimes M) = \sum_{x \in E} \left[\sum_{i=1}^L f(x, x_i) M(x, x_i) \right] \pi(x).$$

En adaptant la démonstration du théorème 16.57, on démontre que, pour toute fonction sur $E \times E$ (ici l'intégrabilité est automatique), on a

$$\frac{1}{n} \sum_{l=0}^{n-1} f(X_l, X_{l+1}) \xrightarrow{P_X - p.s.} \int f d\pi \otimes M. \quad (16.85)$$

En effet, pour $y \in E$ quelconque, si on définit

$$Z_p = \begin{cases} \mathbf{1}_y^{T_y^p + 1} \\ \sum_{n=T_y^p} f(X_n, X_{n+1}) & \text{sur } (T_y^p < +\infty) \\ 0 & \text{sur } (T_y^p = +\infty). \end{cases}$$

par un calcul analogue à celui de la démonstration du théorème 16.57, on a, pour tout borélien B de \mathbb{R} ,

$$E_x \mathbf{1}_y^{T_y^p} [\mathbf{1}_{(Z_p \in B)}] = P_y(Z_0 \in B) :$$

ceci permet encore de démontrer que les Z_p , $p \in \mathbb{N}^*$, ont même loi sous P_x que Z_0 sous P_y et que les Z_p sont P_x -indépendantes.

Reste à calculer $E_x(Z_1)$. Par une démarche analogue à celle de la démonstration du théorème 16.57, la propriété de Markov forte permet de montrer que l'on a

$$E_x(Z_1) = E_x \left[\sum_{n=0}^{T_y^1 - 1} f(X_n, X_{n+1}) \right].$$

Remarquant que $(n < T_y^1) \in \mathcal{A}_n$, il résulte de la propriété de Markov simple que l'on a

$$E_x(Z_1) = \sum_{n=0}^{+\infty} E_y \left[\mathbf{1}_{(n < T_y^1)} E_y^{\mathcal{A}_n} f(X_n, X_{n+1}) \right] = \sum_{n=0}^{+\infty} E_y \left[\mathbf{1}_{(n < T_y^1)} E_{X_n} f(X_n, X_{n+1}) \right],$$

ce qui s'écrit encore

$$E_x(Z_1) = \sum_{n=0}^{+\infty} E_y \left[\mathbf{1}_{(n < T_y^1)} \left(\sum_{i=1}^L f(X_n, x_i) M(X_n, x_i) \right) \right],$$

ou encore

$$E_x(Z_1) = E_y \left[\sum_{n=0}^{T_y^1 - 1} \left(\sum_{i=1}^L f(X_n, x_i) M(X_n, x_i) \right) \right].$$

Les formules (16.82) et l'égalité $\mu = E_y(T_y^1)\pi$ (cf. théorème 16.57) permettent alors d'écrire

$$E_x(Z_1) = \sum_{x \in E} \mu(x) \left[\sum_{i=1}^L f(x, x_i) M(x, x_i) \right] = E_y(T_y^1) \int f d(\pi \otimes M).$$

La convergence annoncée en (16.85) s'obtient ensuite comme à la fin de la démonstration du théorème 16.57 et de la remarque qui le suit.

En prenant pour f la fonction définie par

$$f(x, y) = \mathbf{1}_{\{x, i\}}(x) \mathbf{1}_{\{x, i\}}(y),$$

il vient alors que

$$N_{i,j}^n \xrightarrow{P_x\text{-p.s.}} \int f d(\pi \otimes M) = \pi(x_i) M(x_i, x_j).$$

Tenant compte de (16.84), la proposition en résulte. \square

Exercices

Sauf mention contraire, les variables aléatoires introduites sont définies sur un espace probabilisé (Ω, \mathcal{A}, P) , et les processus sur une base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{Z}^+}, P)$ adéquats.

Exercice 16.1. Gestion de stock. Un vendeur d'appareils photos a remarqué que le nombre A_t d'acheteurs d'un appareil d'un certain type, pendant la semaine t , était indépendant du nombre d'acheteurs de cet appareil durant les semaines précédentes, et que sa loi était donnée par

$$\begin{aligned} P(A_t = 0) &= 0.4, & P(A_t = 1) &= 0.4, & P(A_t = 2) &= 0.15, \\ P(A_t = 3) &= 0.05, & P(A_t > 3) &= 0. \end{aligned}$$

Le vendeur fait ses commandes en fin de semaine, et n'en fait que s'il n'a plus d'appareil en stock en fin de semaine; dans ce cas, il décide d'en commander deux

(qu'il reçoit dès le premier jour d'ouverture de la semaine suivante). On note X_t le nombre d'appareils en stock à la fin de la semaine t . Démontrer que $X = (X_t)_{t \in \mathbb{N}}$ est une chaîne de Markov homogène (relativement à sa filtration naturelle $(\mathcal{A}_t)_{t \in \mathbb{N}}$), à valeurs dans $E = \{0, 1, 2\}$, de matrice de transition M donnée par

$$M = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.6 & 0.4 & 0 \\ 0.2 & 0.4 & 0.4 \end{pmatrix} \end{matrix}$$

Solution. Si f est la fonction définie sur E^2 par

$$f(x, y) = \mathbf{1}_{(x \geq 1) \cap (x \geq y)} \times (x - y) + \mathbf{1}_{(x=0) \cap (2 \geq y)} \times (2 - y),$$

on a¹⁶ $X_{t+1} = f(X_t, A_{t+1})$. Ainsi, la variable aléatoire (X_0, X_1, \dots, X_t) est fonction (mesurable) de (A_1, A_2, \dots, A_t) ; puisque les variables aléatoires $A_t, t \in \mathbb{N}$, sont indépendantes, les variables aléatoires (X_0, X_1, \dots, X_t) et A_{t+1} le sont donc aussi. Alors, pour toute fonction g sur E , on a, pour tout $(x_0, x_1, \dots, x_t) \in E^{t+1}$,

$$\begin{aligned} \mathbb{E}^{(X_0, X_1, \dots, X_t) = (x_0, x_1, \dots, x_t)} [g(X_{t+1})] \\ = \mathbb{E}^{(X_0, X_1, \dots, X_t) = (x_0, x_1, \dots, x_t)} [g \circ f(x_t, A_{t+1})] = \mathbb{E} [g \circ f(x_t, A_{t+1})]. \end{aligned}$$

En notant, pour tout $x \in E$, $M(x, g) = \mathbb{E} [g \circ f(x, A_{t+1})]$, on a montré que

$$\mathbb{E}^{\mathcal{A}_t} [g(X_{t+1})] = M(X_t, g),$$

ce qui prouve que X est une chaîne de Markov homogène de matrice de transition la matrice M d'entrées données par

$$M(x, y) = M(x, \mathbf{1}_{\{y\}}) = \mathbb{P}[f(x, A_{t+1}) = y].$$

La loi de A_{t+1} conduit alors à la matrice M annoncée. En effet on a successivement

$$f(0, A_{t+1}) = (2 - A_{t+1})\mathbf{1}_{(A_{t+1} \leq 2)} = 2\mathbf{1}_{(A_{t+1}=0)} + \mathbf{1}_{(A_{t+1}=1)},$$

ce qui donne

$$\begin{aligned} M(0, 0) &= \mathbb{P}(A_{t+1} \geq 2) = 0.15 + 0.05 = 0.2, \\ M(0, 1) &= \mathbb{P}(A_{t+1} = 1) = 0.4, \\ M(0, 2) &= \mathbb{P}(A_{t+1} = 0) = 0.4; \end{aligned}$$

$$f(1, A_{t+1}) = (1 - A_{t+1})\mathbf{1}_{(1 \geq A_{t+1})} = \mathbf{1}_{(A_{t+1}=0)},$$

ce qui donne

$$\begin{aligned} M(1, 0) &= \mathbb{P}(A_{t+1} \geq 1) = 0.4 + 0.15 + 0.05 = 0.6, \\ M(1, 1) &= \mathbb{P}(A_{t+1} = 0) = 0.4, \\ M(1, 2) &= \mathbb{P}(\emptyset) = 0; \end{aligned}$$

16. Il s'agit donc d'un processus **auto-régressif** (cf. l'exemple 16.6).

$$f(2, A_{t+1}) = (2 - A_{t+1})\mathbf{1}_{(2 \geq A_{t+1})} = 2\mathbf{1}_{(A_{t+1}=0)} + \mathbf{1}_{(A_{t+1}=1)},$$

ce qui donne

$$\begin{aligned} M(2, 0) &= P(A_{t+1} \geq 2) = 0.15 + 0.05 = 0.2, \\ M(2, 1) &= P(A_{t+1} = 1) = 0.4, \\ M(2, 2) &= P(A_{t+1} = 0) = 0.4. \end{aligned}$$

Exercice 16.2. Fiabilité; temps de vie. Le temps est compté de manière discrète (par exemple en secondes) et est donc indexé sur \mathbb{N} . Une machine en fonctionnement permanent possède une pièce critique qui casse facilement. Dès qu'elle est cassée, elle est remplacée instantanément par une pièce identique. On note X_n le temps aléatoire s'écoulant entre le temps n et la prochaine panne après n ; la suite de ces temps vérifie la relation

$$X_{n+1} = \begin{cases} X_n - 1 & \text{si } X_n \geq 1, \\ Z_n - 1 & \text{si } X_n = 0, \end{cases}$$

où Z_n est le temps de vie de la pièce changée à l'instant n . On suppose que les Z_n , $n \in \mathbb{N}^*$, sont des variables aléatoires indépendantes de même loi μ sur \mathbb{N}^* engendrée par le germe $(p_k)_{k \in \mathbb{N}^*}$. Démontrer que le processus $X = (X_n)_{n \in \mathbb{N}^*}$ est une chaîne de Markov homogène (relativement à sa filtration naturelle $(\mathcal{A}_n)_{n \in \mathbb{N}^*}$), à valeurs dans \mathbb{N} , de matrice de transition M à déterminer.

Solution. Si f est la fonction définie sur \mathbb{N}^2 par

$$f(x, y) = \mathbf{1}_{(x \geq 1)} \times (x - 1) + \mathbf{1}_{(x=0)} \times (y - 1),$$

on a $X_{n+1} = f(X_n, Z_n)$. En particulier, puisque les variables aléatoires Z_n , $n \in \mathbb{N}$, sont indépendantes, les variables aléatoires (X_1, \dots, X_n) et Z_n le sont aussi. Alors, pour toute fonction bornée g sur \mathbb{N} , on a, pour tout $(x_1, \dots, x_n) \in \mathbb{N}^n$,

$$\begin{aligned} E^{(X_1, \dots, X_n) = (x_1, \dots, x_n)} [g(X_{n+1})] \\ = E^{(X_1, \dots, X_n) = (x_1, \dots, x_n)} [g \circ f(x_n, Z_n)] = E[g \circ f(x_n, Z_n)]. \end{aligned}$$

En notant, pour tout $x \in \mathbb{N}$, $M(x, g) = E[g \circ f(x, Z_n)] = \int_{\mathbb{R}} g \circ f(x, z) d\mu(z)$, on a montré que

$$E^{\mathcal{A}_n} [g(X_{n+1})] = M(X_n, g),$$

ce qui prouve que X est une chaîne de Markov homogène de matrice de transition M , d'entrées données par

$$M(x, y) = M(x, \mathbf{1}_{\{y\}}) = P[f(x, Z_n) = y] = \mu(\{z \mid f(x, z) = y\}).$$

On peut d'ailleurs les obtenir ainsi :

- si $x \in \mathbb{N}^*$ et $y \in \mathbb{N}$, on a

$$M(x, y) = P(X_{n+1} = y \mid X_n = x) = P(X_n - 1 = y \mid X_n = x),$$

et donc

$$M(x, y) = P(X_n = y + 1 \mid X_n = x) = \mathbf{1}_{(y=x-1)};$$

– si $x = 0$ et $y \in \mathbb{N}$, on a

$$M(0, y) = P(X_{n+1} = y \mid X_n = 0) = P(Z_n - 1 = y \mid X_n = 0),$$

et donc, puisque Z_n et X_n sont indépendantes,

$$M(0, y) = P(Z_n = y + 1) = p_{y+1}.$$

Ainsi, la matrice infinie M a la forme

$$\begin{array}{c} \mathbf{0} \quad \mathbf{1} \quad \mathbf{2} \quad \mathbf{3} \quad \mathbf{4} \quad \dots \\ \mathbf{0} \begin{pmatrix} p_1 & p_2 & p_3 & p_4 & \dots & \dots \\ \mathbf{1} \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & \dots \\ \mathbf{2} \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & \dots \\ \mathbf{3} \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & \dots \\ \mathbf{4} \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & \dots \\ \vdots \begin{pmatrix} \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix} \end{pmatrix} \end{pmatrix} \end{pmatrix} \end{pmatrix} \end{pmatrix} \end{pmatrix} \end{array}$$

Exercice 16.3. Propriétés de Markov simple et forte. Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans $E = \{1, 2, 3\}$ qui, pour tout $i \in E$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_i)$, de matrice de transition

$$M = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

1. Soit f la fonctionnelle sur $E^{\mathbb{N}}$ définie, pour tout $x \in E^{\mathbb{N}}$, par

$$f(x) = \mathbf{1}_{(\sum_{j=1}^3 \mathbf{1}_{\{1\}}(x_j) = 1)};$$

autrement dit, $f(x)$ vaut 1 si, parmi les trois composantes de x d'indices 1 à 3, une et une seule vaut 1; $f(x)$ vaut 0 sinon. Calculer $E_i^{\mathcal{A}_n} [f(\theta_n(X))]$ pour tout $i \in E$.

2. Soit $T = \inf(n \in \mathbb{N}^* \mid X_n = 1)$, avec $\inf \emptyset = +\infty$. Justifier le fait que, pour tout $i \in E$, $P_i(T < +\infty) = 1$ et calculer l'espérance conditionnelle $E_i^{\mathcal{A}_T} [f(\theta_T(X))]$ pour tout $i \in E$.

Solution.

1. La propriété de Markov simple assure que

$$E_i^{\mathcal{A}_n} [f(\theta_n(X))] = E_{X_n} [f(X)].$$

On calcule donc, pour tout $i \in E$, $E_i [f(X)]$. On a $E_i [f(X)] = P_i(\sum_{j=1}^3 \mathbf{1}_{\{1\}}(X_j) = 1)$, et par conséquent,

$$\begin{aligned} E_i [f(X)] &= P_i(X_1 = 1, X_2 \neq 1, X_3 \neq 1) + P_i(X_1 \neq 1, X_2 = 1, X_3 \neq 1) \\ &\quad + P_i(X_1 \neq 1, X_2 \neq 1, X_3 = 1); \end{aligned}$$

or

$$\begin{aligned}
 P_i(X_1 = 1, X_2 \neq 1, X_3 \neq 1) &= \sum_{j,k=2,3} P_i(X_1 = 1, X_2 = j, X_3 = k) \\
 &= \sum_{j,k=2,3} M(i, 1)M(1, j)M(j, k) \\
 &= M(i, 1) \left[\sum_{j=2,3} M(1, j) \sum_{k=2,3} M(j, k) \right],
 \end{aligned}$$

et, puisque $\sum_{k=2,3} M(j, k) = 1 - M(j, 1)$,

$$\begin{aligned}
 P_i(X_1 = 1, X_2 \neq 1, X_3 \neq 1) &= M(i, 1) \left[\sum_{j=2,3} M(1, j) - \sum_{j=2,3} M(1, j)M(j, 1) \right] \\
 &= M(i, 1) \left[1 - \frac{1}{3} \frac{1}{2} - \frac{2}{3} \frac{1}{2} \right],
 \end{aligned}$$

soit

$$P_i(X_1 = 1, X_2 \neq 1, X_3 \neq 1) = \frac{1}{2} M(i, 1). \quad (16.86)$$

De même, on a

$$\begin{aligned}
 P_i(X_1 \neq 1, X_2 = 1, X_3 \neq 1) &= \sum_{j,k=2,3} P_i(X_1 = j, X_2 = 1, X_3 = k) \\
 &= \sum_{j,k=2,3} M(i, j)M(j, 1)M(1, k) \\
 &= \left[\sum_{j=2,3} M(i, j)M(j, 1) \right] \left[\sum_{k=2,3} M(1, k) \right].
 \end{aligned}$$

et, puisque $\sum_{k=2,3} M(1, k) = 1$,

$$P_i(X_1 \neq 1, X_2 = 1, X_3 \neq 1) = \frac{1}{2} [M(i, 2) + M(i, 3)]. \quad (16.87)$$

Enfin, on a de même

$$\begin{aligned}
 P_i(X_1 \neq 1, X_2 \neq 1, X_3 = 1) &= \sum_{j,k=2,3} P_i(X_1 = j, X_2 = k, X_3 = 1) \\
 &= \sum_{j,k=2,3} M(i, j)M(j, k)M(k, 1),
 \end{aligned}$$

soit

$$\begin{aligned}
 P_i(X_1 \neq 1, X_2 \neq 1, X_3 = 1) &= M(i, 2) \left[\sum_{k=2,3} M(2, k)M(k, 1) \right] \\
 &\quad + M(i, 3) \left[\sum_{k=2,3} M(3, k)M(k, 1) \right],
 \end{aligned}$$

ou encore

$$P_i(X_1 \neq 1, X_2 \neq 1, X_3 = 1) = M(i, 2) \left[\frac{1}{4} \frac{1}{2} + \frac{1}{4} \frac{1}{2} \right] + M(i, 3) \left[\frac{1}{2} \frac{1}{2} + 0 \times \frac{1}{2} \right];$$

ainsi, on a

$$P_i(X_1 \neq 1, X_2 \neq 1, X_3 = 1) = \frac{1}{4} [M(i, 2) + M(i, 3)]. \quad (16.88)$$

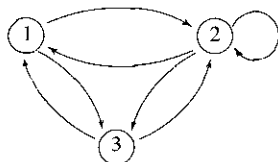
En rassemblant les égalités (16.86), (16.87) et (16.88), on obtient

$$\begin{aligned} E_i[f(X)] &= \frac{1}{2}M(i, 1) + \frac{3}{4}[M(i, 2) + M(i, 3)] \\ &= \frac{1}{2}M(i, 1) + \frac{3}{4}[1 - M(i, 1)] = \frac{3}{4} - \frac{1}{4}M(i, 1), \end{aligned}$$

et donc

$$E_i^{A_n} [f(\theta_n(X))] = \frac{3}{4} - \frac{1}{4}M(X_n, 1).$$

2. Le graphe associé à cette chaîne est



et permet d'assurer que tous les points communiquent; la chaîne est donc irréductible finie, donc récurrente positive. En particulier, on a, pour tout $i \in E$, $P_i(T < +\infty) = 1$. Il résulte alors de la propriété de Markov forte et de la question précédente que

$$E_i^{A_T} [f(\theta_T(X))] = E_{X_T} [f(X)] = \frac{3}{4} - \frac{1}{4}M(X_T, 1).$$

Puisque $X_T = 1$ et que $M(1, 1) = 0$, on a

$$E_i^{A_T} [f(\theta_T(X))] = \frac{3}{4}.$$

Exercice 16.4. Processus de naissance et de mort discret, marche aléatoire sur \mathbb{N} avec barrières élastiques, problème de Dirichlet. Il s'agit de modéliser l'évolution de la taille d'une population (d'individus, de particules physiques ou tout autres choses) dans laquelle, à chaque instant n , peut apparaître, ou disparaître, un élément, et ceci avec une probabilité qui dépend de la taille actuelle X_n de la population.

Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans \mathbb{N} qui, pour tout $x \in \mathbb{N}$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$, de matrice de transition M donnée par, pour tout $x \in \mathbb{N}$,

$$M(x, x+1) = p_x \quad M(x, x-1) = q_x \quad M(x, x) = r_x,$$

où

$$p_x, q_x, r_x \in [0, 1] \quad \text{et} \quad p_x + q_x + r_x = 1.$$

Soient $a, b \in \mathbb{N}$ tels que $0 \leq a < b$. On note T_x le temps d'entrée en x , c'est-à-dire $T_x = \inf(n \in \mathbb{N}^* \mid X_n = x)$, avec $\inf \emptyset = +\infty$.

1. Démontrer que pour tout x , on a

$$P_x(T_a < T_b) = M(x, a) + \sum_{z \neq a, b} P_z(T_a < T_b) M(x, z) \quad (16.89)$$

2. On suppose que $a = 0$ et que a et b sont des **barrières**, c'est-à-dire que $q_0 = 0$ et $p_b = 0$: elles sont **réfléchissantes** dans la mesure où $r_a > 0$ et $r_b > 0$ (si, à un moment donné, tous les éléments de la population ont disparu, on peut remettre un élément à l'instant suivant, et inversement, si la population atteint la taille b , on peut retirer un élément à l'instant suivant, mais on ne peut pas en rajouter). On suppose de plus que $p_x > 0$, pour tout $x \in]0, b[$.

On note, pour tout $x \in]0, b[$, $f(x) = P_x(T_0 < T_b)$ (c'est la probabilité que, partant de la taille x , la population s'éteigne avant d'atteindre la taille b). Démontrer que f est solution d'une équation de récurrence du second ordre avec conditions aux limites (**problème de Dirichlet**). Calculer explicitement $P_x(T_0 < T_b)$ pour $x \in]0, b[$ en fonction de la suite de terme général a_x donnée par

$$a_0 = 1 \quad a_x = \frac{q_1 q_2 \cdots q_x}{p_1 p_2 \cdots p_x} \quad \text{si } x \in]0, b[.$$

Solution.

1. On a, si θ est l'opérateur de translation sur $\mathbb{N}^{\mathbb{N}}$,

$$P_x(T_a < T_b) = E_x \mathbf{1}_{(X_1=a)} + E_x \left[\mathbf{1}_{(X_1 \neq a, b)} \mathbf{1}_{\{\tau_a(\theta(X)) < \tau_b(\theta(X))\}} \right]:$$

en conditionnant par rapport \mathcal{A}_1 , la propriété de Markov simple donne

$$\begin{aligned} P_x(T_a < T_b) &= M(x, a) + E_x \left[\mathbf{1}_{(X_1 \neq a, b)} E_x^{\mathcal{A}_1} \left(\mathbf{1}_{\{\tau_a(\theta(X)) < \tau_b(\theta(X))\}} \right) \right] \\ &= M(x, a) + E_x \left[\mathbf{1}_{(X_1 \neq a, b)} E_{X_1} \left(\mathbf{1}_{\{\tau_a(X) < \tau_b(X)\}} \right) \right], \end{aligned}$$

soit

$$P_x(T_a < T_b) = M(x, a) + E_x \left[\sum_{z \neq a, b} \mathbf{1}_{(X_1=z)} E_z(\mathbf{1}_{(T_a < T_b)}) \right],$$

et donc, en factorisant,

$$P_x(T_a < T_b) = M(x, a) + \sum_{z \neq a, b} [E_x(\mathbf{1}_{(X_1=z)}) E_z(\mathbf{1}_{(T_a < T_b)})],$$

ce qui donne (16.89).

2. La fonction f est alors solution du système

$$\begin{cases} f(x) = p_x f(x+1) + q_x f(x-1) + r_x f(x) & \text{si } 1 < x < b, \\ f(1) = p_1 f(2) + q_1 + r_1 f(1) \\ f(b-1) = q_{b-1} f(b-2) + r_{b-1} f(b-1), \end{cases}$$

ce qui s'écrit, en prolongeant la fonction f par les égalités $f(0) = 1$ et $f(b) = 0$ (mais il ne faut pas interpréter de manière probabiliste ce prolongement),

$$f(x+1) - f(x) = \frac{q_x}{p_x} [f(x) - f(x-1)],$$

ce qui conduit, par itération, à

$$f(x+1) - f(x) = \frac{q_x q_{x-1} \cdots q_1}{p_x p_{x-1} \cdots p_1} [f(1) - f(0)].$$

On a donc, si $0 \leq x < b$.

$$f(x) - f(x+1) = a_x [f(0) - f(1)]; \quad (16.90)$$

en sommant sur x à partir de 0, on identifie $f(0) - f(1)$:

$$\left[\sum_{x=0}^{b-1} a_x \right] [f(0) - f(1)] = \sum_{x=0}^{b-1} [f(x) - f(x+1)] = f(0) - f(b) = 1;$$

reportant dans (16.90), on a alors

$$f(x) - f(x+1) = \frac{a_x}{\sum_{y=0}^{b-1} a_y},$$

et donc, en sommant encore les accroissements de f (mais, cette fois, à partir de x).

$$f(x) = \sum_{y=x}^{b-1} [f(y) - f(y+1)] = \frac{\sum_{y=x}^{b-1} a_y}{\sum_{y=0}^{b-1} a_y};$$

on a ainsi, pour tout $x \in]0, b[$

$$P_x(T_0 < T_b) = \frac{\sum_{y=x}^{b-1} a_y}{\sum_{y=0}^{b-1} a_y}.$$

Exercice 16.5. Loi du premier temps de passage en un point. Soient $E = \{1, 2, 3\}$ et $X = (X_n)_{n \in \mathbb{N}}$ un processus qui, pour tout $x \in E$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$, de matrice de transition M donnée par

$$M = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{5}{5} & \frac{1}{15} \end{pmatrix}.$$

1. Déterminer le graphe associé à cette chaîne et spécifier les classes de communication.

2. On étudie, sous la probabilité P_x , la loi du premier temps de passage en 3, soit $T_3 = \inf(n \in \mathbb{N}^* \mid X_n = 3)$, avec $\inf \emptyset = +\infty$. Pour cela, on note, pour $x \in E$, $f_k(x) = P_x(T_3 = k) \equiv F_k(x, 3)$. Démontrer que la suite $(f_k)_{k \in \mathbb{N}^*}$ de vecteurs de \mathbb{R}^3 est solution de l'équation de récurrence

$$f_k = Q f_{k-1}, \quad (16.91)$$

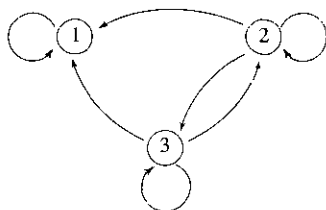
où Q est une matrice 3×3 à déterminer. Calculer alors f_k , pour tout $k \in \mathbb{N}^*$ (on pourra avantageusement utiliser le théorème de Cayley-Hamilton).

3. Calculer, pour tout $x \in E$, la probabilité $P_x(T_3 = +\infty)$.

4. On note, pour tout $y \in E$, $N_y = \sum_{j \in \mathbb{N}} \mathbf{1}_{(X_j = y)}$ le nombre de passages en y . Déterminer $E_3(N_3)$ et $E_2(N_3)$.

Solution.

1. Le graphe associé à cette chaîne est



ce qui met en évidence les deux classes de communication $\{1\}$ et $\{2, 3\}$. Le point 1 est récurrent et même absorbant ; la classe $\{2, 3\}$ est transitoire.

2. D'après la proposition 16.25, et plus précisément l'égalité (16.47), on a

$$\begin{cases} F_1(x, 3) = M(x, 3), \\ F_k(x, 3) = \sum_{z \in E \setminus \{3\}} M(x, z) F_{k-1}(z, 3) & \text{si } k \geq 2. \end{cases}$$

Il en résulte que la suite des vecteurs f_k est solution de (16.91), où f_1 est la dernière colonne de M , et Q est la matrice M dont on a annulé la dernière colonne, c'est-à-dire

$$f_1 = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 1 \\ 15 \end{pmatrix} \quad \text{et} \quad Q = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{6} & 0 \\ \frac{1}{3} & \frac{3}{5} & 0 \end{pmatrix}.$$

On a donc $f_k = Q^{k-1} f_1$. Plutôt que de calculer séparément les puissances de Q , on donne un calcul de f_k s'appuyant sur la méthode de calcul des puissances de Q à l'aide du théorème de Cayley-Hamilton : Q admettant les valeurs propres simples $1, \frac{1}{6}, 0$, annule son polynôme caractéristique :

$$Q(Q-1)(Q-\frac{1}{6}) = 0,$$

ou encore, après développement,

$$Q^3 - Q^2 = \frac{1}{6}(Q^2 - Q).$$

Il en résulte que, pour tout $n \geq 3$, on a

$$Q^n - Q^{n-1} = \frac{1}{6}(Q^{n-1} - Q^{n-2}),$$

et donc, par sommation et application à f_1 , que

$$\sum_{j=3}^n (Q^j - Q^{j-1}) f_1 = \frac{1}{6} \sum_{j=3}^n (Q^{j-1} - Q^{j-2}) f_1,$$

ou encore

$$Q^n f_1 - Q^2 f_1 = \frac{1}{6}(Q^{n-1} f_1 - Q f_1).$$

On a donc, pour tout $n \geq 3$,

$$f_{n+1} = \frac{1}{6} f_n + \left(Q^2 f_1 - \frac{1}{6} Q f_1 \right).$$

Il en résulte que

$$\begin{aligned} f_{n+1} &= \frac{1}{6^{n-2}} f_3 + \left(1 + \frac{1}{6} + \frac{1}{6^2} + \cdots + \frac{1}{6^{n-3}} \right) \left(Q^2 f_1 - \frac{1}{6} Q f_1 \right) \\ &= \frac{1}{6^{n-2}} Q^2 f_1 + \frac{1 - \frac{1}{6^{n-2}}}{1 - \frac{1}{6}} \left(Q^2 f_1 - \frac{1}{6} Q f_1 \right) \\ &= \frac{6}{5} \left(1 - \frac{1}{6^{n-1}} \right) Q^2 f_1 - \frac{1}{5} \left(1 - \frac{1}{6^{n-2}} \right) Q f_1; \end{aligned}$$

puisque l'on a

$$f_2 \equiv Qf_1 = \begin{pmatrix} 0 \\ 1 \\ 18 \\ 1 \\ 5 \end{pmatrix} \quad \text{et} \quad f_3 \equiv Q^2 f_1 = \frac{1}{6} \begin{pmatrix} 0 \\ 1 \\ 18 \\ 1 \\ 5 \end{pmatrix} = \frac{1}{6} Qf_1,$$

il vient, après simplification, si $n \geq 3$,

$$f_{n+1} = \frac{1}{6^{n-1}} \begin{pmatrix} 0 \\ 1 \\ 18 \\ 1 \\ 5 \end{pmatrix}.$$

On remarque, qu'en fait, cette formule est encore valable pour $n = 1, 2$.

3. On a

$$P_x(T_3 = +\infty) = 1 - P_x(T_3 < +\infty) = 1 - \sum_{k \in \mathbb{N}^*} P_x(T_3 = k) = 1 - \sum_{k \in \mathbb{N}^*} f_k(x):$$

on a donc bien sûr $P_1(T_3 = +\infty) = 1$. Par ailleurs, en remarquant que, pour tout $n \in \mathbb{N}^*$, on a

$$f_n(2) = \frac{1}{3} \frac{1}{6^{n-1}},$$

il vient

$$P_2(T_3 = +\infty) = 1 - \sum_{k \in \mathbb{N}^*} \frac{1}{3} \frac{1}{6^{k-1}} = \frac{3}{5};$$

enfin, puisque, pour tout $n \geq 2$, on a $f_n(3) = \frac{1}{5} \frac{1}{6^{n-2}}$, et que $f_1(3) = \frac{1}{15}$, on a

$$P_3(T_3 = +\infty) = 1 - \left[\frac{1}{15} + \sum_{k \geq 2} \frac{1}{5} \frac{1}{6^{k-2}} \right] = \frac{52}{75}.$$

4. Le nombre moyen $R(x, y)$ de passages en y par la chaîne qui part de x au temps 0 est donné par les égalités (16.49), ce qui donne ici

$$E_3(N_3) = R(3, 3) = \frac{1}{1 - P_3(T_3 < +\infty)} = \frac{75}{52},$$

et

$$E_2(N_3) = R(2, 3) = F(2, 3)R(3, 3) = \frac{15}{26}.$$

Exercice 16.6. Probabilité invariante, temps moyen de retour en un point. Soient $E = \{1, 2, 3, 4, 5\}$ et $X = (X_n)_{n \in \mathbb{N}}$ un processus qui, pour tout $x \in E$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$, de matrice de transition M donnée par

$$M = \begin{pmatrix} 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

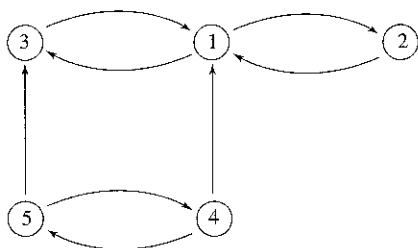
1. Déterminer le graphe associé à cette chaîne et spécifier les classes de communication. Spécifier la période et la nature de ces classes.
2. Justifier, sans calcul, l'existence d'une unique probabilité invariante ν . La calculer et en déduire, pour tout $x \in \{1, 2, 3\}$, $E_x(T_x)$, où T_x est le premier temps de passage en x , soit $T_x = \inf\{n \in \mathbb{N}^* \mid X_n = x\}$, avec $\inf \emptyset = +\infty$.
3. Déterminer la matrice potentiel R .
4. Calculer les probabilités $P_4(T_5 < +\infty)$ et $P_5(T_5 < +\infty)$.
5. On note, pour tout $y \in E$, $N_y = \sum_{j \in \mathbb{N}} \mathbf{1}_{(X_j = y)}$ le nombre de passages en y . Calculer, pour tout $m \in \mathbb{N}$, les probabilités $P_4(N_5 = m)$ et $P_5(N_5 = m)$.
6. Démontrer que l'on a, pour tout $x, y \in E$,

$$E_x[T_y \mathbf{1}_{(T_y < +\infty)}] = M(x, y) + \sum_{z \neq y} M(x, z) [E_z(T_y \mathbf{1}_{(T_y < +\infty)}) + P_z(T_y < +\infty)]. \quad (16.92)$$

En déduire numériquement, pour tout $x \in E$, $E_x(T_1)$.

Solution.

1. Le graphe associé à cette chaîne est



ce qui met en évidence les deux classes de communication $C = \{1, 2, 3\}$ et $D = \{4, 5\}$.

On a $M(4, 4) = 0$ et $M^2(4, 4) \geq M(4, 5)M(5, 4) > 0$; une étude facile (récurrence) montre que l'on a, pour tout $n \in \mathbb{N}^*$, $M^{2n}(4, 4) > 0$ et $M^{2n+1}(4, 4) = 0$. Il en résulte que 4 (et donc 5) est de période 2; c'est la période de la classe D . La même étude montre que 2 (et donc 1 et 3) est de période 2; c'est la période de la classe C . Le point 5 conduit à 3, mais 3 ne conduit pas à 5; 5 n'est donc pas récurrent, il est transitoire. Ainsi, la classe D est transitoire. Puisque E est fini et que C est une classe fermée, elle est récurrente positive.

2. L'existence d'une unique probabilité invariante ν est assurée par l'existence d'une unique classe récurrent positive. Notant $\nu = (a, b, c, d, e)$ un vecteur ligne associé à une **mesure** invariante sur E , ν est solution du système

$$(a, b, c, d, e) \begin{pmatrix} 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} = (a, b, c, d, e),$$

soit

$$\begin{cases} b + c + \frac{d}{2} = a \\ \frac{2}{3} = b \\ \frac{a}{3} + \frac{e}{2} = c \\ \frac{e}{2} = d \\ \frac{d}{2} = e \end{cases} \iff \begin{cases} d = e = 0 \\ b = \frac{2a}{3} \\ c = \frac{a}{3} \end{cases}$$

La probabilité invariante est telle que $a + b + c + d + e = 1$, ce qui donne $a = \frac{1}{2}$. **La probabilité invariante ν est donc** $\nu = (\frac{1}{2}, \frac{1}{3}, \frac{1}{6}, 0, 0)$. On sait que l'on a, pour tout $x \in C$, $\nu(x) = \frac{1}{E_x(T_x)}$, ce qui donne

$$\boxed{E_1(T_1) = 2 \quad E_2(T_2) = 3 \quad E_3(T_3) = 6.}$$

3. La matrice potentiel R a la structure de bloc

$$R = \begin{matrix} & \begin{matrix} C & D \end{matrix} \\ \begin{matrix} C \\ D \end{matrix} & \begin{pmatrix} +\infty & \mathbf{0} \\ +\infty & S \end{pmatrix} \end{matrix},$$

où $Q = M_{|D \times D}$ et $S = \sum_{n=0}^{+\infty} Q^n$, soit encore $S = (I - Q)^{-1}$. On a

$$Q = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \quad \text{et donc} \quad I - Q = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix};$$

ainsi

$$S = \begin{pmatrix} 4 & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{pmatrix}.$$

4. On déduit alors des égalités (16.49) que

$$P_4(T_5 < +\infty) = \frac{R(4, 5)}{R(5, 5)} = \frac{\frac{2}{3}}{\frac{4}{3}} = \frac{1}{2},$$

et

$$P_5(T_5 < +\infty) = 1 - \frac{1}{R(5, 5)} = 1 - \frac{1}{4} = \frac{3}{4}.$$

5. La loi de N_5 sous P_4 est alors donnée (cf la proposition 16.26) par

$$P_4(N_5 = m) = \begin{cases} P_4(T_5 = +\infty) & \text{si } m=0 \\ P_4(T_5 < +\infty)[P_5(T_5 < +\infty)]^{m-1} P_5(T_5 = +\infty) & \text{si } m \in \mathbb{N}^*, \end{cases}$$

soit

$$P_4(N_5 = m) = \begin{cases} \frac{1}{2} & \text{si } m = 0, \\ \frac{3}{8} \left(\frac{1}{4}\right)^{m-1} & \text{si } m \in \mathbb{N}^*. \end{cases}$$

De même, la loi de N_5 sous P_5 est donnée par, pour tout $m \in \mathbb{N}^*$,

$$P_5(N_5 = m) = [P_5(T_5 < +\infty)]^{m-1} P_5(T_5 = +\infty),$$

soit

$$P_5(N_5 = m) = \frac{3}{4} \left(\frac{1}{4}\right)^{m-1};$$

c'est la loi géométrique $\mathcal{G}_{\mathbb{N}^*}(\frac{3}{4})$.

6. On note $F_k(x, y) = P_x(T_y = k)$. D'après la proposition 16.25, et plus précisément l'égalité (16.47), on a

$$\begin{cases} F_1(x, y) = M(x, y), \\ F_k(x, y) = \sum_{z \neq y} M(x, z) F_{k-1}(z, y) & \text{si } k \geq 2. \end{cases}$$

Il en résulte que

$$\begin{aligned} E_x[T_y \mathbf{1}_{(T_y < +\infty)}] &= \sum_{k \in \mathbb{N}^*} k F_k(x, y) \\ &= M(x, y) + \sum_{k=2}^{+\infty} k \left[\sum_{z \neq y} M(x, z) F_{k-1}(z, y) \right] \\ &= M(x, y) + \left[\sum_{z \neq y} M(x, z) \sum_{l=1}^{+\infty} (l+1) F_l(z, y) \right] \\ &= M(x, y) + \sum_{z \neq y} M(x, z) [E_z(T_y \mathbf{1}_{(T_y < +\infty)}) + P_z(T_y < +\infty)], \end{aligned}$$

ce qui est l'égalité (16.92) (tous les termes étant positifs, les permutations de sommes sont toutes licites).

En particulier, puisque pour tout $x \in E$, $P_x(T_1 < +\infty) = 1$, il en résulte que

$$E_x [T_1] = M(x, 1) + \sum_{z \neq 1} M(x, z) E_z(T_1) + \sum_{z \neq 1} M(x, z),$$

soit

$$E_x [T_1] = 1 + \sum_{z \neq 1} M(x, z) E_z(T_1).$$

Ainsi, si on note g le vecteur de composantes $g(x) = E_x [T_1]$, $x \in E$, g est solution de l'équation

$$g = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} g,$$

ou, ce qui est équivalent, ses composantes sont solutions du système

$$\begin{cases} g_1 = 1 + \frac{2}{3}g_2 + \frac{1}{3}g_3 \\ g_2 = 1 \\ g_3 = 1 \\ g_4 = 1 + \frac{1}{2}g_5 \\ g_5 = 1 + \frac{1}{2}g_3 + \frac{1}{2}g_4 \end{cases} \iff \begin{cases} g_1 = 2 \\ g_2 = 1 \\ g_3 = 1 \\ g_4 = \frac{7}{8} \\ g_5 = \frac{8}{3} \end{cases}$$

c'est-à-dire que

$$E_1[T_1] = 2 \quad E_2[T_1] = 1 \quad E_3[T_1] = 1 \quad E_4[T_1] = \frac{7}{8} \quad E_5[T_1] = \frac{8}{3}.$$

Exercice 16.7. Premier temps de passage dans un ensemble et procédé de fabrication. Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans E qui, pour tout $x \in E$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$, de matrice de transition M .

Notations. Pour toute partie C de E , on définit

$$T_C = \inf(n \in \mathbb{N}^* \mid X_n \in C),$$

premier temps de passage dans C ; τ_C est défini sur $E^{\mathbb{N}}$ par, pour tout $u \in E^{\mathbb{N}}$, $\tau_C(u) = \inf(n \in \mathbb{N}^* \mid u_n \in C)$; on pose $\inf \emptyset = +\infty$. Pour tous $x, y \in E$, on note $F(x, y) = P_x(T_y < +\infty)$. On introduit aussi les opérateurs de translation θ_p sur $E^{\mathbb{N}}$, définis, pour tout $u \in E^{\mathbb{N}}$ et tout $n \in \mathbb{N}$, par $[\theta_p(u)]_n = u_{n+p}$.

1. On note, pour tout $x, y \in E$, $G(x, y) = E_x [T_y \mathbf{1}_{(T_y < +\infty)}]$. Démontrer que l'on a

$$G(x, y) = P_x(T_y < +\infty) + \sum_{z \neq y} M(x, z) G(z, y). \quad (16.93)$$

2. Soient C et D deux parties disjointes de E; on note

$$\phi_{C,D}(x) = P_x[(T_C < +\infty) \cap (\tau_D(\theta_{T_C}(X)) = +\infty)]$$

la P_x -probabilité de l'ensemble des trajectoires qui atteignent C en un temps fini et qui ne passent pas dans D après être passées dans C. En utilisant un conditionnement par rapport à \mathcal{A}_1 , démontrer que $\phi_{C,D}$ est solution du système d'équations

$$\phi_{C,D}(x) = \sum_{y \in C} M(x, y) P_y(T_D = +\infty) + \sum_{y \notin C} M(x, y) \phi_{C,D}(y) \quad \forall x \in E. \quad (16.94)$$

Application. Le procédé de fabrication d'une pièce nécessite trois étapes successives, notées 1, 2, 3. Après l'étape i , la pièce est testée; si elle est bonne (ce qui se produit avec la probabilité r_i), elle franchit l'étape $i + 1$; si elle est irrémédiablement défectueuse (ce qui se produit avec la probabilité p_i), elle est jetée (c'est l'état 5); si elle est seulement légèrement défectueuse (ce qui se produit avec la probabilité q_i), elle re franchit l'étape i , l'état 4 étant celui d'une bonne pièce ayant franchi les trois étapes de fabrication avec succès. On suppose que $p_i + q_i + r_i = 1$, pour tout $i = 1, 2, 3$.

On modélise ce procédé de fabrication par une chaîne de Markov homogène X à valeurs dans $E = \{1, 2, 3, 4, 5\}$ (X_n représentant l'état de la pièce à l'étape n) de matrice de transition M, où

$$M = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} q_1 & r_1 & 0 & 0 & p_1 \\ 0 & q_2 & r_2 & 0 & p_2 \\ 0 & 0 & q_3 & r_3 & p_3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}.$$

On garde les notations introduites aux questions précédentes.

3. Déterminer le graphe associé à cette chaîne et spécifier les classes de communication. Spécifier la nature de ces classes.

4. Justifier l'égalité $F(5, 4) = 0$. Si K est le vecteur de \mathbb{R}^3 de composantes $F(i, 4)$, $i = 1, 2, 3$, démontrer que K est solution d'une équation de la forme $K = b + TK$, où $b \in \mathbb{R}^3$ et T est une matrice 3×3 à préciser. En déduire la valeur de K. Déterminer la probabilité $P_1(T_4 < +\infty)$.

5. Écrire la matrice potentiel R de X.

6. Soit, pour tout $x \in E$, $H(x) = E_x[T_4 \mathbf{1}_{(T_4 < +\infty)}]$. Démontrer que $H(5) = 0$; calculer $H(x)$ pour $x = 1, 2, 3$, et interpréter $H(1)$.

7. On pose $C = \{2\}$ et $D = \{4\}$. Calculer, pour tout $x \in E$, $\phi(x) \equiv \phi_{C,D}(x)$; interpréter $\phi(1)$.

Solution.

1. Cette question est la même que la sixième de l'exercice 6, à une ligne de calcul supplémentaire près; nous y renvoyons donc (les calculs sont valides, que E soit

fini ou non). L'égalité (16.92) s'écrit, avec les notations de ce présent exercice, et en regroupant différemment les termes,

$$G(x, y) = M(x, y) + \sum_{z \neq y} M(x, z) P_z(T_y < +\infty) + \sum_{z \neq y} M(x, z) G(z, y);$$

la proposition 16.25, et plus précisément l'égalité (16.48) donne alors l'égalité (16.93).

2. On a

$$\begin{aligned} \phi_{C,D}(x) &= E_x \left[\mathbf{1}_{(X_1 \in C)} \mathbf{1}_{(\tau_D(\theta_1(X)) = +\infty)} \right] \\ &\quad + E_x \left[\mathbf{1}_{(X_1 \notin C)} \mathbf{1}_{(\tau_C(\theta_1(X)) < +\infty)} \mathbf{1}_{(\tau_D(\theta_{\tau_C(\theta_1(X))}(X)) = +\infty)} \right]. \end{aligned}$$

soit, puisque X_1 est \mathcal{A}_1 -mesurable,

$$\begin{aligned} \phi_{C,D}(x) &= E_x \left[\mathbf{1}_{(X_1 \in C)} E_x^{\mathcal{A}_1} (\mathbf{1}_{(\tau_D(\theta_1(X)) = +\infty)}) \right] \\ &\quad + E_x \left[\mathbf{1}_{(X_1 \notin C)} E_x^{\mathcal{A}_1} (\mathbf{1}_{(\tau_C(\theta_1(X)) < +\infty)} \mathbf{1}_{(\tau_D(\theta_{\tau_C(\theta_1(X))}(X)) = +\infty)}) \right]. \end{aligned}$$

et, par la propriété de Markov simple,

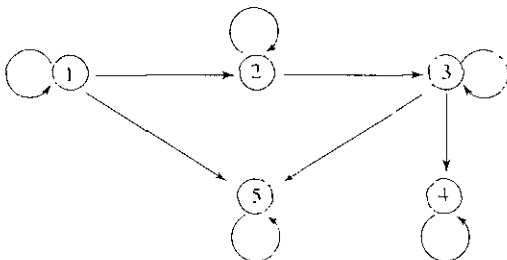
$$\begin{aligned} \phi_{C,D}(x) &= E_x \left[\mathbf{1}_{(X_1 \in C)} E_{X_1} (\mathbf{1}_{(T_D = +\infty)}) \right] \\ &\quad + E_x \left[\mathbf{1}_{(X_1 \notin C)} E_{X_1} (\mathbf{1}_{(T_C < +\infty)} \mathbf{1}_{(\tau_D(\theta_{T_C}(X)) = +\infty)}) \right]. \end{aligned}$$

soit encore

$$\begin{aligned} \phi_{C,D}(x) &= \sum_{y \in C} M(x, y) P_y(T_D = +\infty) \\ &\quad + \sum_{y \notin C} M(x, y) P_y \left[(T_C < +\infty) \cap (\tau_D(\theta_{T_C}(X)) = +\infty) \right], \end{aligned}$$

ce qui est l'égalité (16.94).

3. Le graphe associé à cette chaîne est



ce qui montre qu'aucun point ne communique avec aucun autre : ainsi, il y a cinq classes de communication constituées des singletons $\{i\}$, $1 \leq i \leq 5$. Les points 1, 2, 3 sont transitoires, les points 4, 5 absorbants (puisque les classes associées sont fermées).

4. Les points 4 et 5 ne conduisent pas l'un à l'autre ; en particulier, on a $F(5, 4) = 0$. D'après la proposition 16.25, et plus précisément l'égalité (16.48), on a, si $1 \leq i \leq 5$,

$$F(i, 4) = M(i, 4) + \sum_{j \neq 4} M(i, j)F(j, 4),$$

soit, puisque $F(5, 4) = 0$,

$$F(i, 4) = M(i, 4) + \sum_{j=1}^3 M(i, j)F(j, 4).$$

Ainsi, on a

$$K = b + TK \quad \text{où} \quad b = \begin{pmatrix} 0 \\ 0 \\ r_3 \end{pmatrix} \quad \text{et} \quad T = \begin{pmatrix} q_1 & r_1 & 0 \\ 0 & q_2 & r_2 \\ 0 & 0 & q_3 \end{pmatrix},$$

équation équivalente à

$$\begin{pmatrix} 1 - q_1 & -r_1 & 0 \\ 0 & 1 - q_2 & -r_2 \\ 0 & 0 & 1 - q_3 \end{pmatrix} K = \begin{pmatrix} 0 \\ 0 \\ r_3 \end{pmatrix} \quad \text{ou} \quad \begin{cases} (1 - q_1)k_1 - r_1 k_2 = 0 \\ (1 - q_2)k_2 - r_2 k_3 = 0 \\ (1 - q_3)k_3 - r_3 = 0 \end{cases}.$$

ce qui donne

$$k_3 = \frac{r_3}{1 - q_3} \quad k_2 = \frac{r_2 r_3}{(1 - q_2)(1 - q_3)} \quad k_1 = \frac{r_1 r_2 r_3}{(1 - q_1)(1 - q_2)(1 - q_3)}.$$

Il en résulte que

$$P_1(T_4 < +\infty) = \frac{r_1 r_2 r_3}{(1 - q_1)(1 - q_2)(1 - q_3)}.$$

5. Notant $\text{Tr} = \{1, 2, 3\}$ l'ensemble ordonné des points transitoires et $A = \{4, 5\}$ l'ensemble ordonné des points absorbants (donc récurrents positifs, puisque E est fini), la matrice potentiel R a la structure de blocs

$$R = \begin{matrix} & \text{Tr} & A \\ \begin{matrix} \text{Tr} \\ A \end{matrix} & \begin{pmatrix} (I - T)^{-1} & +\infty \\ \mathbf{0} & +\infty \end{pmatrix} \end{matrix}.$$

où T est la matrice introduite à la question précédente. La matrice $I - T$ étant triangulaire, s'inverse par exemple en résolvant le système associé, à savoir

$$\begin{cases} (1 - q_1)x_1 - r_1 x_2 = a \\ (1 - q_2)x_2 - r_2 x_3 = b \\ (1 - q_3)x_3 = c \end{cases}$$

ce qui donne

$$\begin{aligned} x_3 &= \frac{1}{1-q_3}c & x_2 &= \frac{1}{1-q_2} \left[b + \frac{r_2}{1-q_3}c \right] \\ x_1 &= \frac{1}{1-q_1} \left[a + \frac{r_1}{1-q_2}b + \frac{r_1r_2}{(1-q_2)(1-q_3)}c \right]. \end{aligned}$$

soit, pour l'inverse cherché,

$$(I-T)^{-1} = \begin{pmatrix} \frac{1}{1-q_1} & \frac{r_1}{(1-q_1)(1-q_2)} & \frac{r_1r_2}{(1-q_1)(1-q_2)(1-q_3)} \\ 0 & \frac{1}{1-q_2} & \frac{r_2}{(1-q_2)(1-q_3)} \\ 0 & 0 & \frac{1}{1-q_3} \end{pmatrix}.$$

6. Puisque $P_5(T_4 < +\infty) = 0$, on a $H(5) = 0$. Par ailleurs, avec les notations du début, on a $H(x) = G(x, 4)$; il résulte alors de (16.93) que l'on a, si $1 \leq x \leq 5$,

$$H(x) = P_x(T_4 < +\infty) + \sum_{z \neq 4} M(x, z)H(z),$$

soit, sous forme vectorielle, puisque $H(5) = 0$,

$$\hat{H} = K + T\hat{H},$$

où \hat{H} est le vecteur de \mathbb{R}^3 de composantes $H(i)$, $i = 1, 2, 3$. On a donc $(I-T)^{-1}\hat{H} = K$; en utilisant les résultats de la question précédente, un calcul matriciel montre alors que les composantes de \hat{H} sont données par

$$\begin{aligned} \hat{H}_1 &= \frac{r_1r_2r_3}{(1-q_1)^2(1-q_2)(1-q_3)} + \frac{r_1r_2r_3}{(1-q_1)(1-q_2)^2(1-q_3)} \\ &\quad + \frac{r_1r_2r_3}{(1-q_1)(1-q_2)(1-q_3)^2}, \\ \hat{H}_2 &= \frac{r_2r_3}{(1-q_2)^2(1-q_3)} + \frac{r_2r_3}{(1-q_2)(1-q_3)^3}, \\ \hat{H}_3 &= \frac{r_3}{(1-q_3)^2}. \end{aligned}$$

Le temps moyen pour fabriquer une bonne pièce est $H(1) = E_1[T_4 I_{(T_4 < +\infty)}]$ et vaut donc

$$H(1) = \frac{r_1r_2r_3}{(1-q_1)(1-q_2)(1-q_3)} \left[\frac{1}{1-q_1} + \frac{1}{1-q_2} + \frac{1}{1-q_3} \right].$$

7. Si $C = \{2\}$ et $D = \{4\}$, $\phi\{1\}$ est la **probabilité qu'une pièce franchisse le deuxième stade de fabrication et ne soit jamais bonne**. Il résulte de (16.94) que

$$\phi = P_2(T_4 = +\infty) \begin{pmatrix} r_1 \\ q_2 \\ 0 \\ 0 \\ 0 \end{pmatrix} + Q\phi \quad (16.95)$$

où Q est la matrice M dans laquelle on a annulé la seconde colonne. Puisque $P_x(T_2 < +\infty) = 0$ si $x = 3, 4, 5$, il résulte de la définition de ϕ que $\phi(x) = 0$ si $x = 3, 4, 5$. L'équation (16.95) donne alors

$$\begin{cases} (1 - q_1)\phi(1) = r_1 P_2(T_4 = +\infty) \\ \phi(2) = q_2 P_2(T_4 = +\infty); \end{cases}$$

ainsi, puisqu'il résulte de la question 4 que

$$P_2(T_4 = +\infty) = 1 - K(2) = 1 - \frac{r_2 r_3}{(1 - q_2)(1 - q_3)},$$

on a

$$\begin{cases} \phi(1) = \frac{r_1}{1 - q_1} \left[1 - \frac{r_2 r_3}{(1 - q_2)(1 - q_3)} \right], \\ \phi(2) = q_2 \left[1 - \frac{r_2 r_3}{(1 - q_2)(1 - q_3)} \right]. \end{cases}$$

Exercice 16.8. Jeu de pile ou face et chaîne de Markov. On effectue une suite de jets d'une pièce non nécessairement équilibrée, et on s'intéresse aux résultats obtenus lors de deux jets consécutifs; en particulier, on étudie la variable aléatoire donnant le nombre de coups nécessaires pour obtenir pile, par exemple, lors de deux jets consécutifs.

Soit $(X_n)_{n \in \mathbb{N}}$ un processus de **Bernoulli**, c'est-à-dire une suite de variables aléatoires définies sur un espace probabilisé (Ω, \mathcal{A}, P) à valeurs 0 ou 1, indépendantes de même loi de Bernoulli de paramètre p , telles que $P(X_n = 1) = p$, $P(X_n = 0) = q$, où $p + q = 1$. Pour tout $n \in \mathbb{N}$, on définit la variable aléatoire $Y_n = (X_n, X_{n+1})$ et la tribu $\mathcal{A}_n = \sigma(X_j \mid 0 \leq j \leq n + 1)$. On note E l'ensemble des quatre points

$$\alpha_1 = (1, 0) \quad \alpha_2 = (1, 1) \quad \alpha_3 = (0, 0) \quad \alpha_4 = (0, 1).$$

1. Calculer, pour toute fonction réelle f sur E et tout $n \in \mathbb{N}$, l'espérance conditionnelle $E^{\mathcal{A}_n} [f(Y_{n+1})]$ et en déduire que le processus $Y = (Y_n)_{n \in \mathbb{N}}$ est une chaîne de Markov homogène à valeurs dans E de matrice de transition M à déterminer. Démontrer que l'on a, pour tout $n \geq 2$, $M^n = M^2$.

2. Déterminer le graphe associé à cette chaîne et spécifier les classes de communication. Déterminer la nature et la période des points de E .

Pour tout $x \in E$, on note P_x la probabilité $P_x = P(\cdot \mid Y_0 = x)$.

3. On étudie, sous la probabilité P_x , la loi du premier temps de passage T de la chaîne Y en α_2 , soit $T = \inf(n \in \mathbb{N}^* \mid Y_n = \alpha_2)$, avec $\inf \emptyset = +\infty$. Pour cela, on note, pour $x = 1, 2, 3, 4$, $f_k(i) = P_{\alpha_i}(T = k) \equiv F_k(\alpha_i, \alpha_2)$. Démontrer que la suite $(f_k)_{k \in \mathbb{N}^*}$ de vecteurs de \mathbb{R}^4 est solution d'une équation de récurrence d'ordre un; en déduire que la suite des probabilités $(f_k(1))_{k \in \mathbb{N}^*}$ est solution de l'équation de récurrence d'ordre deux

$$x_k = qx_{k-1} + pqx_{k-2} \quad , \quad (16.96)$$

pour des conditions initiales à spécifier. Déterminer alors deux solutions particulières de la forme $(\lambda^k)_{k \geq 2}$ pour deux valeurs de λ , λ_1 et λ_2 , exprimées en fonction de p et q et en déduire la valeur de la probabilité $f_k(1)$ en fonction de λ_1 et λ_2 .

4. Calculer le temps moyen $E_{\alpha_1}(T)$ (l'exprimer uniquement en fonction de p).
5. Justifier l'existence d'une unique probabilité invariante ν et la calculer; en déduire la valeur du temps moyen $E_{\alpha_2}(T)$ et le comparer à $E_{\alpha_1}(T)$.

Solution.

1. Pour toute fonction f sur E et tout $(x_0, \dots, x_{n+1}) \in \{0, 1\}^{n+2}$, on a

$$\begin{aligned} E^{(X_0, \dots, X_{n+1})=(x_0, \dots, x_{n+1})} [f(Y_{n+1})] \\ &= E^{(X_0, \dots, X_{n+1})=(x_0, \dots, x_{n+1})} [f(X_{n+1}, X_{n+2})] \\ &= E^{(X_0, \dots, X_{n+1})=(x_0, \dots, x_{n+1})} [f(x_{n+1}, X_{n+2})], \end{aligned}$$

soit, puisque les variables aléatoires (X_0, \dots, X_{n+1}) et $f(x_{n+1}, X_{n+2})$ sont indépendantes,

$$\begin{aligned} E^{(X_0, \dots, X_{n+1})=(x_0, \dots, x_{n+1})} [f(Y_{n+1})] &= E[f(x_{n+1}, X_{n+2})] \\ &= pf(x_{n+1}, 1) + qf(x_{n+1}, 0). \end{aligned}$$

Il en résulte que

$$E^{A_n} [f(Y_{n+1})] = M(Y_n, f),$$

où, $M(\cdot, f)$ est définie par, pour tout $y \in E$,

$$M(y, f) = pf(\pi_2(y), 1) + qf(\pi_2(y), 0),$$

π_2 désignant la deuxième projection de $E = \{0, 1\}^2$ sur $\{0, 1\}$ définie par $\pi_2(y_1, y_2) = y_2$. Ceci démontre que Y est une chaîne de Markov homogène de matrice de transition M , d'entrées $M(\alpha_i, \alpha_j) = M(\alpha_i, \mathbf{1}_{\{\alpha_j\}})$. On a donc

$$M = \begin{matrix} & \begin{matrix} .1 & .2 & .3 & .4 \end{matrix} \\ \begin{matrix} .1 \\ .2 \\ .3 \\ .4 \end{matrix} & \begin{pmatrix} 0 & 0 & q & p \\ q & p & 0 & 0 \\ 0 & 0 & q & p \\ q & p & 0 & 0 \end{pmatrix} \end{matrix}.$$

On a

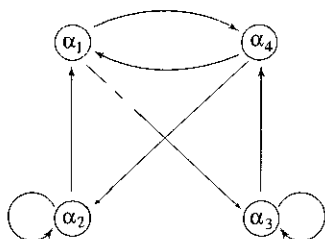
$$M^2 = \begin{pmatrix} pq & p^2 & q^2 & pq \\ pq & p^2 & q^2 & pq \\ pq & p^2 & q^2 & pq \\ pq & p^2 & q^2 & pq \end{pmatrix};$$

si on suppose que, pour $n \geq 2$, on a $M^n = M^2$, on a alors

$$M^{n+1} = \begin{pmatrix} p^2q + pq^2 & p^3 + p^2q & pq^2 + q^3 & p^2q + pq^2 \\ p^2q + pq^2 & p^3 + p^2q & pq^2 + q^3 & p^2q + pq^2 \\ p^2q + pq^2 & p^3 + p^2q & pq^2 + q^3 & p^2q + pq^2 \\ p^2q + pq^2 & p^3 + p^2q & pq^2 + q^3 & p^2q + pq^2 \end{pmatrix},$$

ce qui, en tenant compte de l'égalité $p + q = 1$, démontre que $M^{n+1} = M^2$; on a donc bien pour tout $n \geq 2$, $M^n = M^2$.

2. Le graphe associé à cette chaîne est



ce qui montre que tous les points communiquent; ainsi, il y a une seule classe de communication et la chaîne est **irréductible**, et donc récurrente positive, puisque E est fini. Par ailleurs, puisque $M(\alpha_2, \alpha_2) > 0$, α_2 est **apériodique**: il en est de même de la chaîne Y .

3. D'après la proposition 16.25, et plus précisément l'égalité (16.47), on a, pour tout $x \in E$,

$$\begin{cases} f_1(x) = M(x, \alpha_2), \\ f_k(x) = \sum_{z \in E \setminus \{\alpha_2\}} M(x, z) f_{k-1}(z) & \text{si } k \geq 2. \end{cases}$$

Il en résulte que la suite des vecteurs f_k est solution de l'équation $f_k = Q f_{k-1}$, $k \geq 2$, où f_1 est la deuxième colonne de M , et Q est la matrice M dont on a annulé la deuxième colonne, c'est-à-dire

$$f_1 = \begin{pmatrix} 0 \\ p \\ 0 \\ p \end{pmatrix} \quad \text{et} \quad Q = \begin{pmatrix} 0 & 0 & q & p \\ q & 0 & 0 & 0 \\ 0 & 0 & q & p \\ q & 0 & 0 & 0 \end{pmatrix}.$$

Autrement dit, les composantes de la suite $(f_k)_{k \in \mathbb{N}^*}$ sont solutions du système

$$\begin{cases} f_k(1) = q f_{k-1}(3) + p f_{k-1}(4) \\ f_k(2) = q f_{k-1}(1) \\ f_k(3) = q f_{k-1}(3) + p f_{k-1}(4) \\ f_k(4) = q f_{k-1}(1), \end{cases}$$

système équivalent au système

$$\begin{cases} f_k(1) = f_k(3) = q f_{k-1}(3) + p f_{k-1}(4) \\ f_k(2) = f_k(4) = q f_{k-1}(1) \end{cases},$$

ce qui implique que, si $k \geq 3$,

$$f_k(1) = q f_{k-1}(1) + p q f_{k-2}(1);$$

ainsi la suite des probabilités $(f_k(1))_{k \in \mathbb{N}^*}$ est solution de l'équation de récurrence d'ordre deux

$$x_k = q x_{k-1} + p q x_{k-2} \quad k \geq 3, \quad (16.97)$$

pour des conditions initiales

$$\boxed{x_1 = 0, \quad x_2 = p^2,} \quad (16.98)$$

puisque l'on a

$$x_1 = f_1(1) = M(\alpha_1, \alpha_2) = 0,$$

et que, du fait de l'indépendance des variables aléatoires X_n , on a

$$\begin{aligned} x_2 = f_2(1) &= P_{\alpha_1}(Y_1 \neq \alpha_2, Y_2 = \alpha_2) = \frac{P(Y_0 = \alpha_1, Y_1 \neq \alpha_2, Y_2 = \alpha_2)}{P(Y_0 = \alpha_1)} \\ &= \frac{P(X_0 = 1, X_1 = 0, X_2 = 1, X_3 = 1)}{P(X_0 = 1, X_1 = 0)} = p^2. \end{aligned}$$

L'équation caractéristique associée à (16.97) est

$$\lambda^2 - q\lambda - pq = 0;$$

son discriminant est $\Delta = q^2 + 4pq = q(1 + 3p)$ et elle admet les deux racines

$$\lambda_1 = \frac{q - \sqrt{q(1 + 3p)}}{2} \quad \text{et} \quad \lambda_2 = \frac{q + \sqrt{q(1 + 3p)}}{2}.$$

La solution générale de (16.97) est donc de la forme $x_k = \beta_1 \lambda_1^k + \beta_2 \lambda_2^k$ et la suite $(f_k(1))_{k \in \mathbb{N}^*}$ est la solution de conditions initiales données par (16.98); elle correspond aux constantes β_1 et β_2 qui vérifie

$$\begin{cases} \beta_1 \lambda_1 + \beta_2 \lambda_2 = 0 \\ \beta_1 \lambda_1^2 + \beta_2 \lambda_2^2 = p^2 \end{cases} \iff \begin{cases} \beta_1 = \frac{p^2}{\lambda_1(\lambda_1 - \lambda_2)} \\ \beta_2 = \frac{p^2}{\lambda_2(\lambda_2 - \lambda_1)} \end{cases},$$

ce qui donne, après avoir remarqué que la formule obtenue est valable pour $k = 2$,

$$\boxed{P_{\alpha_1}(T = k) = f_k(1) = \begin{cases} 0 & \text{si } k = 1, \\ \frac{p^2}{\lambda_2 - \lambda_1} [\lambda_2^{k-1} - \lambda_1^{k-1}] & \text{si } k \geq 2. \end{cases}}$$

4. On a alors

$$E_{\alpha_1}(T) = \sum_{k=2}^{+\infty} k P_{\alpha_1}(T = k) = \frac{p^2}{\lambda_2 - \lambda_1} \sum_{k=2}^{+\infty} k [\lambda_2^{k-1} - \lambda_1^{k-1}].$$

Par un calcul classique de dérivation de séries entières, on a si $|x| < 1$,

$$\sum_{k=2}^{+\infty} k x^{k-1} = \left(\sum_{k=0}^{+\infty} x^k \right)' - 1 = \frac{1}{(1-x)^2} - 1 = \frac{x(2-x)}{(1-x)^2};$$

un calcul facile conduit alors, en tenant compte des valeurs de λ_1 et λ_2 (on a bien $|\lambda_1| < 1$ et $|\lambda_2| < 1$), à

$$E_{\alpha_1}(T) = \frac{2-q}{p^2} = \frac{1+p}{p^2}.$$

Remarque. On donne, à titre indicatif, un tableau de valeurs numériques suivant certaines valeurs de p :

p	0.1	0.5	0.8	0.9
$E_{\alpha_1}(T)$	110	6	2.81	2.34

5. La chaîne Y étant irréductible et récurrente positive, il existe une unique probabilité invariante ν . Si ν est une **mesure** invariante, on a en particulier $\nu M^2 = \nu$, soit, pour tout $y \in E$,

$$\sum_{j=1}^4 \nu(\alpha_j) M^2(\alpha_j, y) = \nu(y);$$

mais, ici, $M^2(\alpha_j, y)$ est indépendant de j . Il en résulte que, pour tout $y \in E$,

$$M^2(\alpha_j, y) = \frac{\nu(y)}{\sum_{j=1}^4 \nu(\alpha_j)}.$$

Ainsi, $\nu = (pq, p^2, q^2, pq)$ est la **probabilité** invariante. Il en résulte que

$$E_{\alpha_2}(T) = \frac{1}{\nu(\alpha_2)} = \frac{1}{p^2}.$$

Il est clair que $E_{\alpha_2}(T) < E_{\alpha_1}(T)$.

Remarque. On donne, à titre indicatif, un tableau de valeurs numériques suivant certaines valeurs de p :

p	0.1	0.5	0.8	0.9
$E_{\alpha_2}(T)$	100	4	1.56	1.23

Exercice 16.9. Marche aléatoire sur l'intervalle d'entiers $\{0, 1, 2, \dots, N\}$ avec barrières réfléchissantes; probabilité invariante. Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans $E = \{0, 1, 2, \dots, N\}$ qui, pour tout $x \in E$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, \mathbf{P}_x)$, de matrice de transition M donnée par,

$$\begin{cases} M(x, x+1) = p & M(x, x-1) = q, & \text{si } 1 \leq x \leq N-1, \\ M(0, 1) = 1, & M(N, N-1) = 1, \end{cases}$$

où $p, q \in]0, 1[$ et $p + q = 1$. Pour tout $x \in E$, on note T_x le temps d'entrée en x , c'est-à-dire $T_x = \inf(n \in \mathbb{N}^* \mid X_n = x)$, avec $\inf \emptyset = +\infty$.

1. Spécifier la ou les classes de communication.
2. Justifier l'existence d'une unique probabilité invariante ν et la calculer en fonction de p, q, N ; en déduire la valeur du temps moyen $E_0(T_0)$.

Solution.

1. Le graphe associé à cette chaîne est



Tous les points communiquent; il n'y a qu'une classe de communication et la chaîne est irréductible. Puisque E est fini, cette chaîne est donc récurrente positive.

2. Il en résulte qu'il existe une unique probabilité invariante ν . Écrivons que c'est une mesure invariante; elle vérifie, pour tout $x \in E$,

$$\nu(x) = \sum_{z \in E} \nu(z)M(z, x),$$

ce qui, compte tenu de la valeur de M , est équivalent au système

$$\begin{cases} \nu(x) = p\nu(x-1) + q\nu(x+1) & \text{si } 2 \leq x \leq N-2, \\ \nu(1) = \nu(0) + q\nu(2) \\ \nu(0) = q\nu(1) \\ \nu(N) = p\nu(N-1) \\ \nu(N-1) = p\nu(N-2) + \nu(N). \end{cases}$$

On a alors, si $2 \leq y \leq N-2$,

$$\nu(y+1) - \nu(y) = \frac{p}{q} [\nu(y) - \nu(y-1)];$$

en sommant en y , on a donc, si $2 \leq x \leq N-2$,

$$\nu(x+1) - \nu(2) = \frac{p}{q} \sum_{y=2}^x [\nu(y) - \nu(y-1)] = \frac{p}{q} [\nu(x) - \nu(1)],$$

soit

$$\nu(x+1) = \frac{1}{q} \left(\frac{1}{q} - 1 \right) \nu(0) + \frac{p}{q} \nu(x) - \frac{p}{q^2} \nu(0),$$

ce qui se simplifie en

$$\nu(x+1) = \frac{p}{q} \nu(x) \quad \text{si } 2 \leq x \leq N-2. \quad (16.99)$$

De plus, on a

$$q\nu(2) = \nu(1) - \nu(0) = \nu(1) - q\nu(1) = p\nu(1),$$

ce qui montre que l'égalité (16.99) est encore vraie pour $x = 1$. Il résulte alors de l'égalité (16.99) que l'on a (attention au décalage d'indices), si $2 \leq x \leq N-1$,

$$v(x) = \left(\frac{p}{q}\right)^{x-1} v(1). \quad (16.100)$$

Calculons la masse de cette mesure ; deux cas se présentent :

• si $p \neq q$, on a

$$\begin{aligned} \sum_{x=0}^N v(x) &= v(0) + \frac{1}{q} v(0) + \sum_{x=2}^{N-1} \left[\left(\frac{p}{q}\right)^{x-1} \frac{1}{q} v(0) \right] + \left(\frac{p}{q}\right)^{N-1} v(0) \\ &= v(0) \left[1 + \frac{1}{q} + \frac{1}{q} \frac{p}{q} \frac{1 - \left(\frac{p}{q}\right)^{N-2}}{1 - \frac{p}{q}} + \left(\frac{p}{q}\right)^{N-1} \right] \\ &= v(0) \left[1 + \frac{1}{q} \left(1 + \frac{p}{q-p} \right) + \left(\frac{p}{q}\right)^{N-1} \left(1 - \frac{1}{q-p} \right) \right] \\ &= v(0) \left[1 + \frac{1}{q-p} - \frac{2p}{q-p} \left(\frac{p}{q}\right)^{N-1} \right], \end{aligned}$$

soit encore

$$\sum_{x=0}^N v(x) = v(0) \frac{2q}{q-p} \left[1 - \left(\frac{p}{q}\right)^N \right].$$

Pour obtenir l'unique probabilité invariante, il suffit alors de normaliser cette mesure ; elle est donc donnée, compte tenu de ce qui précède, par

$$\left\{ \begin{array}{l} v(0) = \frac{1 - \frac{p}{q}}{2 \left[1 - \left(\frac{p}{q}\right)^N \right]}, \quad v(1) = \frac{1 - \frac{p}{q}}{2q \left[1 - \left(\frac{p}{q}\right)^N \right]}, \\ v(x) = \frac{1}{2q} \frac{1 - \frac{p}{q}}{1 - \left(\frac{p}{q}\right)^N} \left(\frac{p}{q}\right)^{x-1} \quad \text{si } 2 \leq x \leq N-1, \\ v(N) = \frac{1}{2} \frac{1 - \frac{p}{q}}{1 - \left(\frac{p}{q}\right)^N} \left(\frac{p}{q}\right)^{N-1}. \end{array} \right.$$

Le temps moyen de retour à 0, $E_0(T_0)$, vaut alors $\frac{1}{v(0)}$, soit

$$E_0(T_0) = \frac{2 \left[1 - \left(\frac{p}{q}\right)^N \right]}{1 - \frac{p}{q}}.$$

• si $p = q = \frac{1}{2}$, l'égalité (16.99) étant vraie dès que $1 \leq x \leq N-2$, on a

$$v(x) = v(1), \quad \text{si } 2 \leq x \leq N-1.$$

Puisque de plus on a, dans ce cas,

$$v(0) = \frac{1}{2} v(1) \quad \text{et} \quad v(N) = \frac{1}{2} v(N-1),$$

on a en fait

$$v(0) = \frac{1}{2} v(1) \quad \text{et} \quad v(N) = \frac{1}{2} v(N-1) = \frac{1}{2} v(1);$$

la masse de v est donc $v(E) = (N-1)v(1)$. Il en résulte que l'unique probabilité invariante est donnée par

$$v(x) = \begin{cases} \frac{1}{N-1} & \text{si } 2 \leq x \leq N-1. \\ \frac{1}{2(N-1)} & \text{si } x = 0 \text{ ou } 2. \end{cases}$$

Le temps moyen de retour à 0, $E_0(T_0)$, vaut alors $\frac{1}{v(0)}$, soit

$$E_0(T_0) = 2(N-1).$$

Exercice 16.10. Marche aléatoire sur \mathbb{N} avec barrière de type quelconque ; mesure et probabilité invariante, probabilité limite. Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans \mathbb{N} qui, pour tout $x \in \mathbb{N}$, est une chaîne de Markov homogène sur la base de processus $(\Omega, \mathcal{A}, (\mathcal{A}_n)_{n \in \mathbb{N}}, P_x)$, de matrice de transition M donnée par,

$$\begin{cases} M(x, x+1) = p & M(x, x-1) = q, & \text{si } x \in \mathbb{N}^*, \\ M(0, 0) = \alpha, & M(0, 1) = 1 - \alpha, \end{cases}$$

où $p, q \in]0, 1[$, $p + q = 1$, et $\alpha \in [0, 1]$. Si $\alpha = 0$, le point 0 est appelé barrière réfléchissante, si $\alpha \in]0, 1[$, le point 0 est appelé barrière élastique, si $\alpha = 1$, le point 0 est appelé barrière absorbante. Pour tout $x \in \mathbb{N}$, on note T_x le temps d'entrée en x , c'est-à-dire $T_x = \inf(n \in \mathbb{N}^* \mid X_n = x)$, avec $\inf \emptyset = +\infty$.

1. Spécifier la ou les classes de communication. Étudier la périodicité des points.

On étudie d'abord le cas où $\alpha \in [0, 1]$.

2. Démontrer, par le calcul, l'existence d'une mesure invariante v . Étudier, suivant les valeurs respectives de p et q , le problème d'existence et d'unicité d'une probabilité invariante et la calculer dans le cas d'existence et d'unicité. En déduire, dans le cas où $\alpha \in [0, 1[$ et $p < q$, la nature des points de \mathbb{N} , et donner, pour tout $x \in \mathbb{N}$, la valeur $E_x(T_x)$ du temps moyen de retour à x .

3. Dans le cas où $p \geq q$, étudier la nature des points de \mathbb{N} et si $p > q$, calculer, pour tout $x \in \mathbb{N}^*$, la probabilité $P_x(T_0 = +\infty)$.

4. Dans le cas où $p < q$, justifier la P_x -p.s. convergence de la suite de terme général $\frac{1}{n} \sum_{j=1}^n \exp(-aX_j)$, pour $a > 0$ quelconque.

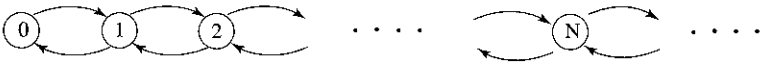
On étudie le cas où $\alpha = 1$.

5. Calculer $P_0(T_0 < +\infty)$ et $E_0(T_0)$; en déduire la nature du point 0.
6. Calculer, pour tout $x \in \mathbb{N}^*$, la probabilité $P_x(T_0 < +\infty)$. Déterminer la nature des points de \mathbb{N}^* .
7. Étudier la convergence de la suite de terme général $M^n(x, y)$ (et préciser, s'il y a lieu, sa limite), lorsque $x \in \mathbb{N}$ et $y \in \mathbb{N}^*$, puis lorsque $x \in \mathbb{N}^*$ et $y = 0$.

Solution.

1. On étudie le graphe associé à cette chaîne.

- Si $\alpha \in [0, 1[$, le graphe associé à cette chaîne est



Tous les points communiquent; il n'y a qu'une classe de communication et la chaîne est **irréductible**. Puisque \mathbb{N} est infini, on ne peut a priori pas dire la nature de cette chaîne; c'est en particulier l'objet de la suite du problème.

- Si $\alpha = 1$, le graphe associé à cette chaîne est



Tous les points de \mathbb{N}^* communiquent, mais 0 et 1 ne communiquent pas. En effet, $M(0, 0) = 1$ et, si on suppose que $M^n(0, 0) = 1$, on a, puisque $M(z, 0) = 0$ dès que $z \geq 2$,

$$\begin{aligned} M^{n+1}(0, 0) &= \sum_{z \in \mathbb{N}} M^n(0, z)M(z, 0) \\ &= M^n(0, 0)M(0, 0) + M^n(0, 1)M(1, 0) = 1, \end{aligned}$$

ce qui démontre que $M^n(0, 0) = 1$ pour tout $n \in \mathbb{N}^*$, et donc que 0 ne conduit pas à 1. Il y a, dans ce cas, deux classes de communication, $\{0\}$ et \mathbb{N}^* .

Pour ce qui est de la périodicité,

- Si $\alpha = 0$, on a $M(0, 0) = 0$ et, par une récurrence facile, pour tout $n \in \mathbb{N}^*$, $M^{2n}(0, 0) > 0$ et $M^{2n+1}(0, 0) = 0$; 0 est de période 2 et donc aussi tous les points de \mathbb{N} .

- Si $0 < \alpha < 1$, on a $M(0, 0) > 0$, ce qui implique que 0 est apériodique; il en est de même de tous les points de \mathbb{N} .

- Si $\alpha = 1$, on a $M(0, 0) > 0$, ce qui implique que la classe réduite au point 0 est de apériodique. Par ailleurs, on a $M(1, 1) = 0$ et, par une récurrence facile, pour tout $n \in \mathbb{N}^*$, $M^{2n}(1, 1) > 0$ et $M^{2n+1}(1, 1) = 0$; 1 est de période 2 et donc aussi tous les points de la classe \mathbb{N}^* .

2. Une mesure invariante ν vérifie, pour tout $x \in \mathbb{N}$,

$$\nu(x) = \sum_{z \in \mathbb{N}} \nu(z) M(z, x),$$

ce qui, compte tenu de la valeur de M , est équivalent au système

$$\begin{cases} \nu(x) = p\nu(x-1) + q\nu(x+1) & \text{si } 2 \leq x. \\ \nu(1) = (1-\alpha)\nu(0) + q\nu(2) \\ \nu(0) = \alpha\nu(0) + q\nu(1). \end{cases} \quad (16.101)$$

On a alors

$$\nu(1) = \frac{1-\alpha}{q} \nu(0),$$

et donc

$$q\nu(2) = \frac{1-\alpha}{q} \nu(0) - (1-\alpha)\nu(0),$$

ce qui donne, après simplifications,

$$\nu(2) = (1-\alpha) \frac{p}{q^2} \nu(0).$$

Il en résulte, en reportant dans le système (16.101) que

$$q\nu(3) = (1-\alpha)p\nu(0) \left[\frac{1}{q^2} - \frac{1}{q} \right] = (1-\alpha) \frac{p^2}{q^2} \nu(0),$$

et donc,

$$\nu(3) = (1-\alpha) \frac{p^2}{q^3} \nu(0).$$

Si on suppose alors que l'on a, pour tout i tel que $1 \leq i \leq j$,

$$\nu(i) = \frac{1-\alpha}{q} \left(\frac{p}{q} \right)^{i-1} \nu(0), \quad (16.102)$$

on a, par un calcul identique,

$$\begin{aligned} q\nu(j+1) &= \frac{1-\alpha}{q} \left(\frac{p}{q} \right)^{j-1} \nu(0) - p \frac{1-\alpha}{q} \left(\frac{p}{q} \right)^{j-2} \nu(0) \\ &= \frac{1-\alpha}{q} \left(\frac{p}{q} \right)^{j-2} p \left[\frac{1}{q} - 1 \right] \nu(0) = (1-\alpha) \frac{p^j}{q^j} \nu(0), \end{aligned}$$

et donc,

$$\nu(j+1) = \frac{1-\alpha}{q} \left(\frac{p}{q} \right)^j \nu(0).$$

Il en résulte que (16.102) est vraie pour tout $j \in \mathbb{N}^*$. Il existe donc des mesures invariantes ν ; elles sont toutes proportionnelles et définies à un facteur multiplicatif près par l'égalité (16.102).

– Si $\alpha = 1$, la mesure de Dirac est l'unique probabilité invariante.

- Si $\alpha \in [0, 1[$, ces mesures invariantes n'ont une masse finie que si $p < q$. Dans ce cas, leur masse est donnée par

$$\begin{aligned} \nu(\mathbb{N}) &= \nu(0) + \sum_{j \in \mathbb{N}^*} \nu(j) = \nu(0) + \sum_{j \in \mathbb{N}^*} \frac{1-\alpha}{q} \left(\frac{p}{q}\right)^{j-1} \nu(0) \\ &= \left[1 + \frac{1-\alpha}{q} \frac{1}{1-\frac{p}{q}} \right] \nu(0). \end{aligned}$$

soit

$$\nu(\mathbb{N}) = \frac{2q-\alpha}{q-p} \nu(0);$$

l'unique probabilité invariante ν est alors donnée par

$$\nu(j) = \begin{cases} \frac{q-p}{2q-\alpha} & \text{si } j = 0, \\ \frac{q-p}{2q-\alpha} \frac{1-\alpha}{q} \left(\frac{p}{q}\right)^{j-1} & \text{si } j \in \mathbb{N}^*. \end{cases} \quad (16.103)$$

Ainsi, si $\alpha \in [0, 1[$ et si $p < q$, la chaîne est irréductible, récurrente positive et on a $E_j(T_j) = \frac{1}{\nu(j)}$, soit

$$E_j(T_j) = \begin{cases} \frac{2q-\alpha}{q-p} & \text{si } j = 0, \\ \frac{2q-\alpha}{1-\alpha} \frac{q}{q-p} \left(\frac{q}{p}\right)^{j-1} & \text{si } j \in \mathbb{N}^*. \end{cases}$$

3. Si $\alpha \in [0, 1[$ et si $p \geq q$, il n'existe pas de probabilité invariante, et la chaîne est irréductible transitoire ou récurrente nulle. Pour décider de la nature de cette chaîne, on applique le critère analytique de récurrence (corollaire 16.48) : la chaîne est récurrente si et seulement si le système

$$h = Qh \quad 0 \leq h \leq 1. \quad (16.104)$$

a pour unique solution $h = 0$, où Q est la restriction de M à $\mathbb{N}^* \times \mathbb{N}^*$. On a

$$Qh(x) = \sum_{y \in \mathbb{N}^*} Q(x, y)h(y) = \begin{cases} qh(x-1) + ph(x+1) & \text{si } x \geq 2, \\ ph(2) & \text{si } x = 1; \end{cases}$$

Le système (16.104) s'écrit donc

$$\begin{cases} h(1) = ph(2) \\ h(x) = qh(x-1) + ph(x+1) & \text{si } x \geq 2. \\ 0 \leq h(x) \leq 1; \end{cases}$$

la deuxième équation donne, si $x \geq 2$,

$$q[h(x) - h(x-1)] = p[h(x+1) - h(x)],$$

soit, par itération rétrograde,

$$h(x+1) - h(x) = \left(\frac{q}{p}\right)^{x-1} [h(2) - h(1)] = \left(\frac{q}{p}\right)^x h(1);$$

on remarque que cette égalité est encore vraie pour $x = 1$. Il en résulte que, par sommation, on a, pour tout $x \in \mathbb{N}^*$,

$$h(x+1) - h(1) = \sum_{j=1}^x \left(\frac{q}{p}\right)^j h(1).$$

ou encore

$$h(x+1) = h(1) \sum_{j=0}^x \left(\frac{q}{p}\right)^j. \quad (16.105)$$

- Si $p = q$, on a $h(x) = xh(1)$ et la seule solution au système (16.104) est $h \equiv 0$; dans ce cas, **la chaîne est récurrente nulle**.
- Si $p > q$, on a, pour tout $x \geq 2$,

$$h(x) = h(1) \frac{1 - \left(\frac{q}{p}\right)^x}{1 - \frac{q}{p}} = h(1) \frac{p}{p-q} \left[1 - \left(\frac{q}{p}\right)^x\right];$$

ainsi, si on prend $h(1) = \frac{p-q}{p} = 1 - \frac{q}{p} < 1$, la fonction h définie, pour tout $x \in \mathbb{N}^*$, par

$$h(x) = 1 - \left(\frac{q}{p}\right)^x$$

est solution non nulle du système (16.104); ainsi, dans ce cas, **la chaîne est transitoire**. La probabilité $P_x(T_0 = +\infty) = P_x[\bigcap_{j \in \mathbb{N}^*} (X_j \in \mathbb{N}^*)]$ vaut $f(x)$, où f est la solution maximale du système (16.104) (cf proposition 16.47); c'est la fonction h elle-même, puisqu'elle correspond à la plus grande valeur possible pour $h(1)$ relativement à la contrainte $0 \leq h \leq 1$. Ainsi, pour tout $x \in \mathbb{N}^*$, on a

$$P_x(T_0 = +\infty) = 1 - \left(\frac{q}{p}\right)^x.$$

4. Si $\alpha \in [0, 1[$ et si $p < q$, la chaîne est irréductible, récurrente positive et la loi forte des grands nombres pour les chaînes de Markov homogènes s'applique à toute fonction intégrable par rapport à la probabilité invariante ν , ce qui est le cas ici, puisque la fonction considérée est bornée. On a, d'après (16.103),

$$\begin{aligned} \int_{\mathbb{R}} \exp(-ax) d\nu(x) &= \frac{q-p}{2q-\alpha} + \frac{q-p}{2q-\alpha} \frac{1-\alpha}{q} \sum_{j \in \mathbb{N}^*} \exp(-aj) \left(\frac{p}{q}\right)^{j-1} \\ &= \frac{q-p}{2q-\alpha} \left[1 + \frac{1-\alpha}{q} \exp(-a) \sum_{j \in \mathbb{N}^*} \left[\frac{p}{q} \exp(-a)\right]^{j-1}\right] \\ &= \frac{q-p}{2q-\alpha} \left[1 + \frac{1-\alpha}{q} \frac{\exp(-a)}{1 - \frac{p}{q} \exp(-a)}\right]. \end{aligned}$$

soit

$$\int_{\mathbb{R}} \exp(-ax) d\nu(x) = \frac{q-p}{2q-\alpha} \left[1 + (1-\alpha) \frac{\exp(-a)}{q-p \exp(-a)}\right].$$

La loi forte des grands nombres démontre la P_x -p.s. convergence de la suite de terme général $\frac{1}{n} \sum_{j=1}^n \exp(-aX_j)$ et assure que l'on a

$$P_x\text{-p.s.} \quad \lim_n \frac{1}{n} \sum_{j=1}^n \exp(-aX_j) = \frac{q-p}{2q-\alpha} \left[1 + (1-\alpha) \frac{\exp(-a)}{q-p \exp(-a)} \right].$$

5. Si $\alpha = 1$, on a vu que la mesure de Dirac est l'unique probabilité invariante ; 0 est donc récurrent et on a alors $P_0(T_0 < +\infty) = 1$; un argument plus élémentaire est de dire que l'on a

$$P_0(T_0 < +\infty) \geq P_0(X_1 = 0) = M(0, 0) = 1.$$

On a donc $T_0 = 1$ P_0 -p.s., et par conséquent $E_0(T_0) = 1$; ainsi, le point 0 est récurrent positif.

6. La probabilité $P_x(T_0 = +\infty) = P_x[\bigcap_{j \in \mathbb{N}^*} (X_j \in \mathbb{N}^*)]$ vaut $f(x)$, où f est la solution maximale du système (16.104) (cf proposition 16.47) ; sa solution générale h est encore de la forme (16.105) :

- si $p \leq q$, et si $h(1) > 0$, on a alors $\lim_x h(x) = +\infty$; la seule solution du système (16.104) est donc la solution nulle, ce qui implique que, pour tout $x \in \mathbb{N}^*$, $P_x(T_0 < +\infty) = 1$.
- si $p > q$, la solution maximale du système (16.104) est la même que dans le cas $\alpha < 1$ et on a encore, pour tout $x \in \mathbb{N}^*$, $P_x(T_0 = +\infty) = 1 - \left(\frac{q}{p}\right)^x$, ou encore

$$P_x(T_0 < +\infty) = \left(\frac{q}{p}\right)^x.$$

On a vu que si $\alpha = 1$, 0 ne conduit pas à 1 ; le point 0 étant récurrent, 1 est transitoire, puisque, sinon, les points 0 et 1 communiquerait. Ainsi, **lorsque $\alpha = 1$, la classe \mathbb{N}^* est transitoire.**

7. Si $\alpha = 1$, on a donc, si $x \in \mathbb{N}$ et $y \in \mathbb{N}^*$, $\lim_n M^n(x, y) = 0$. Par ailleurs, 0 étant récurrent a périodique, on a si $x \in \mathbb{N}^*$,

$$\lim_n M^n(x, 0) = \frac{P_x(T_0 < +\infty)}{E_0(T_0)} = P_x(T_0 < +\infty) :$$

ainsi

$$\lim_n M^n(x, 0) = \begin{cases} \left(\frac{q}{p}\right)^x & \text{si } p > q, \\ 1 & \text{si } p \leq q. \end{cases}$$

Exercice 16.11. Processus de Galton-Watson et martingales ; évolution de la taille d'une population. Il s'agit d'étudier l'évolution de la taille des générations successives d'une population d'individus qui donnent chacun naissance à un nombre aléatoire de descendants suivant la même loi de probabilité μ ; en particulier, il s'agit d'évaluer la probabilité d'extinction de la population. Le modèle est le suivant.

On considère une loi de probabilité μ sur \mathbb{N} telle que $0 < \mu(\{0\}) < 1$ et telle que $0 < m < +\infty$, où m désigne la moyenne de μ , définie par $\mu = \sum_{n=0}^{+\infty} n\mu(\{n\})$. Enfin, on note g la fonction génératrice de μ définie sur $[0, 1]$ par

$$g(u) = \sum_{n=0}^{+\infty} u^n \mu(\{n\}).$$

On considère une famille, indexée sur $\mathbb{N} \times \mathbb{N}^*$, de variables aléatoires $Y_{n,i}$, définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{N} , indépendantes et de même loi μ ($Y_{n,i}$ représente le nombre de descendants directs du i -ième individu de la n -ième génération). On se donne de plus un entier $a \geq 1$ (nombre d'individus dans la population initiale). Le processus X , dit **processus de branchement** (les individus peuvent être identifiés aux **sommets** d'un arbre, au sens mathématique, ou aux **nœuds** de ramification, au sens arbre généalogique) ou **processus de Galton-Watson**, est défini par

$$X_0 = a \quad \text{et pour tout } n \in \mathbb{N}, \quad X_{n+1} = \mathbf{1}_{\{X_n \geq 1\}} \sum_{j=1}^{X_n} Y_{n,j},$$

avec la convention d'écriture $\sum_{j=1}^0 Y_{n,j} = 0$ (X_n est le nombre d'individus de n -ième génération).

La filtration naturelle $(\mathcal{A}_n)_{n \in \mathbb{N}}$ du processus X sera la seule filtration considérée par la suite.

- Démontrer que X est une chaîne de Markov homogène de matrice de transition M à déterminer.
- Démontrer que X est une martingale, une sur ou sous-martingale, suivant les valeurs de m .
- On définit, pour tout $n \in \mathbb{N}$, la variable aléatoire $Y_n = \frac{X_n}{m^n}$; démontrer que Y est une martingale positive.
- Si $m > 1$, on admettra qu'il existe un unique réel $s \in]0, 1[$ tel que $g(s) = s$. On définit, pour tout $n \in \mathbb{N}$, la variable aléatoire $Z_n = s^{X_n}$. Démontrer que Z est une martingale équi-intégrable.
- Démontrer que la suite $(X_n)_{n \in \mathbb{N}}$ converge P-p.s. vers une variable aléatoire X_∞ (étudier séparément les cas $0 < m \leq 1$ et $m > 1$). Identifier la limite X_∞ dans le cas où $0 < m < 1$.
- Soit $j \in \mathbb{N}^*$; calculer, pour tout $k > \mathbb{N}$, la probabilité $P\left[\bigcap_{n=0}^k (X_n = j)\right]$ en fonction de $M(j, j)$ et de $P(X_N = j)$; en déduire que $P[\liminf_n (X_n = j)] = 0$.
- Démontrer alors que l'on a, pour tout $j \in \mathbb{N}^*$, $P(X_\infty = j) = 0$, et qu'en conséquence $X_\infty \in \{0, +\infty\}$ P-p.s. Justifier le fait que tous les points de \mathbb{N}^* sont transitoires. Si $m > 1$, déduire de la quatrième question que l'on a

$$P(X_\infty = 0) = s^a \quad \text{et} \quad P(X_\infty = +\infty) = 1 - s^a.$$

8. On note T le **temps d'extinction** du processus X , c'est-à-dire le temps d'entrée en 0 défini par $T = \inf(n \in \mathbb{N}^* \mid X_n = 0)$, avec $\inf \emptyset = +\infty$. Vérifier que l'on a P-p.s.

$$(X_\infty = +\infty) = \liminf_n (X_n \neq 0) = (T = +\infty),$$

et en déduire la valeur de la probabilité $P(T < +\infty)$ pour les différentes valeurs de m (le temps d'arrêt T est la **date d'extinction** de la population).

Solution.

1. Pour toute fonction bornée f sur \mathbb{N} , pour tout $n \in \mathbb{N}^*$ et tout $(x_1, \dots, x_n) \in \mathbb{N}^n$, on a

$$\begin{aligned} E^{(X_0, X_1, \dots, X_n) = (a, x_1, \dots, x_n)} [f(X_{n+1})] \\ = E^{(X_0, X_1, \dots, X_n) = (a, x_1, \dots, x_n)} \left[f \left(\mathbf{1}_{(x_n \geq 1)} \sum_{j=1}^{x_n} Y_{n,j} \right) \right], \end{aligned}$$

soit, puisque les variables aléatoires (X_0, \dots, X_n) et $Y_{n,j}$, $j \in \mathbb{N}^*$, sont indépendantes,

$$E^{(X_0, X_1, \dots, X_n) = (a, x_1, \dots, x_n)} [f(X_{n+1})] = E \left[f \left(\mathbf{1}_{(x_n \geq 1)} \sum_{j=1}^{x_n} Y_{n,j} \right) \right].$$

Les variables aléatoires $Y_{n,j}$, $j \in \mathbb{N}^*$ étant indépendantes de même loi μ , on a alors

$$E^{(X_0, X_1, \dots, X_n) = (a, x_1, \dots, x_n)} [f(X_{n+1})] = \begin{cases} f(0) & \text{si } x_n = 0, \\ \int f d\mu^{*x_n} & \text{si } x_n \in \mathbb{N}^*. \end{cases}$$

Il en résulte que

$$E^{\delta_n} [f(X_{n+1})] = M(X_n, f). \quad (16.106)$$

où, δ_0 désignant la mesure de Dirac en 0, $M(\cdot, f)$ est définie par, pour tout $x \in \mathbb{N}$,

$$M(x, f) = \mathbf{1}_{\{0\}}(x) \int f d\delta_0 + \mathbf{1}_{\mathbb{N}^*}(x) \int f d\mu^{*x}.$$

Ceci démontre que X est une chaîne de Markov homogène de matrice de transition M , d'entrées $M(x, y) = M(x, \mathbf{1}_{\{y\}})$, soit

$$M(x, y) = \begin{cases} \mathbf{1}_{\{0\}}(y) & \text{si } x = 0, \\ \mu^{*x}(\{y\}) & \text{si } x \in \mathbb{N}^*. \end{cases}$$

2. Le processus X est adapté et positif. Le même calcul de moyenne conditionnelle pour des variables aléatoires positives (non nécessairement intégrables) donne, pour tout $n \in \mathbb{N}^*$ et tout $(x_1, \dots, x_n) \in \mathbb{N}^n$,

$$E^{(X_0, X_1, \dots, X_n) = (a, x_1, \dots, x_n)} (X_{n+1}) = E^{(X_0, X_1, \dots, X_n) = (a, x_1, \dots, x_n)} \left(\mathbf{1}_{(x_n \geq 1)} \sum_{j=1}^{x_n} Y_{n,j} \right)$$

soit, puisque les variables aléatoires (X_0, \dots, X_n) et $Y_{n,j}$, $j \in \mathbb{N}^*$, sont indépendantes,

$$E^{(X_0, X_1, \dots, X_n) = (a, x_1, \dots, x_n)}(X_{n+1}) = \mathbf{1}_{(x_n \geq 1)} E \left[\sum_{j=1}^{x_n} Y_{n,j} \right].$$

Les variables aléatoires $Y_{n,j}$, $j \in \mathbb{N}^*$, étant de même loi μ , de moyenne m , on a

$$\begin{aligned} E^{(X_0, X_1, \dots, X_n) = (a, x_1, \dots, x_n)}(X_{n+1}) &= \mathbf{1}_{(x_n \geq 1)} \sum_{j=1}^{x_n} E(Y_{n,j}) \\ &= \mathbf{1}_{\mathbb{N}^*}(x_n) m x_n = m x_n; \end{aligned}$$

il en résulte que

$$\boxed{E^{\mathcal{A}_n}(X_{n+1}) = m X_n.} \quad (16.107)$$

Le processus X est une martingale (intégrable, puisqu'alors $E(X_{n+1}) = E(X_0) = a$) si $m = 1$, une surmartingale positive si $m < 1$, une sous-martingale positive si $m > 1$.

3. Divisant les deux membres de l'égalité (16.107) par m^{n+1} , on a

$$E^{\mathcal{A}_n} \left(\frac{X_{n+1}}{m^{n+1}} \right) = \frac{X_n}{m^n},$$

ce qui démontre que Y est une martingale positive.

4. Puisque $s \in]0, 1[$, la fonction $p : x \mapsto s^x$ est bornée par 1 sur \mathbb{N} ; il en résulte immédiatement que la suite des variables aléatoires Z_n est équi-intégrable. De plus l'égalité (16.106) donne, pour cette fonction,

$$E^{\mathcal{A}_n} [s^{X_{n+1}}] = M(X_n, p),$$

où, pour tout $x \in \mathbb{N}$, on a

$$M(x, p) = \mathbf{1}_{\{0\}}(x) s^0 + \mathbf{1}_{\mathbb{N}^*}(x) \int s^y d\mu^{*x}(y).$$

Mais on a (résultat classique sur la convolution de mesures)

$$\int s^y d\mu^{*x}(y) = \left[\int s^y d\mu(y) \right]^x = [g(s)]^x;$$

puisque $g(s) = s$, il en résulte que

$$M(x, p) = \mathbf{1}_{\{0\}}(x) s^0 + \mathbf{1}_{\mathbb{N}^*}(x) s^x = s^x,$$

et donc que

$$E^{\mathcal{A}_n} [s^{X_{n+1}}] = s^{X_n}.$$

Ainsi, Z est une martingale équi-intégrable.

5. On distingue les trois cas suivants.

- Si $m \leq 1$, X est une surmartingale positive : elle converge donc P-p.s. vers une variable aléatoire X_∞ telle que, pour tout $n \in \mathbb{N}$, $0 \leq E^{A_n}(X_\infty) \leq X_n$; de plus, X_n est intégrable puisque l'on a $0 \leq E^{A_0}(X_n) \leq X_0 = a$. Il en résulte que X_∞ est intégrable et que l'on a $0 \leq E(X_\infty) \leq EX_n$. Par ailleurs, il résulte de l'égalité (16.107) que

$$E(X_{n+1}) = mE(X_n) = m^{n+1}a;$$

si $m < 1$, on a alors $\lim_n E(X_n) = 0$ et donc, par le lemme de Fatou, $0 \leq E(X_\infty) \leq \liminf_n E(X_n) = 0$; puisque X_∞ est positive, on a $X_\infty = 0$ P-p.s.

Remarque. Si $m = 1$, on ne peut rien dire, a priori, par ce dernier argument quant à la finitude de X_∞ .

- Si $m = 1$, X est une martingale positive telle que $E(X_n) = a$; autrement dit, elle est bornée dans L^1 et est donc convergente P-p.s. vers une variable aléatoire positive X_∞ P-p.s. finie.
- Si $m > 1$, $Z = s^X$ est une martingale équi-intégrable (donc bornée dans L^1); elle converge P-p.s. et dans L^1 vers une variable aléatoire positive U_∞ P-p.s. finie. Il en résulte que la suite $(X_n)_{n \in \mathbb{N}}$ converge P-p.s. vers une variable aléatoire positive X_∞ . Elle vérifie

$$\left[\lim_n s^{X_n} \right] \mathbf{1}_{(X_\infty = +\infty)} = 0. \quad (16.108)$$

6. Soit $j \in \mathbb{N}^*$; X étant une chaîne de Markov homogène de matrice de transition M , on a, pour tout $k > N$,

$$\begin{aligned} P \left[\bigcap_{n=N}^k (X_n = j) \right] &= P(X_k = j \mid X_{k-1} = j) \\ &\quad \times P(X_{k-1} = j \mid X_{k-2} = j) \dots P(X_N = j), \end{aligned}$$

soit

$$P \left[\bigcap_{n=N}^k (X_n = j) \right] = [M(j, j)]^{k-N} P(X_N = j).$$

Mais, puisque $\mu(\{0\}) > 0$, on a $\mu^{*j}(\{0\}) > 0$, et donc $M(j, j) = \mu^{*j}(\{j\}) < 1$; il en résulte que l'on a, pour tout $N \in \mathbb{N}^*$,

$$\lim_k P \left[\bigcap_{n=N}^k (X_n = j) \right] = P \left[\bigcap_{n \geq N} (X_n = j) \right] = 0,$$

et, de façon immédiate, que $P[\liminf_n (X_n = j)] = 0$.

7. La suite $(X_n)_{n \in \mathbb{N}}$ étant à valeurs entières et convergeant P-p.s. dans $\overline{\mathbb{N}}$, on a, pour tout $j \in \mathbb{N}^*$,

$$\text{P-p.s. } (X_\infty = j) \subset \liminf_n (X_n = j),$$

ce qui, d'après la question précédente, démontre que l'on a, pour tout $j \in \mathbb{N}^*$, $P(X_\infty = j) = 0$, et donc que $X_\infty \in \{0, +\infty\}$ P-p.s. Si $j \in \mathbb{N}^*$ était récurrent, on aurait $P[\limsup_n (X_n = j)] = 1$, et, la suite $(X_n)_{n \in \mathbb{N}}$ convergeant vers X_∞ sur $\limsup_n (X_n = j)$, on aurait $X_\infty = j$ P-p.s., et il y aurait contradiction; donc tous les points de \mathbb{N}^* sont transitoires.

Si $m > 1$, on a vu que Z est une martingale équi-intégrable et que la suite $(Z_n)_{n \in \mathbb{N}}$ converge dans L^1 ; on a donc

$$s^a = E(s^{X_0}) = \lim_n E(s^{X_n}) = E(\lim_n s^{X_n});$$

par conséquent, puisque $X_\infty \in \{0, +\infty\}$ P-p.s., il résulte de l'égalité (16.108) que

$$s^a = E(s^{X_\infty} \mathbf{1}_{\{X_\infty=0\}}) = P(X_\infty = 0).$$

On a donc

$$\boxed{P(X_\infty = 0) = s^a \quad \text{et} \quad P(X_\infty = +\infty) = 1 - s^a.}$$

8. Puisque l'on a l'implication

$$X_n(\omega) = 0 \implies (\forall p \geq n, X_p(\omega) = 0),$$

on a

$$\liminf_n (X_n \neq 0) \subset \bigcap_{n \in \mathbb{N}^*} (X_n \geq 1) = (T = +\infty),$$

l'inclusion inverse étant évidente, on a

$$\liminf_n (X_n \neq 0) = (T = +\infty).$$

Par ailleurs, on a démontré que P-p.s.,

$$\liminf_n (X_n \neq 0) = (X_\infty = +\infty).$$

Il en résulte que $P(T = +\infty) = P(X_\infty = +\infty)$, et donc $P(T < +\infty) = P(X_\infty = 0)$.

– Si $m > 1$, on a

$$\boxed{P(T < +\infty) = s^a \quad \text{et} \quad P(T = +\infty) = 1 - s^a.}$$

– Si $m < 1$, on a vu que $X_\infty = 0$ P-p.s.; on a donc $P(T < +\infty) = 1$.

– Si $m = 1$, X est une martingale et on a vu que $P(X_\infty < +\infty) = 1$; on a donc $P(X_\infty = 0) = 1$ et, par conséquent, $P(T < +\infty) = 1$.

Exercice 16.12. Modèle de diffusion de maladies contagieuses de Pólya (suite).

Il s'agit, dans cet exercice, de compléter l'étude du modèle de Pólya (cf. 16.3 et 16.8), décrit sous forme de tirages de boules dans une urne, et plus précisément, de démontrer que la loi de la variable aléatoire Y_∞ , limite P-p.s. de la suite des proportions Y_n de boules blanches contenues dans l'urne **après** le n -ième tirage, **et après** avoir rajouté la boule tirée et les c boules de la couleur de la boule tirée¹⁷, est la loi **bêta** $\beta(\frac{b}{c}, \frac{z}{c})$ de première espèce sur $[0, 1]$. La méthode est de calculer les moments de tous ordres de Y_∞ .

On garde toutes les notations de (16.8).

17. On rappelle que le processus Y est une chaîne de Markov **non** homogène et une martingale.

1. Soit un entier quelconque $l \geq 1$; on définit le processus $Z = (Z_n)_{n \in \mathbb{N}^*}$ par

$$Z_n = \prod_{j=0}^{l-1} Y_{n+j} = \frac{B_n(B_n + c) \dots [B_n + (l-1)c]}{k_n k_{n+1} \dots k_{n+(l-1)}}.$$

Démontrer que Z est une martingale bornée et que la suite $(Z_n)_{n \in \mathbb{N}^*}$ converge P-p.s. et dans \mathcal{L}^1 .

2. En déduire la valeur de $E(Y_\infty^l)$; l'exprimer à l'aide de la fonction Γ .

3. Si U est une variable aléatoire réelle de loi bêta $\beta(a, b)$ de première espèce sur $[0, 1]$, calculer son moment d'ordre l .

4. Démontrer, à l'aide des résultats des questions 2. et 3., que la loi de Y_∞ est la loi bêta $\beta(\frac{b}{c}, \frac{a}{c})$ de première espèce sur $[0, 1]$.

Solution.

1. Remarquons que

$$B_{n+1} = B_n + cX_{n+1}.$$

Cela permet d'expliciter Z_n sur chacun des événements $(X_{n+1} = 0)$ et $(X_{n+1} = 1)$.

• Sur $(X_{n+1} = 0)$, on a $B_{n+1} = B_n$ et donc

$$Z_{n+1} = \frac{B_n(B_n + c) \dots [B_n + (l-1)c]}{k_{n+1} k_{n+2} \dots k_{n+(l-1)} k_{n+1}} = \frac{k_n}{k_{n+1}} Z_n.$$

• Sur $(X_{n+1} = 1)$, on a $B_{n+1} = B_n + c$ et donc

$$Z_{n+1} = \frac{(B_n + c)(B_n + 2c) \dots [B_n + c + (l-1)c]}{k_{n+1} k_{n+2} \dots k_{n+(l-1)} k_{n+1}} = \frac{k_n}{k_{n+1}} Z_n \frac{B_n + lc}{B_n}.$$

Ainsi, on a, en tenant compte de l'adaptation des processus Z et B ,

$$\begin{aligned} E^{\mathcal{A}_n}(Z_{n+1}) &= E^{\mathcal{A}_n}[\mathbf{1}_{(X_{n+1}=0)} Z_{n+1}] + E^{\mathcal{A}_n}[\mathbf{1}_{(X_{n+1}=1)} Z_{n+1}] \\ &= \frac{k_n}{k_{n+1}} Z_n E^{\mathcal{A}_n}[\mathbf{1}_{(X_{n+1}=0)}] + \frac{k_n}{k_{n+1}} Z_n \frac{B_n + lc}{B_n} E^{\mathcal{A}_n}[\mathbf{1}_{(X_{n+1}=1)}], \end{aligned}$$

soit, d'après les égalités (16.21) de l'exemple (16.8) et l'égalité $B_n = k_n Y_n$,

$$E^{\mathcal{A}_n}(Z_{n+1}) = \frac{k_n}{k_{n+1}} Z_n \left[(1 - Y_n) + \frac{B_n + lc}{B_n} Y_n \right] = \frac{k_n}{k_{n+1}} Z_n \left[1 + \frac{lc}{B_n} Y_n \right].$$

En remarquant que $B_n = k_n Y_n$, et que $k_{n+1} = k_n + lc$, on a

$$\frac{k_n}{k_{n+1}} \left[1 + \frac{lc}{k_n} \right] = 1,$$

et donc

$$\boxed{E^{\mathcal{A}_n}(Z_{n+1}) = Z_n.}$$

Ainsi, Z est une martingale. Puisque $|Y_n| \leq 1$, pour tout n , on a aussi $|Z_n| \leq 1$ pour tout n : la martingale Z est donc **équi-intégrable** et converge P-p.s. et dans \mathcal{L}^1 vers une variable aléatoire intégrable Z_∞ .

2. Par ailleurs, puisque, pour tout $n \in \mathbb{N}^*$, on a

$$Z_n = \prod_{j=0}^{l-1} Y_{n+j},$$

et que la suite $(Y_n)_{n \in \mathbb{N}^*}$ converge P-p.s. vers Y_∞ , on a

$$Z_\infty = Y_\infty^l \quad \text{P-p.s.}$$

La convergence de la martingale Z ayant lieu aussi dans \mathcal{L}^1 , on a alors

$$\boxed{E(Z_1) = E(Z_\infty) = E(Y_\infty^l)}. \quad (16.109)$$

Il reste donc à calculer $E(Z_1)$. Puisque $B_1 = b + cX_1$, on a

$$Z_1 = \frac{[b + cX_1][b + c(1 + X_1)] \dots [b + c(l-1 + X_1)]}{k_1 k_2 \dots k_l}.$$

Mais la loi de X_1 étant

$$P_{X_1} = \frac{b}{b+r} \delta_1 + \frac{r}{b+r} \delta_0.$$

on a

$$\begin{aligned} E\left[[b + cX_1][b + c(1 + X_1)] \dots [b + c(l-1 + X_1)] \right] \\ = \frac{b}{b+r} [(b+c)(b+2c) \dots (b+lc)] + \frac{r}{b+r} [b(b+c) \dots (b+(l-1)c)] \\ = \frac{b}{b+r} (b+c)(b+2c) \dots (b+(l-1)c) [b+lc+r]. \end{aligned}$$

De plus

$$k_1 k_2 \dots k_l = (b+r+c)(b+r+2c) \dots (b+r+lc).$$

ce qui donne

$$E(Z_1) = \frac{b}{b+r} \frac{(b+c)(b+2c) \dots (b+(l-1)c) [b+lc+r]}{(b+r+c)(b+r+2c) \dots (b+r+(l-1)c)(b+r+lc)},$$

et après simplifications,

$$\boxed{E(Y_\infty^l) = E(Z_1) = \frac{\Gamma(l + \frac{b}{c}) \Gamma(\frac{b+r}{c})}{\Gamma(l + \frac{b+r}{c}) \Gamma(\frac{b}{c})}}.$$

3. Si U est une variable aléatoire réelle de loi bêta $\beta(a, b)$ de première espèce sur $[0, 1]$, on a

$$E(U^l) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^{a+l-1} (1-x)^{b-1} dx = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \beta(a+l, b),$$

et donc

$$\boxed{E(U^l) = \frac{\Gamma(l+a)\Gamma(a+b)}{\Gamma(l+a+b)\Gamma(a)}}.$$

4. On remarque donc que Y_∞ a mêmes moments qu'une variable aléatoire U de loi bêta $\beta(\frac{b}{c}, \frac{r}{c})$ de première espèce sur $[0, 1]$. Puisque ces variables aléatoires sont **bornées**, leurs **fonctions caractéristiques sont analytiques** sur \mathbb{R} et coïncident partout (cf. chapitre 12, prop. 12.16). Le théorème de Lévy assure alors que Y_∞ **suit la loi bêta $\beta(\frac{b}{c}, \frac{r}{c})$ de première espèce sur $[0, 1]$** .

Remarque. En particulier, si $b = r = c$, la loi de Y_∞ est la **loi uniforme** sur $[0, 1]$.

Appendice A

Résumé de théorie de la mesure

On donne les grandes lignes et les énoncés des principaux théorèmes de théorie de la mesure et de l'intégration, de manière à avoir sous la main les théorèmes essentiels. Pour une étude approfondie, nous renvoyons aux livres de théorie de la mesure ou de probabilité comme ceux de Durrett (dans lequel il existe un résumé de théorie de la mesure assez détaillé), de Gramain, Métivier ou Neveu.

A.1. Mesure et probabilité

Définition A.1. Une famille \mathcal{A} de parties d'un ensemble Ω est

- une **algèbre** (ou un **anneau**) si elle est stable par union (finie) et différence.
- une **algèbre** (ou un **anneau**) **unitaire** si c'est une algèbre qui contient Ω . (exemple : l'ensemble des unions finies d'intervalles de \mathbb{R}).
- une **semi-algèbre** (ou un **semi-anneau unitaire**) si Ω et $\emptyset \in \mathcal{A}$, si elle est stable par intersection (finie) et si, pour tout $A \in \mathcal{A}$, A^c est union d'un nombre fini d'éléments de \mathcal{A} disjoints deux à deux (exemple : l'ensemble des pavés de \mathbb{R}^n de la forme $\prod_{i=1}^n]a_i, b_i[$ où $-\infty \leq a_i < b_i < +\infty$).
- une **tribu** ou **σ -algèbre** si c'est une algèbre unitaire stable par réunion dénombrable, c'est-à-dire que pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , on a $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Remarque. Si \mathcal{A} est une tribu, alors $\Omega \in \mathcal{A}$; de plus, \mathcal{A} est stable par complémentarité, c'est-à-dire que si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$, et stable par intersection dénombrable, c'est-à-dire que pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , on a $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Soit $\{\mathcal{A}_i, i \in I\}$ une famille d'anneaux sur Ω (resp. de tribus); la famille $\bigcap_{i \in I} \mathcal{A}_i$ est encore un anneau (resp. une tribu); attention la réunion de tribus n'est pas une tribu. L'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω est un anneau et une tribu; on définit alors l'**anneau engendré** (resp. la **tribu engendrée**) par une famille quelconque \mathcal{E} de parties de Ω comme l'intersection de tous les anneaux (resp. toutes les tribus) contenant \mathcal{E} . La tribu engendrée par \mathcal{E} est souvent notée $\sigma(\mathcal{E})$ et \mathcal{E} est appelée **système générateur** de la tribu $\sigma(\mathcal{E})$; la **tribu engendrée** par une partie A de Ω

est la famille $\{A, A^c, \Omega, \emptyset\}$. La tribu $\{\Omega, \emptyset\}$ est appelée **tribu triviale**. La **tribu borélienne** de \mathbb{R}^n , notée $\mathcal{B}_{\mathbb{R}^n}$, est la tribu engendrée par la famille des ouverts de \mathbb{R}^n ; elle est aussi engendrée, par exemple, par la famille des pavés de \mathbb{R}^n de la forme $\prod_{i=1}^n [a_i, b_i]$ où $-\infty \leq a_i < b_i < +\infty$. La **tribu borélienne** de $\overline{\mathbb{R}}$ est la tribu engendrée par la famille des ouverts de \mathbb{R} à laquelle on adjoint les singletons $\{-\infty\}$ et $\{+\infty\}$; elle est notée $\mathcal{B}_{\overline{\mathbb{R}}}$. Elle est aussi engendrée, par exemple, par la famille des intervalles de la forme $[a, b]$ où $-\infty \leq a < b \leq +\infty$. Soit $\{\mathcal{A}_i, i \in I\}$ une famille de tribus sur Ω ; la tribu engendrée par la réunion des $\mathcal{A}_i, i \in I$, est notée $\bigvee_{i \in I} \mathcal{A}_i$. Si \mathcal{A} est une tribu, le couple (Ω, \mathcal{A}) est appelé **espace mesurable** ou **probabilisable**.

Définition A.2. Soit f une **application** de E dans F , ensembles munis respectivement des tribus \mathcal{E} et \mathcal{F} . On dit que f est **mesurable** ou est une **variable aléatoire** si, pour tout $A \in \mathcal{F}$, l'image inverse $f^{-1}(A) \in \mathcal{E}$. On rappelle que $f^{-1}(A) = \{x \in E \mid f(x) \in A\}$.

Proposition A.3. (a) La composée de deux applications mesurables est mesurable.

(b) Soient f une **application** de E dans F et \mathcal{C} une famille de parties de F . On a l'égalité des tribus $f^{-1}(\sigma(\mathcal{C})) = \sigma[f^{-1}(\mathcal{C})]$ (pour une famille quelconque \mathcal{D} de parties de F , $f^{-1}(\mathcal{D})$ désigne la famille des $f^{-1}(A)$ lorsque A parcourt \mathcal{D}). En particulier, si \mathcal{F} est une tribu, la famille $f^{-1}(\mathcal{F})$ l'est aussi; elle est appelée **tribu engendrée** par f .

(c) Soit f une **application** de E dans F munis respectivement des tribus \mathcal{E} et \mathcal{F} . Si \mathcal{F} est engendrée par une famille \mathcal{C} de parties de F , pour que f soit mesurable il faut et il suffit que $f^{-1}(\mathcal{C}) \subset \mathcal{E}$.

Définition A.4. Soit $f_i, i \in I$, une famille d'**applications** de E dans F_i muni de la tribu \mathcal{F}_i . La tribu engendrée par la réunion des tribus $f_i^{-1}(\mathcal{F}_i)$ est appelée **tribu engendrée** par les f_i et notée $\sigma(f_i; i \in I)$; c'est la plus petite tribu sur E rendant mesurables toutes les f_i .

Proposition A.5. Soit $\{f_n; n \in \mathbb{N}\}$ une suite d'**applications** mesurables de l'espace mesurable (E, \mathcal{E}) dans \mathbb{R} (resp. $\overline{\mathbb{R}}$) muni de sa tribu borélienne. Quand elles sont définies¹, les applications $f_1 + f_2, f_1 f_2, f_1^+, f_1^-, \sup_{n \in \mathbb{N}} f_n, \inf_{n \in \mathbb{N}} f_n, \limsup_n f_n, \liminf_n f_n$ sont mesurables.

Une **application** continue de \mathbb{R}^n dans \mathbb{R}^p est **borélienne**, c'est-à-dire mesurable par rapport aux tribus boréliennes.

Par la suite, il est sous-entendu que les espaces $\mathbb{R}, \overline{\mathbb{R}}$ et \mathbb{R}^n sont munis de leur tribu borélienne.

1. Conventions : pour tout $a \in \mathbb{R}$, on a $a + \infty + a = +\infty, -\infty + a = -\infty, +\infty + (+\infty) = +\infty, -\infty + (-\infty) = -\infty, 0 \times (\pm\infty) = 0$ et pour tout $a \in \mathbb{R}^*$, $a \times (+\infty) = \text{sign}(a)\infty, a \times (-\infty) = -\text{sign}(a)\infty$. Enfin, la somme $+\infty + (-\infty)$ n'est pas définie.

Définition A.6. Une fonction f définie sur l'espace mesurable (E, \mathcal{E}) à valeurs dans \mathbb{R} (resp. $\overline{\mathbb{R}}$) est **étagée** si elle est mesurable et ne prend qu'un nombre fini de valeurs, **toutes finies**. Elle s'écrit $f = \sum_{j=1}^n f_j \mathbf{1}_{A_j}$ où les A_j appartiennent à \mathcal{E} sont disjoints deux à deux et où $f_j \in \mathbb{R}$.

Lemme A.7. (Fondamental.) Toute fonction mesurable définie sur l'espace mesurable (E, \mathcal{E}) , à valeurs dans $\overline{\mathbb{R}}^+$, est limite simple d'une suite **croissante** de fonctions étagées à valeurs dans \mathbb{R}^+ .

Toute fonction mesurable définie sur l'espace mesurable (E, \mathcal{E}) à valeurs dans \mathbb{R} ou $\overline{\mathbb{R}}$ est limite simple d'une suite de fonctions étagées.

Définition A.8. Soient \mathcal{F} une famille de parties de Ω et μ une application de \mathcal{F} dans $\overline{\mathbb{R}}^+$. On dit que μ possède la propriété d'**additivité finie** (resp. d'**additivité dénombrable**); on dit encore dans ce cas que μ est **σ -additive** si, pour toute famille finie (resp. dénombrable) A_i , $i \in I$, d'éléments de \mathcal{F} disjoints deux à deux dont la réunion appartient à \mathcal{F} , on a

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i).$$

Exemple A.1. L'application μ définie par $\mu(A) = 0$ si $|A| < +\infty$ et $\mu(A) = +\infty$ sinon est additive mais non σ -additive.

La fonction μ définie sur la famille \mathcal{I} des intervalles de \mathbb{R} par $\mu(A) =$ longueur (A) est σ -additive.

Définition A.9. Une **mesure** μ sur l'espace mesurable (Ω, \mathcal{A}) est une application σ -additive de \mathcal{A} dans $\overline{\mathbb{R}}^+$ telle que $\mu(\emptyset) = 0$. Une mesure μ est **finie** si elle est à valeurs dans \mathbb{R}^+ . La **masse** d'une mesure est $\mu(\Omega)$. Une mesure μ est **σ -finie** s'il existe un recouvrement dénombrable de Ω par une famille A_n , $n \in \mathbb{N}$, d'éléments de \mathcal{A} de mesure finie. Une **probabilité** P sur (Ω, \mathcal{A}) est une mesure de masse 1. La **mesure de Dirac** en $\omega \in \Omega$ est la mesure δ_ω définie par $\delta_\omega(A) = 1$ si $\omega \in A$, $\delta_\omega(A) = 0$ sinon. Une mesure est **discrète** si elle est de la forme $\mu = \sum_{\omega \in D} \alpha_\omega \delta_\omega$ où D est une partie dénombrable de Ω et $\alpha_\omega \in \mathbb{R}^+$.

Proposition A.10. Soit μ une mesure sur l'espace mesurable (Ω, \mathcal{A}) non identique à $+\infty$. On a les propriétés suivantes :

- Pour tout $A, B \in \mathcal{A}$ disjoints, $\mu(A \cup B) = \mu(A) + \mu(B)$.
- Pour tous $A, B \in \mathcal{A}$ tels que $A \subset B$, on a $\mu(A) \leq \mu(B)$.
- Pour tout $A, B \in \mathcal{A}$, $\mu(A \cup B) \leq \mu(A) + \mu(B)$ (sous σ -additivité).
- Si $(A_n)_{n \in \mathbb{N}}$ est une suite croissante d'éléments de \mathcal{A} , on a

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_n \mu(A_n).$$

(e) Si $(A_n)_{n \in \mathbb{N}}$ est une suite décroissante d'éléments de \mathcal{A} telle qu'existe n_0 pour lequel on ait $\mu(A_{n_0}) < +\infty$, on a

$$\mu\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_n \mu(A_n).$$

Génération d'une mesure

Théorème A.11 (Théorème de Carathéodory). Une fonction μ , σ -additive sur une algèbre unitaire \mathcal{A} et vérifiant $\mu(\emptyset) = 0$, se prolonge de manière unique en une mesure μ sur la tribu engendrée par \mathcal{A} .

Théorème A.12. Soit \mathcal{S} une semi-algèbre sur Ω ; l'algèbre $\overline{\mathcal{S}}$ engendrée par \mathcal{S} est la famille des réunions finies d'éléments disjoints de \mathcal{S} .

Soit une fonction μ , additive sur la semi-algèbre \mathcal{S} , vérifiant $\mu(\emptyset) = 0$, et sous σ -additive sur \mathcal{S} , c'est-à-dire telle pour toute famille dénombrable A_i , $i \in I$, d'éléments de \mathcal{S} disjoints deux à deux dont la réunion appartient à \mathcal{S} , on ait $\mu(\biguplus_{i \in I} A_i) \leq \sum_{i \in I} \mu(A_i)$. Alors μ se prolonge de manière unique en une fonction σ -additive sur $\overline{\mathcal{S}}$ et, en conséquence, en une mesure unique μ sur la tribu engendrée par \mathcal{S} .

Exemple A.2. Soient \mathcal{J}_d la semi-algèbre sur \mathbb{R} des intervalles de la forme $]a, b]$ et μ la fonction longueur définie sur \mathcal{J}_d ; μ est σ -additive sur \mathcal{J}_d . L'unique prolongement de μ en une mesure sur la tribu borélienne est la **mesure de Borel**² sur \mathbb{R} . Plus généralement, si F est une fonction réelle définie sur \mathbb{R} , croissante et continue à droite, il existe une unique mesure μ sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ telle que $\mu(]a, b]) = F(b) - F(a)$ pour tout a, b ; elle est appelée mesure de **Borel-Stieltjes** associée à F et est σ -finie. De même, si \mathcal{P} est la semi-algèbre sur \mathbb{R}^n des pavés de la forme $\prod_{i=1}^n]a_i, b_i]$, la fonction volume μ définie sur \mathcal{P} par $\mu(\prod_{i=1}^n]a_i, b_i]) = \prod_{i=1}^n (b_i - a_i)$ est σ -additive sur \mathcal{P} . L'unique prolongement de μ en une mesure sur la tribu borélienne est la **mesure de Borel** sur \mathbb{R}^n .

Définition A.13. Un ensemble A est μ -négligeable s'il est contenu dans un ensemble $B \in \mathcal{A}$ de mesure nulle. On dit que l'espace mesuré $(\Omega, \mathcal{A}, \mu)$ est **complet**, ou que la mesure μ est complète, si tout ensemble μ -négligeable appartient à \mathcal{A} .

2. Émile Borel (1871-1956), né à Saint-Affrique, a été professeur à l'École normale supérieure, puis à la Sorbonne. Ses travaux de recherche portent d'abord sur la théorie de la mesure (c'est lui qui introduit la notion d'ensemble de mesure nulle), des fonctions de variables réelles et de sommation de séries. Il se tourne ensuite vers la théorie des probabilités, la théorie des jeux, et la physique mathématique. On lui doit aussi une approche probabiliste de la théorie cinétique des gaz.

Proposition A.14. Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$ et soit la famille de parties de Ω

$$\mathcal{A}^\mu = \{X \mid \exists B_1, B_2 \in \mathcal{A} \text{ tels que } B_1 \subset X \subset B_2 \text{ et } \mu(B_2 - B_1) = 0\};$$

\mathcal{A}^μ est une tribu et μ se prolonge de manière unique en une mesure $\hat{\mu}$ sur \mathcal{A}^μ et $\hat{\mu}$ est complète.

Exemple A.3. La complétée de la mesure de Borel sur \mathbb{R}^n est la **mesure de Lebesgue** sur \mathbb{R}^n .

Définition A.15. Une propriété \mathcal{P} dépendant de $\omega \in \Omega$ est dite vraie μ -presque partout si l'ensemble des ω où la propriété $\mathcal{P}(\omega)$ est fautive est μ -négligeable. On abrège μ -presque partout en μ -p.p.

Exemple A.4. Dire qu'une suite de fonctions mesurables $(f_n)_{n \in \mathbb{N}^*}$ converge μ -presque-partout est dire que l'ensemble des ω où la suite $(f_n(\omega))_{n \in \mathbb{N}^*}$ ne converge pas est de mesure nulle.

A.2. Intégrale

Soient (Ω, \mathcal{A}) un espace mesurable et \mathcal{M}^+ l'ensemble des fonctions à valeurs dans $\overline{\mathbb{R}}^+$ et \mathcal{A} -mesurables.

Définition A.16. On appelle **intégrale** sur (Ω, \mathcal{A}) toute application \mathcal{I} de \mathcal{M}^+ dans $\overline{\mathbb{R}}^+$ telle que $\mathcal{I}(0) = 0$ et qui soit σ -additive sur \mathcal{M}^+ , c'est-à-dire telle que $\mathcal{I}(\sum_{n \in \mathbb{N}} f_n) = \sum_{n \in \mathbb{N}} \mathcal{I}(f_n)$, pour toute suite $(f_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{M}^+ .

Proposition A.17. Soit \mathcal{I} une intégrale sur (Ω, \mathcal{A}) et soient f, g et $f_n, n \in \mathbb{N}$, des éléments de \mathcal{M}^+ . On a, dans $\overline{\mathbb{R}}^+$:

- (a) $\mathcal{I}(f + g) = \mathcal{I}(f) + \mathcal{I}(g)$;
- (b) si $f \leq g$, alors $\mathcal{I}(f) \leq \mathcal{I}(g)$;
- (c) si $f_n \nearrow f$ alors $\mathcal{I}(f_n) \nearrow \mathcal{I}(f)$ (propriété de **Beppo Levi**³);
- (d) si $f_n \searrow f$ et si il existe n_0 tel que $\mathcal{I}(f_{n_0}) < +\infty$, alors $\mathcal{I}(f_n) \searrow \mathcal{I}(f)$;
- (e) pour tout $a \in \mathbb{R}^+$, $\mathcal{I}(af) = a \mathcal{I}(f)$.

Lien entre intégrale et mesure

Théorème A.18. Soit \mathcal{I} une intégrale sur (Ω, \mathcal{A}) . L'application $A \mapsto \mathcal{I}(\mathbf{1}_A)$ est une mesure sur (Ω, \mathcal{A}) .

3. Beppo Levi (1875-1961), né à Turin, s'installe en Argentine en 1939. Ses travaux portent sur la théorie des fonctions intégrables et la mécanique quantique.

Inversement, soit une mesure μ sur (Ω, \mathcal{A}) ; il existe une intégrale unique \mathcal{I}_μ sur (Ω, \mathcal{A}) telle que l'on ait, pour tout $A \in \mathcal{A}$, $\mathcal{I}_\mu(\mathbf{1}_A) = \mu(A)$. De plus, si $f \in \mathcal{M}^+$, $\mathcal{I}_\mu(f)$ est donnée par

$$\mathcal{I}_\mu(f) = \begin{cases} \sum_{x \in f(\Omega)} x \mu(f = x) & \text{si } f \text{ est étagée,} \\ \sup \{ \mathcal{I}_\mu(g) \mid g \leq f, g \text{ étagée} \} & \text{dans le cas général.} \end{cases}$$

Notation. $\mathcal{I}_\mu(f)$, élément de $\overline{\mathbb{R}}^+$, est noté indifféremment

$$\int_{\Omega} f \, d\mu, \quad \int_{\Omega} f(\omega) \, d\mu(\omega) \quad \text{ou} \quad \int_{\Omega} f(\omega) \, \mu(d\omega),$$

et appelé **intégrale de f par rapport à μ** .

Lemme A.19 (Lemme d'unicité). Deux intégrales \mathcal{I} et \mathcal{I}' sur (Ω, \mathcal{A}) qui sont telles que $\mathcal{I}(\mathbf{1}_A) = \mathcal{I}'(\mathbf{1}_A)$ pour tout $A \in \mathcal{A}$ sont égales.

Intégration de fonctions de signe quelconque

Définition A.20. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Une fonction à valeurs dans $\overline{\mathbb{R}}$ est dite μ -**intégrable** (resp. μ -**semi-intégrable**) si elle est \mathcal{A} -mesurable, et si $\int_{\Omega} |f| \, d\mu < +\infty$, ou de manière équivalente, $\int_{\Omega} f^+ \, d\mu < +\infty$ et $\int_{\Omega} f^- \, d\mu < +\infty$ (resp. et si $\int_{\Omega} f^+ \, d\mu < +\infty$ ou $\int_{\Omega} f^- \, d\mu < +\infty$).

Si f est μ -semi-intégrable (resp. μ -intégrable), l'élément de $\overline{\mathbb{R}}$ (resp. \mathbb{R}), $\int_{\Omega} f \, d\mu = \int_{\Omega} f^+ \, d\mu - \int_{\Omega} f^- \, d\mu$, noté indifféremment $\int_{\Omega} f \, d\mu$, $\int_{\Omega} f(\omega) \, d\mu(\omega)$ ou $\int_{\Omega} f(\omega) \mu(d\omega)$ est appelé **intégrale de f par rapport à μ** .

Proposition A.21. (a) Si f et g sont semi-intégrables on a

$$\int_{\Omega} f \, d\mu \leq \int_{\Omega} g \, d\mu.$$

(b) Si f est mesurable, si g est μ -intégrable et si on a $|f| \leq g$, alors f est μ -intégrable.

(c) Si f est μ -intégrable, on a $\left| \int_{\Omega} f \, d\mu \right| \leq \int_{\Omega} |f| \, d\mu$.

(d) L'ensemble $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ des fonctions à valeurs dans \mathbb{R} et μ -intégrables est un espace vectoriel et l'application $f \mapsto \int_{\Omega} f \, d\mu$ est linéaire de $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ dans \mathbb{R} .

Exemple A.5. • Soit δ_{ω_0} la mesure de Dirac en $\omega_0 \in \Omega$. L'application \mathcal{I} de \mathcal{M}^+ dans $\overline{\mathbb{R}}^+$, $f \mapsto f(\omega_0)$, est une intégrale; puisque, pour tout $A \in \mathcal{A}$, on a $\mathcal{I}(\mathbf{1}_A) = \mathbf{1}_A(\omega_0) = \delta_{\omega_0}(A)$, il résulte du lemme A.19 que l'on a, pour tout $f \in \mathcal{M}^+$, $\int_{\Omega} f \, d\delta_{\omega_0} = f(\omega_0)$. De plus, si f est \mathcal{A} -mesurable de signe

quelconque, pour que f soit μ -intégrable (resp. μ -semi-intégrable) il faut et il suffit que $f^+(\omega_0) < +\infty$ et $f^-(\omega_0) < +\infty$ (resp. $f^+(\omega_0) < +\infty$ ou $f^-(\omega_0) < +\infty$); dans ce cas, on a encore $\int_{\Omega} f d\delta_{\omega_0} = f(\omega_0)$.

• Le même raisonnement montre que si μ est la mesure discrète $\mu = \sum_{n=1}^{+\infty} \alpha_n \delta_{\omega_n}$ où $\alpha_n \in \mathbb{R}^+$ et $\omega_n \in \Omega$, pour tout $f \in \mathcal{M}^+$, on a $\int_{\Omega} f d\mu = \sum_{n=1}^{+\infty} \alpha_n f(\omega_n)$. Si f est \mathcal{A} -mesurable de signe quelconque, pour que f soit μ -intégrable il faut et il suffit que $\sum_{n=1}^{+\infty} \alpha_n |f(\omega_n)| < +\infty$; dans ce cas, on a encore $\int_{\Omega} f d\mu = \sum_{n=1}^{+\infty} \alpha_n f(\omega_n)$.

• **Intégrale de Lebesgue⁴ d'une fonction Riemann-intégrable sur $[a, b]$.** Si \mathcal{P} est une partition finie de $[a, b]$ en intervalles, et si f est une fonction définie sur $[a, b]$, de signe quelconque et bornée, on note, pour tout intervalle $P \in \mathcal{P}$,

$$\underline{f}_P = \inf\{f(x) \mid x \in P\} \quad \text{et} \quad \overline{f}_P = \sup\{f(x) \mid x \in P\};$$

si $|\mathcal{P}|$ désigne la longueur de P , on définit les sommes de Darboux,

$$s_P = \sum_{P \in \mathcal{P}} \underline{f}_P |P| \quad \text{et} \quad S_P = \sum_{P \in \mathcal{P}} \overline{f}_P |P|.$$

Par définition, la fonction f est Riemann-intégrable sur $[a, b]$ si, pour toute suite $(\mathcal{P}_n)_{n \in \mathbb{N}}$ de subdivisions emboîtées dont le pas tend vers 0, les suites $(s_{\mathcal{P}_n})_{n \in \mathbb{N}}$ et $(S_{\mathcal{P}_n})_{n \in \mathbb{N}}$ sont convergentes et ont même limite, celle-ci étant, par définition l'intégrale de Riemann $\int_a^b f(x) dx$ de f sur $[a, b]$. Si f est Riemann-intégrable et **positive** sur $[a, b]$, la fonction $f \mathbf{1}_{[a,b]}$ est Lebesgue-intégrable (c'est-à-dire intégrable par rapport à la mesure de Lebesgue λ sur \mathbb{R}) et on a $\int_{\mathbb{R}} f \mathbf{1}_{[a,b]} d\lambda = \int_a^b f(x) dx$.

A.3. Trois théorèmes de convergence

Lemme A.22 (Lemme de Fatou). Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f_n , $n \in \mathbb{N}$, des éléments de \mathcal{M}^+ . On a l'inégalité dans $\overline{\mathbb{R}}^+$:

$$\int_{\Omega} \liminf_n f_n d\mu \leq \liminf_n \int_{\Omega} f_n d\mu.$$

Remarque. Ce lemme sert essentiellement à démontrer l'intégrabilité d'une fonction qui est limite simple d'une suite de fonctions.

4. Henri Lebesgue (1875-1941), né à Beauvais, a suivi les cours d'Émile Borel. Les travaux de ce dernier, ainsi que ceux de Jordan et Peano, l'ont conduit à élaborer (1901 et 1902) sa théorie de l'intégrale, qui généralise celle de Riemann. Il a montré en 1904 qu'une fonction bornée est intégrable au sens de Riemann si et seulement si l'ensemble de ses points de discontinuité est de mesure nulle. Il a étudié les fonctions de plusieurs variables, les séries de fonctions et les séries de Fourier à l'aide de sa théorie de l'intégration.

Théorème A.23 (Théorème de convergence monotone). Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et soit $(f_n)_{n \in \mathbb{N}}$ une suite **monotone** de fonctions mesurables à valeurs dans $\overline{\mathbb{R}}$, de limite f .

1. Si la suite est croissante et s'il existe n_0 tel que $\int_{\Omega} f_{n_0}^- d\mu < +\infty$, on a l'égalité dans $\overline{\mathbb{R}}$:

$$\lim_n \nearrow \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu.$$

2. Si la suite est décroissante et s'il existe n_0 tel que $\int_{\Omega} f_{n_0}^+ d\mu < +\infty$, on a l'égalité dans $\overline{\mathbb{R}}$:

$$\lim_n \searrow \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu.$$

Théorème A.24 (Théorème de convergence dominée, première version).

Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables à valeurs dans $\overline{\mathbb{R}}$, convergeant simplement vers f . Si $\sup |f_n|$ est μ -intégrable, les fonctions f_n et f sont μ -intégrables et la suite des intégrales $\int_{\Omega} f_n d\mu$ est convergente (dans \mathbb{R}). De plus, on a

$$\lim_n \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu \quad \text{et} \quad \lim_n \int_{\Omega} |f_n - f| d\mu = 0.$$

Remarque. L'hypothèse « $\sup |f_n|$ est μ -intégrable » est équivalente à l'hypothèse souvent formulée, et qui donne son nom au théorème, à savoir : « il existe une fonction g μ -intégrable telle que l'on ait, pour tout $n \in \mathbb{N}$ et tout $\omega \in \Omega$, $|f_n(\omega)| \leq g(\omega)$ ».

En application, on montre le résultat fondamental : si f est Riemann-intégrable de signe quelconque sur $[a, b]$ (donc bornée), la fonction $f \cdot \mathbf{1}_{[a,b]}$ est Lebesgue-intégrable et on a $\int_{\mathbb{R}} f \cdot \mathbf{1}_{[a,b]} d\lambda = \int_a^b f(x) dx$.

Définition A.25. Une fonction f définie sur un intervalle ouvert ou semi-ouvert quelconque $I = (a, b)$, où $-\infty \leq a < b \leq +\infty$, est **localement Riemann-intégrable** sur I si elle est Riemann-intégrable sur tout intervalle fermé borné contenu dans I .

Proposition A.26. Soit f une fonction définie sur un intervalle ouvert ou semi-ouvert quelconque $I = (a, b)$, où $-\infty \leq a < b \leq +\infty$, **localement Riemann-intégrable** sur I . La fonction $\mathbf{1}_I \cdot f$ est Lebesgue-intégrable si et seulement si l'intégrale de Riemann généralisée $\int_a^b f(x) dx$ est absolument convergente et, dans ce cas, on a

$$\int_{\mathbb{R}} \mathbf{1}_I \cdot f d\lambda = \int_a^b f(x) dx.$$

Intégrale sur un ensemble ; intégrale d'une fonction définie μ -presque partout

Définition A.27. Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f une fonction définie sur Ω , à valeurs dans $\overline{\mathbb{R}}$, μ -semi-intégrable (resp. μ -intégrable). Pour tout $A \in \mathcal{A}$, $\mathbf{1}_A \cdot f$ est μ -semi-intégrable (resp. μ -intégrable). L'intégrale $\int_{\Omega} \mathbf{1}_A \cdot f \, d\mu$ est notée $\int_A f \, d\mu$ et appelée intégrale de f sur A .

Proposition A.28. Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $f \in \mathcal{M}^+$. Pour que $\int_{\Omega} f \, d\mu = 0$, il faut et il suffit que $f = 0$ μ -p.p.

Proposition A.29. Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f et g des fonctions définies sur Ω , à valeurs dans $\overline{\mathbb{R}}$, \mathcal{A} -mesurables et telles que $f = g$ μ -p.p.

(a) Si f est positive, g est positive μ -p.p. et on a $\int_{\Omega} f \, d\mu = \int_{\Omega} g \, d\mu$.

(b) Si f et g sont de signes quelconques et si f est μ -intégrable, g est aussi μ -intégrable et on a $\int_{\Omega} f \, d\mu = \int_{\Omega} g \, d\mu$.

Définition A.30. Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $B \in \mathcal{A}$. La famille de parties $B \cap A$ constituée des ensembles $B \cap A$ lorsque A parcourt \mathcal{A} est une tribu appelée **tribu trace** de \mathcal{A} sur B . On définit alors l'espace mesuré $(B, B \cap \mathcal{A}, \mu|_B)$, appelé **espace mesuré trace** sur B de $(\Omega, \mathcal{A}, \mu)$, où $\mu|_B$ est la **mesure trace**, restriction de μ à $B \cap \mathcal{A}$ (elle est encore souvent notée μ).

Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f une fonction définie sur $\Omega_f \subset \Omega$. Pour toute fonction mesurable g sur $(\Omega, \mathcal{A}, \mu)$, on définit un prolongement f_g de f à Ω par, pour tout $\omega \in \Omega$,

$$f_g(\omega) = \begin{cases} f(\omega) & \text{si } \omega \in \Omega_f, \\ g(\omega) & \text{sinon.} \end{cases}$$

Si $\Omega_f \in \mathcal{A}$, et si f est mesurable relativement à l'espace mesuré trace sur Ω_f , f_g est mesurable ; on dit que f_g est un **prolongement mesurable** de f . Si de plus $\mu(\Omega_f) = 0$, on dit que f est définie μ -p.p. ; deux prolongements mesurables de f sont alors égaux μ -p.p.

D'après la proposition précédente, si f est définie μ -p.p. et admet un prolongement mesurable μ -intégrable, tout autre prolongement mesurable de f est aussi μ -intégrable et leurs intégrales sont égales. ce qui permet de définir l'intégrale de f comme intégrale d'un prolongement mesurable quelconque de f . On dit encore dans ce cas que f est μ -intégrable et on note $\int_{\Omega} f \, d\mu$ son intégrale.

Lemme A.31. Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f une fonction à valeurs dans $\overline{\mathbb{R}}$, μ -intégrable. Alors f est fini μ -p.p.

Théorème A.32 (Théorème de convergence dominée, deuxième version). Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables à valeurs dans $\overline{\mathbb{R}}$, convergeant vers f μ -p.p. On suppose qu'il existe

une fonction g μ -intégrable telle que l'on ait, pour tout $n \in \mathbb{N}$, $|f_n| \leq g$ μ -p.p.; alors les fonctions f_n et f sont μ -intégrables et la suite des intégrales $\int_{\Omega} f_n d\mu$ est convergente (dans \mathbb{R}). De plus, on a

$$\lim_n \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu \quad \text{et} \quad \lim_n \int_{\Omega} |f_n - f| d\mu = 0.$$

Remarque. L'hypothèse de domination est équivalente à l'hypothèse « il existe une fonction g μ -intégrable telle que l'on ait μ -p.p., pour tout $n \in \mathbb{N}$, $|f_n| \leq g$ ». Cette équivalence vient du fait que, toute réunion dénombrable d'ensembles de mesure nulle étant de mesure nulle, on peut intervertir les conditions « μ -p.p. » et « pour tout $n \in \mathbb{N}$ ».

Corollaire A.33. Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables à valeurs dans $\overline{\mathbb{R}}$ telle que

$$\sum_{n=0}^{+\infty} \int_{\Omega} |f_n| d\mu < +\infty.$$

Alors la série de fonctions $\sum f_n$ est μ -p.p. absolument convergente, sa somme est μ -intégrable et on a

$$\int_{\Omega} \sum_{n=0}^{+\infty} f_n d\mu = \sum_{n=0}^{+\infty} \int_{\Omega} f_n d\mu.$$

A.4. Mesure produit et théorème de Fubini

Soient deux espaces mesurés $(\Omega_1, \mathcal{A}_1, \mu_1)$ et $(\Omega_2, \mathcal{A}_2, \mu_2)$. On note Π_1 et Π_2 les projections canoniques de $\Omega_1 \times \Omega_2$ sur Ω_1 et Ω_2 .

Définition A.34. Sur $\Omega_1 \times \Omega_2$, la tribu engendrée par la semi-algèbre des pavés $A_1 \times A_2$, où $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$, est appelée **tribu produit** de \mathcal{A}_1 et \mathcal{A}_2 et notée $\mathcal{A}_1 \otimes \mathcal{A}_2$. C'est la plus petite tribu rendant mesurables les projections canoniques.

Le théorème de Carathéodory permet de démontrer l'existence et l'unicité d'une mesure produit :

Proposition A.35. Si μ_1 et μ_2 sont σ -finies, il existe une mesure unique μ sur l'espace mesurable produit $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ qui satisfasse la relation

$$\forall A_1 \in \mathcal{A}_1, \forall A_2 \in \mathcal{A}_2 \quad \mu(A_1 \times A_2) = \mu_1(A_1) \mu_2(A_2);$$

cette mesure est appelée **mesure produit** et notée $\mu_1 \otimes \mu_2$.

Sections d'ensembles. Si A est une partie de $\Omega_1 \times \Omega_2$, pour tout $\omega_2 \in \Omega_2$, on définit la **section** de A en ω_2 , éventuellement vide,

$$A_{\omega_2}^1 = \{\omega_1 \in \Omega_1 \mid (\omega_1, \omega_2) \in A\}$$

et, pour tout $\omega_1 \in \Omega_1$, on définit la section de A en ω_1 , éventuellement vide, $A_{\omega_1}^2 = \{\omega_2 \in \Omega_2 \mid (\omega_1, \omega_2) \in A\}$.

Si f est une fonction de $\Omega_1 \times \Omega_2$ dans $\overline{\mathbb{R}}$, pour tout $\omega_2 \in \Omega_2$, on définit l'application partielle de f en ω_2 , notée $f_{\omega_2}^1$, de Ω_1 dans $\overline{\mathbb{R}}$ par $\omega_1 \mapsto f(\omega_1, \omega_2)$ et, pour tout $\omega_1 \in \Omega_1$, on définit l'application partielle de f en ω_1 , notée $f_{\omega_1}^2$, de Ω_2 dans $\overline{\mathbb{R}}$ par $\omega_2 \mapsto f(\omega_1, \omega_2)$.

Lemme A.36. (a) Si $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, pour tout $\omega_2 \in \Omega_2$, $A_{\omega_2}^1 \in \mathcal{A}_1$, et pour tout $\omega_1 \in \Omega_1$, $A_{\omega_1}^2 \in \mathcal{A}_2$.

(b) Si f est une fonction de $\Omega_1 \times \Omega_2$ dans $\overline{\mathbb{R}}$, $\mathcal{A}_1 \otimes \mathcal{A}_2$ -mesurable, pour tout $\omega_2 \in \Omega_2$, l'application partielle $f_{\omega_2}^1$ est \mathcal{A}_1 -mesurable, et, pour tout $\omega_1 \in \Omega_1$, l'application partielle $f_{\omega_1}^2$ est \mathcal{A}_2 -mesurable.

Théorème A.37 (Théorème de Fubini⁵). Soient deux espaces mesurés $(\Omega_1, \mathcal{A}_1, \mu_1)$ et $(\Omega_2, \mathcal{A}_2, \mu_2)$ tels que μ_1 et μ_2 soient σ -finies. Soit f une fonction de $\Omega_1 \times \Omega_2$ dans $\overline{\mathbb{R}}$, $\mathcal{A}_1 \otimes \mathcal{A}_2$ -mesurable.

1. Si f est positive,

– l'application $\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}^2(\omega_2) d\mu_2(\omega_2)$ est mesurable de $(\Omega_1, \mathcal{A}_1)$ dans $(\overline{\mathbb{R}}^+, \mathcal{B}_{\overline{\mathbb{R}}^+})$,

– l'application $\omega_2 \mapsto \int_{\Omega_1} f_{\omega_2}^1(\omega_1) d\mu_1(\omega_1)$ est mesurable de $(\Omega_2, \mathcal{A}_2)$ dans $(\overline{\mathbb{R}}^+, \mathcal{B}_{\overline{\mathbb{R}}^+})$,

et on a le calcul par intégrales itérées de l'intégrale de f par l'une des formules suivantes

$$\int_{\Omega_1 \wedge \Omega_2} f d\mu_1 \otimes \mu_2 = \int_{\Omega_1} \left[\int_{\Omega_2} f_{\omega_1}^2(\omega_2) d\mu_2(\omega_2) \right] d\mu_1(\omega_1) \quad (\text{A.1})$$

ou

$$\int_{\Omega_1 \times \Omega_2} f d\mu_1 \otimes \mu_2 = \int_{\Omega_2} \left[\int_{\Omega_1} f_{\omega_2}^1(\omega_1) d\mu_1(\omega_1) \right] d\mu_2(\omega_2) \quad (\text{A.2})$$

2. Si f est de signe quelconque, et si f est $\mu_1 \otimes \mu_2$ -intégrable,

– pour μ_1 presque tout $\omega_1 \in \Omega_1$, l'application partielle $f_{\omega_1}^2$ est μ_2 -intégrable ;

5. Guido Fubini (1879-1943), né à Venise, suit les cours d'Ulisse Dini à Pise, ville où il devient professeur, avant de le devenir au Politecnico de Turin. Interdit d'enseignement par les fascistes, il s'exile à Paris, puis s'installe à l'université de Princeton où il termine sa carrière. Ses travaux portent sur l'analyse fonctionnelle, la théorie de l'intégrale de Lebesgue, la géométrie différentielle et la géométrie projective ; en particulier, le théorème ci dessus, ramenant le calcul d'intégrales doubles à celui d'intégrales simples date de 1907.

- pour μ_2 presque tout $\omega_2 \in \Omega_2$, l'application partielle $f_{\omega_2}^1$ est μ_1 -intégrable ;
- l'application $\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}^2(\omega_2) d\mu_2(\omega_2)$ est définie μ_1 -p.p. et μ_1 -intégrable ;
- l'application $\omega_2 \mapsto \int_{\Omega_1} f_{\omega_2}^1(\omega_1) d\mu_1(\omega_1)$ est définie μ_2 -p.p. et μ_2 -intégrable, et on a encore le calcul de l'intégrale de f par intégrales itérées selon l'une des formules (A.1) ou (A.2).

Remarque. On s'assure souvent de la $\mu_1 \otimes \mu_2$ -intégrabilité de f en calculant l'intégrale $\int_{\Omega_1 \times \Omega_2} |f| d\mu_1 \otimes \mu_2$ à l'aide de la première partie du théorème.

Il est facile de définir le produit de n mesures σ -finies et de vérifier que le produit est associatif. Le théorème de **Kolmogorov** permet de construire des **probabilités** sur l'espace produit $\mathbb{R}^{\mathbb{N}}$. Ce théorème sert en particulier à construire un espace probabilisé supportant une famille de variables aléatoires réelles indépendantes de lois arbitraires données.

Définition A.38. Soit une suite d'espaces mesurables $(\Omega_n, \mathcal{A}_n)$, $n \in \mathbb{N}$. On note Π_j la projection canonique du produit cartésien infini $\prod_{n \in \mathbb{N}} \Omega_n$ sur Ω_j , c'est-à-dire l'application $\omega \mapsto \omega_j$ où ω_j est la j -ième coordonnée de la suite ω . Sur $\prod_{n \in \mathbb{N}} \Omega_n$, on définit la **tribu produit** des \mathcal{A}_n , tribu engendrée par la semi-algèbre des **cylindres à base finie**, c'est-à-dire des produits cartésiens infinis de la forme $\prod_{n \in \mathbb{N}} A_n$, où $A_n \in \mathcal{A}_n$ et où $A_n = \Omega_n$, sauf pour un nombre fini d'indices. La tribu produit est notée $\otimes_{n \in \mathbb{N}} \mathcal{A}_n$. C'est la plus petite tribu rendant mesurables toutes les projections canoniques.

On munit $\mathbb{R}^{\mathbb{N}}$ de la tribu produit des tribus boréliennes, notée brièvement $\mathcal{B}_{\mathbb{R}^{\mathbb{N}}}$.

Théorème A.39 (Théorème de prolongement de Kolmogorov⁶). Soit une suite **consistante** de probabilités, c'est-à-dire une suite $(P_n)_{n \in \mathbb{N}}$ telle que pour tout $n \in \mathbb{N}$, P_n est une probabilité sur $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ et telle que, pour tout pavé $\prod_{j=1}^n]a_j, b_j]$, on ait

$$P_{n+1} \left(\prod_{j=1}^n]a_j, b_j] \times \mathbb{R} \right) = P_n \left(\prod_{j=1}^n]a_j, b_j] \right).$$

Il existe une unique probabilité P sur l'espace mesurable $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_{\mathbb{R}^{\mathbb{N}}})$ telle que l'on ait, pour tout $n \in \mathbb{N}$,

$$P \left[\left(\omega \in \mathbb{R}^{\mathbb{N}} \mid (\omega_1, \dots, \omega_n) \in \prod_{j=1}^n]a_j, b_j] \right) \right] = P_n \left(\prod_{j=1}^n]a_j, b_j] \right).$$

6. Andreï Kolmogorov (1903-1987), nommé professeur à l'université de Moscou en 1931, y était nommé, deux ans plus tard, directeur de l'Institut de mathématiques. Il s'est intéressé à de nombreux domaines des mathématiques : son apport le plus important a été à la théorie des probabilités. C'est à lui que l'on doit l'axiomatisation de la théorie moderne des probabilités à l'aide de la théorie de la mesure. Dans les années 1930, il a travaillé sur les processus de Markov et les processus stationnaires ; il s'est ensuite intéressé à la théorie de l'information.

Corollaire A.40. Soit une suite de probabilités $(\mu_n)_{n \in \mathbb{N}^*}$ sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ et soit, pour tout $n \in \mathbb{N}^*$, la probabilité produit $P_n = \otimes_{j=1}^n \mu_j$ sur $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, unique probabilité telle que l'on ait, pour tout pavé $\prod_{j=1}^n]a_j, b_j]$,

$$P_n \left(\prod_{j=1}^n]a_j, b_j] \right) = \prod_{j=1}^n \mu_j (]a_j, b_j]).$$

La suite de probabilités $(P_n)_{n \in \mathbb{N}}$ est consistante et il existe une unique probabilité P sur l'espace mesurable $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_{\mathbb{R}^{\mathbb{N}}})$ telle que l'on ait, pour tout $n \in \mathbb{N}$,

$$P \left[\left(\omega \in \mathbb{R}^{\mathbb{N}} \mid (\omega_1, \dots, \omega_n) \in \prod_{j=1}^n]a_j, b_j] \right) \right] = \prod_{j=1}^n \mu_j (]a_j, b_j]).$$

Cette probabilité est appelée **probabilité produit (infini)** des probabilités μ_n , $n \in \mathbb{N}^*$.

Index

A

- absorbant
 - point, 446, 448, 449, 470, 485, 493, 494
- absorbante
 - classe de communication, 446
- adapté (processus), 351
- algèbre, 517
- algèbre métrique, 94
- apériodique
 - chaîne, 438
 - classe, 438
- asymptotique
 - tribu, 48
- auto-covariance, 22
- auto-régressif (processus), 408, 412

B

- barrière
 - absorbante, 503
 - élastique, 482, 503
 - réfléchissante, 439, 483, 500, 503
- Beppo Levi, 6, 521
- Bernoulli (théorème de), 104
- Bernoulli-Laplace
 - modèle de diffusion, 397
- Bernstein (théorème de), 271
- bêta (loi), 28, 33, 72, 78, 172, 416, 513, 514
- Borel-Cantelli (lemme de)
 - application, 81
 - énoncé, 50
 - et convergence p.s., 82

C

- Cauchy (loi de), 13, 223
- Cauchy (suite de)
 - en probabilité, 90
- Cesàro (lemme de), 102
- Chaîne de Markov
 - définition, 405
 - homogène, 412
- changement (de variables), 8
- Chapman-Kolmogorov, 410, 411, 416, 421
- chi-deux
 - distance du, 318
 - loi du, 28, 68, 261, 262, 265, 284, 287, 316, 318
 - test du, 318

classe

- cyclique, 438
- de communication, 436
- fermée, 446
- Cochran (théorème de), 281
- conditionnelle
 - densité, 140
 - espérance (d'une classe de v.a. de L^1), 154
 - espérance (d'une classe de v.a. de L^2), 151
 - loi, 140
 - moyenne, 147
 - variance, 147
 - version de l'espérance, 151
- confiance
 - intervalle de, 263, 264
 - niveau de, 263
- conjugués (réels), 16
- convergence
 - en loi, 332–334, 338, 340, 343
 - en loi et fonctions de répartition, 308, 309, 314, 329, 339
 - en moyenne, 95
 - en moyenne quadratique, 95
 - en probabilité, 87, 116, 122, 125
 - étroite, 327, 330
 - L^p , 95
 - presque sûre, 87, 125
- convergence de martingale
 - bornée dans L^1 , 371
 - bornée dans L^2 , 362
- convergence en loi
 - vers une probabilité, 303
 - vers une v.a., 303
- convolution
 - d'une fonction et d'une mesure, 193
 - de fonctions, 62
 - de mesures, 61
- corrélation (coefficient de), 25
- covariance
 - de deux variables aléatoires, 20
 - matrice de, 22
- critère de récurrence, 459

D

- dégénérée (variable aléatoire), 225
- densité
 - d'une variable aléatoire, 9
 - mesure à densité, 7
- développement décimal, 338
- Dirichlet
 - loi de, 77, 169
 - problème de, 482, 483
 - théorème de, 213
- discrète (variable aléatoire), 10
- Doob
 - décomposition de, 366
 - deuxième théorème d'arrêt, 378
 - inégalité de, 361
 - inégalité maximale de, 360
 - premier théorème d'arrêt, 360

E

- écart-type (d'une variable aléatoire), 19
- échantillon, 253, 318
- effectif
 - observé, 318
 - théorique, 318
- Ehrenfest
 - modèle de diffusion, 399, 439, 465, 467
- empirique
 - coefficient de corrélation, 253
 - échantillon, 111, 286, 318
 - fonction de répartition, 111
 - mesure, 254
 - moyenne, 101, 253
 - variance, 253
- ensemble de μ -continuité, 295
- équi-continuité, 94
- équi-intégrabilité, 93, 94, 118
- estimateur
 - consistant, 320
 - de variance minimum, 255
 - du maximum de vraisemblance, 259, 321
 - linéaire, 255
 - sans biais, 255, 320
- estimation (paramétrique), 320
- événements antérieurs (tribu des), 355, 379
- exponentielle (loi), 31
 - caractérisation, 73
 - maximum, 66

F

- fermée
 - classe de communication, 446
- filtration, 344, 351
 - naturelle, 351
- Fisher (information de), 323
- fonction caractéristique
 - définition, 191
 - développement limité, 209
- fonction cumulative d'une mesure, 198
- fonction de répartition, 9
- forme quadratique
 - de v.a. gaussienne, 281
- Fubini (théorème de)
 - généralisé, 138

G

- Galton-Watson (processus de), 508, 509
- gamma (loi), 11, 28, 62, 72, 174
- Gauss-Markov (théorème de), 255
- gaussienne
 - loi, 11, 63, 65, 118, 177
 - loi de paramètres m et C , 237
 - loi sur \mathbb{R} , 235
 - loi sur un euclidien, 236
 - v.a. à valeurs dans un euclidien, 236
 - v.a. réelle, 236
- Glivenko-Cantelli (théorème de), 111
- grandes déviations
 - inégalité de Hoeffding, 127

H

- Hölder (inégalité de), 16, 34
- harmonique, 417
- Hotelling (loi de), 31
- hypergéométrique
 - approximation de la loi, 328
 - loi, 329

I

- inégalité
 - d'Ottaviani, 125
 - de Bienaymé-Tchebitchev, 24
 - de Doob, 361
 - de Hölder, 16, 34
 - de Hoeffding, 127
 - de Jensen, 162
 - de Kolmogorov, 98
 - de Markov, 23
 - de Minkowski, 17
 - de Schwarz, 16
 - maximale de Doob, 360

- indépendance
 asymptotique, 347
 conditionnelle, 402
 d'événements, 39
 de familles d'événements, 39
 de variables aléatoires, 40
- inessentiel, 447
- intercovariance, 35, 245, 252
- invariante
 mesure, 457
 probabilité, 457
- irréductible
 chaîne, 436
 classe, 436
- K**
- Kolmogorov
 inégalité de, 98
 test de Kolmogorov-Smirnov, 114
 théorème de prolongement, 47, 528
- Kronecker (lemme de), 102
- L**
- Lévy (théorème de)
 pour les suites de mesures, 299
 pour les suites de v.a., 304
 séries aléatoires, 125
 séries aléatoires convergence en loi,
 341
- λ -système, 3
- Laplace-Gauss (loi de), 235
- lemme
 d'unicité, 522
 de Borel-Cantelli, 50
 de Cesàro, 102
 de Fatou, 523
 de Fatou conditionnel, 159
 de Kronecker, 102
 de Schetté, 307
 de Slutsky, 327, 335
 maximal, 360
- loi (d'une variable aléatoire), 9
- loi faible des grands nombres
 ordre deux, 103
 ordre un (Khintchine), 105
- loi forte des grands nombres
 martingale, 367
 ordre deux, 107
 ordre un (Khintchine), 109
- M**
- marche aléatoire, 482, 500, 503
 conditionnelle, 412
 sur \mathbb{N} avec barrières, 482
 sur \mathbb{Z}^n , 350
 sur \mathbb{Z} , 405
- marginale (d'une variable aléatoire), 13
- Markov
 chaîne de, 182, 405
 inégalité de, 23, 363
 propriété de, 405, 418, 425
- martingale, 344
 arrêtée, 359
 bornée dans L^1 , 352
 bornée dans L^2 , 352
 dans L^2 , 352
 discrète, 352
 fermée, 352
 fermable, 352
 intégrable, 351
 sous-martingale, 351
 surmartingale, 351
- matrice
 de transition, 409, 412
 potentiel, 434
 stochastique, 409
- mesure
 absolument continue, 7
 égalité de deux mesures, 5
 mesure image, 8
 mesures étrangères, 7
- Minkowski (inégalité de), 17
- modèle génétique
 caractère dominant, 468
 martingale, 448
- modèle linéaire
 d'ordre deux, 254
 gaussien, 259
- moment, 33, 217
 d'une variable aléatoire réelle, 19
 moment centré d'une v.a.r., 19
- Monte-Carlo (méthode de), 123
- moyenne
 d'une variable aléatoire réelle, 18
 d'une variable aléatoire vectorielle, 21
- empirique, 101, 275, 282, 322
- multinomiale
 fonction caractéristique, 202
 loi, 79, 80, 164, 203, 316

N

- normale (loi), 11
- noyau
 - de probabilité, 135
 - gaussien, 193

O

- Ottaviani (inégalité d'), 124

P

- période
 - d'un point, 438
 - d'une classe, 438
- π -système, 3
- pile ou face, 380, 384, 387
- Poisson
 - processus de, 166, 167, 170, 172
 - théorème de, 310
 - théorème des événements rares, 311
- Pólya
 - modèle de diffusion, 400, 414, 513
 - processus de, 400
- potentiel, 440
- prédiction, 264
- probabilité
 - de transition, 135
 - invariante, 487
 - limité, 467
- processus
 - auto-régressif, 408
 - croissant prévisible, 366
 - de renouvellement, 350, 352
 - des accroissements, 365
 - prévisible, 366
 - processus auto-régressif, 412
- prolongement (par mesurabilité), 6

R

- Radon-Nikodym, 7
- récurrent, 440
 - nul, 440
 - positif, 440, 447
- régression
 - droite de régression estimée, 253
 - droite de régression linéaire, 251
 - linéaire, 26, 35, 149, 250
 - surface de régression linéaire, 37, 252
- rejet (zone de), 262, 319
- relation
 - de communication, 436
 - de conduction, 435

- risque d'erreur, 319
- ruine du joueur, 380

S

- Scheffé (lemme de), 307
 - réciroque (contre-exemple), 309
- Schwarz (inégalité de), 16
- section, 138
- semi-algèbre, 517
- σ -algèbre, 517
- signal (théorie du), 279, 280
- simulation, 29, 58
 - de la loi exponentielle, 31
 - de lois discrètes, 30
 - de variables gaussiennes, 65, 242
 - par rejet, 182
- Slutsky (lemme de), 327, 335
- statistiques d'ordre, 77
- Student
 - loi de, 68, 262, 263, 267, 287
 - test de, 262, 286, 287
- symétrisation (d'une v.a.), 84

T

- Tchebitchev (inégalité de), 24
- temps d'arrêt, 354
- temps d'attente, 141
- tendue (suite de mesures), 299
- test
 - de Kolmogorov-Smirnov, 114
 - de Student, 262, 286, 287
 - du chi-deux, 318, 319
- théorème
 - d'arrêt de Doob, 360, 378
 - d'existence de mesure gaussienne, 239
 - d'unicité des mesures, 5
 - de Bernoulli, 104
 - de Bernstein, 271
 - de Carathéodory, 520
 - de Chacon-Orstein, 470
 - de changement de variables, 8
 - de classification des états, 441
 - de Cochran, 281
 - de convergence dominée, 524, 525
 - de convergence L^2 , 362
 - de convergence monotone, 524
 - de Fubini, 527
 - de Fubini généralisé, 138
 - de Gauss-Markov, 255
 - de Glivenko-Cantelli, 111
 - de Jirina, 143

- de Karl Pearson, 315
- de Khintchine, 105
- de Lévy, 299, 304
- de Poisson, 310, 311
- de prolongement de Kolmogorov, 47, 528
- de Radon-Nikodym, 8
- de Riesz, 293
- de transfert, 8
- de transfert conditionnel, 144
- des trois séries, 121
- fondamental de la statistique, 111
- limite central, 81, 314
- topologie
 - étroite, 290
 - faible, 290
 - vague, 290
- Tout ou rien (loi du), 49
- transfert (théorème de), 8
- transfert conditionnel (théorème de), 144
- transformée de Fourier
 - d'une convolution, 201
 - d'une fonction, 193
 - d'une mesure, 191
 - injectivité, 193
- transitoire, 440
- tribu, 517
 - du futur, 405
 - du passé, 405
 - du présent, 405
 - engendrée par une application, 40

U

- uniforme (loi)
 - convolution, 213
 - en dimension d , 124
 - sur la sphère, 284
 - sur un intervalle, 28, 29, 54

V

- variable aléatoire
 - classe, 93
 - définie presque sûrement, 92
- variance, 35
 - d'une variable aléatoire réelle, 19
 - d'une variable aléatoire vectorielle, 22
 - empirique, 275, 282, 322
- variation quadratique, 369
- vraisemblance
 - équation de, 321
 - équation de log-, 321
 - fonction de, 259, 321



Achévé d'imprimer sur les presses de l'Imprimerie BARNÉOUD
53960 BONCHAMP-LÈS-LAVAL

Dépôt légal : novembre 2009 - N° d'imprimeur : 910085

Imprimé en France

Voici un ouvrage important, unique en son genre en français, qui présente l'ensemble de la théorie des probabilités telle qu'on l'enseigne au niveau du master et dans les préparations à l'agrégation : compléments de théorie de la mesure ; lois et moments de variables aléatoires ; indépendance de tribus et de variables aléatoires ; convergences, lois des grands nombres ; espérance conditionnelle ; transformation de Fourier et fonctions caractéristiques ; variables aléatoires gaussiennes ; convergence de mesures, convergence en loi ; processus discrets, martingales ; chaînes de Markov.

La lecture de ce livre ne suppose que des connaissances élémentaires en probabilités ; celles-ci sont exposées dans le tome I, où la théorie de la mesure n'est pas utilisée.

Le travail du lecteur sera facilité par la présence d'un grand nombre d'exercices, résolus de façon détaillée. Certains d'entre eux apportent au cours des compléments substantiels.

Conçu pour les candidats à l'agrégation, ce manuel sera aussi un instrument utile pour les étudiants de première année de master, ainsi que pour les étudiants plus avancés désireux d'approfondir leurs bases en probabilités.

Collection enseignement des mathématiques

l'essentiel **en théorie des** **probabilités**

Jean Jacod
Philip Protter

C A S S I N I

L'ESSENTIEL EN THÉORIE DES PROBABILITÉS

Enseignement des mathématiques

1. J.-Y. Oувrard, *Probabilités I*
2. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques*
3. M. Cottrell, V. Genon-Catalot, Ch. Duhamel, Th. Meyre, *Exercices de probabilités*
4. F. Rouvière, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*
5. J.-Y. Oувrard, *Probabilités II*
6. G. Zémor, *Cours de cryptographie*
7. A. Szpirglas, *Exercices d'algèbre*
8. B. Perrin-Riou, *Algèbre, arithmétique et Maple*
9. V. I. Arnold, *Leçons sur les équations aux dérivées partielles* (en préparation)
- 10-14. S. Francinou, H. Gianella, S. Nicolas, *Exercices de mathématiques — Oraux X-ENS*
15. H. Krivine, *Exercices de mathématiques pour physiciens*
16. J. Jacod, Ph. Protter, *L'essentiel en théorie des probabilités*
17. M. Willem, *Analyse fonctionnelle élémentaire*
18. É. Amar, É. Matheron, *Analyse complexe*
19. B. Randé, *Problèmes corrigés. Concours 2002 et 2003 (MP)* (en préparation)

JEAN JACOD PHILIP PROTTER

L'essentiel
en théorie des probabilités

CASSINI

JEAN JACOD est professeur à l'université Paris VI-Pierre et Marie Curie. Il est l'auteur de *Calcul stochastique et problèmes de martingales* (Springer, 1979, Lecture notes in mathematics 714), de *Limit theorems for stochastic processes*, en coll. avec Albert N. Shiryaev (Springer, 1987, Grundlehren der mathematischen Wissenschaften 288) et de *Malliavin calculus for processes with jumps*, en coll. avec Klaus Bichteler et Jean-Bernard Gravereaux (Gordon and Breach, 1987).

PHILIP PROTTER est professeur à Cornell University (Ithaca, NY, États-Unis). Il est l'auteur de *Stochastic integration and differential equations : a new approach* (Springer, 1990, Applications of mathematics 21) et de *Calculus with analytic geometry*, en coll. avec M.H. Protter (Jones and Bartlett, Boston, 1998).

ISBN 2-84225-050-8

© Cassini, Paris, 2003 (pour l'édition française)

Édition anglaise : *Probability Essentials*, 2nd ed., Springer, 2003, ISBN 3-540-43871-8

à

Diane et Sylvie

et à

Rachel, Margot,

Olivier, Martin, Serge, Thomas et Vincent

Préface

Nous présentons ici le contenu d'un cours de probabilité destiné aux étudiants de licence ou de master de mathématiques ou mathématiques appliquées : un tel cours peut raisonnablement être constitué des chapitres 1 à 23. Les derniers chapitres 24 à 28, en revanche, sont sans doute trop difficiles pour un tel cours, mais peuvent servir d'introduction à la théorie des processus stochastiques.

L'origine de ce livre se trouve dans les notes d'un cours enseigné à l'université Paris 6 par le premier auteur, puis reprises et largement modifiées par le second dans un cours enseigné d'abord à Pérouse en 1997, puis à l'université de Purdue. Nous tenons à remercier les étudiants de tous ces cours pour leurs commentaires. Nous remercions également notre éditeur Catriona Byrne pour l'édition anglaise, ainsi que Nick Bingham pour ses remarques et suggestions constructives, ainsi qu'un referee anonyme qui a contribué à rendre le texte plus lisible, et enfin Judy Mitchell pour la réalisation matérielle.

La version française est une traduction assez fidèle, mais améliorée (du moins l'espérons-nous) sur certains points. La plupart de ces améliorations ont été suggérées par André Bellaïche que nous remercions chaleureusement. La bibliographie a été élargie de façon à citer les ouvrages de référence en français, mais les références en langue anglaise ont été conservées.

Jean Jacod, Paris
Philip Protter, Ithaca

Table des matières

1. Introduction - Phénomènes aléatoires	1
2. Axiomes des probabilités	7
3. Probabilités conditionnelles et indépendance	15
4. Probabilités sur un espace fini ou dénombrable	23
5. Variables aléatoires sur un espace fini ou dénombrable	29
6. Construction d'une mesure de probabilité	37
7. Probabilités sur \mathbb{R} et fonctions de répartition	41
8. Variables aléatoires	51
9. Intégration par rapport à une mesure de probabilité	55
10. Variables aléatoires indépendantes	71
11. Loix de probabilité sur \mathbb{R}	83
12. Probabilités sur \mathbb{R}^n	93
13. Fonctions caractéristiques	109
14. Propriétés des fonctions caractéristiques	117
15. Sommes de variables aléatoires indépendantes	123
16. Variables aléatoires gaussiennes	131
17. Convergence des variables aléatoires	147
18. Convergence en loi	157
19. Convergence en loi et fonctions caractéristiques	173
20. La loi des grands nombres	179
21. Le théorème-limite central	187
22. L^2 et les espaces de Hilbert	195
23. Espérance conditionnelle	203
24. Martingales	217
25. Surmartingales et sous-martingales	225
26. Les inégalités de martingales	229
27. Les théorèmes de convergence de martingales	235
28. Le théorème de Radon-Nikodym	249
Bibliographie	255
Index des notations	257
Index	259

Chapitre 1

Introduction - Phénomènes aléatoires

Chacun est maintenant familier avec le concept de probabilité : on nous indique (en Amérique du Nord au moins) chaque jour la probabilité pour qu'il pleuve le lendemain ; on évalue la probabilité de gagner au loto ou de survivre à un accident d'avion. Les compagnies d'assurances calculent les probabilités qu'un homme ou une femme ayant atteint 20 ans soit encore vivant(e) à 80 ans. Les supermarchés calculent le nombre de caisses à ouvrir en fonction de la probabilité du nombre de clients. Les banques estiment la probabilité qu'un prêt ne puisse pas être honoré. En médecine on évalue les probabilités de succès de divers traitements. Un succès récent de la théorie des probabilités concerne la gestion de portefeuille par les compagnies boursières, en utilisant des modèles d'évolution aléatoire des cours de la bourse très sophistiqués. On pourrait poursuivre cette liste presque indéfiniment : les probabilités sont omniprésentes dans la société moderne, et dans l'ensemble des sciences.

La théorie des probabilités est un sujet relativement ancien. Les premières publications sur les jeux de hasard remontent à J. Cardan (1501-1576) avec son livre *De Ludo Aleae* [5], ou à Kepler (1571-1630) et Galilée (1564-1642). Cependant les historiens des sciences s'accordent à dater le véritable début de la théorie aux travaux de Pascal (1623-1662) et Fermat (1601-1665). Ces derniers ont échangé une correspondance fournie, résolvant certains « paradoxes » posés par le Chevalier de Méré pour des jeux de cartes ou de dés. Plus tard le mathématicien hollandais Christian Huygens (1629-1695) écrivit un livre [16] ayant eu une grande influence et développant les idées de Pascal et de Fermat. Finalement en 1685 Jacques Bernoulli (1654-1705) proposa de nombreux problèmes probabilistes (dans le « Journal des Sçavans » ; voir aussi [3]) nécessitant le développement d'une théorie élaborée. Après les travaux de J. Bernoulli et de son contemporain A. De Moivre (1667-1754) [9], de nombreux mathématiciens renommés de cette époque travaillèrent sur des questions de probabilité, notamment Daniel Bernoulli (1700-1782), Euler (1707-1803), Gauss (1777-1855), et Laplace (1749-1827) : on peut se référer à [24] pour une histoire documentée des probabilités à cette époque. Au cours du vingtième siècle, le mérite de la reconnaissance des rapports entre les

probabilités et la « théorie de la mesure » de Borel et Lebesgue revient à Kolmogorov (1903-1987). Après le travail fondamental de Kolmogorov, Paul Lévy a introduit nombre d'idées modernes sur les « processus aléatoires » et sur les fonctions caractéristiques et les théorèmes-limite.

Nous concevons la théorie des probabilités comme un modèle mathématique pour le « hasard », ou les « événements aléatoires ». L'idée consiste à partir d'un petit nombre de principes de base décrivant comment un phénomène aléatoire se comporte ; ces principes doivent être assez simples pour qu'on puisse légitimement considérer qu'ils décrivent de manière adéquate la réalité, ou la « nature ». Ils doivent aussi permettre l'élaboration d'une théorie mathématique suffisamment complexe pour résoudre des problèmes plus compliqués. C'est là le but de ce livre.

Décrivons maintenant l'approche adoptée. En premier lieu nous décrivons les principaux éléments des probabilités discrètes de façon à fonder les idées essentielles de la théorie. Ces idées sont ensuite étendues au cas des probabilités sur l'ensemble des réels : tout ceci occupe les chapitres 2 à 7. La notion de variable aléatoire est traitée de manière analogue, d'abord dans le cas discret (fini ou dénombrable), puis dans le cas général. Puis l'espérance mathématique des variables aléatoires est introduite et les rapports entre l'espérance et l'intégrale de Lebesgue clarifiés, dans un cadre « abstrait », puis dans les cas particuliers de \mathbb{R} (où \mathbb{R} est l'ensemble des réels) et de \mathbb{R}^n , ce qui nous amène au chapitre 12.

Les chapitres 13 à 21 sont consacrés aux théorèmes-limite, en commençant par les outils analytiques nécessaires et notamment les fonctions caractéristiques (ou transformées de Fourier) et par une étude assez complète des variables aléatoires gaussiennes (normales) multidimensionnelles.

Nous introduisons ensuite l'espérance conditionnelle, en utilisant la méthode des projections dans l'espace de Hilbert des variables de carré intégrable, ce qui nous amène à consacrer le chapitre 22 à un exposé rapide des propriétés de base des espaces de Hilbert. Nous prolongeons l'espérance conditionnelle à toutes les variables intégrables ou positives, dans le chapitre 23. Enfin, dans les chapitres 24 à 28 un aperçu de la théorie des martingales est esquissé, avec une application au théorème de Radon-Nikodym. Ces cinq derniers chapitres sont sans doute « hors sujet » dans un livre consacré aux bases des probabilités ; nous les avons néanmoins inclus car la plupart des applications actuelles des probabilités utilisent la théorie des martingales, et aussi parce que cette théorie est une introduction naturelle au sujet des processus stochastiques.

Quelques exercices proposent la démonstration de propriétés énoncées mais non démontrées dans le corps du texte ; beaucoup sont des

applications simples, permettant au lecteur de vérifier si les concepts sont bien assimilés ; enfin, certains proposent des extensions (en général assez simples) des résultats du texte lui-même. Les exercices flanqués d'un astérisque sont à notre avis les plus difficiles. Nous ne prétendons pour ces exercices à aucune originalité : la plupart se trouvent déjà dans d'autres livres ; un nombre appréciable d'entre eux est tiré du livre [14] de Allan Gut. Signalons aussi que la présentation de la théorie des martingales suit d'assez près celle du livre [1] de Richard Bass.

Aucune connaissance préalable des probabilités n'est requise, mais le lecteur est supposé maîtriser l'analyse élémentaire (séries, intégrale de Riemann, intégrales multiples) et un minimum d'algèbre linéaire.

Nous terminons cette introduction par une présentation générale – non mathématique – de la problématique et des principaux concepts des probabilités.

Les phénomènes aléatoires

Les phénomènes aléatoires sont des phénomènes dont on ne peut prévoir le résultat à l'avance, mais qui par « répétition » présentent un certain caractère de régularité. Un exemple typique est constitué par le jeu de pile ou face : on ne peut prévoir *a priori* le résultat d'un tirage particulier, mais si on fait un grand nombre de tirages on obtiendra une moyenne d'à peu près 50% de « pile », si la pièce n'est pas truquée.

La théorie des probabilités vise à fournir un modèle mathématique pour décrire ces phénomènes. Cette théorie contient trois ingrédients essentiels :

a) **L'espace d'états** : C'est l'ensemble, noté habituellement Ω , de tous les résultats possibles de l'expérience (aléatoire) qu'on réalise.

Exemples.

- 1) Un tirage à pile ou face : $\Omega = \{p, f\}$.
- 2) Deux tirages successifs à pile ou face : $\Omega = \{pp, pf, fp, ff\}$.
- 3) Tirage de deux dés : $\Omega = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$.
- 4) Mesure d'une longueur L , avec une erreur de mesure : $\Omega = \mathbb{R}_+$; $\omega \in \Omega$ désigne le résultat de la mesure, et $\omega - L$ est l'erreur de mesure.
- 5) Durée de vie d'une ampoule électrique : $\Omega = \mathbb{R}_+$.

b) Les événements : Un événement est une propriété dont on peut dire si elle est vraie ou non, une fois l'expérience réalisée : lors du tirage de deux dés, par exemple, « le second dé donne 4 ou 5 », ou « le résultat du second dé est inférieur à celui du premier » sont des événements. En termes mathématiques, un événement est représenté par l'ensemble des résultats pour lesquels il est réalisé, c'est-à-dire par une partie de Ω . Si A et B sont deux événements,

- l'événement **contraire** de A est représenté par le complémentaire A^c ;
- l'événement « **A ou B** » est représenté par la réunion $A \cup B$;
- l'événement « **A et B** » est représenté par l'intersection $A \cap B$;
- l'événement **certain** est Ω ;
- l'événement **impossible** est l'ensemble vide \emptyset ;
- un **événement élémentaire** est un « singleton », i.e. une partie $\{\omega\}$ ne contenant qu'un seul point ω de Ω .

On note \mathcal{A} l'ensemble de tous les événements. Souvent (mais pas toujours : on verra pourquoi plus loin) \mathcal{A} est l'ensemble, noté $\mathcal{P}(\Omega)$ ou 2^Ω , de toutes les parties de Ω . En tous cas, \mathcal{A} doit être « stable » par les opérations logiques décrites ci-dessus : si $A, B \in \mathcal{A}$, alors on doit avoir $A^c \in \mathcal{A}$, $A \cap B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$, et aussi $\Omega \in \mathcal{A}$ et $\emptyset \in \mathcal{A}$.

c) La probabilité : À chaque événement A on associe un nombre, noté $P(A)$ et appelé « probabilité de A ». Ce nombre mesure le degré de vraisemblance qu'on accorde *a priori* à A , avant la réalisation de l'expérience. Il est choisi entre 0 et 1, et il est d'autant plus près de 1 que l'événement est jugé plus vraisemblable.

Pour avoir une idée des propriétés de ces nombres, on peut imaginer la probabilité comme limite de « fréquences » : répétons n fois la même expérience ; les n résultats obtenus peuvent bien sûr être différents (penser à n jets successifs d'un même dé, par exemple). Notons $f_n(A)$ la fréquence de réalisation de l'événement A (i.e. le nombre de fois où il est réalisé, divisé par n). Alors, « intuitivement » on a :

$$P(A) = \text{limite de } f_n(A) \text{ quand } n \uparrow +\infty. \quad (1.1)$$

(on donnera un sens précis à cette « limite » plus tard). Des propriétés évidentes des fréquences, on déduit immédiatement que :

1. $0 \leq P(A) \leq 1, -$
2. $P(\Omega) = 1, -$

3. $P(A \cup B) = P(A) + P(B)$ si $A \cap B = \emptyset$. ✕

Un *espace de probabilité*, ou *espace probabilisé*, est donc un triplet (Ω, \mathcal{A}, P) , constitué de l'espace Ω , de l'ensemble des événements \mathcal{A} , et de la famille des $P(A)$ pour $A \in \mathcal{A}$: on peut ainsi considérer P comme une application de \mathcal{A} dans $[0, 1]$, qui vérifie au moins les propriétés (2) et (3) ci-dessus (plus une propriété supplémentaire, plus délicate à appréhender, et qui sera expliquée au chapitre suivant).

Une quatrième notion, importante également, quoique moins fondamentale, est celle de :

d) Variable aléatoire : Il s'agit là d'une grandeur qui dépend du résultat de l'expérience. En termes mathématiques, c'est une application de Ω dans un espace E , souvent $E = \mathbb{R}$ ou $E = \mathbb{R}^d$. **Attention :** cette terminologie, consacrée par l'usage, est assez malencontreuse : une « variable » aléatoire n'est pas une variable (au sens de l'analyse), mais une fonction ! c'est en fait une terminologie apparentée à la notion de variable en physique ou en sciences humaines, domaines où on désigne volontiers par « variable » la valeur prise par telle ou telle fonction de l'état du système.

Soit X une telle variable aléatoire, qui applique Ω dans E . On peut alors « transporter » la structure probabiliste sur l'espace d'arrivée E , en posant

$$P_X(B) = P(X^{-1}(B)) \quad \text{pour } B \subset E,$$

où $X^{-1}(B)$ désigne l'image réciproque de B par X , c'est-à-dire l'ensemble des $\omega \in \Omega$ tels que $X(\omega) \in B$. L'événement $X^{-1}(B)$ correspond donc exactement à la propriété « $X \in B$ », et on écrit en général $P(X \in B)$ au lieu de $P(X^{-1}(B))$. Cette formule définit une nouvelle probabilité, notée P_X , cette fois-ci sur l'espace E (au lieu de Ω). Cette probabilité P_X s'appelle la **loi de la variable** X .

Exemple (tirage de deux dés). On a vu que $\Omega = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$, et il est naturel de prendre ici $\mathcal{A} = 2^\Omega$, et

$$P(A) = \frac{\text{card}(A)}{36} \quad \text{si } A \subset \Omega,$$

où $\text{card}(A)$ désigne le nombre de points contenus dans A . On vérifie aisément les propriétés (1), (2), (3) ci-dessus, et on a $P(\{\omega\}) = \frac{1}{36}$ pour chaque singleton. L'application $X : \Omega \rightarrow \mathbb{N}$ (où \mathbb{N} est l'ensemble des entiers positifs ou nul) définie par $X(i, j) = i + j$ est la variable aléatoire « somme des deux dés », de loi

$$P_X(B) = \frac{\text{nombre de couple } (i, j) \text{ tels que } i + j \in B}{36}$$

(par exemple $P_X(\{2\}) = P_X(\{12\}) = \frac{1}{36}$, $P_X(\{3\}) = \frac{2}{36}$, etc.).

Nous allons formaliser la notion d'espace de probabilité au chapitre 2, tandis que les variables aléatoires seront introduites rigoureusement dans les chapitres 5 et 8.

Chapitre 2

Axiomes des probabilités

Nous commençons par présenter ci-dessous les propriétés minimales qui nous sont nécessaires pour définir une probabilité. Nous espérons que le lecteur se convaincra que les deux axiomes dans la définition 2.3 ci-dessous sont raisonnables, notamment au vu de l'approche par les fréquences évoquée lors de l'introduction de la formule (1.1), et la théorie toute entière découlera de ces deux axiomes simples. Toutefois, avant de présenter nous devons introduire le concept de tribu.

Soit Ω un espace « abstrait », c'est-à-dire sans structure particulière. Rappelons que $\mathcal{P}(\Omega)$ ou 2^Ω désigne l'ensemble de tous les sous-ensembles de Ω , y compris le sous-ensemble vide noté \emptyset . Étant donnée une partie \mathcal{A} de 2^Ω , on considère les propriétés suivantes :

- (1) $\emptyset \in \mathcal{A}$ et $\Omega \in \mathcal{A}$;
- (2) \mathcal{A} est *stable par complémentation* : si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$, où A^c désigne le complémentaire de A (sous-entendu : dans Ω) ;
- (3) \mathcal{A} est *stable pour les réunions finies et les intersections finies* : si les A_i sont tous dans \mathcal{A} , alors $\bigcup_{i=1}^n A_i$ et $\bigcap_{i=1}^n A_i$ sont aussi dans \mathcal{A} (il suffit pour cela que \mathcal{A} soit stable par réunion et intersection de deux éléments quelconques) ;
- (4) \mathcal{A} est stable pour les réunions dénombrables et les intersections dénombrables : si les A_i sont tous dans \mathcal{A} , alors $\bigcup_{i=1}^{\infty} A_i$ et $\bigcap_{i=1}^{\infty} A_i$ sont aussi dans \mathcal{A} :

Définition 2.1. \mathcal{A} est une algèbre si elle vérifie (1), (2) et (3) ci-dessus. C'est une tribu (ou une σ -algèbre) si elle vérifie (1), (2) et (4) ci-dessus.

Noter que si on a (2), alors (1) peut être remplacé soit par (1') : $\emptyset \in \mathcal{A}$, soit par (1'') : $\Omega \in \mathcal{A}$. Noter aussi que (1)+(4) implique (3), de sorte que toute tribu est une algèbre (mais une algèbre peut ne pas être une tribu : voir l'exercice 17).

Définition 2.2. Si $\mathcal{C} \subset 2^\Omega$, la tribu engendrée par \mathcal{C} et notée $\sigma(\mathcal{C})$ est la plus petite tribu contenant \mathcal{C} . (Elle existe toujours ; en effet l'intersection d'une famille quelconque de tribus est encore une tribu : voir l'exercice 2 ; on applique alors ce résultat à la famille de toutes les tribus contenant \mathcal{C} , famille qui contient au moins la tribu 2^Ω .)

Exemples.

- (i) $\mathcal{A} = \{\emptyset, \Omega\}$ (la tribu « triviale »).
- (ii) Si $A \subset \Omega$, alors $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$.
- (iii) Si $\Omega = \mathbb{R}$ (ou plus généralement si Ω est un espace topologique, un cas que nous traiterons au chapitre 8), la *tribu borélienne*, ou de Borel, est la tribu \mathcal{B} engendrée par la classe des ouverts (ou de manière équivalente, par la classe des fermés); une partie de \mathbb{R} qui est dans la tribu borélienne s'appelle un *borélien*.

Théorème 2.1. *La tribu borélienne \mathcal{B} de \mathbb{R} est engendrée par les intervalles de la forme $] - \infty, a]$, où $a \in \mathbb{Q}$ (\mathbb{Q} = ensemble des rationnels).*

Preuve. Soit \mathcal{C} la classe de tous les intervalles ouverts. Tout ouvert A de \mathbb{R} est réunion dénombrable d'intervalles ouverts : en effet pour tout $x \in A$ on note $I_x =]a_x, b_x[$ le plus grand intervalle ouvert contenant x et inclus dans A (on prend $b_x = \sup\{y : y > x, [x, y] \subset A\}$ et $a_x = \inf\{y : y < x,]y, x] \subset A\}$); alors A est la réunion des I_x lorsque x décrit l'ensemble des rationnels contenus dans A . Cette propriété, jointe à l'axiome (4), entraîne que $\sigma(\mathcal{C}) = \mathcal{B}$, tribu borélienne de \mathbb{R} .

Soit \mathcal{D} la classe des intervalles de la forme $] - \infty, a]$, où $a \in \mathbb{Q}$. Soit $]a, b[\in \mathcal{C}$: il existe des rationnels $(a_n)_{n \geq 1}$ décroissant vers a et des rationnels $(b_n)_{n \geq 1}$ croissant strictement vers b , de sorte que

$$\begin{aligned}]a, b[&= \bigcup_{n=1}^{\infty}]a_n, b_n[\\ &= \bigcup_{n=1}^{\infty} (] - \infty, b_n] \cap] - \infty, a_n]^c). \end{aligned}$$

Par suite $\mathcal{C} \subset \sigma(\mathcal{D})$, donc $\sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$. Par ailleurs chaque élément de \mathcal{D} est un fermé, donc un borélien, ce qui entraîne $\sigma(\mathcal{D}) \subset \mathcal{B}$. Le résultat découle alors de :

$$\mathcal{B} = \sigma(\mathcal{C}) \subset \sigma(\mathcal{D}) \subset \mathcal{B}. \quad \blacksquare$$

Sur l'espace d'états Ω , la classe des événements sera toujours une tribu \mathcal{A} : les axiomes (1), (2) et (3) correspondent aux opérations « logiques » décrites dans le chapitre 1. L'axiome (4) est nécessaire pour assurer la cohérence interne de la théorie (par exemple, pour donner un sens mathématique précis à la convergence (1.1)), et aussi pour assurer sa richesse mathématique (par exemple si une suite (A_n) d'événements

« converge » vers une limite A , en un sens précisé plus loin, alors on veut que A soit aussi un événement).

On peut maintenant définir la notion de probabilité :

Définition 2.3. Une mesure de probabilité, ou simplement une probabilité, définie sur une tribu \mathcal{A} d'un espace Ω , est une application $P : \mathcal{A} \rightarrow [0, 1]$ qui vérifie :

1. $P(\Omega) = 1$.
2. Pour toute suite A_1, A_2, A_3, \dots d'éléments de \mathcal{A} qui sont deux à deux disjoints (i.e. $A_n \cap A_m = \emptyset$ si $n \neq m$), on a

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Le nombre $P(A)$ s'appelle la *probabilité* de l'événement A . L'axiome 2 ci-dessus s'appelle *σ -additivité*; voir l'exercice 9 pour un affaiblissement de l'axiome 1.

Dans cette définition, on pourrait imaginer une condition plus simple que 2, à savoir que

$$A, B \in \mathcal{A}, A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B). \quad (2.1)$$

Cette condition s'appelle *l'additivité* et, par une récurrence élémentaire, on voit qu'elle entraîne que pour toute famille finie A_1, A_2, \dots, A_m d'éléments deux à deux disjoints de \mathcal{A} , on a

$$P\left(\bigcup_{n=1}^m A_n\right) = \sum_{n=1}^m P(A_n).$$

Théorème 2.2. Si P est une probabilité sur la tribu \mathcal{A} , alors :

- (i) on a $P(\emptyset) = 0$;
- (ii) P est additive.

Preuve. Si dans l'axiome 2 on prend $A_n = \emptyset$ pour tout n , on voit que le nombre $a = P(\emptyset)$ est égal à $a + a + a + \dots$; comme par ailleurs $0 \leq a \leq 1$, ce n'est possible que si $a = 0$, et on a (i). Pour (ii) il suffit d'appliquer l'axiome 2 avec $A_1 = A$, $A_2 = B$ et $A_3 = A_4 = \dots = \emptyset$, plus le fait que $P(\emptyset) = 0$, pour obtenir (2.1). ■

À l'inverse, la σ -additivité n'est pas impliquée par l'additivité. En fait l'additivité, malgré son caractère intuitif, n'est pas suffisante pour

traiter mathématiquement les problèmes de la théorie, même dans des cas aussi simples qu'une suite de jets de dés, comme nous le verrons plus loin.

Le théorème ci-dessous exprime ce qu'il faut exactement ajouter à l'additivité pour obtenir la σ -additivité. Afin de rendre son énoncé plus clair, nous présentons d'abord deux conséquences immédiates de l'additivité et de l'axiome $P(\Omega) = 1$:

$$A \in \mathcal{A} \Rightarrow P(A^c) = 1 - P(A) ;$$

$$A, B \in \mathcal{A}, A \subset B \Rightarrow P(A) \leq P(B)$$

(pour la première, appliquer (2.1) à A et $B = A^c$, et utiliser $P(\Omega) = 1$; pour la seconde appliquer (2.1) à A et $C = A^c \cap B$, donc $A \cap C = \emptyset$ et $A \cup C = B$).

Ci-dessous on note $A_n \downarrow A$ et on dit que A_n décroît (ou « tend en décroissant ») vers A si $A_n \supset A_{n+1}$ pour tout n et si $\bigcap_n A_n = A$. On note $A_n \uparrow A$ et on dit que A_n croît (ou « tend en croissant ») vers A si $A_n \subset A_{n+1}$ pour tout n et si $\bigcup_n A_n = A$. Noter que si les A_n sont dans la tribu \mathcal{A} , il en est de même de A dans les deux cas ci-dessus.

Théorème 2.3. Soit \mathcal{A} une tribu de Ω . Supposons que $P : \mathcal{A} \rightarrow [0, 1]$ satisfasse l'axiome 1 et soit additive. Les assertions suivantes sont alors équivalentes :

- (i) on a l'axiome 2 (σ -additivité) ;
- (ii) si $A_n \in \mathcal{A}$ et $A_n \downarrow \emptyset$, alors $P(A_n) \downarrow 0$;
- (iii) si $A_n \in \mathcal{A}$ et $A_n \downarrow A$, alors $P(A_n) \downarrow P(A)$;
- (iv) si $A_n \in \mathcal{A}$ et $A_n \uparrow \Omega$, alors $P(A_n) \uparrow 1$;
- (v) si $A_n \in \mathcal{A}$ et $A_n \uparrow A$, alors $P(A_n) \uparrow P(A)$.

Preuve. Observons que $A_n \downarrow A$ implique $A_n^c \uparrow A^c$. Comme $P(A_n^c) = 1 - P(A_n)$ et $P(A^c) = 1 - P(A)$, on voit que (ii) et (iv) sont équivalents, ainsi que (iii) et (v). L'implication (v) \Rightarrow (iv) est évidente.

Supposons maintenant (iv). Soit $A_n \in \mathcal{A}$ avec $A_n \uparrow A$. Posons $B_n = A_n \cup A^c$. Alors B_n croît vers Ω , de sorte que $P(B_n) \uparrow 1$. Comme $A_n \subset A$, on a $A_n \cap A^c = \emptyset$, donc $P(A_n \cup A^c) = P(A_n) + P(A^c)$ et il vient

$$1 = \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} \{P(A_n) + P(A^c)\}.$$

Par suite $\lim_{n \rightarrow \infty} P(A_n) = 1 - P(A^c) = P(A)$, et on a (v).

Il reste à montrer que (i) et (v) sont équivalents. Supposons d'abord (v). Soit des $A_n \in \mathcal{A}$ deux à deux disjoints, et définissons $B_n =$

$\bigcup_{1 \leq p \leq n} A_p$ et $B = \bigcup_{n=1}^{\infty} A_n$. L'additivité implique $P(B_n) = \sum_{p=1}^n P(A_p)$, qui croît vers $\sum_{n=1}^{\infty} P(A_n)$ quand $n \rightarrow \infty$, et croît aussi vers $P(B)$ par (v), donc

$$P(B) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n),$$

et (i) est prouvé.

Finalement, supposons (i). Soit $A_n \in \mathcal{A}$, avec A_n croissant vers A . On construit une nouvelle suite d'éléments de \mathcal{A} ainsi :

$$\begin{aligned} B_1 &= A_1, \\ B_2 &= A_2 \setminus A_1 = A_2 \cap (A_1^c), \\ &\vdots \\ B_n &= A_n \setminus A_{n-1}. \end{aligned}$$

On a $\bigcup_{n=1}^{\infty} B_n = A$ et les $(B_n)_{n \geq 1}$ sont deux à deux disjoints. Donc (i) entraîne

$$P(A) = \lim_{n \rightarrow \infty} \sum_{p=1}^n P(B_p).$$

On a aussi $\sum_{p=1}^n P(B_p) = P(A_n)$, donc $\lim_{n \rightarrow \infty} P(A_n) = P(A)$ et on a (v). ■

Si A est une partie de Ω , on définit sa *fonction indicatrice* par

$$1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

On dit que A_n converge vers A (et on écrit $A_n \rightarrow A$) si $\lim_{n \rightarrow \infty} 1_{A_n}(\omega) = 1_A(\omega)$ pour tout $\omega \in \Omega$. Noter que si la suite A_n croît (resp. décroît) vers A , elle tend aussi vers A au sens ci-dessus.

On associe aussi à toute suite A_n de parties de Ω les ensembles suivants :

$$\left. \begin{aligned} \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m, \\ \liminf_{n \rightarrow \infty} A_n &= \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m. \end{aligned} \right\} \quad (2.2)$$

Par suite $A_n \rightarrow A$ si et seulement si $A = \limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n$: en effet, l'indicatrice d'une réunion (resp. d'une intersection) d'ensembles est le sup (resp. l'inf) des indicatrices, donc

$$1_{\limsup_{n \rightarrow \infty} A_n} = \limsup_{n \rightarrow \infty} 1_{A_n}, \quad 1_{\liminf_{n \rightarrow \infty} A_n} = \liminf_{n \rightarrow \infty} 1_{A_n},$$

et par ailleurs on sait qu'une suite de réels u_n converge vers une limite u si et seulement si $u = \limsup_{n \rightarrow \infty} u_n = \liminf_{n \rightarrow \infty} u_n$.

Noter que $\limsup_{n \rightarrow \infty} A_n$ est l'ensemble des ω qui appartiennent à une infinité d'ensembles A_n , tandis que $\liminf_{n \rightarrow \infty} A_n$ est l'ensemble des ω qui appartiennent à tous les A_n sauf au plus à un nombre fini d'entre eux.

Théorème 2.4. Soit P une probabilité et A_n une suite d'éléments de \mathcal{A} qui converge vers A . Alors $A \in \mathcal{A}$ et $\lim_{n \rightarrow \infty} P(A_n) = P(A)$.

Preuve. \mathcal{A} étant une tribu, on a $\limsup_{n \rightarrow \infty} A_n \in \mathcal{A}$ et $\liminf_{n \rightarrow \infty} A_n \in \mathcal{A}$ (voir l'exercice 4). Comme $A_n \rightarrow A$ on a donc aussi $A \in \mathcal{A}$.

Ensuite, soit $B_n = \bigcap_{m \geq n} A_m$ et $C_n = \bigcup_{m \geq n} A_m$. Alors B_n croît vers A et C_n décroît vers A , donc $\lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} P(C_n) = P(A)$ d'après le théorème 2.3. Mais $B_n \subset A_n \subset C_n$, donc $P(B_n) \leq P(A_n) \leq P(C_n)$, de sorte que $\lim_{n \rightarrow \infty} P(A_n) = P(A)$. ■

Exercices

1. Soit Ω un ensemble fini. Montrer que la tribu 2^Ω de toutes les parties de Ω est également finie.
2. Soit $(\mathcal{G}_\alpha)_{\alpha \in A}$ une famille quelconque de tribus sur l'ensemble Ω . Montrer que $\mathcal{H} = \bigcap_{\alpha \in A} \mathcal{G}_\alpha$ est aussi une tribu.
3. Soit $(A_n)_{n \geq 1}$ une suite de parties de Ω . Montrer que
 - a) $(\bigcup_{n=1}^{\infty} A_n)^c = \bigcap_{n=1}^{\infty} A_n^c$
 - b) $(\bigcap_{n=1}^{\infty} A_n)^c = \bigcup_{n=1}^{\infty} A_n^c$.
4. Soit \mathcal{A} une tribu et $(A_n)_{n \geq 1}$ une suite d'éléments de \mathcal{A} . Avec les notations (2.2), montrer que

$$\liminf_{n \rightarrow \infty} A_n \in \mathcal{A}; \quad \limsup_{n \rightarrow \infty} A_n \in \mathcal{A}; \quad \liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n.$$

5. Soit $(A_n)_{n \geq 1}$ une suite de parties de Ω . Montrer que

$$\limsup_{n \rightarrow \infty} 1_{A_n} - \liminf_{n \rightarrow \infty} 1_{A_n} = 1_{\{\limsup_n A_n \setminus \liminf_n A_n\}}$$

(avec la notation $A \setminus B = A \cap B^c$ si $B \subset A$).

6. Soit \mathcal{A} une tribu sur Ω et $B \in \mathcal{A}$. Montrer que $\mathcal{F} = \{A \cap B : A \in \mathcal{A}\}$ est une tribu sur l'ensemble B (appelée la « tribu trace » de \mathcal{A} sur B). Est-ce encore vrai si B est une partie de Ω qui n'appartient pas à \mathcal{A} ?
7. Soit f une application de Ω dans un autre espace E muni lui aussi d'une tribu \mathcal{E} . Soit $\mathcal{A} = \{A \subset \Omega : \text{il existe } B \in \mathcal{E} \text{ avec } A = f^{-1}(B)\}$. Montrer que \mathcal{A} est une tribu de Ω .
8. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue, et soit $\mathcal{A} = \{A \subset \mathbb{R} : \text{il existe } B \in \mathcal{B} \text{ avec } A = f^{-1}(B)\}$, où \mathcal{B} est la tribu borélienne de l'espace d'arrivée \mathbb{R} . Montrer que $\mathcal{A} \subset \mathcal{B}$, tribu borélienne de l'espace de départ \mathbb{R} .

Pour les problèmes 9 à 16 nous supposons donnés un espace Ω , une tribu \mathcal{A} de cet espace, et une probabilité P sur cette tribu; les parties A, B, A_i , etc. sont toutes supposées appartenir à \mathcal{A} .

9. Si $A \cap B = \emptyset$, montrer que $P(A \cup B) = P(A) + P(B)$.
10. Montrer que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
11. Montrer que $P(A) = 1 - P(A^c)$.
12. Montrer que $P(A \cap B^c) = P(A) - P(A \cap B)$.
13. (Identité de Poincaré.) Montrer que

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_i P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n)$$

où, par exemple, $\sum_{i < j}$ signifie la somme pour toutes les paires d'indices (i, j) avec $1 \leq i < j \leq n$.

14. Montrer que si $P(A) = \frac{3}{4}$ et $P(B) = \frac{1}{3}$, on a nécessairement $\frac{1}{12} \leq P(A \cap B) \leq \frac{1}{3}$.
15. (Sous-additivité.) Montrer que

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i),$$

pour chaque n , et aussi que

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i).$$

16. (Inégalités de Bonferroni.) Montrer que

$$\text{a) } P\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j),$$

$$\text{b) } P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k).$$

17. Soit Ω un ensemble infini (dénombrable ou non) et \mathcal{A} l'ensemble des parties de Ω qui sont finies, ou de complémentaire fini. Montrer que \mathcal{A} est une algèbre, mais pas une tribu.

Chapitre 3

Probabilités conditionnelles et indépendance

Soit A et B deux événements d'un espace probabilisé; rappelons « l'approche par les fréquences » : la fréquence $f_n(A)$ de réalisation de l'événement A lorsqu'on répète n fois la même expérience converge (en un sens à préciser !) vers la probabilité $P(A)$, qui quantifie « les chances de voir A réalisé ».

Supposons maintenant qu'on sache que l'événement B est réalisé. Les chances de voir A réalisé vont changer et être quantifiées par un nouveau nombre $P(A|B)$, la « probabilité de A sachant B » : on peut à nouveau considérer la fréquence de A lorsque l'expérience est répétée n fois, sauf qu'il faut calculer cette fréquence sur l'ensemble des expériences où B se réalise. Il est ainsi naturel de considérer le nombre de fois où A et B sont réalisés, c'est-à-dire $nf_n(A \cap B)$; pour obtenir une fréquence, il convient de diviser ce nombre par le nombre d'occurrences de B , i.e. $nf_n(B)$, de sorte qu'en définitive on doit avoir

$$P(A|B) \approx \frac{nf_n(A \cap B)}{nf_n(B)} = \frac{f_n(A \cap B)}{f_n(B)}.$$

La définition 3.2 ci-après est tirée de cette relation en « prenant la limite en n ».

En particulier, il se peut que le fait de savoir que B est réalisé ne donne aucune information sur la réalisation de A : on dit alors que A est « indépendant » de B , et cela se traduit par le fait que $P(A|B) = P(A)$. Compte tenu de la définition 3.2, cela se traduit par

$$\frac{P(A \cap B)}{P(B)} = P(A), \quad \text{donc} \quad P(A \cap B) = P(A)P(B).$$

Nous allons maintenant introduire les définitions proprement dites, en commençant par l'indépendance.

Définition 3.1. (a) Deux événements A et B sont dits indépendants s'ils vérifient $P(A \cap B) = P(A)P(B)$.

(b) Soit une famille (finie ou infinie) d'événements $(A_i)_{i \in I}$. Ces événements sont dits indépendants, ou parfois « mutuellement indépendants », si pour toute partie finie J de I on a

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i).$$

Attention : si les $(A_i)_{i \in I}$ sont indépendants, ils sont aussi deux à deux indépendants, ce qui signifie que A_i et A_j sont indépendants pour tous i, j avec $i \neq j$, mais la réciproque est fautive.

Théorème 3.1. Si A et B sont indépendants, il en est de même de A et B^c , de A^c et B , et de A^c et B^c .

Preuve. Pour A et B^c , on a

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) \\ &= P(A)P(B^c). \end{aligned}$$

Les autres propriétés se montrent de manière analogue. ■

Exemples.

- On jette 3 fois une pièce. Si A_i est l'événement « le i -ième jet donne pile » il est habituel de choisir la probabilité de sorte que les événements A_1, A_2 et A_3 soient indépendants.
- On choisit une carte au hasard parmi 52 cartes. $A = \{\text{la carte est un cœur}\}$, et $B = \{\text{la carte est une dame}\}$. Un modèle naturel pour cette expérience consiste à affecter la probabilité $\frac{1}{52}$ au choix de chacune des cartes. Par additivité, $P(A) = \frac{13}{52}$ et $P(B) = \frac{4}{52}$ et $P(A \cap B) = \frac{1}{52}$, donc A et B sont indépendants.
- Soit $\Omega = \{1, 2, 3, 4\}$ et $\mathcal{A} = 2^\Omega$. Soit $P(i) = \frac{1}{4}$ pour $i = 1, 2, 3, 4$. Soit enfin $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{2, 3\}$. Alors A, B, C sont deux à deux indépendants, mais pas (mutuellement) indépendants.

Définition 3.2. Soit A, B deux événements. avec $P(B) > 0$. La probabilité conditionnelle de A sachant B est $P(A|B) = P(A \cap B)/P(B)$.

Théorème 3.2. Supposons que $P(B) > 0$.

- A et B sont indépendants si et seulement si $P(A|B) = P(A)$.
- L'application $A \rightarrow P(A|B)$ de \mathcal{A} dans $[0, 1]$ définit une nouvelle mesure de probabilité sur \mathcal{A} , appelée la « probabilité conditionnelle sachant B ».

Preuve. (a) résulte d'un calcul immédiat. Pour (b), posons $Q(A) = P(A|B)$, avec B fixé. On doit montrer que Q vérifie (1) et (2) de la définition 2.3. Mais (1) provient de

$$Q(\Omega) = P(\Omega | B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Quant à (2), on observe que si $(A_n)_{n \geq 1}$ est une suite d'éléments de \mathcal{A} qui sont deux à deux disjoints, alors

$$Q\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} A_n \mid B\right) = \frac{P\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{n=1}^{\infty} (A_n \cap B)\right)}{P(B)}$$

et les $(A_n \cap B)_{n \geq 1}$ sont également deux à deux disjoints; donc

$$= \sum_{n=1}^{\infty} \frac{P(A_n \cap B)}{P(B)} = \sum_{n=1}^{\infty} P(A_n | B) = \sum_{n=1}^{\infty} Q(A_n). \quad \blacksquare$$

Les trois résultats suivants, quoique élémentaires, sont extrêmement utiles, et en particulier le second d'entre eux.

Théorème 3.3. (Théorème des probabilités composées.) Si $A_1, \dots, A_n \in \mathcal{A}$ et si $P(A_1 \cap \dots \cap A_{n-1}) > 0$, alors

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \dots P(A_n | A_1 \cap \dots \cap A_{n-1}).$$

Preuve. La preuve se fait par récurrence. Pour $n = 2$, le résultat est simplement la définition 3.2. Supposons le résultat correct pour $n - 1$ événements. Soit $B = A_1 \cap \dots \cap A_{n-1}$. La définition 3.2 donne $P(B \cap A_n) = P(A_n | B)P(B)$; puis on remplace $P(B)$ par sa valeur, donnée par l'hypothèse de récurrence :

$$P(B) = P(A_1)P(A_2 | A_1) \dots P(A_{n-1} | A_1 \cap \dots \cap A_{n-2}),$$

et le résultat s'ensuit. \blacksquare

Une famille $(E_i)_{i \in I}$ de parties de Ω est appelée une *partition* si les E_i sont deux à deux disjointes et si $\bigcup_{i \in I} E_i = \Omega$.

Théorème 3.4. (Formule des probabilités totales.) Soit (E_n) une partition finie ou dénombrable de Ω , constituée d'éléments de la tribu \mathcal{A} . Pour tout $A \in \mathcal{A}$ on a

$$P(A) = \sum_n P(A | E_n)P(E_n).$$

Preuve. Observons que

$$A = A \cap \Omega = A \cap \left(\bigcup_n E_n \right) = \bigcup_n (A \cap E_n).$$

Les E_n étant deux à deux disjoints, il en est de même des $(A \cap E_n)_{n \neq 1}$, donc

$$P(A) = P\left(\bigcup_n (A \cap E_n)\right) = \sum_n P(A \cap E_n) = \sum_n P(A | E_n)P(E_n). \quad \blacksquare$$

Exemple. On dispose de deux urnes, la première contenant 1 boule blanche et 3 noires, la seconde contenant 2 boules blanches et 2 noires. L'expérience consiste à choisir d'abord « au hasard » une urne, puis à tirer « au hasard » une boule dans l'urne choisie. On cherche la probabilité de l'événement $A =$ « on tire une boule noire ».

L'espace Ω est $\{(i, j) : i = 1, 2; j = 1, 2, 3, 4\}$, où i est le numéro de l'urne et j le numéro de la boule dans l'urne. On considère la partition (E_1, E_2) de Ω définie par $E_i = \{(i, j) : j = 1, 2, 3, 4\}$. Le mode opératoire conduit naturellement à supposer que $P(E_i) = 1/2$ pour $i = 1, 2$, et aussi que $P(\{(i, j)\} | E_{i'})$ vaut $1/4$ si $i = i'$ et 0 sinon. Si on convient que dans chaque urne les boules noires ont les premiers numéros, on a $A = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2)\}$. Comme la probabilité conditionnelle est aussi une mesure de probabilité (théorème 3.2(b)), donc est additive, on a $P(A | E_1) = 3/4$ et $P(A | E_2) = 1/2$. D'après la formule des probabilités totales il vient alors

$$P(A) = P(A | E_1)P(E_1) + P(A | E_2)P(E_2) = \frac{3}{4} \times \frac{1}{2} + \frac{2}{4} \times \frac{1}{2} = \frac{5}{8}.$$

Théorème 3.5. (Théorème de Bayes, ou de « probabilité des causes ».)

Soit (E_n) une partition finie ou dénombrable de Ω , constituée d'éléments de la tribu \mathcal{A} . Pour tout $A \in \mathcal{A}$ tel que $P(A) > 0$ on a

$$P(E_n | A) = \frac{P(A | E_n)P(E_n)}{\sum_m P(A | E_m)P(E_m)}.$$

Preuve. Au vu du théorème 3.4 le dénominateur vaut

$$\sum_m P(A | E_m)P(E_m) = P(A).$$

La formule devient donc

$$\frac{P(A | E_n)P(E_n)}{P(A)} = \frac{P(A \cap E_n)}{P(A)} = P(E_n | A). \quad \blacksquare$$

Ce théorème a des conséquences profondes en probabilités et en statistique : voir par exemple l'exercice 6.

Exercices

Dans tous les exercices l'espace de probabilité est fixé, et A , B , A_n , etc. sont des événements.

1. Montrer que si $A \cap B = \emptyset$, alors A et B ne sont pas indépendants, sauf si $P(A) = 0$ ou $P(B) = 0$.
2. Si $P(C) > 0$, montrer que $P(A \cup B | C) = P(A | C) + P(B | C) - P(A \cap B | C)$.
3. Si $P(C) > 0$ et si les A_i sont deux à deux disjoints, montrer que

$$P\left(\bigcup_{i=1}^n A_i | C\right) = \sum_{i=1}^n P(A_i | C).$$

4. Si $P(B) > 0$, montrer que $P(A \cap B) = P(A | B)P(B)$.
5. Si $0 < P(B) < 1$, montrer que

$$P(A) = P(A | B)P(B) + P(A | B^c)P(B^c).$$

6. Chaque don de sang est soumis à un test du SIDA. On suppose que ce test a une efficacité de 99% (= probabilité que le test soit positif pour une personne atteinte du SIDA), et une probabilité de fausse alarme de 5% (= probabilité que le test soit positif pour une personne non atteinte). Enfin, on suppose la fréquence de séropositivité est 1/10000. Quelle est la probabilité pour qu'une personne obtenant un test positif soit atteinte du SIDA ?
7. Soit $A_n \rightarrow A$ et $B_n \rightarrow B$ (voir avant le théorème 2.4). Supposons aussi que $P(B) > 0$ et $P(B_n) > 0$ pour tout n . Montrer que
 - a) $\lim_{n \rightarrow \infty} P(A_n | B) = P(A | B)$;
 - b) $\lim_{n \rightarrow \infty} P(A | B_n) = P(A | B)$;
 - c) $\lim_{n \rightarrow \infty} P(A_n | B_n) = P(A | B)$.
8. On modélise le jet d'une pièce avec deux résultats possibles, p = Pile et f = Face, chacun avec la probabilité 1/2. On jette deux fois cette pièce, de sorte que l'ensemble des résultats possibles Ω contient les quatre points : pp , pf , fp , ff . On suppose que les deux jets sont indépendants.

- a) Quelle est la probabilité pour obtenir deux fois Face, sachant que le premier jet donne Face? [Rép. : $\frac{1}{2}$].
- b) Quelle est la probabilité pour obtenir deux fois Face, sachant que l'un des deux jets au moins donne Face? [Rép. : $\frac{1}{3}$].
9. Si A, B, C sont indépendants et $P(A \cap B) \neq 0$, montrer que $P(C | A \cap B) = P(C)$.
10. Une urne contient r boules rouges et n boules noires. Une boule est choisie au hasard (chaque boule a la même probabilité d'être tirée), et une seconde boule est ensuite choisie au hasard parmi les $r + n - 1$ boules restantes. Trouver la probabilité pour que
- a) Les deux boules soient rouges [Rép. : $\frac{r(r-1)}{(r+n)(r+n-1)}$].
- b) La première boule soit rouge et la seconde noire [Rép. : $\frac{r n}{(r+n)(r+n-1)}$].
11. (Urne de Pólya.) Une urne contient r boules rouges et n boules noires. Une boule est choisie au hasard, on note sa couleur, et on la remet avec d boules supplémentaires de la même couleur. Puis on recommence la même procédure aussi souvent que nécessaire. Trouver la probabilité pour que
- a) La seconde boule tirée soit noire [Rép. : $\frac{n}{n+r}$].
- b) La première boule est noire, sachant que la seconde est noire [Rép. : $\frac{n+d}{n+r+d}$].
12. Dans la situation de l'exercice 11, on note B_m l'événement selon lequel la m -ième boule tirée est noire. Montrer que $P(B_m) = P(B_1)$.
13. Toujours dans la situation de l'exercice 11, trouver la probabilité pour que la première boule soit noire, sachant que les $m - 1$ suivantes sont noires; trouver la limite de cette probabilité lorsque $m \rightarrow \infty$ [Rép. : $\frac{n+d}{n+r+d}$; la limite est 1].
14. Une compagnie d'assurance assure un nombre égal de conducteurs masculins et féminins. Tous les conducteurs (masculins) ont, chaque année, la probabilité α d'avoir un accident, et ceci indépendamment des autres années, et des autres conducteurs. Il en est de même des conductrices, avec une probabilité d'accident égale à β . La compagnie sélectionne un/une conducteur/trice au hasard.
- a) Quelle est la probabilité pour que la personne sélectionnée ait un accident cette année? [Rép. : $\frac{\alpha + \beta}{2}$].

- b) Quelle est la probabilité pour que la personne sélectionnée ait un accident deux années consécutives? [Rép. : $\frac{\alpha^2 + \beta^2}{2}$].
15. Dans la situation de l'exercice 14, soit A_i l'événement « la personne sélectionnée a un accident l'année numéro i ». Montrer que $P(A_2 | A_1) \geq P(A_1)$. [Rép. : $P(A_2 | A_1) - P(A_1) = \frac{(\alpha - \beta)^2}{2(\alpha + \beta)}$].
16. Toujours dans la même situation, trouver la probabilité pour que, une année donnée, une personne sélectionnée au hasard parmi celles ayant eu un accident soit une conductrice. [Rép. : $\frac{1}{\alpha + \beta}$].
17. Si A_1, A_2, \dots, A_n sont des événements indépendants, montrer que la probabilité pour qu'aucun des A_i ne soit réalisé est au plus égale à $\exp(-\sum_{i=1}^n P(A_i))$.
18. Si $P(A) > 0$, montrer que $P(A \cap B | A \cup B) \leq P(A \cap B | A)$.

Chapitre 4

Probabilités sur un espace fini ou dénombrable

Dans tout le chapitre 4 nous supposons que l'espace d'états Ω est fini ou dénombrable, et nous choisissons pour tribu la classe $\mathcal{A} = 2^\Omega$ de toutes les parties de Ω .

Théorème 4.1. (a) Une probabilité sur l'ensemble fini ou dénombrable Ω est caractérisée par ses valeurs sur les singletons : $p_\omega = P(\{\omega\})$, $\omega \in \Omega$.

(b) Si $(p_\omega)_{\omega \in \Omega}$ est une famille de réels indexée par l'ensemble fini ou dénombrable Ω , il existe une probabilité P (nécessairement unique par (a)) sur Ω telle que $P(\{\omega\}) = p_\omega$ pour tout $\omega \in \Omega$ si et seulement si $p_\omega \geq 0$ et $\sum_{\omega \in \Omega} p_\omega = 1$.

Lorsque Ω est infini dénombrable, $\sum_{\omega} p_\omega$ est la somme d'une infinité de termes, qui *a priori* ne sont pas ordonnés (bien qu'on puisse énumérer les points de Ω , cette énumération est en fait arbitraire); on n'a donc pas à proprement parler une série, mais une « famille sommable ». Nous donnons en appendice à ce chapitre un résumé des propriétés des familles sommables qui nous sont utiles dans ce chapitre et dans le suivant.

Preuve. Soit $A \in \mathcal{A}$; on a $A = \bigcup_{\omega \in A} \{\omega\}$, réunion finie ou dénombrable de singletons deux à deux disjoints. Si P est une probabilité sur Ω , la σ -additivité implique

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p_\omega.$$

On a donc (a).

Pour la condition nécessaire de (b), on remarque que si $P(\{\omega\}) = p_\omega$, alors par définition $p_\omega \geq 0$, et aussi

$$1 = P(\Omega) = P\left(\bigcup_{\omega \in \Omega} \{\omega\}\right) = \sum_{\omega \in \Omega} P(\omega) = \sum_{\omega \in \Omega} p_\omega.$$

Supposons inversement que les p_ω vérifient $p_\omega \geq 0$ et $\sum_{\omega} p_\omega = 1$. On pose $P(A) \equiv \sum_{\omega \in A} p_\omega$, avec la convention qu'une somme « vide » est

nulle. On a donc $P(\Omega) = \sum_{\omega} p_{\omega} = 1$. Pour la σ -additivité, elle est évidente lorsque Ω est fini, et lorsque Ω est dénombrable elle provient de ce qu'on peut « sommer par paquets » pour obtenir que $\sum_{i \in I} \sum_{\omega \in A_i} p_{\omega} = \sum_{\omega \in \bigcup_{i \in I} A_i} p_{\omega}$ si les A_i sont deux à deux disjoints. ■

Supposons d'abord Ω fini. N'importe quelle famille de réels positifs indicée par Ω et de somme 1 constitue un exemple de probabilité sur Ω , mais parmi tous ces exemples le suivant est particulièrement important :

Définition 4.1. On dit que la probabilité P sur l'espace fini Ω est uniforme si $p_{\omega} = P(\{\omega\})$ ne dépend pas de ω .

Dans ce cas il est immédiat que

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)},$$

où $\text{card}(A)$ désigne le « cardinal », ou nombre de points, de A . Dans ce cas le calcul des probabilités se ramène à des dénombrements : on est dans le cas de la *combinatoire*. Sur un espace fini donné Ω il existe une et une seule probabilité uniforme.

Nous allons maintenant donner deux exemples de probabilités très importantes pour les applications, puis examiner les rapports entre ces deux exemples.

a) La loi hypergéométrique. Une urne contient N boules blanches et M boules noires. On tire n boules (sans remettre les boules tirées dans l'urne, donc $n \leq N + M$). Parmi les boules tirées, il y en a X blanches et $n - X$ noires. On cherche la probabilité pour que $X = x$, où x est un entier (arbitraire) fixé.

Il s'agit d'une épreuve aléatoire, dans la mesure où on ne connaît pas *a priori* le résultat. Comme il s'agit d'un tirage sans remise, on peut supposer qu'on tire *simultanément* les n boules. Ainsi, il est naturel de considérer qu'un résultat est une partie à n éléments de l'ensemble $\{1, 2, \dots, N + M\}$ des $N + M$ boules (qu'on peut supposer numérotées de 1 à $N + M$). Donc Ω est l'ensemble de toutes les parties à n éléments, et $\text{card}(\Omega) = C_{N+M}^n = \frac{(N+M)!}{n!(N+M-n)!}$ (rappelons que la factorielle d'un entier p est $p! = 1 \cdot 2 \cdot \dots \cdot (p-1) \cdot p$).

Ensuite, il est également naturel de considérer que tous les tirages possibles sont équiprobables, donc P est la probabilité uniforme sur Ω . La quantité X est une « variable aléatoire » car si on connaît le tirage ω , on connaît aussi le nombre $X(\omega)$ de boules blanches qu'il contient. L'ensemble $X^{-1}(\{x\})$, noté aussi $\{X = x\}$, contient $C_N^x C_M^{n-x}$ éléments si

$x \leq N$ et $n - x \leq M$, et est vide sinon. Donc

$$P(X = x) = \begin{cases} \frac{C_N^x C_M^{n-x}}{C_{N+M}^n} & \text{si } 0 \leq x \leq N \text{ et } 0 \leq n - x \leq M \\ 0 & \text{sinon.} \end{cases}$$

On a ainsi obtenu, lorsque x varie, la loi de X . Cette loi, appelée loi hypergéométrique, intervient naturellement dans la théorie des sondages : il y a $N + M$ électeurs, dont N ont l'opinion « blanche » et M l'opinion « noire », et on sonde avec un échantillon de taille n (cf. l'exercice 3 pour une généralisation à plus de deux « opinions »).

b) La loi binomiale. De la même urne que ci-dessus, on tire n boules, avec remise après chaque tirage (donc n peut être aussi grand qu'on veut). On cherche encore la probabilité $P(X = x)$, lorsque x est un entier entre 0 et n .

Ici, l'espace d'états naturel est le produit cartésien $\Omega = \{1, 2, \dots, N + M\}^n$, avec encore la probabilité uniforme. On a donc $\text{card}(\Omega) = (N + M)^n$, et un calcul simple montre que le nombre d'éléments $\text{card}(X = x)$ vaut $C_n^x N^x M^{n-x}$. Donc

$$P(X = x) = C_n^x \left(\frac{N}{N + M} \right)^x \left(\frac{M}{N + M} \right)^{n-x} \quad \text{pour } x = 0, 1, \dots, n.$$

On écrit en général le résultat ainsi, en posant $p = \frac{N}{N + M}$:

$$P(X = x) = C_n^x p^x (1 - p)^{n-x} \quad \text{pour } x = 0, 1, \dots, n.$$

Cette formule donne la loi binomiale de taille n et de paramètre p . A priori p est quelconque dans $[0, 1]$ (dans l'exemple ci-dessus, p est bien sûr rationnel). On note $B(p, n)$ cette loi.

c) La loi binomiale comme limite de lois hypergéométriques. Dans la situation a) ci-dessus, on suppose que n est fixé et que N et M tendent vers $+\infty$, de telle sorte que $\frac{N}{N + M}$ tende vers une limite p (nécessairement dans $[0, 1]$). En développant les combinaisons dans (11), on voit facilement que

$$P(X = x) \rightarrow C_n^x p^x (1 - p)^{n-x} \quad \text{pour } x = 0, 1, \dots, n,$$

donc les lois hypergéométriques « convergent » vers la loi $B(p, n)$ (en comparant à b) ci-dessus, on pourra vérifier que le résultat est intuitivement évident).

Exemples avec Ω dénombrable.

1. La loi de Poisson de paramètre $\lambda > 0$ est la probabilité P sur \mathbb{N} définie par

$$p_n = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n = 0, 1, 2, 3, \dots$$

2. La loi géométrique de paramètre $\alpha \in]0, 1[$ est la probabilité P sur \mathbb{N} définie par

$$p_n = (1 - \alpha)\alpha^n, \quad n = 0, 1, 2, 3, \dots$$

Noter que dans le modèle binomial, lorsque n est grand, le nombre $C_n^j p^j (1-p)^{n-j}$ peut se révéler impossible à calculer numériquement. Si p n'est proche ni de 0 ni de 1, il existe une approximation dite « normale » qui sera obtenue ultérieurement, quand on parlera du théorème-limite central. Lorsque p est « petit » il existe une autre approximation que nous décrivons ci-dessous.

Pour rendre compte du fait que p est « petit » et n est « grand », on suppose que p dépend de n et on le note p_n : on fait tendre n vers l'infini, en supposant que $\lim_{n \rightarrow \infty} np_n = \lambda$. On peut alors montrer (cf. exercice 1) que (pour j fixé)

$$\lim_{n \rightarrow \infty} C_n^j (p_n)^j (1 - p_n)^{n-j} = e^{-\lambda} \frac{\lambda^j}{j!},$$

et ainsi on approche la loi binomiale par une loi de Poisson.

Appendice : Quelques résultats utiles sur les séries

Dans cet appendice nous faisons un résumé, essentiellement sans démonstrations, des résultats sur les séries et familles sommables qui sont d'usage constant dans l'étude des probabilités sur un espace dénombrable.

Auparavant, signalons que nous serons amenés très souvent à faire des opérations faisant intervenir $+\infty$ (qu'on écrit plus simplement ∞) ou $-\infty$. Pour que ces opérations aient un sens précis, on fera toujours les conventions suivantes :

$$+\infty + \infty = +\infty, \quad -\infty - \infty = -\infty, \quad a + \infty = +\infty, \quad a - \infty = -\infty \quad \text{si } a \in \mathbb{R}. \quad (4.1)$$

$$0 \times \infty = 0, \quad a \in]0, \infty] \Rightarrow a \times \infty = +\infty, \quad a \in [-\infty, 0[\Rightarrow a \times \infty = -\infty. \quad (4.2)$$

Soit (u_n) une suite numérique, et $S_n = u_1 + \dots + u_n$ la « somme partielle » à l'ordre n .

Sl : La série $\sum_n u_n$ est dite *convergente* si S_n converge vers une limite finie S , notée aussi $S = \sum_n u_n$ (c'est la « somme » de la série).

S2 : La série $\sum_n u_n$ est dite *absolument convergente* si la série $\sum_n |u_n|$ converge.

S3 : Si $u_n \geq 0$ pour tout n , la suite S_n est croissante, donc elle tend toujours vers une limite $S \in [0, \infty]$. On écrit encore $S = \sum_n u_n$, bien que la série converge au sens de (S1) si et seulement si $S < \infty$. Avec les conventions (1.1) ceci s'applique même si les u_n sont à valeurs dans $[0, \infty]$.

En général la convergence d'une série dépend de l'ordre dans lequel on considère les termes. Il existe deux cas importants, signalés ci-dessous, où l'ordre des termes n'a pas d'importance; au lieu de « série » on dit alors plutôt qu'on a une « famille sommable » :

S4 : Lorsque les u_n sont des réels de signe quelconque et lorsque la série est absolument convergente, on peut modifier de manière arbitraire l'ordre des termes sans changer la propriété d'être absolument convergente, ni la somme de la série.

S5 : Si $u_n \in [0, \infty]$ pour tout n , la somme $\sum_n u_n$ (finie ou infinie : cf. (S3) ci-dessus) ne change pas si on change l'ordre de sommation.

Dans ces deux situations, une propriété supplémentaire mérite d'être signalée :

S6 : Si $u_n \in [0, \infty]$, ou si la série est absolument convergente, on peut « sommer par paquets » : soit $(A_i)_{i \in I}$ une partition de \mathbb{N}^* , avec $I = \{1, 2, \dots, N\}$ pour un entier N , ou $I = \mathbb{N}^*$. Pour chaque $i \in I$ on pose $v_i = \sum_{n \in A_i} u_n$: si A_i est fini, c'est une somme ordinaire ; sinon, v_i est elle-même la somme d'une série. On a alors $\sum_n u_n = \sum_{i \in I} v_i$, cette dernière somme est de nouveau la somme d'une série si $I = \mathbb{N}^*$.

Exercices

1. (Approximation poissonnienne de la loi binomiale, suite.) On considère la loi binomiale $B(p, n)$, et on pose $\lambda = pn$. Montrer que

$$P(\{k\}) = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left\{ \binom{n}{k} \binom{n-1}{n-k} \cdots \binom{n-k+1}{n-k+1} \right\} \left(1 - \frac{\lambda}{n}\right)^k.$$

On fait tendre n vers l'infini, et on fait dépendre p de n , de sorte que λ reste constant. Montrer que pour k fixé on a

$$P(\{k\}) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}.$$

[Remarque : pour que cette approximation soit bonne il faut que n soit grand et p petit, de sorte que $\lambda = np$ soit de taille modérée, par exemple $\lambda \leq 20$.]

2. (Approximation poissonnienne de la loi binomiale, suite.) On se place dans la situation de l'exercice 1. Si $p_k = P(\{k\})$, montrer que les $q_k = p_{n-k}$ sont les probabilités des singletons pour une loi binomiale $B(1-p, n)$. En déduire une approximation de la loi binomiale par la loi de Poisson lorsque n est grand et p est proche de 1.
3. On se place dans le cadre de la « loi hypergéométrique », sauf qu'il y a m couleurs et N_i boules de couleur i . On note $N = N_1 + \dots + N_m$, et X_i le nombre de boules de couleur i parmi les n boules tirées. On a bien sûr $X_1 + \dots + X_m = n$. Montrer que

$$P(X_1=x_1, \dots, X_m=x_m) = \begin{cases} \frac{C_{N_1}^{x_1} \dots C_{N_m}^{x_m}}{C_N^n} & \text{si } x_1 + \dots + x_m = n \\ 0 & \text{sinon.} \end{cases}$$

Chapitre 5

Variables aléatoires sur un espace fini ou dénombrable

Dans ce chapitre nous supposons encore l'espace d'états Ω fini ou dénombrable, et nous choisissons pour tribu la classe $\mathcal{A} = 2^\Omega$ de toutes les parties de Ω . Une *variable aléatoire* X est définie comme une application de Ω dans un ensemble \mathbf{E} *a priori* arbitraire. Une variable aléatoire représente une quantité inconnue (d'où la terminologie « variable ») qui dépend de l'issue de l'expérience : ce n'est donc pas une « variable » au sens algébrique du terme, comme x dans la relation $x^2 - 9 = 0$. Avant de réaliser l'expérience, ce qui revient à choisir un point ω dans Ω , on connaît les valeurs que X peut éventuellement prendre, mais on ne sait pas encore quelle valeur elle va effectivement prendre lorsqu'on réalise l'expérience.

Noter que l'espace d'arrivée \mathbf{E} n'est pas nécessairement dénombrable, c'est par exemple \mathbb{R} si l'expérience consiste à choisir une personne dans une salle et si $X(\omega)$ représente la taille de la personne ω ; mais l'image \mathbf{E}' de Ω par X (i.e., tous les points i de \mathbf{E} pour lesquels il existe au moins un $\omega \in \Omega$ avec $X(\omega) = i$) est, quant à elle, nécessairement finie ou dénombrable.

On peut alors définir la *loi* de X , appelée aussi la *distribution* de X , par la formule

$$P_X(A) = P(\{\omega : X(\omega) \in A\}) = P(X^{-1}(A)) = P(X \in A).$$

Le fait que cette formule définisse une mesure de probabilité sur \mathbf{E}' (muni de la tribu de toutes ses parties $2^{\mathbf{E}'}$) est évident. Comme \mathbf{E}' est au plus dénombrable, cette probabilité est complètement déterminée par les nombre suivants :

$$p_j^X = P(X = j) = \sum_{\{\omega: X(\omega)=j\}} p_\omega,$$

et la famille $(p_j^X : j \in \mathbf{E}')$ est aussi appelée parfois la loi de la variable X . On a bien sûr $P_X(A) = \sum_{j \in A} p_j^X$. Si P_X est une loi munie d'un nom,

par exemple une loi de Poisson, on dit aussi que X est une variable de Poisson.

Définition 5.1. Soit X est une variable aléatoire à valeurs dans $E = \mathbb{R}$, définie sur un espace fini ou dénombrable Ω . On appelle espérance, ou espérance mathématique, de X le nombre

$$E(X) = \sum_{\omega} X(\omega)p_{\omega},$$

lorsqu'il est bien défini : c'est le cas si Ω est fini ; c'est aussi le cas, lorsque Ω est infini dénombrable, si la série ci-dessus est absolument convergente, ou si $X(\omega) \geq 0$ pour tout ω (dans ce dernier cas on peut avoir $E(X) = +\infty$).

La motivation de cette définition vient de l'approche par les fréquences : si on répète n fois l'expérience et si on note X_1, \dots, X_n les valeurs successives obtenues pour X , un calcul simple montre que la *moyenne empirique* $\frac{1}{n}(X_1 + \dots + X_n)$ vaut $\sum_{\omega \in \Omega} X(\omega)f_n(\{\omega\})$, et donc « converge » vers la quantité $E(X)$ définie ci-dessus (au moins si Ω est fini).

Notons \mathcal{L}^1 l'espace de toutes les variables aléatoires réelles qui ont une espérance finie. Il est alors immédiat que :

- (i) \mathcal{L}^1 est un espace vectoriel, et la fonctionnelle espérance E est linéaire.
- (ii) La fonctionnelle espérance E est positive : si $X \in \mathcal{L}^1$ et $X \geq 0$, alors $E(X) \geq 0$. Plus généralement si $X, Y \in \mathcal{L}^1$ et $X \leq Y$ alors $E(X) \leq E(Y)$.
- (iii) \mathcal{L}^1 contient toutes les variables aléatoires bornées. Si $X \equiv a$, alors $E(X) = a$.
- (iv) Si $X \in \mathcal{L}^1$, son espérance dépend seulement de sa loi, et si E' est l'image de Ω par X , on a

$$E(X) = \sum_{j \in E'} jP(X = j). \quad (5.1)$$

(Dans la formule de définition de $E(X)$, appliquer la « sommation par paquets », propriété S6 du chapitre précédent, avec les « paquets » $A_j = \{\omega : X(\omega) = j\}$, en remarquant que $\sum_{\omega \in A_j} p_{\omega} = P(X = j)$.)

- (v) Si $X = 1_A$ est l'indicatrice d'un événement A , on a $E(X) = P(A)$.

Théorème 5.1. Soit $h : \mathbb{R} \rightarrow [0, \infty[$ une fonction positive et X une variable aléatoire réelle. On a alors

$$P(\{\omega : h(X(\omega)) \geq a\}) \leq \frac{E(h(X))}{a}$$

pour tout $a > 0$.

Preuve. Comme X est une variable aléatoire, il en est de même de $Y = h(X)$; soit

$$A = Y^{-1}([a, \infty[) = \{\omega : h(X(\omega)) \geq a\} = \{h(X) \geq a\}.$$

On a $h(X) \geq a1_A$, donc

$$E(h(X)) \geq E(a1_A) = aE(1_A) = aP(A)$$

et le résultat suit. ■

Corollaire 5.1. (Inégalité de Markov.) On a

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}.$$

Preuve. Prendre $h(x) = |x|$ dans le théorème 5.1. ■

Pour bien comprendre la définition suivante, on remarquera que si une X est une variable aléatoire réelle on a $|X| \leq 1 + X^2$, de sorte que si X^2 est dans \mathcal{L}^1 alors X est aussi dans \mathcal{L}^1 (cela sera un résultat général pour les variables réelles, voir le théorème 9.3).

Définition 5.2. Soit X une variable aléatoire réelle telle que X^2 soit dans \mathcal{L}^1 . On appelle variance de X le nombre

$$\sigma^2 = \sigma_X^2 \equiv E((X - E(X))^2).$$

L'écart-type de X , soit σ_X , est la racine carrée positive de la variance. La motivation essentielle pour introduire l'écart-type est le fait qu'il est comparable à la variable X elle-même, au sens de la « dimension » : si X est par exemple une longueur, alors σ_X^2 représente une surface, tandis que σ_X représente à nouveau une longueur.

Si $\mu = E(X)$ représente l'espérance de X , ou *moyenne* comme on l'appelle souvent, alors $E(|X - E(X)|) = E(|X - \mu|)$ représente la moyenne de l'écart entre X et sa moyenne ; cette quantité mesure donc comment X diffère de sa moyenne. La variance est la moyenne du carré de l'écart entre

X et sa moyenne : par rapport à $E(|X - \mu|)$ cela diminue l'importance des « petits » écarts et augmente l'importance des grands. Cependant la variance est en général (beaucoup) plus facile à calculer que $E(|X - \mu|)$ (voir par exemple l'exercice 11.) Ainsi, la variance peut également être conçue comme une mesure de la variabilité de la variable aléatoire X .

Corollaire 5.2. (Inégalités de Bienaymé-Chebyshev.) Si X^2 est dans \mathcal{L}^1 , on a

$$(a) P(|X| \geq a) \leq \frac{E(X^2)}{a^2} \quad \text{pour } a > 0,$$

$$(b) P(|X - E(X)| \geq a) \leq \frac{\sigma_X^2}{a^2} \quad \text{pour } a > 0.$$

Preuve. Pour (a), on applique le théorème 5.1 avec $h(x) = x^2$, ce qui donne

$$P(|X| \geq a) = P(h(X) \geq a^2) \leq \frac{E(X^2)}{a^2}.$$

Pour (b), on applique (a) à la variable $Y = |X - E(X)|$. ■

Exemples.

1) X est de Poisson de paramètre λ . Si $A \subset \mathbb{N}$, on a

$$P(X \in A) = \sum_{j \in A} P(X = j) = \sum_{j \in A} \frac{\lambda^j}{j!} e^{-\lambda}.$$

L'espérance de X est

$$\begin{aligned} E(X) &= \sum_{j=0}^{\infty} j P(X = j) = \sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} \\ &= \lambda \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} e^{-\lambda} = \lambda e^{\lambda} e^{-\lambda} = \lambda. \end{aligned}$$

2) X est de Bernoulli, i.e. elle prend seulement les deux valeurs 1 et 0 avec les probabilités respectives p et $q = 1 - p$ (avec $p \in [0, 1]$).

Alors

$$E(X) = 1.P(X = 1) + 0.P(X = 0) = 1.p + 0.q = p.$$

3) X est binomiale $B(p, n)$, i.e. X peut prendre les valeurs 0, 1, 2, ..., n , et

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}.$$

Supposons qu'on réalise n expériences successives pour une variable de Bernoulli (avec le même paramètre p), en notant Y_i le résultat

de la i -ième expérience. La somme $X = Y_1 + \dots + Y_n$ est précisément une variable binomiale $B(p, n)$ (cf. l'exemple 3 du chapitre 4). Donc

$$E(X) = E\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n E(Y_i) = \sum_{i=1}^n p = np.$$

Noter qu'on aurait aussi pu calculer $E(X)$ algébriquement en utilisant la définition :

$$E(X) = \sum_{i=0}^n iP(X=i) = \sum_{i=1}^n iC_n^i p^i (1-p)^{n-i},$$

mais ce calcul n'est pas immédiat.

- 4) Supposons encore qu'on réalise des expériences de Bernoulli avec le même paramètre p , mais au lieu de fixer le nombre n d'expériences *a priori*, on s'arrête dès que le nombre de succès (le nombre de fois où Y_i vaut 1) atteint un entier fixé $r \geq 1$.

Soit d'abord X le nombre d'échecs avant le premier succès (i.e. $r = 1$). Dans ce cas X suit une *loi géométrique* de paramètre $1-p$:

$$P(X=k) = (1-p)^k p, \quad k = 0, 1, 2, 3, \dots$$

et on a

$$E(X) = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} kp(1-p)^k = p(1-p) \frac{1}{p^2} = \frac{1-p}{p}.$$

- 5) Dans la même situation que ci-dessus, X est maintenant le nombre d'échecs avant le r -ième succès, avec $r \geq 1$. Cette variable X suit une *loi de Pascal*, ou *loi binomiale négative*, de paramètres r et p :

$$P(X=j) = C_{j+r-1}^{r-1} p^r (1-p)^j, \quad j = 0, 1, 2, \dots$$

Noter que dans ce cas $X = \sum_{i=1}^r Z_i$, où les Z_i sont des variables géométriques de paramètre p . Par suite

$$E(X) = \sum_{i=1}^r E(Z_i) = \frac{r(1-p)}{p}.$$

- 6) Une loi fréquemment rencontrée en sciences sociales est la *loi de Pareto*, ou *loi Zêta*. Une variable aléatoire X suit cette loi si elle prend ses valeurs dans \mathbb{N}^* et vérifie

$$P(X=j) = c \frac{1}{j^{\alpha+1}}, \quad j = 1, 2, 3, \dots$$

pour un paramètre fixé $\alpha > 0$. La constante c est telle que $c \sum_{j=1}^{\infty} \frac{1}{j^{\alpha+1}} = 1$. La fonction

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}, \quad s > 1,$$

est appelée *fonction Zêta de Riemann*, et elle est tabulée. On a donc $c = \frac{1}{\zeta(\alpha+1)}$, et on peut écrire

$$P(X = j) = \frac{1}{\zeta(\alpha+1)} \frac{1}{j^{\alpha+1}}.$$

On peut facilement calculer l'espérance :

$$\begin{aligned} E(X) &= \sum_{j=1}^{\infty} jP(X = j) = \frac{1}{\zeta(\alpha+1)} \sum_{j=1}^{\infty} \frac{j}{j^{\alpha+1}} \\ &= \frac{1}{\zeta(\alpha+1)} \sum_{j=1}^{\infty} \frac{1}{j^{\alpha}} = \frac{\zeta(\alpha)}{\zeta(\alpha+1)}. \end{aligned}$$

Noter que cette espérance est infinie si $0 < \alpha \leq 1$, et finie si $\alpha > 1$.

- 7) On dit que la variable X est *uniforme* sur $\{1, 2, \dots, n\}$ si sa loi est la loi uniforme sur cet ensemble fini, c'est-à-dire si

$$P(X = j) = \frac{1}{n}, \quad j = 1, 2, \dots, n,$$

Comme $\sum_{i=1}^n i = \frac{n(n+1)}{2}$, on voit que

$$E(X) = \sum_{j=1}^n jP(X = j) = \sum_{j=1}^n j \frac{1}{n} = \frac{1}{n} \sum_{j=1}^n j = \frac{n(n+1)}{n \cdot 2} = \frac{n+1}{2}.$$

Exercices

1. Si $g : [0, \infty[\rightarrow [0, \infty[$ est strictement croissante, montrer que

$$P(|X| \geq a) \leq \frac{E(g(|X|))}{g(a)} \quad \text{pour } a > 0.$$

2. Si $h : \mathbb{R} \rightarrow [0, \alpha]$ montrer que pour $0 \leq a < \alpha$ on a

$$P(h(X) \geq a) \geq \frac{E(h(X)) - a}{\alpha - a}.$$

3. Montrer que $\sigma_X^2 = E(X^2) - E(X)^2$ si les deux espérances existent et sont finies.
4. Montrer que $E(X)^2 \leq E(X^2)$ si les deux espérances existent.
5. Montrer que $\sigma_X^2 = E(X(X-1)) + \mu_X - \mu_X^2$, si $\mu_X = E(X)$ et si toutes les espérances existent et sont finies.
6. Soit X binomiale $B(p, n)$. Quelle valeur de j maximise l'expression $P(X = j)$? (*Indication* : Calculer $\frac{P(X=k)}{P(X=k-1)}$.) [Rép. : $[(n+1)p]$, où $[x]$ est la « partie entière » de x .]
7. Avec X est comme ci-dessus, calculer la probabilité pour que X soit pair. [Rép. : $\frac{1}{2}(1 + (1-2p)^n)$.]
8. Soit X_n binomiale $B(p_n, n)$, avec $\lambda = np_n$ constant. Soit $A_n = \{X_n \geq 1\}$, et soit Y une variable de Poisson de paramètre (λ) . Montrer que $\lim_{n \rightarrow \infty} P(X_n = j | A_n) = P(Y = j | Y \geq 1)$.
9. Soit X une variable de Poisson de paramètre λ . Quelle valeur de j maximise $P(X = j)$? [Voir l'exercice 6. Rép. : $[\lambda]$.]
10. Avec X comme ci-dessus, pour $j > 0$ fixé, quelle valeur de λ maximise $P(X = j)$? [Rép. : j .]
11. Avec X comme ci-dessus, et si λ est un entier, montrer que $E(|X - \lambda|) = \frac{2\lambda^\lambda e^{-\lambda}}{(\lambda - 1)!}$ et que $\sigma_X^2 = \lambda$.
- 12.* Soit X binomiale $B(p, n)$. Montrer que pour tous $\lambda > 0$ et $\varepsilon > 0$ on a

$$P(X - np > n\varepsilon) \leq E(\exp(\lambda(X - np - n\varepsilon))).$$

13. Soit X_n binomiale $B(p, n)$. Montrer que pour tout $b > 0$ fixé on a $P(X_n \leq b) \rightarrow 0$ quand $n \rightarrow \infty$.
14. Soit X_n binomiale $B(p, n)$. Montrer que

$$P\left(\left|\frac{X_n}{n} - p\right| > a\right) \leq \frac{\sqrt{p(1-p)}}{a^2 n} \min\left\{\sqrt{p(1-p)}, a\sqrt{n}\right\},$$

et que $P(|X_n - np| \leq n\varepsilon) \rightarrow 1$ pour tout $\varepsilon > 0$ quand $n \rightarrow \infty$.

15.^{*} Soit X_n binomiale $B(1/2, 2m)$ avec m entier, et soit

$$a(m, k) = \frac{1^m}{C_{2m}^m} P\{X = m + k\}.$$

Montrer que $\lim_{m \rightarrow \infty} (a(m, k))^m = e^{-k^2}$.

16. Soit X géométrique. Montrer que si $i, j > 0$, on a $P\{X > i + j | X > i\} = P\{X > j\}$.

17. Soit X géométrique de paramètre p . Montrer que

$$E\left(\frac{1}{1 + X}\right) = \log((1 - p)^{-\frac{1}{p}}).$$

18. On jette plusieurs fois de suite et indépendamment une pièce de monnaie, la probabilité de tomber sur « pile » étant p . Trouver :

- la probabilité de ne pas avoir de « face » au cours des n premiers jets ;
- la probabilité d'obtenir « face » pour la première fois au n -ième jet ;
- l'espérance du nombre de jets jusqu'à la première obtention de « face ». [Rép. : $\frac{1}{1-p}$.]

19. Étant donnée une suite quelconque d'événements $(A_n)_{n \geq 1}$, montrer que

$$E\left(\sum_{n=1}^{\infty} 1_{A_n}\right) = \sum_{n=1}^{\infty} P(A_n),$$

(∞ étant une valeur possible pour chacun des membres de cette équation).

20. Si X prend ses valeurs dans \mathbb{N} , montrer que

$$E(X) = \sum_{n=0}^{\infty} P(X > n).$$

21. Soit X une variable de Poisson de paramètre λ . Montrer que pour $r = 2, 3, 4, \dots$ on a

$$E(X(X-1) \dots (X-r+1)) = \lambda^r.$$

22. Soit X une variable géométrique de paramètre p . Montrer que pour $r = 1, 2, 3, \dots$, on a

$$E(X(X-1) \dots (X-r+1)) = \frac{r! p^r}{(1-p)^r}.$$

Chapitre 6

Construction d'une mesure de probabilité

Dans ce chapitre Ω est un espace totalement quelconque, muni d'une tribu \mathcal{A} ; le couple (Ω, \mathcal{A}) est appelé un *espace mesurable*. Nous nous proposons de construire, ou de caractériser, une mesure de probabilité sur \mathcal{A} (on dit aussi, sur (Ω, \mathcal{A})) à partir de la connaissance des probabilités $P(A)$ pour les événements A appartenant à une certaine sous-classe de \mathcal{A} , qu'on aimerait être « aussi petite que possible ». Quand Ω est fini ou dénombrable nous avons vu que c'était une chose simple à réaliser, à partir par exemple de la probabilité $p_\omega = P(\{\omega\})$ des singletons. Mais dans le cas général la même méthode ne marche plus : une probabilité « typique » P vérifiera $P(\{\omega\}) = 0$ pour tout ω et donc les nombres $P(\{\omega\})$ ne caractérisent pas P .

Pour construire P il convient d'abord de préciser les ingrédients de base. On verra dans des cas « concrets », et notamment au chapitre suivant, qu'il est souvent aisé de construire une « probabilité » sur une algèbre qui engendre la tribu \mathcal{A} , et le problème qu'on se pose ci-dessous est l'extension de cette probabilité à la tribu elle-même.

Soit donc une algèbre \mathcal{A}_0 qui engendre la tribu \mathcal{A} , et une probabilité P sur \mathcal{A}_0 : par là, on entend une fonction $P : \mathcal{A}_0 \rightarrow [0, 1]$ qui satisfait :

1. $P(\Omega) = 1$,
2. (σ -additivité.) Pour toute suite (A_n) d'éléments de \mathcal{A}_0 deux à deux disjoints et telle que $\bigcup_n A_n \in \mathcal{A}_0$, on a $P(\bigcup_n A_n) = \sum_n P(A_n)$.

Théorème 6.1. *Toute probabilité P définie sur une algèbre \mathcal{A}_0 admet une unique extension (notée aussi P) à la tribu $\mathcal{A} = \sigma(\mathcal{A}_0)$.*

Il importe de souligner qu'on ne peut en général pas étendre la probabilité P à la tribu 2^Ω de toutes les parties de Ω : il n'existe pas de probabilité sur cette tribu qui coïncide avec P sur \mathcal{A}_0 .

Nous allons démontrer ci-dessous l'unicité; quant à l'existence, assez difficile, nous renvoyons à l'un des nombreux livres de théorie de la mesure, par exemple [19], [4] ou [26]. Nous devons d'abord établir un résultat très utile dans de nombreux contextes,

Définition 6.1. Une classe \mathcal{C} de parties de Ω est stable par intersection si pour tous $A, B \in \mathcal{C}$ on a $A \cap B \in \mathcal{C}$ (et alors, si $A_1, \dots, A_n \in \mathcal{C}$, on a $A_1 \cap A_2 \cap \dots \cap A_n \in \mathcal{C}$ pour n arbitraire fini).

Une classe \mathcal{C} est stable par limite croissante si pour toute suite croissante $A_1 \subset A_2 \subset A_3 \subset \dots \subset A_n \subset \dots$ d'éléments de \mathcal{C} on a $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$.

Une classe \mathcal{C} est stable par différence si pour tous $A, B \in \mathcal{C}$ avec $A \subset B$ on a $B \setminus A \in \mathcal{C}$.

Théorème 6.2. (Théorème des classes monotones.) Soit \mathcal{C} une classe de parties de Ω , stable par intersection et contenant Ω . Soit \mathcal{D} la plus petite classe contenant \mathcal{C} et stable par limite croissante et par différence. Alors $\mathcal{D} = \sigma(\mathcal{C})$.

Preuve. Remarquons d'abord que l'intersection d'une famille quelconque de classes qui sont stables par limite croissante et par différence est une classe également stable par limite croissante et par différence. Ainsi, en prenant l'intersection de toutes les classes ayant ces propriétés et contenant \mathcal{C} on voit qu'il existe une plus petite classe contenant \mathcal{C} et stable par limite croissante et par différence.

Pour tout $B \subset \Omega$ notons \mathcal{D}_B la classe de toutes les parties A vérifiant $A \in \mathcal{D}$ et $A \cap B \in \mathcal{D}$. Étant données les propriétés de \mathcal{D} , on voit facilement que \mathcal{D}_B est stable par limite croissante et par différence.

Soit $B \in \mathcal{C}$; pour chaque $C \in \mathcal{C}$ on a $B \cap C \in \mathcal{C} \subset \mathcal{D}$ et $C \in \mathcal{D}$, donc $C \in \mathcal{D}_B$. Par suite $\mathcal{C} \subset \mathcal{D}_B \subset \mathcal{D}$. Donc $\mathcal{D} = \mathcal{D}_B$, grâce aux propriétés de \mathcal{D} et de \mathcal{D}_B .

Maintenant, soit B dans \mathcal{D} . Pour chaque $C \in \mathcal{C}$ on a $B \in \mathcal{D}_C$, et à cause de ce qui précède il vient $B \cap C \in \mathcal{D}$, donc $C \in \mathcal{D}_B$, donc $\mathcal{C} \subset \mathcal{D}_B \subset \mathcal{D}$, donc $\mathcal{D} = \mathcal{D}_B$.

Comme $\mathcal{D} = \mathcal{D}_B$ pour tout $B \in \mathcal{D}$, on conclut que \mathcal{D} est stable par intersection. De plus $\Omega \in \mathcal{D}$, et \mathcal{D} est stable par différence, donc aussi par complémentation. Comme \mathcal{D} est aussi stable par limite croissante, il s'ensuit que \mathcal{D} est une tribu, et c'est clairement la plus petite tribu contenant \mathcal{C} . ■

La preuve de l'unicité dans le théorème 6.1 est une conséquence immédiate du corollaire 6.1 ci-dessous, qui est lui-même une conséquence du théorème 6.2.

Corollaire 6.1. Soit P et Q deux probabilités définies sur \mathcal{A} . Si P et Q coïncident sur une classe $\mathcal{C} \subset \mathcal{A}$ qui est stable par intersection, et si $\sigma(\mathcal{C}) = \mathcal{A}$, alors $P = Q$.

1. $B \setminus A$ désigne $B \cap A^c$.

Preuve. $\Omega \in \mathcal{A}$ parce que \mathcal{A} est une tribu, et comme $P(\Omega) = Q(\Omega) = 1$ on peut supposer sans perte de généralité que $\Omega \in \mathcal{C}$. Soit $\mathcal{B} = \{A \in \mathcal{A} : P(A) = Q(A)\}$. Par définition d'une mesure de probabilité et par le théorème 2.3, \mathcal{B} est stable par différence et par limite croissante. De plus \mathcal{B} contient \mathcal{C} par hypothèse. Par suite, comme $\sigma(\mathcal{C}) = \mathcal{A}$, on a $\mathcal{B} = \mathcal{A}$ par application du théorème des classes monotones 6.2. ■

Il existe une version du théorème 6.2 pour les fonctions. Nous ne l'utiliserons pas dans ce livre, mais c'est un théorème qui est utile pour de nombreuses applications et nous allons l'énoncer sans démonstration, et en anticipant sur le chapitre 8 pour la notion de fonction mesurable : le lecteur pourra trouver une preuve dans [18] ou [22]. Si \mathcal{M} une classe de fonctions de Ω dans \mathbb{R} , on note $\sigma(\mathcal{M})$ la plus petite des tribus de Ω qui rendent toutes les fonctions de \mathcal{M} mesurables.

Théorème 6.3. (Théorème des classes monotones.) *Soit \mathcal{M} une classe de fonctions bornées de Ω dans \mathbb{R} . Supposons que \mathcal{M} soit stable par multiplication : $f, g \in \mathcal{M}$ impliquent $fg \in \mathcal{M}$. Soit $\mathcal{A} = \sigma(\mathcal{M})$ (la plus petite tribu rendant toutes les fonctions de \mathcal{M} mesurables). Soit enfin \mathcal{H} un espace vectoriel de fonctions, contenant \mathcal{M} . Si \mathcal{H} contient les fonctions constantes et est stable par limite croissante (i.e. si les $(f_n)_{n \geq 1}$ sont des fonctions de \mathcal{H} telles que $0 \leq f_1 \leq f_2 \leq f_3 \leq \dots$, et si $f = \lim_{n \rightarrow \infty} f_n$ est bornée, alors $f \in \mathcal{H}$). Alors \mathcal{H} contient toutes les fonctions bornées \mathcal{A} -mesurables.*

Si en général on ne peut pas étendre la probabilité P sur la tribu \mathcal{A} à la tribu de toutes les parties de Ω , on peut néanmoins l'étendre à une tribu un peu plus grande que \mathcal{A} . C'est l'objet de la fin de ce chapitre.

Définition 6.2. *Soit P une probabilité sur \mathcal{A} . Un ensemble négligeable pour P est une partie A de Ω telle qu'il existe un élément B de \mathcal{A} vérifiant $A \subset B$ et $P(B) = 0$.*

On dit qu'une propriété est vraie *presque sûrement* (en abrégé : p.s.) si elle est vraie en dehors d'un ensemble négligeable ; cette notion dépend bien sûr de la probabilité, de sorte qu'on dit parfois P -presque sûrement, ou P -p.s.

Les ensembles négligeables ne sont pas nécessairement dans \mathcal{A} . Néanmoins il est naturel de leur affecter une probabilité nulle, et on peut même étendre la probabilité à la tribu engendrée par \mathcal{A} et les ensembles P -négligeables. C'est ce qu'on fait dans le théorème suivant, qui ne sera pas utilisé dans la suite.

Théorème 6.4. *Soit P une probabilité sur \mathcal{A} et \mathcal{N} la classe de tous les ensembles P -négligeables. Alors $\mathcal{A}' = \{A \cup N : A \in \mathcal{A}, N \in \mathcal{N}\}$ est une*

tribu, appelée *P-complétion de \mathcal{A}* . C'est la plus petite tribu contenant \mathcal{A} et \mathcal{N} , et P s'étend de manière unique en une probabilité (notée aussi P) sur \mathcal{A}' , en posant $P(A \cup N) = P(A)$ pour $A \in \mathcal{A}$ et $N \in \mathcal{N}$.

Preuve. L'unicité de l'extension, si elle existe, est triviale; de même, comme \emptyset appartient à \mathcal{A} et à \mathcal{N} , le fait que \mathcal{A}' , si c'est une tribu, soit la plus petite tribu contenant \mathcal{A} et \mathcal{N} est trivial. Il nous suffit donc de montrer que \mathcal{A}' est une tribu et que si on pose $Q(B) = P(A)$ pour $B = A \cup N$ (avec $A \in \mathcal{A}$ et $N \in \mathcal{N}$), alors $Q(B)$ ne dépend pas de la décomposition $B = A \cup N$ et Q est une probabilité sur \mathcal{A}' .

D'abord, $\Omega \in \mathcal{A} \subset \mathcal{A}'$ et on a déjà vu que $\emptyset \in \mathcal{A}'$. Ensuite comme \mathcal{A} et \mathcal{N} sont stables par les unions dénombrables (pour vérifier ceci pour \mathcal{N} , utiliser l'exercice 16 du chapitre 2), \mathcal{A}' l'est également. Soit $B = A \cup N \in \mathcal{A}'$, avec $A \in \mathcal{A}$ et $N \in \mathcal{N}$. Il existe $C \in \mathcal{A}$ avec $P(C) = 0$ et $N \subset C$. Posons $A' = A^c \cap C^c$ (qui est dans \mathcal{A}) et $N' = N^c \cap A^c \cap C$ (qui est contenu dans C et qui donc appartient à \mathcal{N}); comme $B^c = A' \cup N'$, on a $B^c \in \mathcal{A}'$, et \mathcal{A}' est stable par complémentation : ainsi, \mathcal{A}' est une tribu.

Supposons maintenant que $A_1 \cup N_1 = A_2 \cup N_2$ avec $A_i \in \mathcal{A}$ et $N_i \in \mathcal{N}$. La différence symétrique $A_1 \Delta A_2 = (A_1 \cap A_2^c) \cup (A_1^c \cap A_2)$ est contenue dans $N_1 \cup N_2$, lui-même contenu dans un élément C de \mathcal{A} de probabilité nulle : par suite $P(A_1) = P(A_2)$, ce qui montre que Q est défini sans ambiguïté, et coïncide de manière évidente avec P sur \mathcal{A} . Enfin le fait que Q soit une probabilité sur \mathcal{A}' est à peu près évident. ■

Chapitre 7

Probabilités sur \mathbb{R} et fonctions de répartition

Dans ce chapitre nous traitons un cas particulier – concret – du problème exposé au chapitre précédent, à savoir la caractérisation d'une probabilité P à partir des $P(A)$ pour une classe aussi petite que possible d'événements A . On suppose que $\Omega = \mathbb{R}$, et on munit cet espace de la tribu borélienne \mathcal{B} (c'est-à-dire $\mathcal{B} = \sigma(\mathcal{O})$, où \mathcal{O} désigne la classe de tous les ouverts de \mathbb{R}).

Définition 7.1. Si P est une probabilité sur $(\mathbb{R}, \mathcal{B})$, sa fonction de répartition est la fonction

$$F(x) = P(]-\infty, x]). \quad (7.1)$$

Théorème 7.1. La fonction de répartition F caractérise la probabilité.

Preuve. Nous voulons montrer que si Q est une autre probabilité sur $(\mathbb{R}, \mathcal{B})$, de fonction de répartition

$$G(x) = Q(]-\infty, x])$$

et si $F = G$, alors $P = Q$.

Soit d'abord \mathcal{B}_0 la classe de toutes les unions finies disjointes d'intervalles de la forme $]x, y]$, avec $-\infty \leq x \leq y \leq +\infty$ (avec la convention que $]x, \infty[=]x, \infty[$; noter aussi que $]x, y] = \emptyset$ si $x = y$). Il est facile de vérifier que \mathcal{B}_0 est une algèbre. De plus tout intervalle ouvert $]a, b[$ s'écrit $]a, b[= \bigcup_{n=\mathbb{N}}]a, b - \frac{1}{n}]$ pour un N assez grand, de sorte que $\sigma(\mathcal{B}_0)$ contient tous les intervalles ouverts. Comme tout ouvert de \mathbb{R} est une réunion dénombrable d'intervalles ouverts et comme \mathcal{B} est la tribu engendrée par les ouverts, on en déduit que $\sigma(\mathcal{B}_0) \supset \mathcal{B}$. Enfin $\bigcap_{n=1}^{\infty}]a, b + \frac{1}{n}[=]a, b[$, donc $\mathcal{B}_0 \subset \mathcal{B}$ et par suite $\mathcal{B} = \sigma(\mathcal{B}_0)$.

La relation (7.1) implique

$$P(]x, y]) = F(y) - F(x),$$

et si $A \in \mathcal{B}_0$ s'écrit

$$A = \bigcup_{1 \leq i \leq n}]x_i, y_i], \quad \text{avec } y_i < x_{i+1},$$

alors $P(A) = \sum_{1 \leq i \leq n} \{F(y_i) - F(x_i)\}$. Comme $G = F$, $Q(A)$ est donné par la même expression : par suite $Q(A) = P(A)$ pour toute $A \in \mathcal{B}_0$, et le théorème 6.1 implique que $P = Q$ sur \mathcal{B} , i.e. P et Q sont la même probabilité. ■

La signification du théorème 7.1 est que la probabilité est entièrement connue si on connaît seulement F : on peut donc en principe retrouver $P(A)$ pour n'importe quel borélien A à partir de la fonction F (déterminer « explicitement » $P(A)$ en fonction de F est une tout autre chose !)

Il est donc important de savoir caractériser les fonctions F qui sont des fonctions de répartition : cela permet en principe de caractériser toutes les probabilités sur \mathbb{R} , et cela fait l'objet du théorème suivant. (Rappelons qu'une fonction F est *continue à droite* si $\lim_{y \downarrow x} F(y) = F(x)$, pour tout $x \in \mathbb{R}$.)

Théorème 7.2. *Une fonction F est la fonction de répartition d'une (unique) probabilité sur $(\mathbb{R}, \mathcal{B})$ si et seulement si elle vérifie les trois propriétés suivantes :*

- (i) F est croissante (au sens large) :
- (ii) F est continue à droite :
- (iii) $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Preuve. Supposons que F soit la fonction de répartition de la probabilité P . Si $y > x$ on a $] -\infty, x[\subset] -\infty, y[$, donc $F(x) = P(] -\infty, x[) \leq P(] -\infty, y[) = F(y)$ et on a (i). Ensuite, soit x_n une suite de réels décroissant vers x . On a $\bigcap_{n=1}^{\infty }] -\infty, x_n[=] -\infty, x[$, et la suite d'intervalles $\{] -\infty, x_n[, n \geq 1 \}$ est décroissante. Donc $P(] -\infty, x[) = P(\bigcap_{n=1}^{\infty }] -\infty, x_n[) = \lim_{n \rightarrow \infty } P(] -\infty, x_n[)$ d'après le théorème 2.3, et on a (ii). De la même manière le théorème 2.3 nous donne (iii).

Inversement, supposons qu'on ait (i), (ii) et (iii). En accord avec (iii), on pose $F(-\infty) = 0$ et $F(+\infty) = 1$. Introduisons la même algèbre \mathcal{B}_0 que dans la preuve du théorème 7.1. Définissons ensuite la fonction $P : \mathcal{B}_0 \rightarrow [0, 1]$ comme suit : $P(\emptyset) = 0$ et, pour $A = \bigcup_{1 \leq i \leq n}]x_i, y_i[$ avec $-\infty \leq x_1 < \dots < y_i < x_{i+1} < y_{i+1} < \dots < y_n \leq +\infty$,

$$P(A) = \sum_{1 \leq i \leq n} \{F(y_i) - F(x_i)\}.$$

(Noter que la représentation de chaque $A \in \mathcal{B}_0$ ci-dessus est unique). La condition (iii) entraîne $P(\mathbb{R}) = 1$. Il nous reste à montrer que P est une probabilité sur \mathcal{B}_0 : dans ce cas, le théorème 6.1 impliquera que P admet une unique extension à la tribu \mathcal{B} , dont la fonction de répartition est trivialement F , et le résultat sera démontré.

Il nous reste donc à montrer la σ -additivité de P sur \mathcal{B}_0 : pour toute suite $(A_n) \in \mathcal{B}_0$ d'ensembles deux à deux disjoints, tels que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}_0$, alors $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$. Mais l'additivité de P sur \mathcal{B}_0 (i.e. $P(A \cup B) = P(A) + P(B)$ si $A \cap B = \emptyset$ et $A, B \in \mathcal{B}_0$) est une conséquence immédiate de la définition de P , et le théorème 2.3 donne plusieurs conditions équivalentes pour qu'une mesure « additive » sur une tribu soit σ -additive. Il est immédiat de vérifier que ces conditions restent valables si on remplace la tribu par une algèbre : par suite, il nous suffit de montrer que la condition (ii) est vérifiée, autrement dit que si $A_n \in \mathcal{B}_0$ avec A_n décroissant vers \emptyset , alors $P(A_n) \rightarrow 0$.

L'idée de la preuve consiste, pour chaque $\varepsilon > 0$ fixé, à construire une suite décroissante d'ensembles B_n de \mathcal{B}_0 tels que $\bar{B}_n \subset A_n$ et $P(A_n) - P(B_n) \leq \varepsilon$, et que les \bar{B}_n soient compacts : comme $\bigcap A_n = \emptyset$, on a $\bigcap \bar{B}_n = \emptyset$ et donc $\bar{B}_n = \emptyset$ pour n assez grand.

Soit donc $A_n \in \mathcal{B}_0$ comme ci-dessus. Chaque A_n s'écrit

$$A_n = \bigcup_{1 \leq i \leq k_n}]x_i^n, y_i^n],$$

avec $y_i^n < x_{i+1}^n$. Soit $\varepsilon > 0$. Par (iii) il existe un nombre z tel que $F(\cdot - z) \leq \varepsilon$ et $1 - F(\cdot) \leq \varepsilon$. Pour chaque n et i il existe $a_i^n \in]x_i^n, y_i^n]$ tel que $F(a_i^n) - F(x_i^n) \leq \frac{\varepsilon}{2^{i+n}}$, grâce à (ii). Posons alors

$$B_n' = \bigcup_{1 \leq i \leq k_n} \{]a_i^n, y_i^n] \cap]-z, z]\}, \quad B_n = \bigcap_{m \geq n} B_m'.$$

On remarque que $B_n' \in \mathcal{B}_0$ et $B_n' \subset A_n$, donc $B_n \in \mathcal{B}_0$ et $B_n \subset A_n$ également. De plus $A_n \setminus B_n \subset \bigcup_{m \geq n} (A_m \setminus B_m')$, par suite

$$\begin{aligned} P(A_n) - P(B_n) &\leq P(]-z, z]^c) + \sum_{m=1}^n P((A_m \setminus B_m') \cap]-z, z]) \\ &\leq P(]-z, z]^c) + \sum_{m=1}^n \sum_{i=1}^{k_m} P(]x_i^m, a_i^m]) \\ &\leq F(-z) + 1 - F(z) + \sum_{m=1}^n \sum_{i=1}^{k_m} \{F(a_i^m) - F(x_i^m)\} \\ &\leq 3\varepsilon. \end{aligned} \tag{7.2}$$

De plus $\bar{B}_n \subset A_n$ (où \bar{B}_n est la fermeture de B_n), donc la suite des \bar{B}_n , qui est décroissante, est d'intersection vide par hypothèse. Enfin

$\bar{B}_n \subset [-z, z]$, donc chaque \bar{B}_n est un ensemble compact. Une propriété des compacts, connue sous le nom de « propriété d'intersection finie »¹, est que si l'intersection d'une infinité de compacts est vide, il existe une sous-famille finie parmi ces compacts qui est déjà d'intersection vide : la suite de compacts \bar{B}_n décroissant vers \emptyset , il s'ensuit qu'il existe un m tel que $\bar{B}_m = \emptyset$, donc *a fortiori* $B_n = \emptyset$ et $P(B_n) = 0$, pour tout $n \geq m$. Donc

$$P(A_n) = P(A_n) - P(B_n) \leq 3\varepsilon$$

par (7.2), si $n \geq m$. Comme ε est arbitraire, on en déduit que $P(A_n) \downarrow 0$.

(Noter que cette démonstration — plutôt longue — deviendrait presque triviale si la suite k_n ci-dessus était bornée ; mais, bien que les A_n décroissent vers \emptyset , la suite k_n tend typiquement vers l'infini !). ■

Proposition 7.1. *Soit F la fonction de répartition de la probabilité P sur \mathbb{R} . En notant $F(x-)$ la limite à gauche de F au point x (qui existe puisque F est croissante), pour tous $x < y$ on a*

- (i) $P(]x, y]) = F(y) - F(x)$,
- (ii) $P([x, y]) = F(y) - F(x-)$,
- (iii) $P([x, y[) = F(y-) - F(x-)$,
- (iv) $P(]x, y[) = F(y-) - F(x)$,
- (v) $P(\{x\}) = F(x) - F(x-)$,

et en particulier $P(\{x\}) = 0$ pour tout x si et seulement si la fonction F est continue.

Preuve. (i) a déjà été montré. Pour (ii) on écrit que

$$P\left(\left]x - \frac{1}{n}, y\right]\right) = F(y) - F\left(x - \frac{1}{n}\right)$$

par (i). Le membre de droite converge quand $n \rightarrow \infty$ vers $F(y) - F(x-)$ par définition de la limite à gauche de F ; quant au membre de gauche, il converge vers $P([x, y])$ par application du théorème 2.3 et par le fait que la suite d'intervalles $]x - \frac{1}{n}, y]$ décroît vers $[x, y]$. Les assertions (iii), (iv) et (v) se montrent de manière analogue. ■

Exemples.

Considérons d'abord deux exemples généraux :

1. Voir par exemple [15, p. 81] ou [8].

1. Si f est positive et Riemann-intégrable et $\int_{-\infty}^{\infty} f(x)dx = 1$, la fonction $F(x) = \int_{-\infty}^x f(y)dy$ est la fonction de répartition d'une probabilité P sur \mathbb{R} . La fonction f est appelée sa *densité*. (Il n'est pas vrai que toute probabilité sur \mathbb{R} admet une densité : en effet s'il existe une densité la fonction F est continue. alors qu'il existe beaucoup de fonctions de répartition discontinues, ainsi que le montre l'exemple suivant.)
2. Soit $\alpha \in \mathbb{R}$. La « masse de Dirac » au point α est la probabilité sur \mathbb{R} qui est donné par

$$P(A) = \begin{cases} 1 & \text{si } \alpha \in A, \\ 0 & \text{sinon.} \end{cases}$$

Sa fonction de répartition est

$$F(x) = \begin{cases} 0 & \text{si } x < \alpha, \\ 1 & \text{si } x \geq \alpha. \end{cases}$$

Dans les exemples 3 à 10 ci-dessous nous définissons la probabilité (ou « loi ») P sur \mathbb{R} par sa densité f ; c'est-à-dire que nous définissons la fonction positive f d'intégrale 1, et la fonction de répartition F de P est $F(x) = \int_{-\infty}^x f(u)du$. Dans chaque cas le lecteur vérifiera aisément que f est positive et d'intégrale 1. On commet un léger abus de langage (habituel) en identifiant f avec la loi P correspondante, puisqu'elle détermine en fait de manière unique cette loi.

3. $f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b, \\ 0 & \text{sinon,} \end{cases}$ est appelée la *loi uniforme sur*

$[a, b]$. Cette loi correspond à l'idée que « chaque point de $[a, b]$ est également vraisemblable », et elle est l'analogue « continu » de la loi uniforme sur un ensemble d'entiers $\{p, p+1, \dots, q\}$; elle correspond à une densité constante sur l'intervalle considéré $[a, b]$.

4. $f(x) = \begin{cases} \beta e^{-\beta x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0, \end{cases}$ est appelée la *loi exponentielle* de paramètre $\beta > 0$. Cette loi est souvent utilisée pour modéliser la durée de vie d'un objet qui ne « vieillit » pas : sachant que cette durée de vie dépasse s , la probabilité conditionnelle pour qu'elle dépasse $s+t$ est la même que la probabilité *a priori* pour qu'elle dépasse t . C'est par exemple ce qui se passe (semble-t-il!) pour la

durée de vie des lampes au néon : si on croit en la validité d'un tel modèle, il est stupide de remplacer un tube au néon en état de marche par un tube neuf. Cette propriété de « non vieillissement » caractérise en fait la loi exponentielle : voir les exercices 21 et 22 du chapitre 9.

$$5. f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0, \end{cases} \quad \text{est appelée la loi gamma}$$

de paramètres α et β (avec $0 < \alpha < \infty$ et $0 < \beta < \infty$; Γ désigne la fonction gamma²). Cette loi intervient dans diverses applications. Un exemple en est la fiabilité : si l'élément essentiel d'une machine a une durée de vie exponentielle de paramètre β , on augmente la fiabilité en stockant $n-1$ pièces de rechange. La durée de vie totale suit alors une loi gamma de paramètres (n, β) . (voir l'exercice 17 du chapitre 15). Les lois gamma ont aussi des relations étroites avec la loi de Poisson (voir l'exercice 22 du chapitre 9) et avec les lois du chi-deux (voir l'exemple 6 du chapitre 15).

$$6. f(x) = \begin{cases} \alpha \beta^\alpha x^{\alpha-1} e^{-\beta x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0, \end{cases} \quad \text{est appelée la loi de Weibull}$$

de paramètres α et β (avec $0 < \alpha < \infty$ et $0 < \beta < \infty$). Cette loi apparaît comme une généralisation de la loi exponentielle, dans une direction différente de celle des lois gamma : voir par exemple l'exercice 23 du chapitre 9.

$$7. f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{est appelée la loi normale de paramètres } (\mu, \sigma^2),$$

avec $\mu \in \mathbb{R}$ et $0 < \sigma^2 < \infty$, ou *loi de Gauss*, et elle est notée $N(\mu, \sigma^2)$. Cette famille de lois sera discutée en détail dans les chapitres 16 et 21 : ce sont certainement les lois les plus importantes en probabilités ainsi qu'en statistique.

$$8. \text{ Soit } g_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{la densité normale (cf. ci-dessus). Alors}$$

$$f(x) = \begin{cases} \frac{1}{x} g_{\mu, \sigma^2}(\log x) & \text{si } x > 0, \\ 0 & \text{si } x \leq 0, \end{cases}$$

est appelée la *loi lognormale* de paramètres (μ, σ^2) , (avec $\mu \in \mathbb{R}$ et $0 < \sigma^2 < \infty$). Cette famille de lois est utilisée pour modéliser des

2. La fonction gamma est définie par $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, $\alpha > 0$, ou $\Gamma(\alpha) = (\alpha-1)!$ si $\alpha \in \mathbb{N}^*$, avec la convention $0! = 1$, et $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

quantités aléatoires *positives*. On les appelle aussi lois de Galtou-McAlister, ou de Cobb-Douglas dans un contexte économique.

9. $f(x) = \frac{\beta}{2} e^{-\beta(x-\alpha)}$ est appelée la *loi exponentielle double* de paramètres α et β (avec $\alpha \in \mathbb{R}$ et $0 < \beta < \infty$), ou aussi *loi de Laplace*.
10. $f(x) = \frac{1}{\beta\pi} \frac{1}{1+(x-\alpha)^2/\beta^2}$ est appelée la *loi de Cauchy* de paramètres α et β (avec $\alpha \in \mathbb{R}$ et $0 < \beta < \infty$). Cette famille de lois est souvent utilisée pour fournir des contre-exemples en théorie des probabilités, et a été introduite historiquement pour cette raison précise³, du fait qu'elle présente des queues «épaisses». Elle est utilisée en mécanique et en électricité.

Exercices des chapitres 6 et 7

1. Soit $(A_n)_{n \geq 1}$ une suite d'événements deux à deux disjoints sur un espace de probabilité (Ω, \mathcal{A}, P) . Montrer que $\lim_{n \rightarrow \infty} P(A_n) = 0$.
- 2* Soit $(A_\beta)_{\beta \in B}$ une famille d'événements deux à deux disjoints sur un espace de probabilité (Ω, \mathcal{A}, P) . Montrer que si $P(A_\beta) > 0$ pour tout $\beta \in B$, alors l'ensemble d'indices B est fini ou dénombrable.
3. Montrer que le maximum de la densité de la loi gamma est atteint pour $x = (\alpha-1)^{-1}$, lorsque $\alpha \geq 1$.
4. Montrer que le maximum de la densité de la loi de Weibull est atteint pour $x = \frac{1}{\alpha} (\frac{\alpha-1}{\alpha})^{\frac{1}{\alpha}}$, lorsque $\alpha \geq 1$.
5. Montrer que le maximum de la densité de la loi normale est atteint pour $x = \mu$.
6. Montrer que le maximum de la densité de la loi lognormale est atteint pour $x = e^{\mu - \sigma^2}$.
7. Montrer que le maximum de la densité de la loi exponentielle double est atteint pour $x = \alpha$.
8. Montrer que les lois gamma et de Weibull contiennent comme cas particulier la loi exponentielle.

³ En fait Poisson utilisa ces lois dès 1824 pour exhiber un cas où le théorème-limite central (cf. chapitre 21) est faux. Plus tard ces lois ont joué un rôle central dans une polémique entre Cauchy et Bienayme, d'où leur nom.

9. Montrer que les densités uniforme, normale, exponentielle double et Cauchy sont symétriques par rapport à un point qu'on déterminera.
10. Une loi est dite *unimodale* si elle admet une densité ayant un et un seul maximum (un « mode » est un maximum de la densité, ou pour certains auteurs un maximum local de la densité). Montrer que les lois normale, exponentielle, exponentielle double, de Cauchy, gamma, de Weibull et lognormale sont unimodales.
11. Soit P une probabilité sur \mathbb{R} admettant une densité f . Montrer que $P(\{x\}) = 0$ pour tout x .
12. Soit P comme dans l'exercice 11. Montrer que si B est une partie finie ou dénombrable de \mathbb{R} , alors B est un borélien et $P(B) = 0$.
13. Soit P et B comme dans l'exercice 12. Soit A un borélien vérifiant $P(A) = \frac{1}{2}$. Montrer que $A \cup B$ est un borélien et que $P(A \cup B) = \frac{1}{2}$.
14. Soit A_1, \dots, A_n, \dots une suite d'ensembles négligeables. Montrer que $B = \bigcup_{i=1}^{\infty} A_i$ est également négligeable.
15. Soit X une variable aléatoire définie sur un espace dénombrable. Supposons que $E(|X|) = 0$. Montrer que $X = 0$ sauf peut-être sur un ensemble négligeable. Est-il possible d'en conclure que $X = 0$ partout (i.e., $X(\omega) = 0$ pour tout ω)? [Rép. : Non.]
- 16^{*} Soit F une fonction de répartition. Montrer que F peut avoir une infinité de points de discontinuité, mais que le nombre de ces points est au plus dénombrable.
17. Soit F la fonction de répartition donnée par

$$F(x) = \frac{1}{4} 1_{[0, \infty[}(x) + \frac{1}{2} 1_{[1, \infty[}(x) + \frac{1}{4} 1_{[2, \infty[}(x),$$

et P la probabilité admettant F pour fonction de répartition. Trouver la probabilité des événements suivants :

- a) $A =]-\frac{1}{2}, \frac{1}{2}[$,
 b) $B =]-\frac{1}{2}, \frac{3}{2}[$,
 c) $C =]\frac{2}{3}, \frac{5}{2}[$,
 d) $D =]0, 2[$,
 e) $E =]3, \infty[$.
18. Soit la fonction F donnée par

$$F(x) = \sum_{i=1}^{\infty} \frac{1}{2^i} 1_{[1, \infty[}(x),$$

Montrer que c'est la fonction de répartition d'une probabilité P sur \mathbb{R} . Trouver la probabilité des événements suivants :

a) $A = [1, \infty[$,

b) $B = [\frac{1}{10}, \infty[$,

c) $C = \{0\}$,

d) $D =]0, \frac{1}{2}]$,

e) $E =]-\infty, 0[$,

f) $G =]0, \infty[$.

Chapitre 8

Variables aléatoires

Dans le chapitre 5 nous avons considéré des variables aléatoires définies sur un espace fini ou dénombrable Ω . Nous voulons maintenant considérer un espace Ω quelconque, éventuellement non dénombrable, cet espace étant muni d'une tribu \mathcal{A} d'événements et d'une probabilité P (sur \mathcal{A}).

Comme nous l'avons vu au chapitre 1, une variable aléatoire est une application X de Ω dans un autre espace F . Nous cherchons en premier lieu à déterminer la probabilité pour que X appartienne à une partie donnée A de F : c'est-à-dire, la probabilité de la partie $\{X \in A\} = X^{-1}(A) = \{\omega : X(\omega) \in A\}$ (trois manières équivalentes d'écrire la même chose, la première étant la plus parlante). Pour cela, il faut évidemment que l'ensemble $\{X \in A\}$ soit dans la tribu \mathcal{A} , ce qui *a priori* n'est pas vrai pour une partie arbitraire A de F . C'est la motive la définition suivante.

Définition 8.1. (a) Soit (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables. Une fonction $X : E \rightarrow F$ est appelée mesurable (relativement aux tribus \mathcal{E} et \mathcal{F}) si $X^{-1}(\Lambda) \in \mathcal{E}$ pour tout $\Lambda \in \mathcal{F}$. (On écrit aussi $X^{-1}(\mathcal{F}) \subset \mathcal{E}$: voir l'exercice 1 du chapitre 9.)

(b) Si $(E, \mathcal{E}) = (\Omega, \mathcal{A})$, une fonction mesurable X est appelée une variable aléatoire (en abrégé : v.a.)

(c) Quand $F = \mathbb{R}$ (resp. $F = \mathbb{R}^d$), on prend en général pour \mathcal{F} la tribu borélienne \mathcal{B} de \mathbb{R} (resp. \mathcal{B}^d de \mathbb{R}^d) : sauf mention contraire, ce sera toujours le cas dans la suite.

Théorème 8.1. Soit \mathcal{C} une classe de parties de F telle que $\sigma(\mathcal{C}) = \mathcal{F}$. Pour qu'une application X de E dans F soit mesurable (relativement aux tribus \mathcal{E} et \mathcal{F}), il faut et il suffit que $X^{-1}(\mathcal{C}) \subset \mathcal{E}$, c'est-à-dire $X^{-1}(C) \in \mathcal{E}$ pour tout $C \in \mathcal{C}$.

Preuve. La condition nécessaire est évidente. Pour la condition suffisante, supposons que $X^{-1}(C) \in \mathcal{E}$ pour tout $C \in \mathcal{C}$. Posons $\mathcal{D} = \{A \in \mathcal{F} : X^{-1}(A) \in \mathcal{E}\}$. On a $\mathcal{C} \subset \mathcal{D}$ par hypothèse et, comme $X^{-1}(\bigcup_n \Lambda_n) = \bigcup_n X^{-1}(\Lambda_n)$, $X^{-1}(\bigcap_n \Lambda_n) = \bigcap_n X^{-1}(\Lambda_n)$ et $X^{-1}(\Lambda^c) = (X^{-1}(\Lambda))^c$, ou en d'autres termes X^{-1} « commute » avec les réunions dénombrables, les

intersections dénombrables et le passage au complémentaire, on voit que \mathcal{D} est aussi une tribu. Donc $\mathcal{D} \supset \sigma(\mathcal{C})$ et aussi $\mathcal{F} \supset \mathcal{D}$. Comme $\mathcal{F} = \sigma(\mathcal{C})$ on conclut que $\mathcal{F} = \mathcal{D}$, ce qui donne le résultat. ■

Un cas particulier du théorème précédent permet de caractériser de manière particulièrement simple les v.a. à valeurs réelles (i.e. quand $F = \mathbb{R}$) :

Corollaire 8.1. Soit $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$ et X et X_n des applications de E , espace muni de la tribu \mathcal{E} , dans F .

- (a) X est mesurable si et seulement si $\{X \leq a\} = \{\omega : X(\omega) \leq a\} = X^{-1}(] - \infty, a]) \in \mathcal{E}$ pour tout $a \in \mathbb{R}$, ou même si et seulement si $\{X \leq a\} \in \mathcal{E}$ pour tout $a \in \mathbb{Q}$.
- (b) Si les X_n sont toutes mesurables, les fonctions $\sup_n X_n$, $\inf_n X_n$, $\limsup_{n \rightarrow \infty} X_n$ et $\liminf_{n \rightarrow \infty} X_n$ sont toutes mesurables.
- (c) Si les X_n sont toutes mesurables et si X_n converge simplement vers X , alors X est mesurable.

Preuve. (a) D'après le théorème 2.1 on sait que $\mathcal{B} = \sigma(\mathcal{C}) = \sigma(\mathcal{C}')$, où $\mathcal{C} = \{]-\infty, a]; a \in \mathbb{R}\}$ et $\mathcal{C}' = \{]-\infty, a]; a \in \mathbb{Q}\}$. Il suffit alors d'appliquer le théorème 2.1.

(b) Comme X_n est mesurable, $\{X_n \leq a\} \in \mathcal{E}$. Donc $\{\sup_n X_n \leq a\} = \bigcap_n \{X_n \leq a\} \in \mathcal{E}$ pour tout a , et $\sup_n X_n$ est mesurable par (a). De manière analogue $\{\inf_n X_n < a\} = (\bigcap_n \{X_n \geq a\})^c \in \mathcal{E}$ et donc $\inf_n X_n$ est mesurable. On a $\limsup_{n \rightarrow \infty} X_n = \inf_n \sup_{m \geq n} X_m = \inf_n Y_n$, où $Y_n = \sup_{m \geq n} X_m$. On vient de voir que chaque Y_n est mesurable et donc que $\inf_n Y_n$ est aussi mesurable : donc $\limsup_{n \rightarrow \infty} X_n$ est mesurable. De manière analogue $\liminf_{n \rightarrow \infty} X_n = \sup_n \inf_{m \geq n} X_m$ est mesurable.

(c) Si $X_n \rightarrow X$ simplement, on a $X = \limsup_{n \rightarrow \infty} X_n = \liminf_{n \rightarrow \infty} X_n$, et la mesurabilité de X se déduit de (b). ■

Théorème 8.2. Soit X mesurable de (E, \mathcal{E}) dans (F, \mathcal{F}) , et soit Y mesurable de (F, \mathcal{F}) dans (G, \mathcal{G}) ; alors la composée $Y \circ X$ est mesurable de (E, \mathcal{E}) dans (G, \mathcal{G}) .

Preuve. Soit $A \in \mathcal{G}$. On a $(Y \circ X)^{-1}(A) = X^{-1}(Y^{-1}(A))$. Comme Y est mesurable, $B = Y^{-1}(A) \in \mathcal{F}$. Comme X est mesurable, $X^{-1}(B) \in \mathcal{E}$. ■

Un espace topologique est un ensemble F muni d'une famille d'ouverts¹. Cette famille d'ouverts est aussi appelée la *topologie* de F . La

1. Une « famille d'ouverts » est une famille de sous-ensembles qui est stable par réunion quelconque et par intersection finie, et qui contient l'ensemble F tout entier et l'ensemble vide \emptyset .

définition d'une *application continue* est alors la suivante : étant donnés deux espaces topologiques (E, \mathcal{U}) et (F, \mathcal{V}) (où \mathcal{U} et \mathcal{V} sont les familles d'ouverts de E et F respectivement), alors une application $f : E \rightarrow F$ est *continue* si $f^{-1}(A) \in \mathcal{U}$ pour tout $A \in \mathcal{V}$ (c'est-à-dire si $f^{-1}(\mathcal{V}) \subset \mathcal{U}$). La *tribu borélienne*, ou *de Borel*, d'un espace topologique (E, \mathcal{U}) est $\sigma(\mathcal{U})$: lorsque $E = \mathbb{R}$, on retrouve la tribu \mathcal{B} introduite avant le théorème 2.1 du chapitre 2. (La famille des ensembles ouverts n'est pas elle-même une tribu, car elle n'est pas stable par intersection dénombrable, ni par complémentation.)

Théorème 8.3. *Soit (E, \mathcal{U}) et (F, \mathcal{V}) deux espaces topologiques, et soit \mathcal{E} et \mathcal{F} leurs tribus boréliennes. Chaque application continue X de E dans F est alors mesurable (on dit aussi que X est borélienne).*

Preuve. Comme $\mathcal{F} = \sigma(\mathcal{V})$, d'après le théorème 6.4 il suffit de montrer que $X^{-1}(\mathcal{V}) \subset \mathcal{E}$. Mais comme X est continue on a $X^{-1}(\mathcal{V}) \subset \mathcal{U}$, d'où le résultat. ■

Rappelons que si A est une partie de E , son *indicatrice* (ou « fonction indicatrice »), est définie par

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{si } x \notin A. \end{cases}$$

Ainsi la fonction $x \mapsto 1_A(x)$, écrite 1_A , « indique » si un point x appartient ou non à la partie A .

Théorème 8.4. *Soit $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$, et (E, \mathcal{E}) un espace mesurable quelconque.*

- Une indicatrice 1_A sur E est mesurable si et seulement si $A \in \mathcal{E}$.
- Si X_1, \dots, X_n sont des fonctions réelles mesurables sur (E, \mathcal{E}) et si f est une fonction borélienne réelle sur \mathbb{R}^n , alors $f(X_1, \dots, X_n)$ est mesurable.
- Si X, Y sont des fonctions réelles mesurables sur (E, \mathcal{E}) , il en est de même de $X + Y$, XY , $X \vee Y$ (une manière concise d'écrire $\max(X, Y)$), $X \wedge Y$ (une manière concise d'écrire $\min(X, Y)$), et X/Y (si Y ne s'annule pas).

Preuve. (a) Si $B \subset \mathbb{R}$, on a

$$(1_A)^{-1}(B) = \begin{cases} \emptyset & \text{si } 0 \notin B, 1 \notin B \\ A & \text{si } 0 \notin B, 1 \in B \\ A^c & \text{si } 0 \in B, 1 \notin B \\ E & \text{si } 0 \in B, 1 \in B \end{cases}$$

Le résultat est alors immédiat.

(b) La tribu borélienne \mathcal{B}^n de \mathbb{R}^n est engendrée par les ensembles $\prod_{i \leq n}]-\infty, a_i]$: ceci se montre exactement de la même manière que le théorème 2.1. Si $X : E \rightarrow \mathbb{R}^n$ désigne la fonction « vectorielle » $X = (X_1, \dots, X_n)$, on a $X^{-1}(\prod_{i \leq n}]-\infty, a_i]) = \bigcap_{i \leq n} \{X_i \leq a_i\} \in \mathcal{E}$; donc X est une application mesurable de (E, \mathcal{E}) dans $(\mathbb{R}^n, \mathcal{B}^n)$. L'assertion désirée suit alors du théorème 8.2.

(c) Noter que la fonction $f_1 : \mathbb{R}^2 \mapsto \mathbb{R}$ donnée par $f_1(x, y) = x + y$ est continue, ainsi que les fonctions $f_2(x, y) = xy$, $f_3(x, y) = x \vee y = \max(x, y)$ et $f_4(x, y) = x \wedge y = \min(x, y)$. La fonction $f_5(x, y) = \frac{x}{y}$ est continue de $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$ dans \mathbb{R} . Donc (c) découle de (b) et du théorème 8.3 (les fonctions continues sont boréliennes). ■

Soit maintenant X une v.a. sur un espace de probabilité (Ω, \mathcal{A}, P) , à valeurs dans un espace mesurable (E, \mathcal{E}) . La loi de X est alors définie par

$$P_X(B) = P(X^{-1}(B)) = (P \circ X^{-1})(B) = P(\{\omega : X(\omega) \in B\}) = P(X \in B),$$

pour tout $B \in \mathcal{E}$. Les quatre manières d'écrire $P_X(B)$ du côté droit sont utilisées, mais la plus usuelle est

$$P_X(B) = P(X \in B),$$

ce qui évite de spécifier explicitement l'espace Ω : cet espace est en effet souvent difficile à caractériser mathématiquement, tandis qu'on s'intéresse en fait seulement à une ou plusieurs v.a. définies sur cet espace. Parfois on appelle aussi P_X l'image de P par X .

Comme X^{-1} commute avec les réunions et les intersections (finies, dénombrables ou infinies non dénombrables!), et comme $X^{-1}(E) = \Omega$, on a de manière immédiate :

Théorème 8.5. La loi P_X de X est une mesure de probabilité sur (E, \mathcal{E}) .

Lorsque X est une v.a. réelle, sa loi P_X est une probabilité sur \mathbb{R} , dont on sait qu'elle est entièrement caractérisée par sa fonction de répartition :

$$F_X(x) = P_X(\{]-\infty, x]) = P(X \leq x).$$

La fonction F_X s'appelle la *fonction de répartition de la v.a. X* . Lorsque F_X admet une densité f_X (i.e. $F_X(x) = \int_{-\infty}^x f_X(y) dy$ pour tout $x \in \mathbb{R}$), on dit aussi que la fonction f_X est la *densité de la v.a. X* .

Chapitre 9

Intégration par rapport à une mesure de probabilité

Soit (Ω, \mathcal{A}, P) un espace de probabilité. Nous voulons définir l'espérance, ou ce qui est équivalent, « l'intégrale » d'une v.a. réelle sur cet espace. Nous avons déjà fait cela lorsque Ω est fini ou dénombrable, mais le cas général est plus délicat.

Commençons par des cas particuliers.

Définition 9.1. a) Une v.a. réelle X est appelée simple, ou étagée, si elle ne prend qu'un nombre fini de valeurs, ce qui revient à dire qu'elle se met sous la forme

$$X = \sum_{i=1}^n a_i 1_{A_i}, \quad (9.1)$$

où $a_i \in \mathbb{R}$ et $A_i \in \mathcal{A}$, $1 \leq i \leq n$ (Cette formule définit clairement une v.a., c'est-à-dire une fonction mesurable; inversement si X est mesurable et ne prend que les valeurs a_1, \dots, a_n elle se met sous la forme (9.1) avec $A_i = \{X = a_i\}$: une v.a. simple a bien sûr plusieurs représentations différentes du type (9.1).)

b) Si X est une v.a. simple, son espérance (ou « espérance mathématique », ou « intégrale » par rapport à P) est le nombre

$$E(X) = \sum_{i=1}^n a_i P(A_i). \quad (9.2)$$

(On l'écrit aussi $\int X(\omega)P(d\omega)$, ou même plus simplement $\int X dP$.)

Un calcul élémentaire montre que $E(X)$ ne dépend pas de la représentation (9.1) choisie.

Soit X, Y deux v.a. simples et β un réel. On peut écrire X et Y comme dans (9.1), avec les mêmes événements A_i qui forment une partition de Ω , et avec les nombres a_i pour X et b_i pour Y . Ainsi, βX et $X + Y$ sont encore de la forme (9.1) avec les mêmes A_i et avec les nombres βa_i et $a_i + b_i$ respectivement. Donc $E(\beta X) = \beta E(X)$ et $E(X + Y) = E(X) + E(Y)$.

ce qui signifie que l'espérance est linéaire sur l'espace vectoriel des v.a. (réelles) simples. Si de plus $X \leq Y$ ou $a_i \leq b_i$ pour tout i , donc $E(X) \leq E(Y)$.

Ensuite, nous définissons l'espérance pour les v.a. positives, i.e. les v.a. à valeurs dans $[0, \infty]$, en incluant $+\infty$ comme valeur possible : cette extension anodine est nécessaire pour certains des résultats ultérieurs. On pose alors

$$E(X) = \sup\{E(Y) : Y \text{ v.a. simple avec } 0 \leq Y \leq X\}. \quad (9.3)$$

Ce supremum existe toujours dans $[0, \infty]$. Comme $E(X) \leq E(Y)$ pour des variables simples vérifiant $0 \leq X \leq Y$, il est clair que (9.3) et (9.2) coïncident lorsque X est simple.

Noter que $E(X) \geq 0$, et qu'on peut avoir $E(X) = \infty$, même si X ne prend que des valeurs finies.

Finalement soit X une v.a. réelle arbitraire. Soit $X^+ = \max(X, 0)$ et $X^- = -\min(X, 0)$. On a $X = X^+ - X^-$, et X^+ et X^- sont des v.a. positives, et on a aussi $|X| = X^+ + X^-$.

Définition 9.2. (a) On dit qu'une v.a. X a une espérance finie (ou « est intégrable ») si $E(X^+)$ et $E(X^-)$ sont toutes deux finies (on verra plus tard que cela revient à dire que $E(|X|)$ est finie). Dans ce cas, son espérance est le nombre

$$E(X) = E(X^+) - E(X^-), \quad (9.4)$$

qu'on écrit aussi $\int X(\omega)dP(\omega)$ ou $\int X dP$. (Si $X \geq 0$ on a $X^- = 0$ et $X^+ = X$, donc, comme clairement $E(0) = 0$, cette définition coïncide avec (9.3)).

On note \mathcal{L}^1 l'ensemble de toutes les v.a. réelles intégrables (parfois $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ pour éviter toute ambiguïté).

(b) Une v.a. X admet une espérance si $E(X^+)$ et $E(X^-)$ ne sont pas tous deux simultanément infinis. Dans ce cas, son espérance est encore donnée par (9.4), avec les conventions $+\infty + \alpha = +\infty$ et $-\infty + \alpha = -\infty$ pour $\alpha \in \mathbb{R}$. (Si $X \geq 0$ cette définition coïncide encore avec (9.3); noter que si X admet une espérance, on a $E(X) \in [-\infty, +\infty]$, et X est intégrable si et seulement si son espérance existe et est finie.)

Remarque. Lorsque Ω est fini ou dénombrable, on a donc apparemment deux définitions différentes pour l'espérance d'une v.a. X , celle ci-dessus et celle du chapitre 5. Bien entendu ces deux définitions coïncident : il suffit de le vérifier pour les v.a. simples, et pour celles-ci on observe

que les deux formules (9.2) avec $A_i = \{X = a_i\}$ et (5.1) sont en fait identiques.

Le théorème suivant présente les propriétés les plus importantes de la fonctionnelle espérance. La preuve des parties (d), (e) et (f) est difficile et peut être omise. Dans son énoncé, l'assertion « $X = Y$ presque sûrement », ou en abrégé « p.s. », signifie que l'événement $\{X \neq Y\}$ est négligeable, c'est-à-dire que $P(X \neq Y) = 0$ ou de manière équivalente $P(X = Y) = 1$; de même « $X \geq Y$ p.s. » signifie $P(X \geq Y) = 1$, et « X_n converge p.s. vers X » que l'ensemble des ω pour lesquels la suite numérique $X_n(\omega)$ ne converge pas vers $X(\omega)$ est négligeable.

Théorème 9.1. (a) \mathcal{L}^1 est un espace vectoriel et l'espérance est une forme linéaire sur \mathcal{L}^1 , et aussi une fonctionnelle positive (i.e., $X \geq 0 \Rightarrow E(X) \geq 0$). Si de plus X et Y sont deux v.a. avec $0 \leq X \leq Y$ et $Y \in \mathcal{L}^1$, alors $X \in \mathcal{L}^1$ et $E(X) \leq E(Y)$.

(b) $X \in \mathcal{L}^1$ si et seulement si $|X| \in \mathcal{L}^1$, ou de manière équivalente si $E(|X|) < \infty$, et dans ce cas $|E(X)| \leq E(|X|)$. En particulier toute v.a. bornée est intégrable.

(c) Si $X = Y$ presque sûrement, alors $E(X) = E(Y)$.

(d) (Théorème de convergence monotone.) Si les v.a. X_n sont positives et si la suite (X_n) croît p.s. vers X , alors $\lim_{n \rightarrow \infty} E(X_n) = E(X)$ (même si $E(X) = \infty$).

(e) (Lemme de Fatou.) Si les v.a. X_n vérifient $X_n \geq Y$ p.s., avec Y intégrable, les v.a. X_n et $\liminf_{n \rightarrow \infty} X_n$ admettent une espérance, et on a $E(\liminf_{n \rightarrow \infty} X_n) \leq \liminf_{n \rightarrow \infty} E(X_n)$. Cela est vrai en particulier si $X_n \geq 0$ p.s. pour tout n .

(f) (Théorème de convergence dominée de Lebesgue.) Si les v.a. X_n convergent p.s. vers X et si $|X_n| \leq Y$ p.s. pour tout n , où Y est une v.a. intégrable, alors $X_n \in \mathcal{L}^1$, $X \in \mathcal{L}^1$, et $E(X_n) \rightarrow E(X)$.

L'égalité p.s. entre v.a. est clairement une relation d'équivalence, et deux v.a. équivalentes (i.e. p.s. égales) ont la même espérance : On peut donc définir l'espace L^1 comme « \mathcal{L}^1 modulo cette relation d'équivalence ». En d'autres termes un élément de L^1 est une classe d'équivalence, c'est-à-dire l'ensemble de toutes les v.a. qui sont p.s. égales à une v.a. particulière. Vu (c) ci-dessus, on peut parler de « l'espérance » de cette classe d'équivalence, qui est l'espérance de n'importe laquelle des v.a. appartenant à cette classe. Comme de plus l'addition de deux v.a. et la multiplication d'une v.a. par une constante préservent l'égalité p.s., l'ensemble L^1 est aussi un espace vectoriel. Par suite on commet l'abus

(anodin) d'identifier une v.a. avec sa classe d'équivalence, et on écrit d'habitude $X \in L^1$ au lieu de $X \in \mathcal{L}^1$.

Si $1 < p < \infty$, on définit \mathcal{L}^p comme l'ensemble des v.a. X telles que $|X|^p \in \mathcal{L}^1$; L^p est défini de manière analogue à L^1 : c'est \mathcal{L}^p modulo la relation d'équivalence « égalité p.s. »; ou de manière équivalente, deux éléments de \mathcal{L}^p qui sont p.s. égaux sont considérés comme deux représentants du même élément de L^p . Nous n'utiliserons vraiment dans ce livre que les espaces L^1 et L^2 .

Avant de donner la preuve du théorème 9.1, nous montrons deux résultats auxiliaires.

Résultat 1. *Pour toute v.a. positive X il existe une suite $(X_n)_{n \geq 1}$ de v.a. positives simples qui croît vers X quand $n \rightarrow \infty$. Un exemple d'une telle suite est donné par*

$$X_n(\omega) = \begin{cases} k2^{-n} & \text{si } k2^{-n} \leq X(\omega) < (k+1)2^{-n} \text{ et } 0 \leq k \leq n2^n - 1, \\ n & \text{si } X(\omega) \geq n. \end{cases} \quad (9.5)$$

Résultat 2. *Si X est une v.a. positive et si $(X_n)_{n \geq 1}$ est une suite arbitraire de v.a. positives simples croissant vers X , alors $E(X_n)$ croît vers $E(X)$.*

Pour vérifier ceci, on observe d'abord que la suite $E(X_n)$ croît vers une limite a , qui vérifie $a \leq E(X)$ par (9.3). Pour obtenir qu'en fait $a = E(X)$, et au vu de (9.3) encore, il suffit clairement de montrer que si Y est une v.a. simple telle que $0 \leq Y \leq X$, alors $E(Y) \leq a$.

La variable Y prend k valeurs différentes, disons a_1, \dots, a_k , et on pose $A_k = \{Y = a_k\}$. Choisissons $\varepsilon \in]0, 1[$. La v.a. $Y_{n,\varepsilon} = (1 - \varepsilon)Y \mathbf{1}_{\{(1-\varepsilon)Y \leq X_n\}}$ prend les valeurs $(1 - \varepsilon)a_k$ sur l'ensemble $A_{k,n,\varepsilon} = A_k \cap \{(1 - \varepsilon)Y \leq X_n\}$ et 0 sur l'ensemble $\{(1 - \varepsilon)Y > X_n\}$. De plus il est évident que $Y_{n,\varepsilon} \leq X_n$, donc d'après (9.2) il vient

$$E(Y_{n,\varepsilon}) = (1 - \varepsilon) \sum_{k=1}^k a_k P(A_{k,n,\varepsilon}) \leq E(X_n). \quad (9.6)$$

Maintenant, rappelons que $Y \leq \lim_n X_n$, donc $(1 - \varepsilon)Y < \lim_n X_n$ dès que $Y > 0$, de sorte que $A_{k,n,\varepsilon} \rightarrow A_k$ quand $n \rightarrow \infty$. Une application du théorème 2.4 entraîne $P(A_{k,n,\varepsilon}) \rightarrow P(A_k)$, donc en passant à la limite dans (9.6) on obtient

$$(1 - \varepsilon) \sum_{k=1}^k a_k P(A_k) = (1 - \varepsilon)E(Y) \leq a.$$

Il reste à faire $\varepsilon \rightarrow 0$ ci-dessus pour obtenir $\mathbf{E}(Y) \leq a$, d'où le résultat 2.

Preuve du théorème 9.1. (a) Soit X et Y deux v.a. positives, et $\alpha \in \mathbb{R}_+$. Le résultat 1 permet d'obtenir deux suites (X_n) et (Y_n) de v.a. positives simples, croissant vers X et Y respectivement. Les v.a. $X_n + Y_n$ et αX_n sont simples et croissent vers $X + Y$ et αX respectivement, tandis que $\mathbf{E}(X_n + Y_n) = \mathbf{E}(X_n) + \mathbf{E}(Y_n)$ et $\mathbf{E}(\alpha X_n) = \alpha \mathbf{E}(X_n)$: en appliquant le résultat 2 on obtient alors $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$ et $\mathbf{E}(\alpha X) = \alpha \mathbf{E}(X)$. Si de plus $X \leq Y$, l'inégalité $\mathbf{E}(X) \leq \mathbf{E}(Y)$ est une conséquence immédiate de (9.3).

Ce qui précède entraîne les deux dernières assertions de (a). Comme pour des v.a. réelles quelconques X et Y et pour $\alpha \in \mathbb{R}$ on a $(\alpha X)^+ + (\alpha X)^- \leq \alpha(X^+ + X^-)$ et $(X + Y)^+ + (X + Y)^- \leq X^+ + X^- + Y^+ + Y^-$, on en déduit ensuite que \mathcal{L}^1 est un espace vectoriel. Finalement, comme $\mathbf{E}(X) = \mathbf{E}(X^+) - \mathbf{E}(X^-)$ on déduit que l'espérance est linéaire.

(b) Comme $|X| = X^+ + X^-$ la première assertion est évidente. Si de plus $X \in \mathcal{L}^1$, on a $|\mathbf{E}(X)| \leq \mathbf{E}(X^+) + \mathbf{E}(X^-) = \mathbf{E}(|X|)$. Enfin, comme $\mathbf{E}(Y) = \alpha$ si $Y = \alpha \in [0, \infty[$, la dernière assertion est également évidente.

(c) Supposons que $X = Y$ p.s. et supposons d'abord que $X \geq 0$ et $Y \geq 0$. Si $A = \{\omega : X(\omega) \neq Y(\omega)\} = \{X \neq Y\}$ on a $\mathbf{P}(A) = 0$. De plus

$$\mathbf{E}(Y) = \mathbf{E}(Y1_A + Y1_{A^c}) = \mathbf{E}(Y1_A) + \mathbf{E}(Y1_{A^c}) = \mathbf{E}(Y1_A) + \mathbf{E}(X1_{A^c}).$$

Soit une suite (Y_n) de v.a. simples positives croissant vers Y . Les v.a. Y_n1_A sont aussi simples et croissent vers $Y1_A$. Comme Y_n est simple elle est bornée, disons par N_n , et on a

$$0 \leq \mathbf{E}(Y_n1_A) \leq \mathbf{E}(N_n1_A) = N_n \mathbf{P}(A) = 0.$$

Donc $\mathbf{E}(Y1_A) = \lim_{n \rightarrow \infty} \mathbf{E}(Y_n1_A) = 0$. De même $\mathbf{E}(X1_A) = 0$. Finalement, il vient

$$\begin{aligned} \mathbf{E}(Y) &= \mathbf{E}(Y1_A) + \mathbf{E}(X1_{A^c}) = 0 + \mathbf{E}(X1_{A^c}) = \mathbf{E}(X1_A) + \mathbf{E}(X1_{A^c}) \\ &= \mathbf{E}(X). \end{aligned}$$

Lorsque X et Y sont de signe quelconque, la propriété $X = Y$ p.s. entraîne aussi $Y^+ = X^+$ p.s. et $Y^- = X^-$ p.s., et (c) s'ensuit.

(d) Pour chaque n fixé on choisit une suite $(Y_{n,k})_{k \geq 1}$ de v.a. simples positives croissant vers X_n (quand $k \rightarrow \infty$), d'après le résultat 1. Posons

$$Z_k = \max_{n \leq k} Y_{n,k}.$$

Alors $(Z_k)_{k \geq 1}$ est une suite croissante de v.a. simples positives, admettant donc une limite $Z = \lim_{k \rightarrow \infty} Z_k$. De plus

$$Y_{n,k} \leq Z_k \leq X_k \leq X \quad \text{p.s. pour } n \leq k$$

ce qui entraîne

$$X_n \leq Z \leq X \quad \text{p.s.}$$

En faisant $n \rightarrow \infty$ on obtient $Z = X$ p.s. L'espérance étant une fonctionnelle positive,

$$E(Y_{n,k}) \leq E(Z_k) \leq E(X_k) \quad \text{pour } n \leq k.$$

Fixons n et faisons $k \rightarrow \infty$. Le résultat 2 implique

$$E(X_n) \leq E(Z) \leq \lim_{k \rightarrow \infty} E(X_k).$$

En faisant maintenant $n \rightarrow \infty$ on obtient :

$$\lim_{n \rightarrow \infty} E(X_n) \leq E(Z) \leq \lim_{k \rightarrow \infty} E(X_k),$$

et comme les deux membres extrêmes sont égaux, ils sont aussi égaux à celui du milieu. On déduit alors le résultat de (c) et de l'égalité $X = Z$ p.s.

(c) Posons $\tilde{X}_n = X_n - Y$, de sorte que $\tilde{X}_n \geq 0$. Soit $Z_n = \inf_{k \geq n} \tilde{X}_k$. Les Z_n sont de v.a. positives, croissant vers la v.a. positive $\liminf_{n \rightarrow \infty} \tilde{X}_n$, et on a aussi $Z_n \leq \tilde{X}_n$. Par suite (d) implique que

$$E(\liminf_{n \rightarrow \infty} \tilde{X}_n) = \lim_{n \rightarrow \infty} E(Z_n) = \liminf_{n \rightarrow \infty} E(Z_n) \leq \liminf_{n \rightarrow \infty} E(\tilde{X}_n). \quad (9.7)$$

Par ailleurs si une v.a. U vérifie $U \geq Y$ p.s. avec $Y \in \mathcal{L}^1$ on a $U^- \leq |Y|$ p.s., donc $E(U^-) \leq E(|Y|) < \infty$, de sorte que U admet une espérance : on en déduit que les X_n et $\liminf_{n \rightarrow \infty} X_n$ admettent toutes une espérance. Enfin, on a $E(X_n) = E(\tilde{X}_n) + E(Y)$ et $E(\liminf_{n \rightarrow \infty} X_n) = E(\liminf_{n \rightarrow \infty} \tilde{X}_n) + E(Y)$, donc l'inégalité cherchée se déduit immédiatement de (9.7).

(f) Soit $U = \liminf_{n \rightarrow \infty} X_n$ et $V = \limsup_{n \rightarrow \infty} X_n$. Par hypothèse $U = V = X$ p.s. On a aussi $|X_n| \leq Y$ p.s. et $|X| \leq Y$ p.s., donc $E(|X_n|) \leq E(Y) < \infty$, donc X_n est intégrable, et de même X est intégrable. D'une part, $X_n \geq -Y$ p.s. et $-Y \in \mathcal{L}^1$, donc (e) entraîne

$$E(U) \leq \liminf_{n \rightarrow \infty} E(X_n).$$

On a aussi $-X_n \geq -Y$ p.s. et $-V = \liminf_{n \rightarrow \infty} -X_n$, de sorte que

$$-E(V) = E(-V) \leq \liminf_{n \rightarrow \infty} E(-X_n) = -\limsup_{n \rightarrow \infty} E(X_n).$$

En rassemblant ces deux résultats, et grâce à (c), on arrive à

$$E(X) = E(U) \leq \liminf_{n \rightarrow \infty} E(X_n) \leq \limsup_{n \rightarrow \infty} E(X_n) \leq E(V) = E(X).$$

Les deux membres extrêmes étant égaux, tous les termes de l'inéquation ci-dessus sont égaux, et on déduit le résultat. ■

Une conséquence utile du théorème de convergence dominée de Lebesgue (partie (f) du théorème 9.1) est le résultat suivant, qui permet d'intervertir sommation et espérance. Comme la somme d'une série est la limite des sommes partielles et que l'espérance est en fait aussi une limite, cela revient à changer l'ordre dans lequel on prend les deux limites.

Théorème 9.2. *Soit X_n une suite de v.a.*

(a) *Si les X_n sont toutes positives, alors*

$$E\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E(X_n). \quad (9.8)$$

les deux membres étant simultanément finis ou infinis.

(b) *Si $\sum_{n=1}^{\infty} E(|X_n|) < \infty$, alors la série $\sum_{n=1}^{\infty} X_n$ converge p.s., sa somme est intégrable, et on a (9.8).*

Preuve. Soit $S_n = \sum_{k=1}^n |X_k|$ et $T_n = \sum_{k=1}^n X_k$. On a

$$E(S_n) = E\left(\sum_{k=1}^n |X_k|\right) = \sum_{k=1}^n E(|X_k|).$$

et la suite S_n croît vers la limite $S = \sum_{k=1}^{\infty} |X_k|$ (qui peut être infinie pour certaines valeurs de ω et finie pour d'autres). Donc le théorème de convergence monotone 9.1(d) implique :

$$E(S) = \lim_{n \rightarrow \infty} E(S_n) = \sum_{k=1}^{\infty} E(|X_k|) < \infty.$$

Si les X_n sont toutes positives on a $S_n = T_n$ et on a donc prouvé (a). Si les X_n ne sont pas nécessairement positives, mais si on a $\sum_{n=1}^{\infty} E(|X_n|) < \infty$, on obtient aussi que $E(S) < \infty$.

Maintenant, pour chaque $\varepsilon > 0$ on a $1_{\{S=\infty\}} \leq \varepsilon S$, donc

$$P(S = \infty) = E(1_{\{S=\infty\}}) \leq \varepsilon E(S).$$

Par suite, $E(S) < \infty$ et le fait que ε est arbitrairement proche de 0 entraîne que $P(S = \infty) = 0$; donc $\sum_{k=1}^{\infty} X_k$ est p.s. une série absolument convergente, et sa somme, disons T , est la limite des T_n . De plus

$$|T_n| \leq S_n \leq S$$

et $S \in L^1$. Par suite le théorème de convergence dominée (théorème 9.1(f)) implique

$$E\left(\sum_{k=1}^{\infty} X_k\right) = E\left(\lim_{n \rightarrow \infty} T_n\right) = E(T),$$

ce qui est (9.8). ■

Rappelons que L^1 et L^2 sont les ensembles de classes d'équivalence de v.a. intégrables (resp. de carré intégrable), pour la relation d'équivalence « égalité p.s. ».

Théorème 9.3. (a) Si X et Y sont dans L^2 , on a $XY \in L^1$ et l'inégalité de Cauchy-Schwarz :

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)};$$

(b) On a $L^2 \subset L^1$, et si $X \in L^2$ alors $E(X)^2 \leq E(X^2)$;

(c) L'espace L^2 est un espace vectoriel : si $X, Y \in L^2$ et $\alpha, \beta \in \mathbb{R}$, alors $\alpha X + \beta Y \in L^2$ (nous verrons au chapitre 22 qu'en fait L^2 est un espace de Hilbert).

Preuve. (a) On a $|XY| \leq X^2/2 + Y^2/2$, donc $X, Y \in L^2$ implique $XY \in L^1$. Pour tout $x \in \mathbb{R}$ on a

$$0 \leq E((xX + Y)^2) = x^2 E(X^2) + 2xE(XY) + E(Y^2). \quad (9.9)$$

Le discriminant du polynôme du second degré (en x) dans le membre de droite ci-dessus vaut

$$4 \{ (E(XY))^2 - E(X^2)E(Y^2) \},$$

et comme l'expression (9.9) est toujours positive le discriminant doit être négatif, ce qui donne l'inégalité de Cauchy-Schwarz.

(b) Soit $X \in L^2$. Comme $X = X \cdot 1$ et comme la fonction identiquement égale à 1 appartient à L^2 avec $E(1^2) = 1$, l'assertion se déduit de (a).

(c) Soit $X, Y \in L^2$. Pour toutes constantes α et β on a que $(\alpha X + \beta Y)^2 \leq \alpha^2 X^2/2 + \beta^2 Y^2/2$ est intégrable : donc $\alpha X + \beta Y \in L^2$ et L^2 est un espace vectoriel. ■

Si $X \in L^2$, la *variance* de X , notée aussi $\sigma^2(X)$ ou σ_X^2 , est

$$\text{Var}(X) = \sigma^2(X) = E((X - E(X))^2).$$

(Noter que $X \in L^2 \Rightarrow X \in L^1$, donc $E(X)$ existe.) Si $\mu = E(X)$ on a

$$\begin{aligned} \text{Var}(X) = E((X - \mu)^2) &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2. \end{aligned}$$

On a donc aussi l'égalité suivante, triviale mais très utile :

$$\sigma^2(X) = E(X^2) - E(X)^2.$$

Théorème 9.4. (Inégalité de Bienaymé-Chebyshev.) On a

$$P(|X| \geq a) \leq \frac{E(X^2)}{a^2}.$$

Preuve. Sur l'ensemble $\{|X| \geq a\}$ on a $a^2 1_{\{|X| \geq a\}} = a^2 \leq X^2$, et sur le complémentaire de cet ensemble on a $a^2 1_{\{|X| < a\}} = 0 \leq X^2$. Par suite la variable aléatoire $a^2 1_{\{|X| \geq a\}}$ est majorée partout par la variable aléatoire X^2 , et on a $E(a^2 1_{\{|X| \geq a\}}) \leq E(X^2)$. Donc

$$a^2 P(|X| \geq a) \leq E(X^2);$$

et le résultat s'obtient en divisant par a^2 . ■

Cette inégalité est souvent écrite pour la variable $X - E(X)$, ce qui donne :

$$P(|X - E(X)| \geq a) \leq \frac{\sigma^2(X)}{a^2}.$$

Le théorème suivant est très utile, de même que le corollaire qui le suit, pour les calculs effectifs d'espérance. Il montre que l'espérance d'une variable aléatoire ne dépend que de sa loi, et il s'agit d'une version de la formule « de changement de variable » dans les intégrales, vue sous un aspect « abstrait ».

Théorème 9.5. Soit X une v.a. sur (Ω, \mathcal{A}, P) , à valeurs dans (E, \mathcal{E}) , et de loi P_X . Soit $h : (E, \mathcal{E}) \mapsto (\mathbb{R}, \mathcal{B})$ une application mesurable.

- (a) On a $h(X) \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ si et seulement si $h \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, P_X)$.
 (b) Si h est positive, ou si elle satisfait les conditions équivalentes de (a), on a

$$E(h(X)) = \int h(x)P_X(dx). \quad (9.10)$$

Preuve. En se rappelant que la loi P_X est définie par $P_X(B) = P(X^{-1}(B))$, on voit que

$$E(1_B(X)) = P(X^{-1}(B)) = P^X(B) = \int 1_B(x)P^X(dx).$$

Par suite si h est simple, on a (9.10) par linéarité de l'espérance. Si h est positive on choisit une suite (h_n) de fonctions simples et positives, qui croît vers h . On a

$$\begin{aligned} E(h(X)) &= E\left(\lim_{n \rightarrow \infty} h_n(X)\right) \\ &= \lim_{n \rightarrow \infty} E(h_n(X)) \\ &= \lim_{n \rightarrow \infty} \int h_n(x)P_X(dx) \\ &= \int \lim_{n \rightarrow \infty} h_n(x)P_X(dx) \\ &= \int h(x)P_X(dx) \end{aligned}$$

en utilisant deux fois le théorème de convergence monotone. Cela prouve (b) quand h est positive. En l'appliquant à $|h|$, cela prouve aussi (a) (appelons que $X \in \mathcal{L}^1$ si et seulement si $E(|X|) < \infty$). Enfin en écrivant $h = h^+ - h^-$ on déduit (9.10) pour h de signe quelconque de la même formule pour h positive par soustraction. ■

Le résultat suivant peut se déduire du théorème 9.5, mais nous le prouverons en fait au chapitre 11 (corollaire 11.1) et nous omettons donc la preuve ici.

Corollaire 9.1. Soit X une v.a. de densité f (i.e. $P(X \leq x) = \int_{-\infty}^x f(u)du$ pour tout $x \in \mathbb{R}$). Si $E(|h(X)|) < \infty$ si h est positive, on a $E(h(X)) = \int h(x)f(x)dx$.

Exemples.

1. Soit X de loi exponentielle de paramètre α . Alors

$$E(h(X)) = \int_0^{\infty} h(x)\alpha e^{-\alpha x} dx.$$

En particulier, si $h(x) = x$ il vient

$$E(X) = \int_0^{\infty} \alpha x e^{-\alpha x} dx = \frac{1}{\alpha},$$

par intégration par parties. Ainsi, la moyenne d'une variable exponentielle est $1/\alpha$.

2. Soit X de loi $N(\mu, \sigma^2)$ (loi normale, ou gaussienne). Alors $E(X) = \mu$. Pour voir ceci, on écrit

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} x e^{-(x-\mu)^2/2\sigma^2} dx.$$

et on fait le changement de variable $y = x - \mu$, donc $x = y + \mu$, et il vient

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} (y + \mu) e^{-y^2/2\sigma^2} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} y e^{-y^2/2\sigma^2} dy + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2} dy. \end{aligned}$$

La première intégrale est l'intégrale d'une fonction impaire intégrable, donc est nulle; la seconde égale $\mu \int_{-\infty}^{\infty} f(x) dx = \mu \cdot 1 = \mu$.

3. Soit X une variable de Cauchy de densité $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$. On a $E(|X|) = \infty$, puisque

$$E(|X|) = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} dx = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{x}{1+x^2} dx \geq \frac{2}{\pi} \int_1^{\infty} \frac{1}{2x} dx = \infty$$

(en effet $\frac{x}{1+x^2} \geq 0$ pour $x \geq 0$ et $\frac{x}{1+x^2} \geq \frac{1}{2x}$ pour $x > 1$). Par suite l'espérance $E(X)$ n'existe pas.

Exercices des chapitres 8 et 9

1. Soit $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^k, \mathcal{B})$ une v.a. Posons

$$\mathcal{F} = \{A : A = X^{-1}(B), B \in \mathcal{B}\} = X^{-1}(\mathcal{B}).$$

Montrer que X est mesurable en tant qu'application de (Ω, \mathcal{F}) dans $(\mathbb{R}^k, \mathcal{B})$.

2.* Soit (Ω, \mathcal{A}, P) un espace de probabilité, et soit \mathcal{F} et \mathcal{G} deux tribus sur Ω . Supposons que $\mathcal{F} \subset \mathcal{A}$ et $\mathcal{G} \subset \mathcal{A}$ (on dit alors que \mathcal{F} et \mathcal{G} sont des *sous-tribus* de \mathcal{A}). Les tribus \mathcal{F} et \mathcal{G} sont dites *indépendantes* si pour tous $A \in \mathcal{F}$ et $B \in \mathcal{G}$ on a $P(A \cap B) = P(A)P(B)$. Supposons alors \mathcal{F} et \mathcal{G} indépendantes, et soit $X : \Omega \rightarrow (\mathbb{R}, \mathcal{B})$ une fonction mesurable à la fois pour la tribu \mathcal{F} et pour la tribu \mathcal{G} . Montrer que X est p.s. constante, i.e. $P(X = c) = 1$ pour une certaine constante c .

3.* Soit (Ω, \mathcal{A}, P) un espace de probabilité et $\mathcal{A}' = \{A \cup N : A \in \mathcal{A}, N \in \mathcal{N}\}$, où \mathcal{N} est la classe des ensembles négligeables (comme dans le théorème 6.1). Soit X et Y deux applications de Ω dans \mathbb{R} telles que l'ensemble $\{X \neq Y\}$ appartienne à \mathcal{N} (on écrit encore $X = Y$ p.s.). Montrer que $X : (\Omega, \mathcal{A}') \rightarrow (\mathbb{R}, \mathcal{B})$ est mesurable si et seulement si $Y : (\Omega, \mathcal{A}') \rightarrow (\mathbb{R}, \mathcal{B})$ est mesurable.

4.* Soit $X \in \mathcal{L}^1$ sur (Ω, \mathcal{A}, P) et considérons une suite d'événements (A_n) telle que $\lim_{n \rightarrow \infty} P(A_n) = 0$. Montrer que $\lim_{n \rightarrow \infty} E(X1_{A_n}) = 0$. (*Attention* : nous ne supposons pas que $\lim_{n \rightarrow \infty} X1_{A_n} = 0$ p.s.)

5.* Soit (Ω, \mathcal{A}, P) un espace de probabilité et X une v.a. réelle telle que $X \geq 0$ p.s. et $E(X) = 1$. Définissons $Q : \mathcal{A} \rightarrow \mathbb{R}$ par $Q(A) = E(X1_A)$. Montrer que Q est une mesure de probabilité sur (Ω, \mathcal{A}) .

6. Pour Q comme dans l'exercice 5, montrer que si $P(A) = 0$, alors $Q(A) = 0$. Donner un exemple montrant que $Q(A) = 0$ n'implique pas en général que $P(A) = 0$.

7.* Pour Q comme dans l'exercice 5, et en notant E_Q l'espérance par rapport à Q , montrer que $E_Q(Y) = E_P(YX)$ pour toute v.a. Y positive ou bornée.

8. Soit Q comme dans l'exercice 5, et supposons que $P(X > 0) = 1$.

(a) Montrer que $\frac{1}{X}$ est intégrable pour Q .

(b) Définissons $R : \mathcal{A} \rightarrow \mathbb{R}$ par $R(A) = E_Q(\frac{1}{X}1_A)$. Montrer que R est exactement la probabilité P . (*Indication* : utiliser l'exercice 7.)

9. Soit Q comme dans l'exercice 8. Montrer que $Q(A) = 0$ implique $P(A) = 0$ (comparer avec l'exercice 6).
10. Soit X une v.a. de loi uniforme sur $[a, b]$. Montrer que $E(X) = \frac{a+b}{2}$.
11. Soit X une v.a. de densité f , et $\mu = E(X)$ (qu'on suppose exister et être finie). Montrer que

$$\text{Var}(X) = \sigma^2(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

12. Soit X une v.a. de loi uniforme sur $[a, b]$. Montrer que $\sigma^2(X) = \frac{(b-a)^2}{12}$.
13. Soit X une v.a. de Cauchy de densité $\frac{1}{\pi(\Gamma + (x-a)^2)}$. Montrer que $\sigma^2(X)$ n'est pas définie et que $E(X^2) = \infty$.
14. La fonction bêta est $B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$, où Γ est la fonction gamma. De manière équivalente,

$$B(r, s) = \int_0^1 t^{r-1} (1-t)^{s-1} dt \quad (r > 0, s > 0).$$

Une v.a. X est dite *de loi bêta* de paramètres r et s si elle admet la densité

$$f(x) = \begin{cases} \frac{x^{r-1} (1-x)^{s-1}}{B(r, s)} & \text{si } 0 \leq x \leq 1, \\ 0 & \text{si } x < 0 \text{ ou } x > 1. \end{cases}$$

Montrer alors que

$$E(X^k) = \frac{B(r+k, s)}{B(r, s)} = \frac{\Gamma(r+k)\Gamma(r+s)}{\Gamma(r)\Gamma(r+s+k)},$$

pour $k \geq 0$. En déduire que

$$E(X) = \frac{r}{r+s},$$

$$\sigma^2(X) = \frac{rs}{(r+s)^2(r+s+1)}.$$

La famille des lois bêta est souvent utilisée pour modéliser des proportions aléatoires, puisqu'une v.a. de loi bêta prend ses valeurs entre 0 et 1.

15. Soit X une v.a. de loi lognormale de paramètres (μ, σ^2) . Montrer que

$$E(X^r) = e^{r\mu + \frac{1}{2}\sigma^2 r^2}$$

et déduire que $E(X) = e^{\mu + \frac{1}{2}\sigma^2}$ et que $\sigma_X^2 = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$. (*Indication* : $E(X^r) = \int_0^\infty x^r f(x) dx$ où f est la densité lognormale ; faire le changement de variables $y = \log x - \mu$ pour obtenir

$$E(X^r) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(r\mu + ry - y^2/2\sigma^2)} dy.)$$

16. On considère souvent une sous-famille des lois gamma : on dit qu'une v.a. X admet la *loi gamma standard* de paramètre α si elle admet la densité

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

(Le second paramètre β est pris égal à 1 : rappelons que $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.) Montrer que

$$E(X^k) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)} \quad (k \geq 0).$$

En déduire que X a la moyenne α et la variance α .

- 17.* Soit X une v.a. positive de moyenne μ et de variance σ^2 finies. Montrer que pour tout $b > 0$,

$$P(X \geq \mu + b\sigma) \leq \frac{1}{1 + b^2}.$$

(*Indication* : Considérer la fonction $g(x) = \frac{\{(x - \mu)b + \sigma\}^2}{\sigma^2(1 + b^2)^2}$ et établir que $E(\{(X - \mu)b + \sigma\}^2) = \sigma^2(b^2 + 1)$.)

18. Soit X une v.a. de moyenne μ et de variance σ^2 finies. Montrer que

$$P(\mu - d\sigma < X < \mu + d\sigma) \geq 1 - \frac{1}{d^2}.$$

(Noter que cette inégalité n'a d'intérêt que si $d > 1$.)

19. Soit X une v.a. normale $N(0, 1)$. Montrer que $P(X > x) \leq \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, pour $x > 0$.

20. Soit X une v.a. de loi exponentielle. Montrer que $P(X > s+t | X > s) = P(X > t)$ pour $s > 0, t > 0$. Ceci est connu sous le nom de « propriété de non vieillissement » de la loi exponentielle.
- 21.* Soit X une v.a. strictement positive vérifiant $P(X > s+t | X > s) = P(X > t)$ pour tous $s > 0, t > 0$. Montrer que si $h(t) = P(X > t)$, alors h vérifie l'équation de Cauchy :

$$h(s+t) = h(s)h(t) \quad (s > 0, t > 0)$$

et montrer que X admet une loi exponentielle (*Indication* : utiliser le fait que h est continue à droite, et que l'équation de Cauchy admet pour seules solutions continues à droite les exponentielles $h(t) = e^{at}$ pour $a \in \mathbb{R}$).

22. Soit α un entier strictement positif et supposons que la v.a. X admette la loi gamma de paramètres (α, β) . Montrer que $P(X \leq x) = P(Y \geq \alpha)$, où Y suit une loi de Poisson de paramètre $\lambda = x\beta$. (*Indication* : On a $\Gamma(\alpha) = (\alpha-1)!$; écrire $P(X \leq x)$, et utiliser une intégration par parties avec $u = t^{\alpha-1}$ et $dv = e^{-t\beta} dt$.)
23. L'intensité d'une v.a. strictement positive X est la fonction

$$h_X(t) = \lim_{\varepsilon \rightarrow 0} \frac{P(t \leq X < t + \varepsilon | X \geq t)}{\varepsilon}$$

lorsque la limite existe. Elle représente la probabilité pour qu'un objet de durée de vie X « meure » immédiatement après l'instant t , sachant qu'il est en vie à l'instant t . La propriété de non vieillissement des lois exponentielles implique que l'intensité soit constante. Une v.a. de loi de Weibull est aussi souvent utilisée pour modéliser les durées de vie. Montrer que

- a) Si X est exponentielle (λ), son intensité est $h_X(t) = \lambda$;
 b) Si X suit la loi de Weibull (α, β) , son intensité est $h_X(t) = \alpha\beta^\alpha t^{\alpha-1}$.

24. Une v.a. strictement positive X admet la loi *logistique* si sa fonction de répartition est donnée par

$$F(x) = P(X \leq x) = \frac{1}{1 + e^{-(x-\mu)/\beta}}$$

avec les paramètres $\mu \in \mathbb{R}$ et $\beta > 0$.

- a) Montrer que si $\mu = 0$ et $\beta = 1$, alors X admet la densité

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

- b) Montrer que si les paramètres sont (μ, β) , alors X admet l'intensité

$$h_X(t) = \frac{1}{\beta} F(t).$$

Chapitre 10

Variables aléatoires indépendantes

Rappelons que deux événements A et B sont indépendants si la connaissance du fait que B est réalisé ne change pas la probabilité de A , ce qui revient à dire que $P(A \cap B) = P(A)P(B)$. De même une famille quelconque $(A_i)_{i \in I}$ d'événements est indépendante si $P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$ pour tout sous-ensemble fini J de I (définition 3.1).

De manière analogue, deux v.a. X et Y seront indépendantes si la connaissance de la valeur prise par Y ne modifie pas la loi de X , ce qui, grossièrement, revient à dire que les événements $\{Y \in A\}$ et $\{X \in B\}$ sont indépendants pour tout choix de A et B dans les tribus correspondantes. Cela s'exprime de manière plus naturelle et plus facile à mettre en œuvre mathématiquement sur les tribus engendrées par X et Y : si X est mesurable de (Ω, \mathcal{A}) dans (E, \mathcal{E}) , la tribu engendrée par X est $X^{-1}(\mathcal{E}) = \{X^{-1}(A) : A \in \mathcal{E}\}$.

Définition 10.1. (a) Soit $(\mathcal{A}_i)_{i \in I}$ une famille de sous-tribus de \mathcal{A} . Les tribus \mathcal{A}_i sont indépendantes si pour toute partie finie J de I et tous $A_i \in \mathcal{A}_i$ on a

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i).$$

(b) Soit $(X_i)_{i \in I}$ une famille de variables aléatoires, à valeurs respectives dans les (E_i, \mathcal{E}_i) . Les variables X_i sont indépendantes si les tribus engendrées $X_i^{-1}(\mathcal{E}_i)$ sont indépendantes.

Dans la suite de ce chapitre nous considérons essentiellement des couples de v.a., de façon à garder des notations simples. Tous les résultats s'étendent de manière évidente aux familles finies de v.a.

Noter que X et Y prennent leurs valeurs dans (E, \mathcal{E}) et (F, \mathcal{F}) respectivement : les espaces E et F peuvent être différents.

Théorème 10.1. Pour que X et Y soient indépendantes, il faut et il suffit que l'une des conditions équivalentes ci-dessous soit satisfaite :

(a) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ pour tous $A \in \mathcal{E}$, $B \in \mathcal{F}$;

- (b) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ pour tous $A \in \mathcal{C}$, $B \in \mathcal{D}$, où \mathcal{C} et \mathcal{D} sont des classes de parties de E et F stables par intersection finie et engendrant les tribus \mathcal{E} et \mathcal{F} respectivement ;
- (c) $f(X)$ et $g(Y)$ sont indépendantes pour tout couple (f, g) de fonctions mesurables réelles sur E et F respectivement ;
- (d) $E(f(X)g(Y)) = E(f(X))E(g(Y))$ pour tout couple (f, g) de fonctions mesurables réelles, bornées (ou positives) sur E et F respectivement ;
- (e) Si de plus E et F sont des espaces métriques avec leurs tribus boréliennes respectives \mathcal{E} et \mathcal{F} , alors $E(f(X)g(Y)) = E(f(X))E(g(Y))$ pour tout couple (f, g) de fonctions continues réelles bornées sur E et F respectivement.

Preuve. Comme $X^{-1}(\mathcal{E})$ est l'ensemble des $\{X \in A\}$ pour les $A \in \mathcal{E}$, et de même pour $Y^{-1}(\mathcal{F})$, il est clair que (a) est juste une autre manière d'écrire la définition de l'indépendance de X et Y .

(a) \Rightarrow (b) : Implication triviale.

(b) \Rightarrow (a) : La classe \mathcal{C}_B des ensembles $A \in \mathcal{E}$ vérifiant $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ pour un $B \in \mathcal{D}$ donné est stable par limites croissantes et par différence, et contient \mathcal{C} par hypothèse, tandis que \mathcal{C} est stable par intersection. Le théorème des classes monotones 6.2 entraîne alors que $\mathcal{C}_B = \mathcal{E}$. En d'autres termes on a (b) avec $\mathcal{C} = \mathcal{E}$. De manière analogue (en fixant $A \in \mathcal{E}$) on obtient qu'on a (b) avec de plus $\mathcal{D} = \mathcal{F}$: en d'autres termes, on a (a).

(c) \Rightarrow (a) : Il suffit de prendre $f(x) = x$ et $g(y) = y$ (les applications identité sur E et F respectivement).

(a) \Rightarrow (c) : Étant données f et g , on voit que

$$f(X)^{-1}(\mathcal{E}) = X^{-1}(f^{-1}(\mathcal{E})) \subset X^{-1}(\mathcal{E}),$$

et de même $g(Y)^{-1}(\mathcal{F}) \subset Y^{-1}(\mathcal{F})$: le résultat est alors évident.

(d) \Rightarrow (a) : Prendre $f(x) = 1_A(x)$ et $g(x) = 1_B(x)$.

(a) \Rightarrow (d) : Par hypothèse on a (d) pour f et g fonctions indicatrices, et donc pour f et g fonctions simples (i.e. $f(x) = \sum_{i=1}^k a_i 1_{A_i}(x)$) par linéarité. Si f et g sont positives on peut trouver des suites (f_n) et (g_n) de fonctions simples positives croissant vers f et g respectivement. Noter que les produits $f_n(X)g_n(Y)$ croissent vers $f(X)g(Y)$. En appliquant le théorème de convergence monotone, il vient alors

$$\begin{aligned} E(f(X)g(Y)) &= E\left(\lim_{n \rightarrow \infty} f_n(X)g_n(Y)\right) \\ &= \lim_{n \rightarrow \infty} E(f_n(X)g_n(Y)) \end{aligned}$$

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} E(f_n(X))E(g_n(Y)) \\
&= \lim_{n \rightarrow \infty} E(f_n(X)) \lim_{n \rightarrow \infty} E(g_n(Y)) \\
&= E(f(X))E(g(Y)).
\end{aligned}$$

On a donc le résultat lorsque f et g sont positives. Lorsque f et g sont bornées on écrit $f = f^+ - f^-$ et $g = g^+ - g^-$ et on conclut par linéarité.

(d) \Rightarrow (e) : Évident.

(e) \Rightarrow (b) : Il suffit de montrer (b) lorsque \mathcal{C} et \mathcal{D} sont les classes de tous les fermés de E et F (ces classes sont stables par intersection). Soit par exemple A un fermé de E ; si $f_n(x) = \inf(1, nd(x, A))$, où $d(x, A)$ désigne la distance du point x au fermé A , alors f_n est continue, vérifie $0 \leq f_n \leq 1$ et décroît vers l'indicatrice 1_A . Pour B fermé de F soit de même une suite (g_n) de fonctions continues avec $0 \leq g_n \leq 1$ et convergant vers 1_B . On peut reproduire la fin de la preuve de l'implication (a) \Rightarrow (d) en remplaçant le théorème de convergence monotone par celui de convergence dominée, pour obtenir (b). ■

Exemple. Soit E et F finis ou dénombrables. Alors X et Y sont indépendantes si et seulement si $P(X = i, Y = j) = P(X = i)P(Y = j)$ pour tous $i \in E, j \in F$.

Nous présenterons d'autres exemples dans le chapitre 12.

Nous allons maintenant introduire les fonctions mesurables sur un produit d'espaces. En général si \mathcal{E} et \mathcal{F} sont deux tribus sur les espaces E et F respectivement, le produit cartésien $\mathcal{E} \times \mathcal{F} = \{A \subset E \times F : A = A \times \Gamma, A \subset \mathcal{E} \text{ et } \Gamma \in \mathcal{F}\}$ n'est pas une tribu. La plus petite tribu de $E \times F$ contenant $\mathcal{E} \times \mathcal{F}$, soit $\sigma(\mathcal{E} \times \mathcal{F})$, est très fréquemment rencontrée et est notée ainsi :

$$\mathcal{E} \otimes \mathcal{F} = \sigma(\mathcal{E} \times \mathcal{F}).$$

Théorème 10.2. Soit f mesurable : $(E \times F, \mathcal{E} \otimes \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{R})$. Pour tout $x \in E$ la « section » $y \mapsto f(x, y)$ est une fonction \mathcal{F} -mesurable ; de même pour tout $y \in F$ la fonction $x \mapsto f(x, y)$ est \mathcal{E} -mesurable.

Note. La réciproque du théorème 10.2 est fautive en général.

Preuve. Supposons d'abord f de la forme $f(x, y) = 1_C(x, y)$, pour $C \in \mathcal{E} \otimes \mathcal{F}$. Soit $\mathcal{H} = \{C \in \mathcal{E} \otimes \mathcal{F} : y \mapsto 1_C(x, y) \text{ est } \mathcal{F}\text{-mesurable pour chaque } x \in E \text{ fixé}\}$. Alors \mathcal{H} est une tribu (vérification immédiate) contenant $\mathcal{E} \times \mathcal{F}$, donc $\sigma(\mathcal{E} \times \mathcal{F}) \subset \mathcal{H}$. Comme par construction $\mathcal{H} \subset \sigma(\mathcal{E} \times \mathcal{F})$, il vient $\mathcal{H} = \mathcal{E} \otimes \mathcal{F}$. On a donc montré le résultat pour les fonctions indicatrices, donc par linéarité pour les fonctions mesurables simples. Par limite croissante (toute fonction mesurable positive étant

limite croissante de fonctions mesurables simples, et toute limite croissante de fonctions mesurables étant mesurable), on l'obtient aussi pour les fonctions mesurables positives. Enfin par différence (i.e. en écrivant $f = f^+ - f^-$) on l'obtient pour les fonctions mesurables quelconques. ■

Théorème 10.3. (Tonelli-Fubini.) Soit P et Q des probabilités sur (E, \mathcal{E}) et (F, \mathcal{F}) respectivement.

- (a) Posons $R(A \times B) = P(A)Q(B)$ pour $A \in \mathcal{E}$ et $B \in \mathcal{F}$. Alors R s'étend de manière unique en une probabilité sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$, notée $P \otimes Q$.
- (b) Pour toute fonction réelle f sur $E \times F$ qui est $\mathcal{E} \otimes \mathcal{F}$ -mesurable, et positive ou intégrable par rapport à $P \otimes Q$, la fonction $x \rightarrow \int f(x, y)Q(dy)$ est \mathcal{E} -mesurable, la fonction $y \rightarrow \int f(x, y)P(dx)$ est \mathcal{F} -mesurable, et on a

$$\begin{aligned} \int f d(P \otimes Q) &= \int \left\{ \int f(x, y)Q(dy) \right\} P(dx) \\ &= \int \left\{ \int f(x, y)P(dx) \right\} Q(dy). \end{aligned}$$

Preuve. (a) Soit $C \in \mathcal{E} \otimes \mathcal{F}$, et écrivons $C(x) = \{y : (x, y) \in C\}$. Pour $C = A \times B$, on a $C(x) = B$ si $x \in A$ et $C(x) = \emptyset$ sinon, donc

$$R(C) = P \otimes Q(C) = P(A)Q(B) = \int P(dx)Q[C(x)].$$

Soit $\mathcal{H} = \{C \in \mathcal{E} \otimes \mathcal{F} : x \rightarrow Q[C(x)] \text{ est } \mathcal{E}\text{-mesurable}\}$. Clairement \mathcal{H} est stable par limite croissante et par différence, tandis que $\mathcal{E} \times \mathcal{F} \subset \mathcal{H} \subset \mathcal{E} \otimes \mathcal{F}$, donc d'après le théorème des classes monotones on a $\mathcal{H} = \mathcal{E} \otimes \mathcal{F}$. Pour chaque $C \in \mathcal{H} = \mathcal{E} \otimes \mathcal{F}$, et comme $x \mapsto Q[C(x)]$ est mesurable et positive, on peut alors poser

$$R(C) = \int P(dx)Q[C(x)].$$

Nous devons montrer que R est une probabilité. On a d'abord

$$R(\Omega) = R(E \times F) = \int_E P(dx)Q[F] = 1.$$

Soit $C_n \in \mathcal{E} \otimes \mathcal{F}$ des ensembles deux à deux disjoints, et $C = \bigcup_{n=1}^{\infty} C_n$. Comme Q est une probabilité et que les $C_n(x)$ sont aussi deux à deux

disjoints, on a $Q[C(x)] = \sum_{n=1}^{\infty} Q[C_n(x)]$. Appliquons le théorème 9.2 à P et aux fonctions $f_n(x) = Q[C_n(x)]$:

$$\begin{aligned} \sum_{n=1}^{\infty} R(C_n) &= \sum_{n=1}^{\infty} \int f_n dP \\ &= \int \left(\sum_{n=1}^{\infty} f_n \right) dP \\ &= \int P(dx) Q[C(x)] = R(C). \end{aligned}$$

Donc R est une probabilité. L'unicité de R provient du corollaire 6.1.

(b) Nous avons déjà établi le résultat pour f de la forme $f(x, y) = 1_C(x, y)$, avec $C \in \mathcal{E} \otimes \mathcal{F}$. Il est vrai par linéarité pour les fonctions simples. Si f est positive et $\mathcal{E} \otimes \mathcal{F}$ -mesurable, soit f_n des fonctions mesurables croissant vers f . On a

$$\begin{aligned} \int f dR &= \lim_{n \rightarrow \infty} \int f_n dR = \lim_{n \rightarrow \infty} \int \left\{ \int f_n(x, y) Q(dy) \right\} P(dx) \\ &= \int \left\{ \lim_{n \rightarrow \infty} \int f_n(x, y) Q(dy) \right\} P(dx) \\ &= \int \left\{ \int \lim_{n \rightarrow \infty} f_n(x, y) Q(dy) \right\} P(dx) = \int \left\{ \int f(x, y) Q(dy) \right\} P(dx). \end{aligned}$$

Un argument similaire montre que l'expression ci-dessus égale aussi

$$= \int \left\{ \int f(x, y) P(dx) \right\} Q(dy).$$

Finalement pour une fonction f de signe arbitraire (intégrable par rapport à R) il suffit de décomposer en $f = f^+ - f^-$. ■

Corollaire 10.1. Soit X et Y deux v.a. sur (Ω, \mathcal{A}, P) , à valeurs respectives dans (E, \mathcal{E}) et (F, \mathcal{F}) . Le couple $Z = (X, Y)$ peut être considéré comme une v.a. à valeurs dans $(E \times F, \mathcal{E} \otimes \mathcal{F})$, et les deux v.a. X, Y sont indépendantes si et seulement si la loi $P_{(X, Y)}$ du couple égale le produit $P_X \otimes P_Y$ des lois de X et Y .

Preuve. Il est évident que $Z^{-1}(A \times B) = X^{-1}(A) \cap Y^{-1}(B) \in \mathcal{A}$ si $A \in \mathcal{E}$ et $B \in \mathcal{F}$, de sorte que la mesurabilité de Z découle de la définition de la tribu $\mathcal{E} \otimes \mathcal{F}$ et du théorème 8.1.

L'indépendance de X et Y revient au fait que pour tous $A \in \mathcal{E}$ et $B \in \mathcal{F}$ on ait

$$P((X, Y) \in A \times B) = P(X \in A)P(Y \in B),$$

ce qui équivaut à

$$P_{(X, Y)}(A \times B) = P_X(A)P_Y(B),$$

de sorte que l'unicité de la probabilité produit dans le théorème de Fubini nous donne la seconde assertion. ■

Nous faisons maintenant une légère digression en *construisant un modèle pour des variables aléatoires indépendantes*, ce qui permet en particulier de vérifier que cette notion n'est pas vide.

Soit d'abord μ une probabilité sur un espace mesurable (E, \mathcal{E}) . Il est facile de construire une v.a. X à valeurs dans E et de loi μ : prendre simplement $\Omega = E$, $\mathcal{A} = \mathcal{E}$, $P = \mu$, et prendre pour X l'identité ($X(x) = x$).

Un peu plus compliquée est la construction d'un couple de deux v.a. indépendantes X et Y , à valeurs dans (E, \mathcal{E}) et (F, \mathcal{F}) , et de lois respectives μ et ν (probabilités données *a priori* sur E et F) : on peut prendre $\Omega = E \times F$, $\mathcal{A} = \mathcal{E} \otimes \mathcal{F}$, $P = \mu \otimes \nu$, et $X(x, y) = x$ et $Y(x, y) = y$, où $(x, y) \in E \times F$.

Il est malheureusement bien plus compliqué, mais indispensable pour les applications (la loi des grands nombres par exemple qu'on verra plus tard, ou même l'étude d'une suite infinie de jets de dés, etc.), de construire une *suite infinie de variables indépendantes* de lois données. Plus précisément, pour chaque entier n on se donne une v.a. X_n définie sur un espace de probabilité $(\Omega_n, \mathcal{A}_n, P_n)$, à valeurs dans (E_n, \mathcal{E}_n) et de loi μ_n (pour construire chaque X_n on peut opérer comme ci-dessus). Ensuite, on pose

$$\Omega = \prod_{n=1}^{\infty} \Omega_n \quad (\text{produit cartésien dénombrable})$$

$$\mathcal{A} = \bigotimes_{n=1}^{\infty} \mathcal{A}_n$$

où $\bigotimes_{n=1}^{\infty} \mathcal{A}_n$ désigne la plus petite tribu de Ω à laquelle appartiennent tous les ensembles suivants :

$$A_1 \times A_2 \times \cdots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \cdots, \quad A_i \in \mathcal{A}_i, \quad k = 1, 2, 3, \dots$$

Le théorème suivant constitue un résultat non trivial de la théorie de la mesure, et généralise le théorème de Fubini. Nous l'énonçons sans démonstration.

Théorème 10.4. *Avec les notations ci-dessus, il existe une probabilité \mathbb{P} et une seule sur (Ω, \mathcal{A}) , telle que*

$$\mathbb{P}(A_1 \times \cdots \times A_k \times \Omega_{k+1} \times \cdots) = \prod_{i=1}^k \mathbb{P}_i(A_i)$$

pour tous $k = 1, 2, \dots$ et $A_i \in \mathcal{A}_i$.

Maintenant, on note \tilde{X}_n l'extension naturelle de X_n à Ω , c'est-à-dire que pour chaque $\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) \in \Omega$ avec $\omega_i \in \Omega_i$ on pose

$$\tilde{X}_n(\omega) = X_n(\omega_n).$$

Corollaire 10.2. *Les variables \tilde{X}_n ci-dessus sont indépendantes, et chaque \tilde{X}_n a la loi μ_n .*

Preuve. Soit $B_n \in \mathcal{E}_n$. On a

$$\tilde{X}_n^{-1}(B_n) = \Omega_1 \times \cdots \times \Omega_{n-1} \times X_n^{-1}(B_n) \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots,$$

et le théorème 10.4 donne pour $k = 1, 2, \dots$:

$$\begin{aligned} \mathbb{P}\left(\bigcap_{n=1}^k \tilde{X}_n^{-1}(B_n)\right) &= \mathbb{P}(X_1^{-1}(B_1) \times \cdots \times X_k^{-1}(B_k) \times \Omega_{k+1} \times \cdots) \\ &= \prod_{n=1}^k \mathbb{P}_n(X_n \in B_n) = \prod_{n=1}^k \mu_n(B_n). \end{aligned}$$

En particulier $\mathbb{P}(\tilde{X}_n \in B_n) = \mu_n(B_n)$, et on a les résultats souhaités. ■

Nous allons maintenant discuter quelques propriétés importantes de l'indépendance. Soit d'abord une suite (A_n) d'événements. On rappelle la définition de l'événement $\limsup_{n \rightarrow \infty} A_n$ donnée en (2.2) :

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \left(\bigcup_{m \geq n} A_m \right) = \lim_{n \rightarrow \infty} \left(\bigcup_{m \geq n} A_m \right).$$

Théorème 10.5. (Lemme de Borel-Cantelli.) *Soit (A_n) une suite d'événements sur $(\Omega, \mathcal{A}, \mathbb{P})$.*

- (a) Si $\sum_{n=1}^{\infty} P(A_n) < \infty$, alors $P(\limsup_{n \rightarrow \infty} A_n) = 0$.
 (b) Si $P(\limsup_{n \rightarrow \infty} A_n) = 0$ et si les événements A_n sont indépendants, alors on a $\sum_{n=1}^{\infty} P(A_n) < \infty$.

Remarque. Nous verrons plus bas (comme cas particulier de la « loi zéro-un ») que si les A_n sont indépendants, $P(\limsup_{n \rightarrow \infty} A_n)$ ne peut être égal que à 0 ou à 1. Par suite si la série $\sum_{n=1}^{\infty} P(A_n)$ converge ou a $P(\limsup_{n \rightarrow \infty} A_n) = 0$, et si elle diverge on a $P(\limsup_{n \rightarrow \infty} A_n) = 1$.

Preuve. (a) Soit $a_n = P(A_n) = E(1_{A_n})$. Grâce au théorème 9.2(b), $\sum_{n=1}^{\infty} a_n < \infty$ implique $\sum_{n=1}^{\infty} 1_{A_n} < \infty$ p.s., tandis que $\sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty$ si et seulement si $\omega \in \limsup_{n \rightarrow \infty} A_n$: donc on a (a).

(b) Supposons les A_n indépendants. On a

$$\begin{aligned} P(\limsup_{n \rightarrow \infty} A_n) &= \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} P\left(\bigcup_{m=n}^k A_m\right) \\ &= \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \left(1 - P\left(\bigcap_{m=n}^k A_m^c\right)\right) \\ &= 1 - \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \left(\prod_{m=n}^k (1 - P(A_m))\right) \quad (\text{par indépendance}) \\ &= 1 - \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \prod_{m=n}^k (1 - a_m). \end{aligned}$$

Par hypothèse $P(\limsup_{n \rightarrow \infty} A_n) = 0$, donc $\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \prod_{m=n}^k (1 - a_m) = 1$ et en prenant les logarithmes on arrive à

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \sum_{m=n}^k \log(1 - a_m) = 0,$$

donc

$$\lim_{n \rightarrow \infty} \sum_{m \geq n} \log(1 - a_m) = 0,$$

ce qui signifie que la série $\sum_m \log(1 - a_m)$ converge. Comme $|\log(1 - x)| \geq x$ pour $0 < x < 1$, la série $\sum_m a_m$ converge également. ■

Soit maintenant des v.à. X_n définies sur (Ω, \mathcal{A}, P) . Soit les tribus

$$\begin{aligned}
 \mathcal{B}_n &= \sigma(X_n) \\
 \mathcal{C}_n &= \sigma\left(\bigcup_{p \geq n} \mathcal{B}_p\right) \\
 \mathcal{C}_\infty &= \bigcap_{n=1}^{\infty} \mathcal{C}_n
 \end{aligned}$$

\mathcal{C}_∞ est appelée la *tribu asymptotique*. Un événement est dans cette tribu s'il ne dépend que du comportement de la suite X_n lorsque $n \rightarrow \infty$ (d'où son nom) ; nous donnons des exemples de v.a. \mathcal{C}_∞ -mesurables juste après le théorème suivant.

Théorème 10.6. (Loi zéro-un.) *Soit des v.a. indépendantes définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, et soit \mathcal{C}_∞ la tribu asymptotique associée. Pour tout $C \in \mathcal{C}_\infty$ on a $\mathbb{P}(C) = 0$ ou $\mathbb{P}(C) = 1$.*

Preuve. Soit $\mathcal{D}_n = \sigma(\bigcup_{p < n} \mathcal{B}_p)$. Par hypothèse les tribus \mathcal{C}_n et \mathcal{D}_n sont indépendantes, donc si $A \in \mathcal{C}_n$ et $B \in \mathcal{D}_n$ on a

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (10.1)$$

Si $A \in \mathcal{C}_\infty$ nous avons donc (10.1) pour tout $B \subset \bigcup \mathcal{D}_n$, donc aussi pour $B \in \mathcal{D} = \sigma(\bigcup \mathcal{D}_n)$ par le théorème des classes monotones 6.2. Cependant $\mathcal{C}_\infty \subset \mathcal{D}$, donc on a (10.1) pour $B = A \in \mathcal{C}_\infty$, ce qui implique que $\mathbb{P}(A) = \mathbb{P}(A)\mathbb{P}(A) = \mathbb{P}(A)^2$, donc $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$. ■

Conséquence. Si les v.a. $X_1, X_2, \dots, X_n, \dots$ sont indépendantes, on a

1. $\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) \text{ existe}\} \in \mathcal{C}_\infty$, donc soit la suite X_n converge p.s., soit elle diverge p.s.
2. Chaque v.a. qui est \mathcal{C}_∞ -mesurable est p.s. constante. En particulier,

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} X_n, & \quad \liminf_{n \rightarrow \infty} X_n, \\
 \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{p \leq n} X_p, & \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{p \leq n} X_p
 \end{aligned}$$

sont toutes p.s. constantes.

Exercices

1. Soit $f = (f_1, f_2) : \Omega \rightarrow E \times F$. Montrer que $f : (\Omega, \mathcal{A}) \rightarrow (E \times F, \mathcal{E} \otimes \mathcal{F})$ est mesurable si et seulement si f_1 est mesurable de (Ω, \mathcal{A}) dans (E, \mathcal{E}) et f_2 est mesurable de (Ω, \mathcal{A}) dans (F, \mathcal{F}) .
2. Soit $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, et soit \mathcal{B}^2 la tribu borélienne de \mathbb{R}^2 , tandis que \mathcal{B} désigne la tribu borélienne de \mathbb{R} . Montrer que $\mathcal{B}^2 = \mathcal{B} \otimes \mathcal{B}$.
3. Soit $\Omega = [0, 1]$ et \mathcal{A} la tribu borélienne de Ω , et soit $P(A) = \int 1_A(x) dx$ pour $A \in \mathcal{A}$. Soit $X(x) = x$. Montrer que X admet la loi uniforme sur $[0, 1]$.
4. Soit $\Omega = \mathbb{R}$ et $\mathcal{A} = \mathcal{B}$. Soit P donnée par $P(A) = \frac{1}{\sqrt{2\pi}} \int 1_A(x) e^{-x^2/2} dx$, et $X(x) = x$. Montrer que X admet la loi normale $N(0, 1)$.
5. Construire un exemple montrant que $E(XY) = E(X)E(Y)$ n'implique pas en général que X et Y soient indépendantes (on suppose que X , Y et XY sont intégrables).
6. Soit X et Y des v.a. indépendantes à valeurs dans \mathbb{N} , avec

$$P(X = i) = P(Y = i) = \frac{1}{2^i} \quad (i = 1, 2, \dots).$$

Trouver les probabilités des événements suivants :

- a) $P(\min(X, Y) \leq i)$ [Rép. : $1 - \frac{1}{4^i}$.]
 - b) $P(X = Y)$ [Rép. : $\frac{1}{3}$.]
 - c) $P(Y > X)$ [Rép. : $\sum_{i \geq 0} \frac{1}{2^i(2^{i+1}-1)}$.]
 - d) $P(X \text{ divise } Y)$ [Rép. : $\frac{1}{3}$.]
 - e) $P(X \geq kY)$ pour un entier $k \geq 1$ [Rép. : $\frac{1}{2^{1+k}-1}$.]
7. Soit X et Y deux v.a. indépendantes de loi géométrique de paramètres respectifs λ et μ . Soit $Z = \min(X, Y)$. Montrer que Z suit une loi géométrique et trouver son paramètre. [Rép. : $\lambda\mu$.]
 8. Soit X et Y dans L^2 . On définit la *covariance* de X et Y par

$$\text{Cov}(X, Y) = E((X - \mu)(Y - \nu))$$

où $E(X) = \mu$ et $E(Y) = \nu$. Montrer que

$$\text{Cov}(X, Y) = E(XY) - \mu\nu$$

et que, si de plus X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$.

9. Soit X et Y dans L^1 . Si X et Y sont indépendantes, montrer que $XY \in L^1$. Donner un exemple montrant que XY n'est pas nécessairement dans L^1 en général (i.e., si X et Y ne sont pas indépendantes).
- 10.* Soit n un entier premier plus grand que 2. Soit X et Y deux v.a. indépendantes et uniformément distribuées sur l'ensemble fini $\{0, 1, \dots, n-1\}$ (i.e., $P(X=i) = P(Y=i) = \frac{1}{n}$ pour $i = 0, 1, \dots, n-1$). Pour chaque entier r avec $0 \leq r \leq n-1$, soit $Z_r = X + rY \pmod{n}$.
- Montrer que les v.a. $\{Z_r : r = 0, \dots, n-1\}$ sont deux à deux indépendantes (i.e. Z_r et Z_s sont indépendantes si $r \neq s$).
 - Le même résultat est-il encore vrai si l'entier n n'est pas premier? [Rép. : Non.]
11. Soit X et Y deux v.a. indépendantes de loi $P(X=1) = P(Y=1) = \frac{1}{2}$ et $P(X=-1) = P(Y=-1) = \frac{1}{2}$. Soit $Z = XY$. Montrer que X, Y, Z sont indépendantes deux à deux, mais qu'elles ne sont pas globalement indépendantes.
12. Soit (A_n) une suite d'événements. Montrer que

$$P(\limsup A_n) \geq \limsup P(A_n).$$

13. Une suite (X_n) de v.a. est dite *complètement convergente* vers X si

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty \text{ pour tout } \varepsilon > 0.$$

Montrer que si la suite (X_n) est indépendante, la convergence complète équivaut à la convergence p.s.

En vue des deux exercices suivants, nous introduisons la notion de « mesure » (positive) : une mesure μ sur un espace mesurable (E, \mathcal{E}) est une application de \mathcal{E} dans $[0, \infty]$ qui vérifie l'axiome de σ -additivité (2.) de la définition 2.3. Cette mesure est dite *finie* si $\mu(E) < \infty$ (une probabilité est donc une mesure finie). elle est dite σ -finie s'il existe une suite $(A_n)_{n \geq 1}$ avec $A_n \in \mathcal{E}$ et $\bigcup_{n \geq 1} A_n = E$ et $\mu(A_n) < \infty$ pour tout n (un exemple — fondamental — de mesure σ -finie sur \mathbb{R} est la mesure de Lebesgue, définie dans le chapitre suivant).

14. Soit μ et ν deux mesures finies sur (E, \mathcal{E}) et (F, \mathcal{F}) respectivement. On définit $\lambda = \mu \otimes \nu$ sur $(E \times F, \mathcal{E} \times \mathcal{F})$ par $\lambda(A \times B) = \mu(A)\nu(B)$ pour $A \in \mathcal{E}$ et $B \in \mathcal{F}$.

- a) Montrer que λ s'étend de manière unique en une mesure finie sur $\mathcal{E} \otimes \mathcal{F}$;
- b) Soit $f : E \times F \rightarrow \mathbb{R}$ mesurable relativement à $\mathcal{E} \otimes \mathcal{F}$. Montrer le *théorème de Fubini* : si f est λ -intégrable, alors $x \rightarrow \int f(x, y) \nu(dy)$ et $y \rightarrow \int f(x, y) \mu(dx)$ sont respectivement \mathcal{E} et \mathcal{F} mesurables, et de plus

$$\int f d\lambda = \int \left(\int f(x, y) \mu(dx) \right) \nu(dy) = \int \left(\int f(x, y) \nu(dy) \right) \mu(dx).$$

(*Indication* : Utiliser le théorème 10.3.)

- 15.* Montrer que si μ et ν sont supposées σ -finies, et en supposant que $\lambda = \mu \otimes \nu$ existe (même définition que dans l'exercice précédent), alors

- a) $\lambda = \mu \otimes \nu$ est σ -finie ;
- b) (*Théorème de Fubini*.) Si $f : E \times F \rightarrow \mathbb{R}$ est mesurable relativement à $\mathcal{E} \otimes \mathcal{F}$ et est λ -intégrable, alors $x \rightarrow \int f(x, y) \nu(dy)$ et $y \rightarrow \int f(x, y) \mu(dx)$ sont respectivement \mathcal{E} et \mathcal{F} mesurables, et de plus

$$\int f d\lambda = \int \left(\int f(x, y) \mu(dx) \right) \nu(dy) = \int \left(\int f(x, y) \nu(dy) \right) \mu(dx).$$

(*Indication* : Utiliser l'exercice 14 sur les ensembles $E_j \times F_k$, où $\mu(E_j) < \infty$ et $\nu(F_k) < \infty$.)

- 16.* On jette indéfiniment une pièce de monnaie, avec $P(\text{Face}) = p$. Soit A_k l'événement selon lequel au moins k faces consécutifs apparaissent au cours des jets numérotés $2^k, 2^k + 1, \dots, 2^{k+1} - 1$. Montrer que $P(\limsup_{k \rightarrow \infty} A_k)$ vaut 1 ou 0 selon que $p \geq \frac{1}{2}$ ou que $p < \frac{1}{2}$.
17. Soit X_0, X_1, X_2, \dots une suite de v.a. indépendantes telles que $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$. Soit $Z_n = \prod_{j=0}^n X_j$. Montrer que Z_1, Z_2, Z_3, \dots sont indépendantes.
18. Soit X et Y des v.a. indépendantes, et supposons que $P(X + Y = \alpha) = 1$, où α est une constante. Montrer que X et Y sont p.s. des constantes.

Chapitre 11

Lois de probabilité sur \mathbb{R}

On a déjà vu qu'une probabilité sur $(\mathbb{R}, \mathcal{B})$ (\mathcal{B} est la tribu borélienne) est caractérisée par sa fonction de répartition

$$F(x) = P(\cdot] - \infty, x]).$$

Nous allons maintenant utiliser les outils développés précédemment pour étudier la *mesure de Lebesgue* sur \mathbb{R} .

Définition 11.1. La mesure de Lebesgue est une application $m : \mathcal{B} \rightarrow [0, \infty]$ qui vérifie les propriétés :

- (i) (*σ -additivité*) si A_1, A_2, A_3, \dots sont des boréliens deux à deux disjoints, alors

$$m\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m(A_i) :$$

- (ii) si $a, b \in \mathbb{R}$, $a < b$, alors $m(\cdot]a, b]) = b - a$.

Comme $\emptyset \subset]a, b]$ on a alors $m(\emptyset) \leq b - a$ pour tous $a < b$, donc $m(\emptyset) = 0$.

Théorème 11.1. La mesure de Lebesgue est unique.

Preuve. On suppose que m existe. Fixons $a < b$ dans \mathbb{R} , et posons

$$m_{a,b}(A) = \frac{m(A \cap]a, b])}{b - a}, \quad \forall A \in \mathcal{B}.$$

Alors $m_{a,b}$ est une probabilité sur $(\mathbb{R}, \mathcal{B})$ (la « loi uniforme sur $]a, b]$ »), et sa fonction de répartition est

$$F_{a,b}(x) = m_{a,b}(\cdot] - \infty, x]) = \begin{cases} 0 & \text{si } x < a \\ \frac{x - a}{b - a} & \text{si } a \leq x < b \\ 1 & \text{si } b \leq x. \end{cases} \quad (11.1)$$

Donc $m_{a,b}$ est déterminée de manière unique. De plus comme

$$m(A) = \sum_{n \in \mathbb{Z}} m_{n, n+1}(A), \quad \forall A \in \mathcal{B}. \quad (11.2)$$

(\mathbb{Z} désigne l'ensemble des entiers relatifs) on voit que m est déterminée de manière unique également. ■

Maintenant que nous savons que la mesure de Lebesgue est unique, il nous reste à montrer qu'elle existe!

Théorème 11.2. *La mesure de Lebesgue existe.*

Preuve. Comme $F_{a,b}$ donnée par (11.1) existe et est croissante, continue, et vaut 0 pour x assez petit et 1 pour x assez grand, la mesure $m_{a,b}$ existe d'après le théorème 7.1. Il suffit donc de définir m par (11.2). La vérification de la σ -additivité et de $m([a,b]) = b - a$ est immédiate. ■

La théorie de l'intégration décrite au chapitre 9 reste valide pour la mesure de Lebesgue; la seule différence est que $m(\mathbb{R})$ n'est pas égal à 1, mais à $+\infty$: tous les résultats du chapitre 9 restent vrais, à l'exception du fait que toute fonction borélienne bornée est intégrable (théorème 9.1(b)) et du fait que $L^2 \subset L^1$ (théorème 9.3(b)), qui sont maintenant faux: par exemple la fonction identiquement égale à 1 est borélienne bornée mais pas intégrable; la fonction $f(x) = \frac{1}{x} 1_{[1,\infty[}(x)$ est de carré intégrable mais pas intégrable.

Si f est une fonction borélienne intégrable par rapport à la mesure de Lebesgue, son intégrale est notée $\int f(x)dx$. Rappelons que f est intégrable si et seulement si $\int f^+(x)dx < \infty$ et $\int f^-(x)dx < \infty$, où $f = f^+ - f^-$. Toute fonction borélienne à support compact (i.e., nulle en dehors d'un intervalle $[-N, N]$) et intégrable au sens de Riemann est aussi intégrable au sens de Lebesgue (la réciproque est fautive), et dans ce cas les deux intégrales coïncident.

Définition 11.2. *La densité d'une probabilité P sur $(\mathbb{R}, \mathcal{B})$ est une fonction borélienne positive f qui vérifie pour tout $x \in \mathbb{R}$:*

$$P(]-\infty, x]) = \int_{-\infty}^x f(y)dy = \int f(y)1_{]-\infty, x]}(y)dy. \quad (11.3)$$

Si $P = P_X$, la loi d'une v.a. réelle X , on dit alors que X admet la densité f .

Attention: comme on l'a déjà vu, il n'est pas vrai que toute probabilité sur $(\mathbb{R}, \mathcal{B})$ admette une densité, puisque dans ce cas la fonction de répartition F est continue, ce qui n'est pas vrai de toutes les fonctions de répartition. Il existe même des fonctions de répartition continues qui n'admettent pas de densité.

Théorème 11.3. (a) Une fonction borélienne positive f sur \mathbb{R} est la densité d'une probabilité P sur $(\mathbb{R}, \mathcal{B})$ si et seulement si elle vérifie $\int f(x)dx = 1$. Dans ce cas elle détermine la probabilité de manière unique, et toute autre fonction borélienne positive g vérifiant $m(f \neq g) = 0$ est aussi une densité de P .

(b) Inversement si une probabilité admet une densité, celle-ci est déterminée de manière unique à un ensemble de mesure de Lebesgue nulle près (i.e., si f et g sont deux densités pour la même probabilité, on a $m(f \neq g) = 0$).

Preuve. (a) Soit f la densité de P . Par (11.3) on a $\int_{-\infty}^x f(y)dy = P((-\infty, x])$. Si on fait tendre x vers l'infini, on voit que

$$\int f(y)dy = \int_{-\infty}^{\infty} f(y)dy = \lim_{x \rightarrow -\infty} \int_{-\infty}^x f(y)dy = 1.$$

donc $\int f(x)dx = 1$.

Pour la condition suffisante, on pourrait utiliser la fonction de répartition, mais nous allons faire une démonstration directe, qui s'étend immédiatement au cas des probabilités sur \mathbb{P}^n étudiées au chapitre suivant. Soit donc f une fonction borélienne positive vérifiant $\int f(x)dx = 1$. Pour tout borélien A on pose

$$P(A) = \int_A f(y)dy = \int f(y)1_A(y)dy. \quad (11.4)$$

Cela définit une application $P : \mathcal{B} \rightarrow \mathbb{R}_+$ qui vérifie clairement $P(\mathbb{R}) = 1$. De plus si les $A_n \in \mathcal{B}$ sont deux à deux disjoints, alors

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= \int f(x)1_{(\bigcup_{i=1}^{\infty} A_i)}(x)dx \\ &= \int \left(\sum_{i=1}^{\infty} f(x)1_{A_i}(x)\right)dx \\ &= \sum_{i=1}^{\infty} \int f(x)1_{A_i}(x)dx = \sum_{i=1}^{\infty} P(A_i) \end{aligned}$$

en utilisant le théorème 9.2. Il s'ensuit que P est une probabilité sur $(\mathbb{R}, \mathcal{B})$. Si $A =]-\infty, x]$ il vient

$$P(]-\infty, x]) = \int_{-\infty}^x f(y)dy,$$

donc P admet f pour densité. L'unicité de P admettant f pour densité provient immédiatement du théorème 7.1. Enfin si g est une autre fonction borélienne telle que $m(f \neq g) = 0$, dans (11.4) on peut remplacer f par g sans changer $P(A)$, ce qui montre que g est aussi une densité de P .

(b) Montrons maintenant que si P admet deux densités f et g , alors $m(f \neq g) = 0$. On a alors (11.4) pour f et aussi pour g (pour vérifier ceci, on définit P' par (11.4) avec g au lieu de f , et on observe que P et P' admettent la même fonction de répartition, donc sont égales). Soit alors $n \in \mathbb{N}$ et $A_n = \{x : f(x) + \frac{1}{n} \leq g(x)\}$. On a

$$P(A_n) + \frac{1}{n}m(A_n) = \int \left(f(x) + \frac{1}{n} \right) 1_{A_n}(x) dx \leq \int g(x) 1_{A_n}(x) dx = P(A_n),$$

donc nécessairement $m(A_n) = 0$. Comme A_n croît vers $\{f < g\}$ quand $n \rightarrow \infty$, on obtient $m(\{g < f\}) = 0$ par le théorème de limite monotone. On obtient de même $m(\{g > f\}) = 0$, d'où le résultat. ■

Remarque 11.1. La densité f et la fonction de répartition F étant liées par $F(x) = \int_{-\infty}^x f(y) dy$, on est tenté de conclure que F est dérivable et que sa dérivée vaut $F'(x) = f(x)$. C'est vrai en tout point x où la fonction f est continue. On peut montrer (c'est un résultat — difficile — dû à Lebesgue) que c'est vrai pour m -presque tout x . Comme f n'est définie de manière unique qu'à un ensemble de mesure de Lebesgue nulle près, « concrètement » si on connaît F et si on sait que f existe (par exemple si F est continue partout et dérivable par morceaux), on peut prendre pour f la dérivée de F partout où celle-ci existe, et des valeurs arbitraires (par exemple 0) ailleurs.

Corollaire 11.1. Soit X une v.a. réelle admettant la densité f . Si g est une fonction borélienne sur \mathbb{R} , elle est intégrable (resp. admet une intégrale) par rapport à la loi P_X de X si et seulement si la fonction produit fg est intégrable (resp. admet une intégrale) par rapport à la mesure de Lebesgue, et on a alors

$$E(g(X)) = \int g(x) P_X(dx) = \int g(x) f(x) dx. \quad (11.5)$$

Preuve. L'égalité (11.5) est satisfaite pour les fonctions g indicatrices grâce au théorème 11.3, car elle se réduit alors à (11.4). Par linéarité elle est donc vraie pour les fonctions boréliennes simples, puis par limite croissante pour les fonctions boréliennes positives (les trois termes de (11.5) étant simultanément finis, ou infinis). Pour les fonctions de signe

quelconque, on déduit le résultat par linéarité encore, en utilisant les décompositions $g = g^+ - g^-$ et $fg = fg^+ - fg^-$. ■

Nous avons présenté divers exemples de densités au chapitre 7. Dans tous ces exemples la densité était continue ou continue par morceaux, tandis qu'ici nous considérons le cas plus général où la densité est seulement borélienne. La plupart des exemples concrets donnent lieu à des densités régulières, mais dès qu'on fait des opérations relativement simples sur les v.a. ayant des densités régulières (par exemple si l'on prend des espérances conditionnelles, voir plus loin), on obtient de nouvelles variables pouvant avoir des densités simplement mesurables.

Soit X une v.a. réelle de densité f . Soit $Y = g(X)$ pour une autre fonction borélienne g donnée. Peut-on exprimer la densité de Y , si elle existe, en termes de f et g ? On le peut, dans les « bons cas », comme nous allons le voir maintenant. Commençons par un résultat trivial :

Théorème 11.4. *Soit X une v.a. de densité f_X et g une fonction borélienne. Soit $Y = g(X)$. La fonction de répartition de Y est*

$$F_Y(y) = P(Y \leq y) = \int_{A_y} f_X(u) du \quad \text{où } A_y = \{u : g(u) \leq y\}.$$

Si F_Y est continue partout et dérivable sauf en un nombre fini de points, on peut utiliser le résultat ci-dessus pour obtenir la densité f_Y de Y .

Exemple. Soit X uniforme sur $[0, 1]$ et $Y = -\frac{1}{\lambda} \log X$, où $\lambda > 0$. Alors

$$\begin{aligned} F_Y(y) &= P\left(-\frac{1}{\lambda} \log X \leq y\right) \\ &= P(\log X \geq -\lambda y) \\ &= P(X \geq \exp(-\lambda y)) \\ &= \begin{cases} 1 - e^{-\lambda y} & \text{si } y \geq 0 \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

Donc (cf. la remarque 11.1)

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \lambda e^{-\lambda y} & \text{si } y > 0 \\ 0 & \text{si } y \leq 0 \end{cases}$$

et on voit que Y est exponentielle de paramètre λ .

Attention : cet exemple est très simple car g est injective. Le résultat général pour g injective est donné ci-dessous, le cas non injectif étant donné plus bas.

Corollaire 11.2. *Supposons que X admette une densité continue f_X . Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continûment dérivable, dont la dérivée ne s'annule pas (elle est donc strictement monotone), et soit $h = g^{-1}$ sa fonction réciproque, qui est définie sur l'image $g(\mathbb{R})$ de \mathbb{R} par g et continûment dérivable sur $g(\mathbb{R})$. Alors $Y = g(X)$ admet la densité*

$$f_Y(y) = \begin{cases} f_X(h(y))|h'(y)| & \text{si } y \in g(\mathbb{R}) \\ 0 & \text{sinon.} \end{cases}$$

Preuve. Supposons g croissante. L'image $g(\mathbb{R})$ est un intervalle $]a, b[$, avec éventuellement $a = -\infty$ et/ou $b = +\infty$. Soit $F_Y(y) = P(Y \leq y)$. Si $a < y < b$ on a

$$F_Y(y) = P(g(X) \leq y) = P(h(g(X)) \leq h(y)),$$

car h est strictement croissante. Donc ce qui précède vaut

$$= P(X \leq h(y)) = F_X(h(y)) = \int_{-\infty}^{h(y)} f(x) dx.$$

On sait par ailleurs que la dérivée de h est $h'(x) = \frac{1}{f'(h(x))}$. Donc F_Y est dérivable sur $]a, b[$ et

$$\frac{d}{dy} F_Y(y) = f(h(y))h'(y) = f(h(y))|h'(y)|.$$

Enfin $F_Y(y) = 0$ si $y \leq a$, quand $a > -\infty$, et de même $F_Y(y) = 1$ si $y \geq b$, quand $b < \infty$; donc F_Y est également dérivable, et de dérivée nulle, sur $] -\infty, a[$ et sur $]b, \infty[$. On en déduit le résultat.

Si g est décroissante, pour $y \in g(\mathbb{R})$ le même argument conduit à

$$\frac{d}{dy} F_Y(y) = f(h(y))(-h'(y)) = f(h(y))|h'(y)|.$$

et on conclut comme ci-dessus. ■

Corollaire 11.3. *Supposons que X admette une densité continue par morceaux f_X . Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continûment dérivable et strictement monotone par morceaux, i.e. il existe une partition I_1, \dots, I_n de \mathbb{R} constituée d'intervalles et telle que g soit continûment dérivable avec $g' \neq 0$ sur chaque intervalle ouvert I_i ayant mêmes extrémités que I_i . Pour chaque i on note h_i la fonction réciproque de la restriction de g à I_i , et $\Lambda_i = g(I_i)$ l'image de l'intervalle I_i par g . Alors la v.a. $Y = g(X)$ admet la densité*

$$f_Y(y) = \sum_{i=1}^n f_X(h_i(y)) |h'_i(y)| 1_{\Lambda_i}(y).$$

Remarque. La preuve, analogue à celle du corollaire précédent, est laissée au lecteur. Cette méthode de démonstration utilise la continuité de f_X , mais le résultat reste vrai quand f_X est simplement borélienne.

Exemple. Soit X une v.a. de loi $N(0, 1)$, et $Y = X^2$. On applique le résultat précédent à $g(x) = x^2$, qui n'est pas injective. Prenons $I_1 =]0, \infty[$ et $I_2 =]-\infty, 0[$. Alors g est injective et strictement monotone sur I_1 et sur I_2 , de sorte que $h_1 :]0, \infty[\rightarrow \mathbb{R}$ avec $h_1(y) = \sqrt{y}$ et $h_2 :]-\infty, 0[\rightarrow \mathbb{R}$ avec $h_2(y) = -\sqrt{y}$. On a

$$|h'_i(y)| = \left| \frac{1}{2\sqrt{y}} \right| = \frac{1}{2\sqrt{y}}, \quad \text{pour } i = 1, 2.$$

Donc le corollaire 11.3 entraîne

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} \quad (y > 0) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2} 1_{(0, \infty)}(y). \end{aligned}$$

La v.a. Y est appelée une variable χ^2 (ou, « chi-deux ») à 1 degré de liberté.

L'exemple ci-dessus est suffisamment simple pour pouvoir être traité « à la main », sans le corollaire 11.3. En fait

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}); \end{aligned}$$

et

$$F_X(\sqrt{y}) = \int_{-\infty}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Donc en dérivant on obtient

$$\frac{d}{dy} F_X(\sqrt{y}) = \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} 1_{(y>0)}.$$

De même

$$-F_X(-\sqrt{y}) = - \int_{-\infty}^{-\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

donc

$$\begin{aligned} \frac{d}{dy}(-F(-\sqrt{y})) &= -\frac{1}{\sqrt{2\pi}}e^{-y/2} \frac{-1}{2\sqrt{y}} 1_{(y>0)} \\ &= \frac{1}{\sqrt{2\pi}}e^{-y/2} \frac{1}{2\sqrt{y}} 1_{(y>0)}, \end{aligned}$$

et en additionnant on obtient le même résultat qu'auparavant.

Remarque. Les lois du chi-deux jouent un rôle important en statistique. Soit p un entier supérieur ou égal à 1. La probabilité sur \mathbb{R} de densité

$$f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-\frac{x}{2}}, \quad 0 < x < \infty$$

est appelée *loi du chi-deux à p degrés de liberté*, ou χ_p^2 . Il s'agit d'un cas particulier des lois gamma, celles de paramètres $(\frac{p}{2}, \frac{1}{2})$. On vient de voir que si X admet la loi χ_1^2 , alors $X = Z^2$ où Z est $N(0, 1)$. On verra au chapitre 15 (exemple 6) que si X est χ_p^2 on peut l'écrire comme $X = \sum_{i=1}^p Z_i^2$, où les Z_i sont indépendantes de loi $N(0, 1)$. De telles lois apparaissent en statistique lorsqu'on cherche à estimer la variance (inconnue) d'une famille de v.a. indépendantes de loi normale (voir l'exercice 13 du chapitre 15).

Noter aussi que χ_2^2 est simplement la loi exponentielle de paramètre $\lambda = \frac{1}{2}$.

Exercices

1. Utiliser la densité de la loi du chi-deux pour montrer que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
2. Soit X une v.a. uniforme sur $[-1, 1]$. Trouver la densité de $Y = X^k$ pour k entier ≥ 1 . [Rép. : pour k impair, $f_Y(y) = \frac{1}{2k} y^{\frac{1}{k}-1} 1_{[-1,1]}(y)$; pour k pair, $f_Y(y) = \frac{1}{k} y^{\frac{1}{k}-1} 1_{[0,1]}(y)$.]
3. Soit X une v.a. de fonction de répartition F . Quelle est la fonction de répartition de $Y = |X|$? Quand X admet une densité f_X , montrer que Y admet aussi une densité f_Y , et exprimer f_Y en fonction de f_X .
4. Soit X une v.a. de Cauchy de paramètres α et 1. Soit $Y = \frac{\alpha}{X}$ avec $\alpha \neq 0$. Montrer que Y est aussi de Cauchy et trouver ses paramètres. [Rép. : $\frac{\alpha\alpha}{1+\alpha^2}$, $\sqrt{\frac{|\alpha|}{1+\alpha^2}}$.]

5. Soit X une v.a. de densité f_X , et soit $Y = \frac{a}{X}$ avec $a \neq 0$. Trouver la densité f_Y de Y en fonction de f_X . [Rép. : $f_Y(y) = \frac{|a|}{y^2} f_X(\frac{a}{y})$.]
6. Soit X une v.a. uniforme sur $] -\pi, \pi[$, et $Y = \sin(X + \theta)$, avec $\theta \in \mathbb{R}$. Montrer que Y admet la densité $f_Y(y) = \frac{2}{2\pi\sqrt{1-y^2}} 1_{[-1,1]}(y)$.
- 7* Soit X une v.a. de densité f_X , et $Y = a \sin(X + \theta)$ avec $a > 0$ et $\theta \in \mathbb{R}$. Montrer que Y admet la densité

$$f_Y(y) = \frac{1}{\sqrt{a^2 - y^2}} \sum_{i=-\infty}^{\infty} (f_X(h_i(y)) + f_X(k_i(x))) 1_{[-a,a]}(y)$$

pour des fonctions h_i et k_i qu'on déterminera.

8. Soit X une v.a. uniforme sur $] -\pi, \pi[$, et $Y = a \tan X$ avec $a > 0$. Trouver la densité f_Y de Y . [Rép. : $f_Y(y) = \frac{a/\pi}{a^2 + y^2}$.]
- 9* Soit X une v.a. de densité f_X , et

$$Y = ce^{-\alpha X} 1_{\{X > 0\}}, \quad (\alpha > 0, c > 0).$$

Trouver la densité f_Y de Y en fonction de f_X . [Rép. : cette densité n'existe pas si $P(X \leq 0) > 0$, et sinon elle existe et vaut $f_Y(y) = \frac{f_X(-\frac{1}{\alpha} \ln(\frac{y}{c}))}{\alpha y} 1_{]0,c[}(y)$.]

10. Une densité f est appelée *symétrique* si $f(-x) = f(x)$, i.e. si c'est une fonction paire. Une v.a. X est dite *symétrique* si X et $-X$ ont même loi. Montrer qu'une v.a. X admettant une densité est symétrique si et seulement si elle admet une densité symétrique. Dans ce cas, admet-elle aussi une densité non symétrique? [Rép. : Oui, il suffit de modifier la densité symétrique sur un ensemble de mesure de Lebesgue nulle sur $]0, \infty[$.] [Note. Exemples de densités symétriques : la densité uniforme sur $[-a, a]$; la densité normale $N(0, \sigma^2)$; la densité exponentielle double de paramètres θ et β ; la densité de Cauchy de paramètres θ et β .]
11. Soit X une v.a. positive de densité f , et $Y = \frac{1}{X+1}$. Trouver la densité de Y .
12. Soit X une v.a. normale $N(\mu, \sigma^2)$. Montrer que $Y = e^X$ est de loi lognormale.
13. Soit X une v.a. de fonction de répartition F continue. Montrer que la v.a. $Y = F(X)$ est uniforme sur $[0, 1]$.

14. Soit F une fonction de répartition continue et injective, de sorte que sa fonction réciproque F^{-1} soit bien définie sur $]0, 1[$. Soit U une v.a. uniforme sur $]0, 1[$. Montrer que $X = F^{-1}(U)$ admet F pour fonction de répartition.
- 15.* Soit F une fonction de répartition continue, et U une v.a. uniforme sur $]0, 1[$. Posons $G(u) = \inf\{x : F(x) \geq u\}$. Montrer que $X = G(U)$ admet F pour fonction de répartition.
16. Soit $Y = -\frac{1}{\lambda} \log U$, où U est uniforme sur $]0, 1[$. Montrer que Y est exponentielle de paramètre λ en inversant la fonction de répartition de la loi exponentielle (*Indication* : si U est uniforme sur $]0, 1[$, il en est de même de $1 - U$). Cela donne une méthode pour « simuler » une v.a. exponentielle à partir d'une v.a. uniforme.

Chapitre 12

Probabilités sur \mathbb{R}^n

Dans le chapitre 11 nous avons étudié les lois (probabilités) sur $(\mathbb{R}, \mathcal{B})$. Le cas des probabilités sur $(\mathbb{R}^n, \mathcal{B}^n)$ pour $n \geq 2$ est à la fois similaire et plus compliqué. (\mathcal{B}^n désigne la tribu borélienne de \mathbb{R}^n .)

D'abord, et avec une démonstration essentiellement identique à celle du théorème 2.1, on voit que \mathcal{B}^n est la tribu engendrée par les « rectangles » de la forme

$$\prod_{i=1}^n]-\infty, a_i] : \quad a_i \in \mathbb{Q}.$$

En particulier $\mathcal{B} \otimes \mathcal{B} \otimes \cdots \otimes \mathcal{B} = \mathcal{B}^n$, i.e., \mathcal{B}^n est aussi la tribu engendrée par le produit cartésien $\mathcal{B} \times \cdots \times \mathcal{B}$.

La *fonction de répartition* d'une probabilité sur $(\mathbb{R}^n, \mathcal{B}^n)$ est la fonction sur \mathbb{R}^n définie par

$$F(x_1, \dots, x_n) = P \left(\prod_{i=1}^n]-\infty, x_i] \right).$$

La caractérisation des fonctions de répartition lorsque $n \geq 2$ est nettement plus délicate que dans le cas $n = 1$, aussi ne sont-elles que très rarement utilisées.

Nous avons aussi observé que la densité d'une probabilité sur \mathbb{R} , lorsqu'elle existe, est un outil très efficace. Contrairement aux fonctions de répartition, la notion de densité se généralise au cas $n \geq 2$ en offrant la même facilité d'utilisation.

Définition 12.1. La mesure de Lebesgue m_n sur $(\mathbb{R}^n, \mathcal{B}^n)$ est définie sur les produits cartésiens $A_1 \times A_2 \times \cdots \times A_n$ par

$$m_n \left(\prod_{i=1}^n A_i \right) = \prod_{i=1}^n m(A_i), \quad \forall A_i \in \mathcal{B}, \quad (12.1)$$

où m est la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B})$. Comme dans le théorème 10.3, cette formule permet de définir une unique mesure m_n sur $(\mathbb{R}^n, \mathcal{B}^n)$ (vérifiant (12.1)), et m_n est même caractérisée par le fait qu'on ait

$$m_n \left(\prod_{i=1}^n]a_i, b_i] \right) = \prod_{i=1}^n (b_i - a_i), \quad \forall \infty < a_i < b_i < \infty. \quad (12.2)$$

Si $A \in \mathcal{B}^n$, on peut se représenter $m_n(A)$ comme le « volume » de l'ensemble A , comme le suggère la formule (12.2). On note

$$\int f(x) dx = \int f(x_1, \dots, x_n) dx_1 dx_2 \dots dx_n$$

l'intégrale de f par rapport à m_n , et $\int_A f(x) dx$ celle de $f 1_A$ quand $A \in \mathcal{B}^n$, exactement comme dans le cas uni-dimensionnel.

Définition 12.2. Une probabilité P sur $(\mathbb{R}^n, \mathcal{B}^n)$ admet la densité f si cette fonction est borélienne positive sur \mathbb{R}^n et vérifie

$$\begin{aligned} P(A) &= \int_A f(x) dx = \int f(x) 1_A(x) dx \\ &= \int f(x_1, \dots, x_n) 1_A(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

pour tout $A \in \mathcal{B}^n$.

Une fois de plus, nous insistons sur le fait que certaines probabilités sur $(\mathbb{R}^n, \mathcal{B}^n)$ n'ont pas de densité : par exemple, et exactement comme pour \mathbb{R} , la masse de Dirac en un point $a \in \mathbb{R}^n$ n'admet pas de densité.

Le théorème suivant est l'exact analogue du théorème 11.3, avec une preuve identique, que nous ne répétons pas.

Théorème 12.1. (a) Une fonction borélienne positive f sur \mathbb{R}^n est la densité d'une probabilité P sur $(\mathbb{R}^n, \mathcal{B}^n)$ si et seulement si elle vérifie $\int f(x) dx = 1$. Dans ce cas elle détermine la probabilité de manière unique, et toute autre fonction borélienne positive g vérifiant $m_n(f \neq g) = 0$ est aussi une densité de P .

(b) Inversement si une probabilité admet une densité, celle-ci est déterminée de manière unique à un ensemble de mesure de Lebesgue nulle près (i.e., si f et g sont deux densités pour la même probabilité, on a $m_n(f \neq g) = 0$).

Afin de garder des notations simples le résultat suivant est écrit pour $n = 2$; il s'étend, sans aucune difficulté, au cas $n \geq 3$.

Soit X une v.a. à valeurs dans \mathbb{R}^2 , de composantes Y et Z : i.e., $X = (Y, Z)$. On dit que X admet la densité f si sa loi (donc une probabilité sur \mathbb{R}^2) admet f pour densité.

Théorème 12.2. Si $X = (Y, Z)$ admet la densité f , alors :

(a) Les v.a. réelles Y et Z admettent les densités suivantes :

$$f_Y(y) = \int_{-\infty}^{\infty} f(y, z) dz; \quad f_Z(z) = \int_{-\infty}^{\infty} f(y, z) dy. \quad (12.3)$$

(b) Les v.a. Y et Z sont indépendantes si et seulement si

$$f(y, z) = f_Y(y)f_Z(z) \quad (m_2\text{-p.p.}).$$

(c) Pour chaque y tel que $f_Y(y) \neq 0$, la formule ci-dessous définit une densité sur \mathbb{R} :

$$f_{Y=y}(z) = \frac{f(y, z)}{f_Y(y)}.$$

Dans (b), « $m_2\text{-p.p.}$ » (où « p.p. » abrège « presque partout ») signifie « en dehors d'un ensemble de m_2 -mesure nulle ». Avant de prouver ce théorème, nous allons faire quelques commentaires concernant (a) et (c). D'abord, les densités f_Y et f_Z sont appelées les *densités marginales* de f . Ces deux densités marginales ne permettent pas, à elles seules, de retrouver la densité « jointe » f (il faut pour cela disposer d'une information additionnelle, par exemple que X et Y sont indépendantes dans le cas de (b)).

La fonction $f_{Y=y}(z)$ s'appelle la *densité conditionnelle de Z sachant que $Y = y$* . Cette assertion n'est pas à prendre au sens littéral, puisque $P(Y = y) = 0$ pour chaque y et donc $P(A | Y = y)$ n'a *a priori* pas de sens. Néanmoins cette terminologie admet la justification heuristique suivante : soit Δy et Δz des quantités positives « très petites ». Alors

$$f(y, z)\Delta y\Delta z \approx P(y \leq Y \leq y + \Delta y; z \leq Z \leq z + \Delta z)$$

et

$$f_Y(y)\Delta y \approx P(y \leq Y \leq y + \Delta y);$$

si de plus $P(y \leq Y \leq y + \Delta y) > 0$ (ce qui est vrai si par exemple $f_Y(y) > 0$ et si f_Y est continue en y), on obtient alors :

$$\begin{aligned} f_{Y=y}(z)\Delta z &\approx \frac{P(y \leq Y \leq y + \Delta y; z \leq Z \leq z + \Delta z)}{P(y \leq Y \leq y + \Delta y)} \\ &\approx P(z \leq Z \leq z + \Delta z | Y \approx y). \end{aligned}$$

Preuve du théorème 12.2. (a) Pour tout borélien $A \in \mathcal{B}$ on a

$$\begin{aligned} P(Y \in A) &= P(X \in A \times \mathbb{R}) = \iint_{A \times \mathbb{R}} f(y, z) dy dz \\ &= \int_A dy \int_{-\infty}^{\infty} f(y, z) dz \\ &= \int_A dy f_Y(y). \end{aligned}$$

Comme les densités sur \mathbb{R} sont caractérisées par (11.4), f_Y définies dans (12.1) est une densité de Y . La preuve pour f_Z est la même.

(b) Supposons que $f(y, z) = f_Y(y)f_Z(z)$. On a

$$\begin{aligned} P(Y \in A, Z \in B) &= \iint 1_{A \times B}(y, z) f(y, z) dy dz \\ &= \iint 1_A(y) 1_B(z) f_Y(y) f_Z(z) dy dz \\ &= \int 1_A(y) f_Y(y) dy \int 1_B(z) f_Z(z) dz \\ &= P(Y \in A) P(Z \in B), \end{aligned}$$

et comme A et B sont des boréliens arbitraires on a que Y et Z sont indépendantes.

Supposons inversement Y et Z indépendantes. Soit

$$\mathcal{H} = \left\{ C \in \mathcal{B}^2 : \iint_C f(y, z) dy dz = \iint_C f_Y(y) f_Z(z) dy dz \right\}.$$

À cause de l'indépendance, si $C = A \times B$ avec $A \in \mathcal{B}$ et $B \in \mathcal{B}$, on a

$$\begin{aligned} \iint_C f(y, z) dy dz &= P((Y, Z) \in C) \\ &= P(Y \in A, Z \in B) \\ &= P(Y \in A) P(Z \in B) \\ &= \int_A f_Y(y) dy \int_B f_Z(z) dz \\ &= \iint_C f_Y(y) f_Z(z) dy dz \end{aligned}$$

par le théorème de Fubini 10.3. Donc \mathcal{H} contient la classe de tous les produits $C = A \times B$ avec $A, B \in \mathcal{B}$; cette classe est stable par intersection finie et engendre la tribu \mathcal{B}^2 . Comme de plus \mathcal{H} est stable par

limites croissantes et par différences, le théorème des classes monotones (théorème 6.2) entraîne $\mathcal{H} = \mathcal{B}^2$. Donc

$$P(X \in C) = \int_C f(y, z) dy dz = \int_C f_Y(y) f_Z(z) dy dz$$

pour tout $C \in \mathcal{B}^2$. L'unicité de la densité (théorème 12.1) donne

$$f(y, z) = f_Y(y) f_Z(z), \quad m_2\text{-p.p.}$$

(c) On a

$$\begin{aligned} \int f_{Y=y}(z) dz &= \int_{-\infty}^{\infty} \frac{f(y, z)}{f_Y(y)} dz \\ &= \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(y, z) dz = \frac{1}{f_Y(y)} f_Y(y) = 1. \end{aligned}$$

Comme $f_{Y=y}$ est positive, borélienne, et d'intégrale 1, c'est donc une densité. ■

Définition 12.3. Soit X et Y deux v.a. réelles, ayant toutes deux une variance finie. La covariance de X et Y est définie comme étant le nombre

$$\text{Cov}(X, Y) = E\left((X - E(X))(Y - E(Y))\right) = E(XY) - E(X)E(Y).$$

Noter que $E(XY)$ existe : comme X et Y ont des variances finies elles sont toutes deux dans L^2 , et le théorème 9.3(a) implique que $XY \in L^1$. Remarquer aussi que

$$\text{Cov}(X, X) = \text{Var}(X) = \sigma^2(X).$$

Théorème 12.3. Si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$.

Preuve. X et Y indépendantes implique $E(XY) = E(X)E(Y)$, d'où le résultat. ■

Attention : la réciproque du théorème 12.3 est fautive; on peut avoir $\text{Cov}(X, Y) = 0$ sans que X et Y soient indépendantes.

Définition 12.4. Soit X et Y deux v.a. de variances finies non nulles. Le coefficient de corrélation de X et Y est le nombre

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

$$(\sigma(X) = \sqrt{\sigma^2(X)} \text{ et } \sigma(Y) = \sqrt{\sigma^2(Y)}.)$$

Noter que d'après l'inégalité de Cauchy-Schwarz (théorème 9.3(a)) on a toujours $-1 \leq \rho \leq 1$, et si X et Y sont indépendantes alors $\rho = 0$ (théorème 12.3).

Définition 12.5. Soit $X = (X_1, \dots, X_n)$ une v.a. à valeurs dans \mathbb{R}^n . La matrice de covariance de X est la matrice $n \times n$ dont le terme général est

$$c_{ij} = \text{Cov}(X_i, X_j).$$

Théorème 12.4. Une matrice de covariance est symétrique et semi-définie positive : i.e., $c_{ij} = c_{ji}$ et $\sum a_i a_j c_{ij} \geq 0$ pour tous réels a_i .

Preuve. La symétrie est évidente, puisque $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$. Un calcul simple montre que

$$\sum a_i a_j c_{ij} = \sigma^2 \left(\sum_{i=1}^n a_i X_i \right)^2,$$

et comme une variance est toujours positive on a le résultat. ■

Théorème 12.5. Soit X une v.a. à valeurs dans \mathbb{R}^n , de matrice de covariance C . Soit A une matrice $m \times n$, et posons $Y = AX$. Alors Y est une v.a. à valeurs dans \mathbb{R}^m et sa matrice de covariance est $C' = ACA^t$, où A^t est la transposée de A .

Preuve. Il s'agit d'un simple calcul. ■

Passons maintenant à l'étude des fonctions de v.a. vectorielles, i.e. à valeurs dans \mathbb{R}^n . On s'intéresse au problème suivant : soit $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction borélienne. Étant donnée une v.a. $X = (X_1, \dots, X_n)$ à valeurs dans \mathbb{R}^n , admettant la densité f , quelle est la densité de $Y = g(X)$? et pour commencer, existe-t-elle?

Nous aurons besoin d'un résultat classique et important, mais de démonstration assez difficile, sur les changements de variables dans les intégrales multiples (voir par exemple [6] ou [25, p. 83]). Rappelons d'abord que si g est une fonction différentiable d'un ouvert G de \mathbb{P}^n dans \mathbb{E}^m , sa matrice jacobienne $J_g(x)$ au point $x \in G$ est $J_g(x) = \frac{\partial y}{\partial x}(x)$ (i.e., $J_g(x)_{i,j} = \frac{\partial y_i}{\partial x_j}(x)$, où $y = (y_1, y_2, \dots, y_m)$). Le jacobien de g au point x est le déterminant de la matrice jacobienne $J_g(x)$. Si le jacobien en x est non nul, alors la fonction g est inversible sur un voisinage de x , et le jacobien de l'inverse (ou fonction réciproque) g^{-1} au point $y = g(x)$ est l'inverse du jacobien de g en x .

Théorème 12.6. (Formule de changement de variable.) Soit G un ouvert de \mathbb{R}^n et $g : G \rightarrow \mathbb{R}^n$ une fonction continûment différentiable. Supposons de plus que g soit injective sur G et que son jacobien ne s'annule jamais sur G . Soit f une fonction borélienne telle que la fonction produit $f 1_{g(G)}$ soit positive ou intégrable par rapport à la mesure de Lebesgue m_n , avec

$$g(G) = \{y \in \mathbb{R}^n : \text{il existe } x \in G \text{ avec } g(x) = y\}.$$

On a alors

$$\int_{g(G)} f(y) dy = \int_G f(g(x)) |\det(J_g(x))| dx.$$

Ce qui suit est une simple application du théorème 12.6 aux densités de v.a.

Théorème 12.7. Soit $X = (X_1, \dots, X_n)$ une v.a. à valeurs dans \mathbb{R}^n , admettant une densité f_X . Soit $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction continûment différentiable et injective, dont le jacobien ne s'annule pas. Alors la v.a. $Y = g(X)$ admet la densité

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) |\det J_{g^{-1}}(y)| & \text{si } y \text{ est dans l'image de } g \\ 0 & \text{sinon.} \end{cases}$$

Preuve. On note $G = g(\mathbb{R}^n)$ l'image de g . Les propriétés de g impliquent que G est un ouvert et que la fonction réciproque g^{-1} est bien définie sur G et continûment différentiable avec un jacobien ne s'annulant pas. Soit $B \in \mathcal{B}^n$, et $A = g^{-1}(B)$. On a

$$\begin{aligned} P(X \in A) &= \int_A f_X(x) dx \\ &= \int_{g^{-1}(B)} f_X(x) dx \\ &= \int_B f_X(g^{-1}(x)) |\det J_{g^{-1}}(x)| dx, \end{aligned}$$

par le théorème 12.6 appliqué à g^{-1} . On a aussi $P(Y \in B) = P(X \in A)$, donc

$$P(X \in A) = \int_B f_Y(y) dy.$$

Comme $B \in \mathcal{B}^n$ est arbitraire on en déduit que

$$f_X(g^{-1}(x)) |\det J_{g^{-1}}(x)| = f_Y(x),$$

m_n -p.p., d'où le résultat. ■

De manière analogue au corollaire 11.3 du chapitre 11, on peut aussi traiter le cas où g n'est pas injective mais seulement « injective par morceaux ».

Corollaire 12.1. Soit S_0, S_1, \dots, S_m des boréliens deux à deux disjoints de \mathbb{R}^n , avec $m_n(S_0) = 0$ et S_i ouvert pour $i = 1, \dots, m$. Soit X une v.u. à valeurs dans \mathbb{R}^n , admettant une densité f_X nulle en dehors de $S = \bigcup_{i=0}^m S_i$ (donc X est en fait à valeurs dans S). Soit g une fonction de S dans \mathbb{R}^n telle que pour chaque $i = 1, \dots, m$ la restriction g_i de g à S_i soit continûment différentiable, injective, de jacobien ne s'annulant pas. La v.a. $Y = g(X)$ admet alors la densité

$$f_Y(y) = \sum_{i=1}^m f_X(g_i^{-1}(y)) |\det J_{g_i^{-1}}(y)| 1_{y \in (S_i)}(y).$$

Exemples.

1. Soit X et Y des v.a. indépendantes $N(0, 1)$. Calculons la densité du couple $(U, V) = (X + Y, X - Y)$. On a

$$g(x, y) = (x + y, x - y) = (u, v),$$

$$g^{-1}(u, v) = \left(\frac{u+v}{2}, \frac{u-v}{2} \right).$$

La matrice jacobienne ne dépend pas ici de (u, v) et vaut

$$J_{g^{-1}}(u, v) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

et

$$\det J_{g^{-1}} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = -\frac{1}{2}.$$

Donc

$$\begin{aligned} f_{(U,V)}(u, v) &= f_{(X,Y)} \left(\frac{u+v}{2}, \frac{u-v}{2} \right) |\det J| \\ &= f_X \left(\frac{u+v}{2} \right) f_Y \left(\frac{u-v}{2} \right) |J| \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{u+v}{2} \right)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{u-v}{2} \right)^2} \cdot \frac{1}{2} \\ &= \frac{1}{\sqrt{4\pi}} e^{-\frac{v^2}{4}} \frac{1}{\sqrt{4\pi}} e^{-\frac{u^2}{4}} \end{aligned}$$

pour $u, v \in \mathbb{R}$. On en déduit que les v.a. U et V sont indépendantes et $N(0, 2)$.

2. Soit (X, Y) une v.a. bidimensionnelle de densité f . On veut trouver la densité de $U = XY$. Dans ce cas $h(x, y) = xy$ applique \mathbb{R}^2 dans \mathbb{R}^1 , et on ne peut donc pas utiliser le théorème 12.7. On peut cependant s'y ravouer par une astuce simple. Posons

$$g(x, y) = (xy, x).$$

Soit $S_0 = \{(x, y) : x = 0, y \in \mathbb{R}\}$ et $S_1 = \mathbb{R}^2 \setminus S_0$. Alors $m_2(S_0) = 0$ et g est injective de S_1 dans \mathbb{R}^2 et $g^{-1}(u, v) = (v, \frac{u}{v})$. La matrice jacobienne est

$$J_{g^{-1}}(u, v) = \begin{pmatrix} 0 & \frac{1}{v} \\ 1 & -\frac{u}{v^2} \end{pmatrix},$$

et $\det(J_{g^{-1}}) = -\frac{1}{v}$. Donc si $V = X$ et si on considère la v.a. bidimensionnelle (U, V) , le corollaire 12.1 donne

$$f_{(U, V)}(u, v) = \begin{cases} f\left(v, \frac{u}{v}\right) \frac{1}{|v|} & \text{si } v \neq 0 \\ 0 & \text{si } v = 0. \end{cases}$$

Rappelons qu'on cherche la densité f_U de U , qui est donnée par

$$f_U(u) = \int_{-\infty}^{\infty} f_{(U, V)}(u, v) dv = \int_{-\infty}^{\infty} f\left(v, \frac{u}{v}\right) \frac{1}{|v|} dv.$$

3. Parfois on peut calculer une densité directement, sans passer par le théorème 12.6 ou l'astuce de l'exemple 2 ci-dessus.

Soit par exemple X et Y indépendantes de loi $N(0, 1)$, et $Z = X^2 + Y^2$. Quelle est la densité f_Z de Z ?

Le théorème 12.2(b) entraîne que la densité du couple (X, Y) est

$$f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right),$$

et donc

$$\begin{aligned} \mathbb{P}(Z \in A) &= \mathbb{E}(1_A(Z)) = \mathbb{E}(1_A(X^2 + Y^2)) \\ &= \iint 1_A(x^2 + y^2) f(x, y) dx dy \\ &= \iint 1_A(x^2 + y^2) \frac{1}{2\pi} e^{-\frac{x^2 + y^2}{2}} dx dy \end{aligned}$$

En passant en coordonnées polaires il vient

$$\begin{aligned}
&= \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} 1_A(r^2) e^{-r^2/2} r \, dr \, d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^\infty 1_A(r^2) e^{-r^2/2} r \, dr \\
&= \int_0^\infty 1_A(r^2) e^{-r^2/2} r \, dr
\end{aligned}$$

En posant $z = r^2$, donc $dz = 2r \, dr$:

$$= \int_0^\infty 1_A(z) \frac{1}{2} e^{-(z/2)} dz,$$

et on voit que Z admet la densité $\frac{1}{2}e^{-(z/2)}$ de l'exponentielle de paramètre $\frac{1}{2}$. Noter que la transformation en coordonnées polaires n'est pas injective sur \mathbb{R}^2 , de sorte que pour justifier l'argument précédent il faut utiliser de nouveau le corollaire 12.1 : cette transformation est injective sur $S_1 = \mathbb{R}^2 \setminus \{0\}$, tandis que l'ensemble $S_0 = \{0\}$ est de mesure de Lebesgue nulle. Cet argument sera utilisé sans mention explicite dans la suite.

4. Quand une fonction transforme n v.a. en une seule on peut parfois éviter de combiner le théorème 12.6 et l'astuce de l'exemple 2 en utilisant à la place la fonction de répartition. Plus précisément si $Y = g(X_1, \dots, X_n)$, avec g borélienne de \mathbb{R}^n dans \mathbb{R} , et si f est la densité de (X_1, \dots, X_n) , alors

$$F_Y(y) = P(Y \leq y) = \int_{g(x_1, \dots, x_n) \leq y} f(x) dx.$$

Supposons qu'il existe une fonction $h(y; x_2, \dots, x_n)$ telle que

$$g(x_1, x_2, \dots, x_n) \leq y \quad \text{si et seulement si} \quad x_1 \leq h(y; x_2, \dots, x_n).$$

Alors

$$\begin{aligned}
F_Y(y) &= \int_{\{x: g(x) \leq y\}} f(x) dx \\
&= \int_{-\infty}^\infty \dots \int_{-\infty}^\infty \left\{ \int_{-\infty}^{h(y; x_2, \dots, x_n)} f(x_1, \dots, x_n) dx_1 \right\} dx_2, \dots, dx_n.
\end{aligned}$$

Si nous dérivons les deux membres par rapport à y (ce qui suppose en particulier que h soit continûment dérivable en y et f soit continue), on obtient

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\partial h(y; x_2, \dots, x_n)}{\partial y} \times \quad (12.4)$$

$$\times f(h(y, x_2, \dots, x_n); x_2, \dots, x_n) dx_2 \dots dx_n.$$

Un exemple d'utilisation de cette technique concerne le problème suivant, important en statistique. Soit X et Y des v.a. indépendantes, avec X de loi $N(0, 1)$ et Y de loi gamma de paramètres $\alpha = n/2$ avec n entier non nul et $\beta = \frac{1}{2}$. [Y suit donc la loi du chi-deux à n degrés de liberté.] Quelle est la loi de

$$Z = \frac{X}{\sqrt{Y/n}} ?$$

On a $g(x, y) = \frac{x}{\sqrt{y/n}}$, et comme $\frac{x}{\sqrt{y/n}} \leq z$ si et seulement si $x \leq z\sqrt{y/n}$, on peut prendre $h(z; y) = z\sqrt{y/n}$. À cause de l'indépendance la densité f de (X, Y) est

$$f(x, y) = f_X(x)f_Y(y) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right) \left(\frac{y^{\frac{n}{2}-1} e^{-\frac{1}{2}y}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \right) \mathbb{1}_{]0, \infty[}(y).$$

On peut alors appliquer (12.4) :

$$f_Z(z) = \frac{1}{\sqrt{2\pi n} \Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} \int_0^{\infty} y^{\frac{n}{2}-1} e^{-y \left(1 + \frac{z^2}{n}\right)} dy \quad (12.5)$$

$$= \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2}) \sqrt{\pi n} \left(1 + \frac{z^2}{n}\right)^{\frac{1}{2}(n+1)}}.$$

La loi de densité f_Z dans (12.5) s'appelle la *loi t de Student à n degrés de liberté*.

La loi t de Student, introduite pour l'étude statistique de la moyenne des lois normales lorsque les variances sont inconnues, a été introduite par W. Gosset (1876-1937) qui a utilisé pour cela le pseudonyme « Student ».

5. Soit X et Y des v.a. indépendantes de loi $N(0, \sigma^2)$. Soit $Z = \sqrt{X^2 + Y^2}$ et $W = \frac{X}{Y}$ si $Y \neq 0$ et $W = 0$ si $Y = 0$. On veut trouver la densité $f_{(Z, W)}$ de (Z, W) . Ici,

$$g(x, y) = \left(\sqrt{x^2 + y^2}, \frac{x}{y} \right) = (z, w)$$

et g n'est pas injective. On a

$$g^{-1}(z, w) = \left(\frac{zw}{\sqrt{1+w^2}}, \frac{z}{\sqrt{1+w^2}} \right) = (x, y).$$

Une autre fonction « inverse » serait

$$h^{-1}(z, w) = \left(\frac{-zw}{\sqrt{1+w^2}}, \frac{-z}{\sqrt{1+w^2}} \right).$$

La matrice jacobienne $J_{g^{-1}}$ de g^{-1} est

$$\begin{pmatrix} \frac{w}{\sqrt{1+w^2}} & \frac{1}{\sqrt{1+w^2}} \\ \frac{z}{(1+w^2)^{\frac{3}{2}}} & \frac{-zw}{(1+w^2)^{\frac{3}{2}}} \end{pmatrix}$$

de déterminant $\frac{-z}{1+w^2}$. Le corollaire 12.1 entraîne alors

$$f_{(Z,W)}(z, w) = \frac{z}{1+w^2} \left\{ f_{(X,Y)} \left(\frac{zw}{\sqrt{1+w^2}}, \frac{z}{\sqrt{1+w^2}} \right) + f_{(X,Y)} \left(\frac{-zw}{\sqrt{1+w^2}}, \frac{-z}{\sqrt{1+w^2}} \right) \right\}.$$

La densité normale étant symétrique, on a

$$f_{(X,Y)}(x, y) = f_{(X,Y)}(-x, -y),$$

de sorte que

$$f_{(Z,W)}(z, w) = \frac{2z}{1+w^2} \frac{1}{2\pi\sigma^2} e^{-\frac{z^2}{2\sigma^2}} 1_{(z>0)}.$$

Noter que cette densité se factorise :

$$f_{(Z,W)}(z, w) = \frac{1}{\pi\sigma^2} \left(\frac{1}{1+w^2} \right) \left(ze^{-\frac{z^2}{2\sigma^2}} 1_{(z>0)} \right).$$

On en déduit que Z et W sont indépendantes, ce qui n'est pas *a priori* évident, et on peut lire immédiatement les densités de Z et W : en fait, comme $\frac{1}{\pi(1+w^2)}$ est la densité d'une loi de Cauchy de paramètres $\alpha = 0$ et $\beta = 1$, il vient :

$$f_Z(z) = \frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}} 1_{(z>0)}$$

et

$$f_W(w) = \frac{1}{\pi(1+w^2)},$$

La loi de densité f_Z ci-dessus est connue sous le nom de *loi de Rayleigh* de paramètre $\sigma^2 > 0$. Cet exemple montre aussi que le rapport de deux v.a. indépendantes normales $N(0, \sigma^2)$ suit une loi de Cauchy de paramètres $\alpha = 0$ et $\beta = 1$.

Exercices

1. Montrer que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(x^2+y^2)}{2\sigma^2}} dx dy = 2\pi\sigma^2.$$

et donc que $\frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$ est une densité.

2. Soit (X, Y) une v.a. à valeurs dans \mathbb{R}^2 admettant une densité $f_{(X,Y)}$ qui se factorise : $f_{(X,Y)}(x, y) = g(x)h(y)$. Trouver les densités marginales f_X et f_Y .
3. Soit (X, Y) une v.a. à valeurs dans \mathbb{R}^2 admettant la densité

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times \exp\left(-\frac{1}{2(1-r^2)} \left\{ \frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2r(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right\}\right),$$

où $r \in]-1, 1[$, $\sigma_1, \sigma_2 > 0$ et $\mu_1, \mu_2 \in \mathbb{R}$. Trouver $f_{X=X}(y)$.

[Rép. : $\frac{1}{\sigma_2\sqrt{2\pi(1-r^2)}} \exp(-\frac{1}{2\sigma_2^2(1-r^2)} \{y - \mu_2 - \frac{r\sigma_2}{\sigma_1}(x - \mu_1)\}^2)$.]

4. Soit $\rho_{X,Y}$ le coefficient de corrélation des v.a. réelles X et Y . Soit $a > 0$, $c > 0$ et $b \in \mathbb{R}$. Montrer que

$$\rho_{aX+b, cY+b} = \rho_{X,Y}.$$

(Cela montre le résultat utile suivant : le coefficient de corrélation ne dépend pas des « échelles » avec lesquelles on mesure X et Y .)

5. Pour $a \neq 0$, montrer que

$$\rho_{X, aX+b} = \frac{a}{|a|};$$

donc si $Y = aX + b$ est une fonction affine non constante de X , alors $\rho_{X,Y} = \pm 1$.

6. Soit X, Y des v.a. ayant des variances finies non nulles, et soit

$$Z = \left(\frac{1}{\sigma_Y} \right) Y - \left(\frac{\rho_{X,Y}}{\sigma_X} \right) X.$$

Montrer que $\sigma_Z^2 = 1 - \rho_{X,Y}^2$, et déduire que si $\rho_{X,Y} = \pm 1$, alors Y est une fonction affine non constante de X .

7* (Gut (1995), p. 27.) Soit (X, Y) une v.a. à valeurs dans \mathbb{R}^2 , de loi uniforme sur la boule unité :

$$f_{(X,Y)}(x, y) = \begin{cases} \frac{1}{\pi} & \text{si } x^2 + y^2 \leq 1 \\ 0 & \text{si } x^2 + y^2 > 1. \end{cases}$$

Trouver la loi de $R = \sqrt{X^2 + Y^2}$. (*Indication* : Introduire la v.a. auxiliaire $S = \text{Arctan} \frac{Y}{X}$.) [Rép. : $f_R(r) = 2r1_{[0,1]}(r)$.]

8. Soit (X, Y) une v.a. à valeurs dans \mathbb{R}^2 , de densité f . Trouver la densité de $Z = X + Y$. (*Indication* : Trouver d'abord la densité du couple (Z, W) , où $W = Y$.) [Rép. : $f_Z(z) = \int_{-\infty}^{\infty} f_{(X,Y)}(z-w, w) dw$.]

9. Soit X une v.a. normale $N(0, \sigma^2)$, et soit Θ une v.a. uniforme sur $]0, \pi[$: i.e. Θ a la densité $f(\theta) = \frac{1}{\pi} 1_{]0, \pi[}(\theta)$. Supposons X et Θ indépendantes. Trouver la loi de $Z = X + a \cos \Theta$. (Ceci est utilisé en électricité.) [Rép. : $f_Z(z) = \frac{1}{\pi \sigma \sqrt{2\pi}} \int_0^\pi e^{-(z-a \cos w)^2 / 2\sigma^2} dw$.]

10. Soit X et Y des v.a. réelles indépendantes de densités respectives f_X et f_Y , et soit $Z = g(X)$ et $W = h(Y)$, où g et h sont injectives et dérivables. Trouver la densité $f_{(Z,W)}$ du couple (Z, W) .

11. Soit X et Y deux v.a. indépendantes de loi $N(0, \sigma^2)$. Soit

$$Z = \sqrt{X^2 + Y^2} \quad \text{et} \quad W = \text{Arctan} \frac{X}{Y}, \quad -\frac{\pi}{2} < W \leq \frac{\pi}{2}.$$

Montrer que Z admet une loi de Rayleigh, que W est uniforme sur $] -\frac{\pi}{2}, \frac{\pi}{2}[$, et que Z et W sont indépendantes.

12. Soit (X_1, \dots, X_n) des v.a. réelles. Posons

$$Y_1 = \min(X_i; 1 \leq i \leq n)$$

$$Y_2 = \text{seconde plus petite valeur parmi } X_1, \dots, X_n$$

\vdots

$$Y_n = \max(X_i; 1 \leq i \leq n).$$

Montrer que Y_1, \dots, Y_n sont des v.a. et que $Y_1 \leq Y_2 \leq \dots \leq Y_n$ (on dit que les Y_i forment le *réarrangement croissant*, ou la *statistique d'ordre* des X_i et on les note en général

$$Y_k = X_{(k)}.)$$

Supposons que les X_i sont indépendantes et toutes avec la même densité f . Montrer que la densité de la v.a. n -dimensionnelle (Y_1, \dots, Y_n) est donnée par

$$g(y_1, \dots, y_n) = \begin{cases} n! \prod_{i=1}^n f(y_i) & \text{si } y_1 < y_2 < \dots < y_n \\ 0 & \text{sinon.} \end{cases}$$

13. Soit (X_1, \dots, X_n) des v.a. indépendantes uniformes sur $]0, a[$. Montrer que la statistique d'ordre $(X_{(1)}, \dots, X_{(n)})$ (cf. l'exercice précédent) admet la densité

$$g(y_1, \dots, y_n) = \begin{cases} \frac{n!}{a^n} & \text{si } 0 < y_1 < y_2 < \dots < y_n < a \\ 0 & \text{sinon.} \end{cases}$$

14. Dans la situation de l'exercice 12, en notant f la densité commune des X_i (supposées indépendantes) et F leur fonction de répartition, montrer que $X_{(k)}$ admet la densité

$$f_{(k)}(y) = k C_n^k f(y) (1 - F(y))^{n-k} F(y)^{k-1}.$$

15. (Simulation d'une variable aléatoire normale.) Soit U_1 et U_2 deux v.a. indépendantes uniformes sur $]0, 1[$. Soit $\theta = 2\pi U_1$ et $S = -\log U_2$.

- a) Montrer que S suit une loi exponentielle, et que $\sqrt{2S}$ suit une loi de Rayleigh.
 b) Posons $X = R \cos \theta$, $Y = R \sin \theta$. Montrer que X et Y sont indépendantes normales.

[*Indication* : Pour a), la loi exponentielle est un cas particulier des lois gamma, c'est la loi χ_2^2 . Pour b), utiliser la méthode de l'exercice 11 en sens inverse!]

[*Remarque*. La méthode décrite dans cet exercice pour simuler les v.a. normales centrées est connue sous le nom de méthode de Box-Muller.]

Chapitre 13

Fonctions caractéristiques

Il arrive souvent en mathématiques qu'on puisse résoudre un problème ou obtenir certaines propriétés d'objets mathématiques en les « transformant » dans un autre espace, en résolvant le problème dans cet autre espace, et en revenant ensuite dans l'espace initial. Deux parmi les transformations importantes sont la transformée de Laplace et la transformée de Fourier. Ces transformées sont largement utilisées pour l'étude des équations différentielles ou la « théorie du signal » : elles sont aussi extrêmement utiles en probabilités : elles servent à analyser les variables aléatoires (par exemple, à calculer leurs moments), ou à donner des preuves « élémentaires » et élégantes de certains théorèmes, comme le théorème-limite central (voir le chapitre 21). La transformée de Fourier est la plus sophistiquée des deux, et aussi la plus utile.

Avant de l'introduire il nous faut dire un mot de l'intégrale des fonctions complexes. Soit (E, \mathcal{E}, μ) un espace probabilisé, et f une application de E dans l'ensemble des nombres complexes. On peut l'écrire $f = g + ih$, où $i = \sqrt{-1}$ est la racine carrée de -1 et où g et h sont des fonctions réelles (la *partie réelle* et la *partie imaginaire* de f , respectivement). La fonction module est $|f| = \sqrt{g^2 + h^2}$ (racine carrée positive). On dit alors que f est *mesurable* si les deux fonctions g et h sont mesurables, et *intégrable* si g et h sont toutes deux intégrables. L'intégrale de f est alors le nombre complexe

$$\int f(x)\mu(dx) = \int g(x)\mu(dx) + i \int h(x)\mu(dx). \quad (13.1)$$

Il est alors évident que *tous* les résultats du chapitre 9 sont valides pour les fonctions à valeurs complexes, en prenant séparément la partie réelle et la partie imaginaire. Noter en particulier que la fonction complexe mesurable f est intégrable si et seulement si la fonction réelle $|f|$ l'est : en effet on a $|g| \leq |f|$, $|h| \leq |f|$, et $|f| \leq |g| + |h|$. Il est également important de noter que l'intégrale est linéaire sur le corps des complexes : si a est un nombre complexe, alors $\int (af(x))\mu(dx) = a \int f(x)\mu(dx)$. Le seul résultat un peu délicat est la majoration suivante pour les modules :

$$\left| \int f(x) \mu(dx) \right| \leq \int |f(x)| \mu(dx). \quad (13.2)$$

Pour l'obtenir, on écrit le complexe $\int f(x) \mu(dx)$ sous la forme $\rho e^{i\theta}$; à cause de la linéarité de l'intégrale il vient

$$\left| \int f(x) \mu(dx) \right| = \rho = \int (f(x) e^{-i\theta}) \mu(dx),$$

et l'expression de droite, étant réelle, vaut aussi $\int k(x) \mu(dx)$, où k est la partie réelle de $f e^{-i\theta}$. Mais le théorème 9.1(b) implique $|\int k(x) \mu(dx)| \leq \int |k(x)| \mu(dx)$ et comme on a $|k| \leq |f|$, on obtient (13.2).

Une notation nous sera aussi nécessaire : nous écrivons $\langle x, y \rangle$ pour le produit scalaire de $x, y \in \mathbb{R}^n$. C'est-à-dire, si $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$, alors

$$\langle x, y \rangle = \sum_{j=1}^n x_j y_j.$$

(Le produit scalaire est aussi souvent écrit $x \cdot y$).

Définition 13.1. Soit μ une mesure de probabilité sur \mathbb{P}^n . Sa transformée de Fourier, notée $\hat{\mu}$, est la fonction sur \mathbb{P}^n définie par

$$\hat{\mu}(u) = \int e^{i\langle u, x \rangle} \mu(dx).$$

Ci-dessus on intègre la fonction complexe mesurable $x \mapsto e^{i\langle u, x \rangle}$, de module égal à 1, de sorte que d'après ce qui précède elle est intégrable. D'après (13.1), la transformée de Fourier s'écrit aussi

$$\hat{\mu}(u) = \int \cos(\langle u, x \rangle) \mu(dx) + i \int \sin(\langle u, x \rangle) \mu(dx).$$

Si $n = 1$, on a $\langle u, x \rangle = ux$, donc $\hat{\mu}(u) = \int e^{iux} \mu(dx)$.

Définition 13.2. Soit X une v.a. à valeurs dans \mathbb{R}^n . Sa fonction caractéristique φ_X est la fonction sur \mathbb{R}^n définie par

$$\varphi_X(u) = \mathbb{E}(e^{i\langle u, X \rangle}).$$

En d'autres termes,

$$\varphi_X(u) = \int e^{i\langle u, x \rangle} P_X(dx) = \widehat{P}_X(u) \quad (13.3)$$

où P_X est la loi de X . Ainsi, la fonction caractéristique de X est la transformée de Fourier de sa loi.

Théorème 13.1. Soit μ une probabilité sur \mathbb{R}^n . La fonction $\hat{\mu}$ est bornée (de module inférieur ou égal à 1), continue, et vérifie $\hat{\mu}(0) = 1$.

Preuve. On a déjà vu que $\hat{\mu}$ est bien définie. Comme $|e^{i\langle u, x \rangle}| = 1$ pour tous $u, x \in \mathbb{R}^n$, on a

$$|\hat{\mu}(u)| \leq \int |e^{i\langle u, x \rangle}| \mu(dx) = \int 1 \mu(dx) = 1.$$

De plus

$$\hat{\mu}(0) = \int e^{i\langle 0, x \rangle} \mu(dx) = \int 1 \mu(dx) = 1.$$

Pour la continuité de $\hat{\mu}$ il suffit de montrer que pour toute suite (u_p) convergeant vers une limite u (dans \mathbb{R}^n), on a $\hat{\mu}(u_p) \rightarrow \hat{\mu}(u)$. Pour cela, on remarque que $e^{i\langle u_p, x \rangle} \rightarrow e^{i\langle u, x \rangle}$ pour tout $x \in \mathbb{R}^n$, et comme les fonctions $x \mapsto e^{i\langle u, x \rangle}$ sont bornées en module par la fonction constante égale à 1, le résultat découle du théorème de convergence dominée de Lebesgue (théorème 9.1f) [si on ne veut pas utiliser ce théorème pour les fonctions complexes, on peut l'utiliser séparément pour les parties réelle $x \mapsto \cos(\langle u, x \rangle)$ et imaginaire $x \mapsto \sin(\langle u, x \rangle)$, qui sont bornées en valeur absolue par 1]. ■

On peut même montrer que $\hat{\mu}$ est *uniformément continue*, mais nous n'avons pas besoin de ce résultat ici.

Si $X = (X_1, \dots, X_n)$ est une v.a. à valeurs dans \mathbb{P}^n , et si $E(|X|^m) < \infty$ pour un entier m ($|X|$ désigne le module du vecteur aléatoire X , c'est une v.a. réelle), alors $E(|X|^k) < \infty$ pour $k = 1, 2, \dots, m$, et on dit que X admet des moments jusqu'à l'ordre m . Dans ce cas, pour tout choix d'indices j_1, \dots, j_m entre 1 et n , la v.a. produit (réelle) $X_{j_1} \dots X_{j_m}$ est intégrable (car majorée par $|X|^m$), et l'espérance de cette v.a. s'appelle le (j_1, \dots, j_m) -moment de X . Lorsque $n = 1$, on a évidemment $j_1 = \dots = j_m = 1$, le produit ci-dessus est X^m , et son espérance $E(X^m)$ est le m -ième moment de X .

Théorème 13.2. Soit X une v.a. à valeurs dans \mathbb{R}^n , admettant des moments jusqu'à l'ordre m . La fonction caractéristique φ_X de X est alors m fois continûment différentiable, et

$$\frac{\partial^m}{\partial u_{j_1} \dots \partial u_{j_m}} \varphi_X(u) = i^m E(X_{j_1} \dots X_{j_m} e^{i\langle u, X \rangle}).$$

Preuve. On va en fait prouver la formulation équivalente pour les mesures de probabilité μ sur \mathbb{P}^n (si $\mu = P_X$, rappelons que $\varphi_X = \hat{\mu}$),

L'hypothèse sur μ étant que la fonction $f(x) = |x|^m$ est intégrable par rapport à μ .

On veut montrer que $\widehat{\mu}$ est m fois continûment différentiable et que

$$\frac{\partial^m \widehat{\mu}}{\partial u_{j_1} \dots \partial u_{j_m}}(u) = i^m \int x_{j_1} \dots x_{j_m} e^{i(u, x)} \mu(dx).$$

On donne la preuve pour $m = 1$ seulement, le cas général se montrant en répétant le même raisonnement m fois. Nous voulons donc montrer que pour tout j , la dérivée partielle $\frac{\partial \widehat{\mu}}{\partial u_j}$ existe au point u et vaut

$$\frac{\partial \widehat{\mu}}{\partial u_j}(u) = i \int e^{i(u, x)} x_j \mu(dx), \quad (13.4)$$

et que de plus la fonction ci-dessus est continue en u .

Pour l'existence de $\frac{\partial \widehat{\mu}}{\partial u_j}$ et la formule (13.4), il suffit de montrer que pour toute suite de réels t_p convergant vers 0, si $r = (v_1, \dots, v_n)$ désigne le vecteur unitaire de \mathbb{R}^n dans la direction j (i.e. de coordonnées $v_k = 0$ pour $k \neq j$ et $v_j = 1$), alors la suite

$$\frac{1}{t_p} \{ \widehat{\mu}(u + t_p r) - \widehat{\mu}(u) \} = \int e^{i(u, r)} \frac{e^{it_p r_j} - 1}{t_p} \mu(dr), \quad (13.5)$$

converge vers le membre de droite de (13.4). La suite de fonctions $x \rightarrow \frac{e^{it_p r_j} - 1}{t_p}$ (où x_j est la j ème coordonnée de $x \in \mathbb{R}^n$) converge simplement vers r_j (appliquer le fait que la fonction exponentielle est dérivable). De plus

$$\left| \frac{e^{it_p r_j} - 1}{t_p} \right| \leq 2|x|,$$

et

$$\int 2|x| \mu(dx) < \infty$$

par hypothèse. Donc le théorème de convergence dominée de Lebesgue implique que (13.5) converge vers

$$i \int x_j e^{i(u, x)} \mu(dx).$$

Donc on a (13.4).

La preuve de la continuité en u de la fonction $\partial \widehat{\mu} / \partial u_j$ dans (13.4) est exactement la même que la preuve de la continuité dans le théorème 13.1. ■

Corollaire 13.1. Soit X une v.a. réelle admettant des moments jusqu'à l'ordre m . La fonction caractéristique φ_X de X est alors m fois continûment différentiable, et

$$\frac{\partial^m}{\partial u^m} \varphi_X(u) = i^m E(X^m e^{iuX}).$$

Une application immédiate de ce résultat est l'utilisation de la fonction caractéristique pour calculer les moments d'une v.a. réelle X . Pour les deux premiers moments, de loin les plus importants, on obtient immédiatement :

$$E(X) = i\varphi'_X(0) \quad \text{si} \quad E(|X|) < \infty \quad (13.6)$$

$$E(X^2) = -\varphi''_X(0) \quad \text{si} \quad E(X^2) < \infty. \quad (13.7)$$

Exemples.

1. *Variable de Bernoulli* (p). Si X est une v.a. de Bernoulli de paramètre p , i.e. $P(X = 1) = p$ et $P(X = 0) = 1 - p$, il vient

$$\varphi_X(u) = E(e^{iuX}) = e^{iu0}(1-p) + e^{iu}p = \boxed{pe^{iu} + 1 - p}.$$

2. *Variable binomiale* (n, p). Si X est une v.a. binomiale de paramètres n et p , alors

$$\varphi_X(u) = E(e^{iuX}) = \sum_{j=0}^n C_n^j e^{iuj} p^j (1-p)^{n-j} = \boxed{(pe^{iu} + 1 - p)^n}.$$

On aurait aussi pu remarquer que X s'écrit

$$X = \sum_{j=1}^n Y_j,$$

où les Y_1, \dots, Y_n sont indépendantes de loi de Bernoulli (p). Alors

$$\varphi_X(u) = E(e^{iuX}) = E(e^{iu \sum_{j=1}^n Y_j}) = E\left(\prod_{j=1}^n e^{iuY_j}\right) = \prod_{j=1}^n E(e^{iuY_j})$$

par l'indépendance des Y_j :

$$= \prod_{j=1}^n \varphi_{Y_j}(u) = (pe^{iu} + 1 - p)^n.$$

3. Variable de Poisson (λ).

$$\begin{aligned}\varphi_X(u) &= \mathbb{E}(e^{iuX}) = \sum_{k=0}^{\infty} e^{iuk} \mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty} e^{iuk} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} \frac{(\lambda e^{iu})^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} e^{\lambda e^{iu}} = \boxed{e^{\lambda(e^{iu} - 1)}}.\end{aligned}$$

4. Variable uniforme sur $[-a, a]$.

$$\varphi_X(u) = \mathbb{E}(e^{iuX}) = \frac{1}{2a} \int_{-a}^a e^{iux} dx = \frac{e^{iua} - e^{-iua}}{2ai u};$$

on utilisant les propriétés $e^z = \cos z + i \sin z$ et $\cos(a) = \cos(-a)$, on voit que ceci égale

$$= \frac{2i \sin au}{2ai u} = \boxed{\frac{\sin au}{au}}$$

si $u \neq 0$, et bien sûr $\varphi_X(0) = 1$.

5. Variable normale $N(0, 1)$. Le calcul de la fonction caractéristique d'une v.a. normale est un peu plus difficile. On peut le faire par un calcul d'intégrale le long d'un contour fermé du plan complexe et le théorème des résidus, on par prolongement analytique (voir l'exercice 17 du chapitre 14). Voici une autre méthode, peu intuitive mais élémentaire. On a

$$\begin{aligned}\varphi_X(u) &= \int e^{iux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \int \frac{\cos ux}{\sqrt{2\pi}} e^{-x^2/2} dx + i \int \frac{\sin ux}{\sqrt{2\pi}} e^{-x^2/2} dx.\end{aligned}$$

Comme $x \mapsto \sin ux e^{-x^2/2}$ est une fonction impaire et intégrable, son intégrale est nulle. Par suite

$$\varphi_X(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos ux e^{-x^2/2} dx.$$

Grâce au théorème 13.2 on peut dériver les deux membres par rapport à u , et on obtient

$$\varphi'_X(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -x \sin ux e^{-x^2/2} dx.$$

Ensuite on intègre par parties :

$$= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \cos ux e^{-x^2/2} dx = -u\varphi_X(u).$$

Par suite φ_X satisfait à l'équation différentielle $\varphi_X'(u) = -u\varphi_X(u)$, dont la solution générale est

$$\varphi_X(u) = Ce^{-u^2/2}$$

pour une constante C . Comme $\varphi_X(0) = 1$ on a $C = 1$, de sorte que

$$\boxed{\varphi_X(u) = e^{-u^2/2}}.$$

Théorème 13.3. Soit X une v.a. à valeurs dans \mathbb{R}^n et soit $u \in \mathbb{R}^m$. Soit A une matrice $m \times n$. Alors

$$\varphi_{u+AX}(u) = e^{i(u,a)} \varphi_X(A^*u),$$

pour tout $u \in \mathbb{R}^m$, où A^* désigne la transposée de A .

Preuve. On a

$$e^{i(u,a+AX)} = e^{i(u,a)} e^{i(A^*u,X)},$$

et on obtient le résultat en prenant l'espérance des deux membres. ■

Exemples (suite).

6. *Variable normale* $N(\mu, \sigma^2)$. Si X est $N(\mu, \sigma^2)$, on vérifie facilement (voir exercice 18 du chapitre 14) que $Y = \frac{X-\mu}{\sigma}$ est $N(0, 1)$, et $X = \mu + \sigma Y$. D'après le théorème 13.3 et l'exemple 5, il vient

$$\boxed{\varphi_X = e^{iu\mu - u^2\sigma^2/2}}.$$

7. *Variable exponentielle* (λ). Soit X une v.a. exponentielle de paramètre λ . Alors

$$\varphi_X(u) = \int_0^{\infty} e^{iux} \lambda e^{-\lambda x} dx.$$

Un calcul formel donne

$$= \int_0^{\infty} \lambda e^{(iu-\lambda)x} dx = \boxed{\frac{\lambda}{\lambda - iu}}, \quad (13.8)$$

ce qui n'est pas vraiment rigoureux sur le plan mathématique. Une manière simple de justifier cela consiste à intégrer séparément les fonctions partie réelle et partie imaginaire, i.e. $x \mapsto \lambda e^{-\lambda x} \cos ux$ et $x \mapsto \lambda e^{-\lambda x} \sin ux$, en faisant une intégration par parties : un calcul élémentaire montre que la formule (13.8) est exacte.

8. *Variable gamma* (α, β) . Soit X une v.a. de loi gamma de paramètres α et β . En utilisant une intégration dans le plan complexe et le théorème des résidus, on peut montrer que

$$\varphi_X(u) = \frac{\beta^\alpha}{(\beta - iu)^\alpha}.$$

On peut aussi calculer cette fonction caractéristique par une méthode élémentaire : voir l'exercice 19 du chapitre 14.

Chapitre 14

Propriétés des fonctions caractéristiques

Nous avons vu comment calculer les fonctions caractéristiques sur divers exemples (de manière équivalente : comment calculer des transformées de Fourier de probabilités). Pour que cela soit réellement utile nous avons besoin de savoir que cette transformée de Fourier caractérise la probabilité. C'est ce que nous prouvons dans le théorème ci-dessus. Sa démonstration utilise le théorème de Stone-Weierstrass et est donc un peu difficile pour ce livre ; vu l'importance du résultat, nous la donnons néanmoins.

Théorème 14.1. (Théorème d'unicité.) *La transformée de Fourier $\hat{\mu}$ d'une probabilité μ sur \mathbb{R}^n caractérise μ : si deux probabilités admettent la même transformée de Fourier, elles sont égales.*

Preuve. Soit

$$f_{\sigma}(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-|x|^2/2\sigma^2},$$

et

$$\hat{f}_{\sigma}(u) = e^{-|u|^2\sigma^2/2}.$$

Alors $f_{\sigma}(x)$ est la densité de $X = (X_1, \dots, X_n)$, où les X_j sont $N(0, \sigma^2)$ et indépendantes. L'exemple 6 du chapitre 13 et le théorème de Fubini entraînent :

$$\begin{aligned} \int_{\mathbb{R}^n} f_{\sigma}(x) e^{i\langle u, x \rangle} dx &= \int_{\mathbb{R}^n} \prod_{j=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{(\frac{x_j^2}{2\sigma^2} + iu_j x_j)} dx_1 \dots dx_n \\ &= \prod_{j=1}^n \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} e^{(\frac{-x_j^2}{2\sigma^2} + iu_j x_j)} dx_j \\ &= \prod_{j=1}^n e^{-\frac{u_j^2 \sigma^2}{2}} = \hat{f}_{\sigma}(u). \end{aligned}$$

Donc

$$\begin{aligned} f_{\sigma}(u-v) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \hat{f}_{\sigma} \left(\frac{u-v}{\sigma^2} \right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \int_{\mathbb{R}^n} f_{\sigma}(x) e^{i \langle \frac{u-v}{\sigma^2}, x \rangle} dx. \end{aligned}$$

Supposons maintenant que μ_1 et μ_2 soient deux probabilités sur \mathbb{R}^n ayant même transformée de Fourier $\hat{\mu}_1 = \hat{\mu}_2 = \hat{\mu}$. Alors

$$\begin{aligned} \int f_{\sigma}(u-v) \mu_1(du) &= \int \frac{1}{(2\pi\sigma^2)^{n/2}} \left\{ \int f_{\sigma}(x) e^{i \langle \frac{u-v}{\sigma^2}, x \rangle} dx \right\} \mu_1(du) \\ &= \int f_{\sigma}(x) \frac{1}{(2\pi\sigma^2)^{n/2}} \hat{\mu} \left(\frac{x}{\sigma^2} \right) e^{-i \langle \frac{v, x}{\sigma^2} \rangle} dx, \end{aligned}$$

(le lecteur vérifiera qu'on peut appliquer le théorème de Fubini ci-dessus), et la même égalité est vraie pour μ_2 . Par suite

$$\int g(x) \mu_1(dx) = \int g(x) \mu_2(dx)$$

pour toute fonction $g \in \mathcal{H}$, où \mathcal{H} est l'espace vectoriel engendré par toutes les fonctions de la forme $u \mapsto f(\sigma, u-v)$. On peut alors appliquer le théorème de Stone-Weierstrass¹ pour obtenir que \mathcal{H} est dense dans \mathcal{C}_0 pour la convergence uniforme, où \mathcal{C}_0 est l'ensemble des fonctions continues « nulles à l'infini », i.e. l'ensemble des fonctions continues sur \mathbb{R}^n telles que $\lim_{\|x\| \rightarrow \infty} |f(x)| = 0$. On obtient alors

$$\int_{\mathbb{R}^n} g(x) \mu_1(dx) = \int_{\mathbb{R}^n} g(x) \mu_2(dx)$$

pour tout $g \in \mathcal{C}_0$. Comme la fonction indicatrice d'un ouvert est limite croissante d'une suite de fonctions positives dans \mathcal{C}_0 , le théorème de convergence monotone (théorème 9.1(d)) donne

$$\mu_1(A) = \mu_2(A), \quad \text{pour tout ouvert } A \subset \mathbb{R}^n.$$

Finalement le théorème des classes monotones (théorème 6.2) donne

$$\mu_1(A) = \mu_2(A) \quad \text{pour tout borélien } A \subset \mathbb{R}^n,$$

donc $\mu_1 = \mu_2$. ■

1. Voir par exemple [23, p. 160].

Corollaire 14.1. Soit $X = (X_1, \dots, X_n)$ une v.a. à valeurs dans \mathbb{R}^n . Les variables réelles $(X_j)_{1 \leq j \leq n}$ sont indépendantes si et seulement si pour tout $u = (u_1, \dots, u_n) \in \mathbb{R}^n$ on a

$$\varphi_X(u_1, \dots, u_n) = \prod_{j=1}^n \varphi_{X_j}(u_j) \quad (14.1)$$

Preuve. Si les $(X_j)_{1 \leq j \leq n}$ sont indépendantes, alors

$$\begin{aligned} \varphi_X(u) &= (e^{i\langle u, X \rangle}) = \mathbf{E} \left(e^{i \sum_{j=1}^n u_j X_j} \right) \\ &= \mathbf{E} \left(\prod_{j=1}^n e^{i u_j X_j} \right) = \prod_{j=1}^n \mathbf{E}(e^{i u_j X_j}) \quad (\text{par l'indépendance}) \\ &= \prod_{j=1}^n \varphi_{X_j}(u_j). \end{aligned}$$

Supposons à l'inverse qu'on ait (14.1). Soit μ_X la loi de X sur \mathbb{R}^n et soit μ_{X_j} les lois des X_j sur \mathbb{R} . Alors

$$\widehat{\mu}_X = (\mu_{X_1} \widehat{\otimes} \mu_{X_2} \widehat{\otimes} \dots \widehat{\otimes} \mu_{X_n}),$$

et le théorème 14.1 entraîne que

$$\mu_X = \mu_{X_1} \otimes \mu_{X_2} \otimes \dots \otimes \mu_{X_n},$$

qui équivaut à l'indépendance des X_j . ■

Attention : le fait que $\varphi_X(u_1, u_2, \dots, u_n) = \prod_{j=1}^n \varphi_{X_j}(u_j)$ pour tout $u \in \mathbb{R}^n$ n'est pas suffisant pour que les v.a. X_j soient indépendantes.

Exercices des chapitres 13 et 14

Les trois premiers exercices nécessitent une intégration dans le plan complexe et le théorème des résidus.

1. Soit $f(x) = \frac{1}{\pi(1+x^2)}$ la densité d'une v.a. X de Cauchy. Montrer que

$$\varphi_X(u) = e^{-|u|},$$

en intégrant le long d'un demi-cercle de diamètre $[-R, R]$ sur l'axe réel vers la gauche, puis du point $(-R, 0)$ au point $(R, 0)$ le long de l'axe réel; puis passer à la limite ($R \rightarrow \infty$).

- 2* Soit X une v.a. de loi gamma de paramètres α et β . Montrer que $\varphi_X(u) = \frac{\beta^\alpha}{(\beta - iu)^\alpha}$. [Indication : Intégrer le long du segment horizontal de $(d, 0)$ à $(c, 0)$ (avec $0 < c < d$), puis le long du segment vertical descendant de $(c, 0)$ jusqu'à l'intersection avec la droite d'équation $y = -\frac{ux}{\alpha}$, puis suivre cette droite vers le bas jusqu'au point d'abscisse d , puis remonter verticalement jusqu'au point $(d, 0)$; puis passer à la limite avec $d \rightarrow \infty$ et $c \rightarrow 0$.]
- 3* Soit X une v.a. $N(0, 1)$. Montrer que $\varphi_X(u) = e^{-u^2/2}$ en intégrant le long du segment horizontal de $(R, 0)$ à $(-R, 0)$, puis le long du segment vertical de $(-R, 0)$ à $(-R, -iu)$, puis horizontalement jusqu'à $(R, -iu)$, puis horizontalement jusqu'à $(R, 0)$; puis passer à la limite ($R \rightarrow \infty$).
- 4* Soit X une v.a. réelle telle que $E(X^2) < \infty$ et $E(X) = 0$. Montrer que $\sigma^2 := \text{Var}(X) < \infty$, et que

$$\varphi_X(u) = 1 - \frac{1}{2}u^2\sigma^2 + o(u^2)$$

quand $u \rightarrow 0$. [Rappelons qu'une fonction g est $o(t^n)$ quand $t \rightarrow 0$ si $\lim_{t \rightarrow 0} \frac{|g(t)|}{t^n} = 0$.]

5. Soit $X = (X_1, \dots, X_n)$ une v.a. à valeurs dans \mathbb{R}^n . Montrer que
- $\varphi_X(u, 0, 0, \dots, 0) = \varphi_{X_1}(u) \quad (u \in \mathbb{R})$
 - $\varphi_X(u, u, u, \dots, u) = \varphi_{X_1 + \dots + X_n}(u) \quad (u \in \mathbb{R})$
6. On note \bar{z} le complexe conjugué de z , i.e. si $z = a + ib$ on a $\bar{z} = a - ib$ ($a, b \in \mathbb{R}$). Montrer que si X est une v.a., on a $\overline{\varphi_X(u)} = \varphi_X(-u)$.
7. Soit X une v.a. Montrer que $\varphi_X(u)$ est à valeurs réelles si et seulement si X a une loi symétrique, i.e. $P_X = P_{-X}$. [Indication : Utiliser l'exercice 6 et les théorèmes 13.3 et 14.1.]
8. Montrer que si X et Y sont des v.a. indépendantes et de même loi, alors $Z = X - Y$ a une loi symétrique. (Indication : On peut utiliser l'exercice 7.)
9. Soit X_1, \dots, X_n des v.a. réelles indépendantes, ayant chacune un moment d'ordre 3 fini (i.e. $E(|X_i|^3) < \infty$) et de moyenne nulle (i.e. $E(X_i) = 0$). Montrer que

$$E\left(\left(\sum_{i=1}^n X_i\right)^3\right) = \sum_{i=1}^n E(X_i^3).$$

(Indication : Utiliser les fonctions caractéristiques.)

10. Soit μ_1, \dots, μ_n des probabilités sur \mathbb{R}^n . Soit $\lambda_j \geq 0$ avec $\sum_{j=1}^n \lambda_j = 1$. Posons $\nu = \sum_{j=1}^n \lambda_j \mu_j$. Montrer que ν est aussi une probabilité, et que

$$\widehat{\nu}(u) = \sum_{j=1}^n \lambda_j \widehat{\mu}_j(u).$$

11. Soit X une v.a. de loi exponentielle double (Laplace) de paramètres $\alpha = 0, \beta = 1$, i.e. admettant la densité

$$f_X(x) = \frac{1}{2} e^{-|x|}.$$

Montrer que $\varphi_X(u) = \frac{1}{1+u^2}$. (*Indication* : Utiliser l'exercice 10 avec pour μ_1 la loi d'une v.a. exponentielle Y , pour μ_2 la loi de $-Y$, et $\lambda_1 = \lambda_2 = \frac{1}{2}$.)

- 12* (*Loi triangulaire.*) Soit X une v.a. réelle de densité $f_X(x) = (1 - |x|)1_{[-1,1]}(x)$. Montrer que $\varphi_X(u) = \frac{2(1 - \cos u)}{u^3}$. (*Indication* : Soit U et V des v.a. indépendantes uniformes sur $[-\frac{1}{2}, \frac{1}{2}]$; considérer $U + V$. Observer aussi que $\left(\frac{e^{iu/2} - e^{-iu/2}}{iu}\right)^2 = \left(\frac{2 \sin(u/2)}{u}\right)^2$.)
13. Soit X une v.a. positive. Sa transformée de Mellin est la fonction

$$T_X(\theta) = E(X^\theta)$$

pour toutes les valeurs de θ pour lesquelles l'espérance de X^θ existe.

- a) Montrer que

$$T_X(\theta) = \varphi_{\log X} \left(\frac{\theta}{i} \right)$$

quand les deux membres sont bien définis.

- b) Montrer que si X et Y sont indépendantes et positives, on a

$$T_{XY}(\theta) = T_X(\theta) T_Y(\theta).$$

- c) Montrer que $T_{bX^a}(\theta) = b^\theta T_X(a\theta)$ pour $b > 0$ et $a\theta$ dans le domaine de définition de T_X .

14. Soit X une v.a. lognormale de paramètres (μ, σ^2) . Trouver la transformée de Mellin (cf. exercice 13) $T_X(\theta)$. Utiliser cette transformée et le fait que $T_X(k) = E(X^k)$ pour calculer le k -ième moment de X pour $k = 1, 2, \dots$

15. Soit X une v.a. $N(0, 1)$. Montrer que $E(X^{2n+1}) = 0$ et

$$E(X^{2n}) = \frac{(2n)!}{2^n n!} = (2n-1)(2n-3)\dots 3 \cdot 1.$$

16. Soit X une v.a. $N(0, 1)$. Posons

$$M(s) = E(e^{sX}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(sx - \frac{1}{2}x^2\right) dx.$$

Montrer que $M(s) = e^{s^2/2}$. (*Indication* : Compléter le carré dans l'exponentielle.)

17* Remplacer s par iu dans l'exercice 16 pour obtenir que $\varphi_X(u) = e^{-u^2/2}$; justifier ce calcul par la théorie du prolongement analytique des fonctions de variables complexes.

18. Soit X une v.a. $N(\mu, \sigma^2)$. Montrer que $Y = \frac{X-\mu}{\sigma}$ suit la loi $N(0, 1)$.

19* (Feller [12].) Soit X une v.a. de loi gamma de paramètres $\alpha > 0$ et $\beta = 1$. On peut calculer sa fonction caractéristique sans intégration dans le plan complexe : développer e^{ix} en série, et montrer que

$$\frac{1}{\Gamma(\alpha)} \sum_{n=1}^{\infty} \frac{(iu)^n}{n!} \int_0^{\infty} e^{-x} x^{n+\alpha-1} dx = \sum_{n=0}^{\infty} \frac{\Gamma(n+\alpha)}{n! \Gamma(\alpha)} (iu)^n,$$

puis que la somme de cette dernière série est $\frac{1}{(1-iu)^\alpha}$.

Chapitre 15

Sommes de variables aléatoires indépendantes

Une part importante des applications des probabilités découle des propriétés des sommes de variables aléatoires indépendantes. Un exemple simple apparaît en statistique : si nous répétons n fois la même expérience indépendamment et si nous appelons X_j le résultat de la j -ième expérience, la « valeur moyenne » est donnée par $\bar{x} = \frac{1}{n} \sum_{j=1}^n X_j$. La v.a. \bar{x} est alors appelée un *estimateur* de la moyenne μ de la loi commune des X_j . La statistique étudie quand (et comment) \bar{x} converge vers μ quand $n \rightarrow \infty$. Même une fois qu'on sait que \bar{x} tend vers μ , on a aussi besoin de savoir quelle taille donner à n pour être raisonnablement sûr que \bar{x} est proche de la vraie valeur μ , qui est en général inconnue. D'autres problèmes, plus ardu, se posent aussi naturellement : quelle est la loi de \bar{x} ? Si on ne peut la déterminer, peut-on au moins l'approximer ? Quelle taille doit avoir n pour que cette approximation soit bonne ? Si on dispose d'informations complémentaires sur μ , comment les utiliser pour améliorer l'estimateur \bar{x} ? Bien évidemment, avant de commencer à résoudre les plus simples de ces questions, il nous faut étudier les sommes de v.a. indépendantes.

Théorème 15.1. Soit X et Y deux v.a. réelles. La loi μ_Z de $Z = X + Y$ est le produit de convolution des lois μ_X et μ_Y de X et Y , qui est défini par

$$(\mu_X * \mu_Y)(A) = \iint 1_A(x+y) \mu_X(dx) \mu_Y(dy). \quad (15.1)$$

Preuve. Comme X et Y sont indépendantes, la loi du couple (X, Y) est $\mu_X \otimes \mu_Y$. Donc

$$\mathbb{E}(g(X, Y)) = \iint g(x, y) \mu_X(dx) \mu_Y(dy),$$

et en particulier, avec $g(x, y) = f(x + y)$, on obtient pour toute fonction borélienne f sur \mathbb{R} , positive ou bornée :

$$\mathbb{E}(f(X + Y)) = \iint f(x + y) \mu_X(dx) \mu_Y(dy), \quad (15.2)$$

Il suffit alors de prendre $f(x) = 1_A(x)$. ■

Remarque 15.1. La formule (15.2) montre que pour $f : \mathbb{R} \rightarrow \mathbb{R}$ borélienne et $Z = X + Y$, avec X et Y indépendantes :

$$E(f(Z)) = \int f(z)(\mu_X * \mu_Y)(dz) = \iint f(x+y)\mu_X(dx)\mu_Y(dy).$$

Théorème 15.2. Soit X et Y deux v.a. réelles indépendantes, et $Z = X + Y$. La fonction caractéristique φ_Z est le produit de φ_X et de φ_Y , i.e.,

$$\varphi_Z(u) = \varphi_X(u)\varphi_Y(u).$$

Preuve. Prendre $f(z) = e^{iuz}$ et utiliser (15.2). ■

Attention : si $Z = X + Y$, la propriété que $\varphi_Z(u) = \varphi_X(u)\varphi_Y(u)$ pour tout $u \in \mathbb{R}$ n'implique pas l'indépendance de X et Y .

Théorème 15.3. Soit X et Y deux v.a. réelles indépendantes, et $Z = X + Y$.

(a) Si X admet la densité f_X , alors Z admet la densité f_Z donnée par

$$f_Z(z) = \int f_X(z-y)\mu_Y(dy)$$

(b) Si de plus Y admet la densité f_Y , alors

$$f_Z(z) = \int f_X(z-y)f_Y(y)dy = \int f_X(x)f_Y(z-x)dx.$$

Preuve. (b) : Supposons (a) vrai. On a

$$f_Z(z) = \int f_X(z-y)\mu_Y(dy).$$

Comme de plus $\mu_Y(dy) = f_Y(y)dy$, on obtient la première égalité. La seconde s'obtient en échangeant les rôles de X et Y .

(a) : D'après le théorème 15.1 on a

$$\begin{aligned} \mu_Z(A) &= \iint 1_A(x+y)\mu_X(dx)\mu_Y(dy) \\ &= \int \left\{ \int 1_A(x+y)f_X(x)dx \right\} \mu_Y(dy). \end{aligned}$$

En posant $z = x + y$, donc « $dz = dx$ » :

$$= \int \left\{ \int 1_A(z) f_X(z-y) dz \right\} \mu_Y(dy)$$

et en appliquant le théorème de Fubini :

$$= \int \left\{ \int f(z-y) \mu_Y(dy) \right\} 1_A(z) dz.$$

Comme A est un borélien arbitraire, on a le résultat. ■

Le résultat suivant est trivial, mais extrêmement utile.

Théorème 15.4. *Soit X et Y deux v.a. réelles indépendantes, de carré intégrable (i.e. $E(X^2) < \infty$ et $E(Y^2) < \infty$). On a alors*

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2.$$

Preuve. À cause de l'indépendance, on a $E(XY) = E(X)E(Y)$, donc

$$\sigma_{X+Y}^2 = E(X^2) + 2E(XY) + E(Y^2) - (E(X) + E(Y))^2 = \sigma_X^2 + \sigma_Y^2.$$

Exemples.

1. Soit X_1, \dots, X_n des v.a. de Bernoulli (p). Alors $Y = \sum_{j=1}^n X_j$ est binomiale (n, p). On a vu que

$$E(Y) = E\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n E(X_j) = \sum_{j=1}^n p = np.$$

Noter que

$$\sigma_{X_j}^2 = E(X_j^2) - E(X_j)^2 = p - p^2 = p(1-p).$$

Donc le théorème 15.4 entraîne que

$$\sigma_Y^2 = \sum_{j=1}^n \sigma_{X_j}^2 = np(1-p).$$

Cette méthode de calcul de la variance est préférable à un calcul direct à partir de la loi de Y , qui nécessite de trouver la somme

$$\sigma_Y^2 = \sum_{j=0}^n (j - np)^2 \binom{n}{j} p^j (1-p)^{n-j}.$$

2. Soit X une v.a. de Poisson (λ) et Y une v.a. de Poisson (μ), indépendante de X . Alors $Z = X + Y$ suit aussi une loi de Poisson, de paramètre $(\lambda + \mu)$. Pour le voir, on écrit $\varphi_Z = \varphi_X \varphi_Y$, ce qui donne

$$\varphi_Z(u) = e^{\lambda(e^{iu} - 1)} e^{\mu(e^{iu} - 1)} = e^{(\lambda + \mu)(e^{iu} - 1)},$$

et on reconnaît là la fonction caractéristique de la loi de Poisson de paramètre $(\lambda + \mu)$: donc Z suit la loi de Poisson grâce au théorème d'unicité des fonctions caractéristiques (théorème 14.1).

3. Soit X et Y des v.a. binomiales indépendantes de paramètres respectifs (n, p) et (m, p) (le même p pour les deux v.a.). Soit $Z = X + Y$. Alors

$$\varphi_Z = \varphi_X \varphi_Y,$$

donc

$$\varphi_Z(u) = (pe^{iu} + (1-p))^n (pe^{iu} + (1-p))^m = (pe^{iu} + (1-p))^{n+m},$$

qui est la fonction caractéristique d'une loi binomiale $(m+n, p)$; donc Z est binomiale $(m+n, p)$. Noter que pour ce résultat on n'a pas réellement besoin des fonctions caractéristiques : on peut simplement observer que

$$X = \sum_{j=1}^n U_j \quad \text{et} \quad Y = \sum_{j=1}^m V_j,$$

et donc

$$Z = \sum_{j=1}^n U_j + \sum_{j=1}^m V_j.$$

où les U_j et les V_j sont indépendantes et Bernoulli (p). Donc

$$Z = \sum_{j=1}^{m+n} W_j$$

où les W_j sont aussi indépendantes et Bernoulli (p). (Les n premières W_j sont les U_j , et les m suivantes sont les V_j .)

4. Soit X et Y indépendantes et respectivement de lois $N(\mu, \sigma^2)$ et $N(\nu, \tau^2)$. Alors $Z = X + Y$ suit la loi normale $N(\mu + \nu, \sigma^2 + \tau^2)$: on écrit

$$\varphi_Z = \varphi_X \varphi_Y,$$

ce qui donne

$$\varphi_Z(u) = e^{iu\mu} \cdot u^2 \sigma^2 / 2 \cdot e^{iu\nu - u^2 \tau^2 / 2} = e^{iu(\mu + \nu) - u^2(\sigma^2 + \tau^2) / 2}$$

qui est la fonction caractéristique de $N(\mu + \nu, \sigma^2 + \tau^2)$, et on utilise encore le théorème 14.1.

5. Soit X et Y des v.a. indépendantes de lois gamma de paramètres respectifs (α, β) et (δ, β) (avec le même β), et $Z = X + Y$. On a $\varphi_Z = \varphi_X \varphi_Y$, et donc

$$\varphi_Z(u) = \frac{\beta^\alpha}{(\beta - iu)^\alpha} \frac{\beta^\delta}{(\beta - iu)^\delta} = \frac{\beta^{\alpha+\delta}}{(\beta - iu)^{\alpha+\delta}},$$

donc Z suit la loi gamma de paramètres $(\alpha + \delta, \beta)$.

6. Dans le chapitre 11 nous avons défini la *loi du chi-deux à p degrés de liberté* (χ_p^2), qui est aussi la loi gamma de paramètres $\frac{p}{2}$ et $\frac{1}{2}$. Nous avons observé que si Z est $N(0, 1)$ alors Z^2 suit la loi χ_1^2 . Si maintenant Z_1, \dots, Z_p sont des v.a. indépendantes de loi $N(0, 1)$, les v.a. Z_j^2 sont aussi indépendantes, de sorte que l'exemple 5 implique que $X = \sum_{j=1}^p Z_j^2$ suit une loi gamma de paramètres $\frac{p}{2}$ et $\frac{1}{2}$, i.e. la loi χ_p^2 . En d'autres termes, une v.a. X de loi du chi-deux à p degrés de liberté peut être vue comme la somme des carrés de p v.a. indépendantes de loi $N(0, 1)$.

Exercices

1. Soit X_1, \dots, X_n des v.a. indépendantes telles que $E(X_j) = \mu$ et $\text{Var}(X_j) = \sigma^2 < \infty$, $1 \leq j \leq n$. Posons

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n X_j \quad \text{et} \quad S^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{x})^2.$$

(\bar{x} et S^2 sont aussi des v.a., comme sous le nom de « moyenne empirique » et « variance empirique » respectivement.) Montrer que

- $E(\bar{x}) = \mu$;
- $\text{Var}(\bar{x}) = \frac{\sigma^2}{n}$;
- $E(S^2) = \frac{n-1}{n} \sigma^2$.

2. Soit X_1, \dots, X_n des v.a. indépendantes de variances finies. Posons $S_n = \sum_{j=1}^n X_j$. Montrer que

$$\sigma_{\frac{1}{n}S_n}^2 = \frac{1}{n^2} \sum_{j=1}^n \sigma_{X_j}^2,$$

et en déduire que si $\sigma_{X_j}^2 = \sigma^2$ pour $1 \leq j \leq n$, alors $\sigma_{\frac{1}{n}S_n}^2 = \sigma^2/n$.

3. Montrer que si les v.a. X_1, \dots, X_n sont indépendantes de même loi, alors

$$\varphi_{S_n}(u) = (\varphi_X(u))^n.$$

où $S_n = \sum_{j=1}^n X_j$.

Les exercices 4 à 8 font intervenir la somme d'un nombre aléatoire de v.a. indépendantes. On se donne une suite infinie X_1, X_2, \dots de v.a. réelles indépendantes de même loi, et une autre v.a. N à valeurs entières, indépendante des X_j . On pose aussi

$$S_n = \sum_{i=1}^n X_i, \quad \text{et} \quad S_N = X_1 + X_2 + \dots + X_N,$$

avec la convention que $S_N = 0$ sur l'ensemble où $N = 0$.

4. Si A est un borélien de \mathbb{R} , montrer que

$$P(S_N \in A \mid N = n) = P(S_n \in A).$$

5. Si $E(N) < \infty$ et $E(|X_j|) < \infty$, montrer que

$$E(S_N) = \sum_{n=0}^{\infty} E(S_n)P(N = n).$$

(Indication : Montrez d'abord que

$$E(S_N) = E\left(\sum_{n=0}^{\infty} S_n I_{\{N \geq n\}}\right) = \sum_{n=0}^{\infty} E(S_n I_{\{N \geq n\}}),$$

en justifiant la seconde égalité.)

6. Si $E(N) < \infty$ et $E(|X_j|) < \infty$, montrer que $E(S_N) = E(N)E(X_j)$.
(Indication : Utiliser l'exercice 5.)

7. Si $E(N) < \infty$ et $E(|X_j|) < \infty$, montrer que

$$\varphi_{S_N}(u) = E((\varphi_{X_j}(u))^N).$$

(Indication : Montrer d'abord que $\varphi_{S_N}(u) = \sum_{n=1}^{\infty} E(e^{iuS_n} 1_{\{N=n\}})$.)

8. Démontrer le résultat de l'exercice 6 en utilisant l'exercice 7 et les fonctions caractéristiques.

9. Soit X et Y des v.a. réelles indépendantes. Supposons que X et $X + Y$ ont même loi. Montrer que $Y = 0$ p.s.

10. Soit f et g des fonctions boréliennes sur \mathbb{R} , telles que

$$\int_{-\infty}^{\infty} |f(x)| dx < \infty \quad \text{et} \quad \int_{-\infty}^{\infty} |g(x)| dx < \infty.$$

Montrer que

a) $(f * g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy$ existe pour presque tout x (relativement à la mesure de Lebesgue) ;

b) $(f * g)(x) = (g * f)(x)$ ($f * g$ est appelé le produit de convolution de f et g) ;

c) Si l'une des deux fonctions f ou g au moins est continue, alors $f * g$ est continue.

11. Soit X et Y des v.a. réelles indépendantes de même loi. Supposons de plus que $X + Y$ et $X - Y$ soient indépendantes. Montrer que $\varphi_X(2u) = \varphi_X(u)^2 \varphi_X(-u)$.

12.* Soit X et Y comme dans l'exercice 11, et supposons que $E(X) = 0$ et $E(X^2) = 1$. Montrer que X suit la loi $N(0, 1)$. (Indication : Montrer qu'il existe $a > 0$ tel que $\varphi(u) \neq 0$ pour tout u avec $|u| \leq a$. Soit $\psi(u) = \frac{\varphi(u)}{\varphi(u/2^n)}$ pour $|u| \leq a$. Montrer que $\psi(u) = \{\psi(u/2^n)\}^{2^n}$, puis que $\{\psi(u/2^{2^n})\}^{2^{2^n}} \rightarrow 1$ quand $n \rightarrow \infty$ (cf. exercice 4 du chapitre 14), puis que $\varphi(t) = \{\varphi(t/2^n)\}^{4^n}$, et faire $n \rightarrow \infty$.)

13. Soit X_1, X_2, \dots, X_n des v.a. indépendantes de loi $N(\mu, \sigma^2)$. Soit $\bar{x} = \frac{1}{n} \sum_{j=1}^n X_j$ et $Y_j = X_j - \bar{x}$. Trouver la fonction caractéristique de $(\bar{x}, Y_1, \dots, Y_n)$. En déduire que si $S^2 = \frac{1}{n} \sum_{j=1}^n Y_j^2$ les v.a. \bar{x} et S^2 sont indépendantes.

14. Montrer que $|1 - e^{ix}|^2 = 2(1 - \cos x) \leq x^2$ pour tout $x \in \mathbb{R}$. Montrer que pour toute v.a. réelle X on a $|1 - \varphi_X(u)| \leq E(|uX|)$.

15. Soit $A = [-\frac{1}{u}, \frac{1}{u}]$. Montrer que si μ_X est la loi de X , on a

$$\int_A x^2 \mu_X(dx) \leq \frac{24}{11u^2} \{1 - \operatorname{Re} \varphi_X(u)\}.$$

(Indication : $1 - \cos x \geq 0$ et $1 - \cos x \geq \frac{1}{2}x^2 - \frac{1}{24}x^4$ pour tout $x \in \mathbb{R}$.)

16. Si φ est une fonction caractéristique, montrer que $|\varphi|^2$ en est une aussi (Indication : prendre deux v.a. indépendantes X et Y de fonction caractéristique φ , et considérer $Z = X - Y$.)
17. Soit X_1, \dots, X_α de v.a. indépendantes de loi exponentielle de paramètre $\beta > 0$. Montrer que $Y = \sum_{i=1}^\alpha X_i$ suit une loi gamma de paramètres α et β .

Chapitre 16

Variabes aléatoires gaussiennes

Rappelons que la loi normale $N(\mu, \sigma^2)$, où $\mu \in \mathbb{R}$ et $\sigma^2 > 0$, est la probabilité sur \mathbb{R} de densité

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}. \quad (16.1)$$

Il est commode d'étendre la classe des lois normales de façon à inclure le cas où $\sigma^2 = 0$, de la manière suivante : on note $N(\mu, 0)$ la loi égale à la masse de Dirac au point μ , i.e. la loi d'une v.a. qui est p.s. égale à la constante μ . Cette loi n'admet pas de densité, aussi parle-t-on parfois de loi normale « dégénérée ». La loi $N(0, 1)$ s'appelle la *loi normale réduite*, ou *standard*.

Si X est une v.a. de loi $N(\mu, \sigma^2)$, sa fonction caractéristique φ_X est

$$\varphi_X(u) = e^{iu\mu - \frac{\sigma^2 u^2}{2}}. \quad (16.2)$$

Quand $\sigma^2 > 0$ cela a fait l'objet de l'exemple 13.6, et quand $\sigma^2 = 0$ c'est un résultat trivial. Rappelons aussi qu'on a alors

$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2. \quad (16.3)$$

À première vue il peut sembler étrange d'appeler des v.a. ayant une densité aussi compliquée des variables *normales*. La raison en remonte au début du XVIII^e siècle, quand les premières versions du théorème-limite central apparurent dans les livres de Jacob Bernoulli (1713) et A. de Moivre (1718). Ces versions primitives du théorème-limite central ont été généralisées par P. Laplace et surtout C. F. Gauss : ce sont les travaux de ce dernier qui ont valu à ces lois d'être aussi appelées *lois gaussiennes*, et à leur auteur (et à la fonction représentative de la densité normale, dite « courbe de Gauss ») d'apparaître sur l'ancien billet de 10 marks. La version de Gauss du théorème-limite central peut être paraphrasée en écrivant que les sommes de variables i.i.d. (pour « Indépendantes et Identiquement Distribuées », c'est-à-dire de même loi) sont approximativement gaussiennes. C'est un résultat très profond,

puisqu'il affirme qu'on n'a besoin de connaître que très peu de choses sur la loi des v.a. dont on fait la somme pour conclure que la loi de la somme est approximativement gaussienne. Finalement, soulignons que plus tard Paul Lévy a déterminé les hypothèses minimales pour que le théorème-limite central soit valide. Ce théorème joue un rôle absolument central en statistique, et c'est ce rôle central qui lui a valu son nom et qui explique pourquoi les variables gaussiennes ou normales sont si importantes et présentes partout en probabilité. Ce théorème sera traité dans le chapitre 21 ; ici, nous en établissons les bases, en étudiant les v.a. gaussiennes.

Pour une v.a. réelle X la définition $\mathcal{L}(X) = N(\mu, \sigma^2)$ est claire : c'est une v.a. de densité donnée par (16.1) si $\sigma^2 > 0$, c'est $X \equiv \mu$ si $\sigma^2 = 0$. Pour une v.a. à valeurs dans \mathbb{R}^n la définition est plus subtile ; la raison en est qu'on veut ainsi décrire la classe des variables qui peuvent être obtenues comme limites dans le théorème-limite central, et cette classe est plus compliquée dans \mathbb{R}^n quand $n \geq 2$.

Définition 16.1. Une variable aléatoire $X = (X_1, \dots, X_n)$ à valeurs dans \mathbb{R}^n (ou « vecteur aléatoire ») est dite gaussienne, ou normale multivariée, si toute combinaison linéaire des composantes, soit $\sum_{j=1}^n a_j X_j$ avec des $a_j \in \mathbb{R}$ quelconques, suit une loi normale uni-dimensionnelle (éventuellement dégénérée, par exemple si on prend $a_j = 0$ pour tout j).

Les fonctions caractéristiques sont très utiles dans ce contexte.

Théorème 16.1. X est une variable gaussienne n -dimensionnelle si et seulement si sa fonction caractéristique est de la forme

$$\varphi_X(u) = \exp\{i\langle u, \mu \rangle - \frac{1}{2}\langle u, Qu \rangle\} \quad (16.4)$$

où $\mu \in \mathbb{R}^n$ et où Q est une matrice $n \times n$ symétrique. Dans ce cas le vecteur μ est le vecteur moyenne de X (i.e. $\mu_j = E(X_j)$ pour tout j), et Q est sa matrice de covariance, et est donc semi-définie positive.

Preuve (Condition suffisante). Supposons qu'on ait (16.4). Soit

$$Y = \sum_{j=1}^n a_j X_j = \langle a, X \rangle$$

une combinaison linéaire des composantes de X . Pour tout $v \in \mathbb{R}$ on a

$$\varphi_Y(v) = \varphi_X(va) = \exp\left\{iv\langle a, \mu \rangle - \frac{v^2}{2}\langle a, Qa \rangle\right\}$$

Remarquer que $\langle a, Qa \rangle \geq 0$ puisque $|\varphi_Y(v)| \leq 1$ pour tout v . Grâce à (16.2) on voit alors que $\varphi_Y(v)$ est la fonction caractéristique de la loi $N(\langle a, \mu \rangle, \langle a, Qa \rangle)$, donc par le théorème 14.1 on obtient la normalité de Y : cela montre la condition suffisante.

De plus, un calcul des dérivées $\frac{\partial \varphi_X}{\partial u_j}(0)$ et $\frac{\partial^2 \varphi_X}{\partial u_j \partial u_k}(0)$ et une double utilisation du théorème 13.2 montre que $\mu_j = E(X_j)$ et que $E(X_j X_k) - \mu_j \mu_k + Q_{j,k}$, de sorte que μ est le vecteur moyenne de X et Q est sa matrice de covariance.

(Condition nécessaire.) Supposons le vecteur aléatoire X gaussien, et posons de nouveau

$$Y = \sum_{j=1}^n a_j X_j = \langle a, X \rangle$$

Soit μ le vecteur moyenne et Q la matrice de covariance de X . On a alors par linéarité

$$E(Y) = \langle a, \mu \rangle$$

tandis que le théorème 12.4 donne

$$\sigma^2(Y) = \langle a, Qa \rangle.$$

Comme Y est normale par hypothèse, une nouvelle application de (16.2) donne

$$\varphi_Y(v) = \exp \left\{ i v \langle a, \mu \rangle - \frac{v^2}{2} \langle a, Qa \rangle \right\}.$$

Donc

$$\varphi_Y(1) = \varphi_{\langle a, X \rangle}(1) = E(\exp(i \langle a, X \rangle)) = \varphi_X(a),$$

et on a (16.4). ■

Notation. Pour X comme dans le théorème précédent, on note $N(\mu, Q)$ sa loi. Elle dépend de deux « paramètres » : le *vecteur moyenne* μ et la *matrice de covariance* Q .

Exemple 1.

Si les X_j pour $j = 1, \dots, n$ sont des v.a. réelles *indépendantes* de lois $N(\mu_j, \sigma_j^2)$, le vecteur $X = (X_1, \dots, X_n)$ est gaussien : il suffit de remarquer que

$$\varphi_X(u_1, \dots, u_n) = \prod_{j=1}^n \varphi_{X_j}(u_j)$$

par le corollaire 14.1. donc

$$\begin{aligned} \varphi_X(u) &= \prod_{j=1}^n e^{iu_j \mu_j - \frac{1}{2} u_j^2 \sigma_j^2} \\ &= \exp \left(\sum_{j=1}^n i u_j \mu_j - \frac{1}{2} \sum_{j=1}^n u_j^2 \sigma_j^2 \right) \\ &= e^{i \langle u, \mu \rangle - \frac{1}{2} \langle u, Qu \rangle} \end{aligned}$$

où $\mu = (\mu_1, \dots, \mu_n)$ et Q est la matrice diagonale

$$\begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{pmatrix},$$

et on conclut par le théorème 16.1. Cela peut aussi se voir directement : pour tous réels a_j , les v.a. $a_j X_j$ sont normales et indépendantes, donc $\sum_{j=1}^n a_j X_j$ est normale (exemple 4 du chapitre 15), et le vecteur X est gaussien.

Une forme de réciproque de cet exemple est vraie :

Corollaire 16.1. *Soit $X = (X_1, \dots, X_n)$ une v.a. gaussienne n -dimensionnelle. Ses composantes X_j sont indépendantes si et seulement si la matrice de covariance Q est diagonale.*

Preuve. La nécessité provient du théorème 12.3. Supposons inversement Q diagonale de la forme

$$Q = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{pmatrix}.$$

Dans ce cas l'équation (16.4) implique que φ_X se factorise :

$$\varphi_X(u) = \prod_{j=1}^n \varphi_{X_j}(u_j),$$

où

$$\varphi_{X_j}(u_j) = \exp \left\{ i u_j \mu_j - \frac{1}{2} u_j^2 \sigma_j^2 \right\}.$$

Le corollaire 14.1 entraîne alors l'indépendance des X_j . ■

Le théorème suivant montre que *tous* les vecteurs aléatoires gaussiens peuvent s'obtenir comme transformées linéaires de vecteurs gaussiens dont les composantes sont indépendantes.

Théorème 16.2. *Soit X un vecteur aléatoire n -dimensionnel gaussien de vecteur moyenne μ . Il existe alors des v.a. réelles indépendantes Y_1, \dots, Y_n de lois*

$$\mathcal{L}(Y_j) = N(0, \lambda_j) \quad \text{avec } \lambda_j \geq 0,$$

et une matrice $n \times n$ orthogonale A , telles que $X = \mu + AY$.

Commentaire important. Certains des λ_j ci-dessus peuvent être nuls, auquel cas $Y_j = 0$. Donc le nombre de v.a. normales Y_j nécessaires dans le théorème 16.2 peut être strictement plus petit que n , même si chaque X_j est non constante. En fait le nombre minimal de v.a. Y_j est égal au rang de la matrice de covariance Q de X .

Preuve du théorème 16.2. Comme Q est symétrique semi-définie positive, il existe une matrice orthogonale A et une matrice diagonale Λ à termes positifs, telles que $Q = A\Lambda A^*$. (Rappelons qu'une *matrice orthogonale* est une matrice A dont les vecteurs lignes (resp. colonnes) sont orthogonaux et de norme 1 : dans ce cas la matrice transposée A^* de A est aussi son inverse.)

Posons

$$Y = A^*(X - \mu).$$

Comme X est gaussien par hypothèse, le vecteur Y est aussi gaussien puisque toute combinaison linéaire des composantes de Y est une combinaison linéaire des composantes de X plus une constante, donc est une v.a. réelle normale. De plus la matrice de covariance de Y est $A^*QA = \Lambda$, la matrice diagonale cherchée. Comme $X = \mu + AY$ (parce que $A^{*-1} = A$), on a prouvé le résultat. ■

Corollaire 16.2. *Un vecteur aléatoire gaussien n -dimensionnel admet une densité sur \mathbb{R}^n si et seulement si sa matrice de covariance Q est non dégénérée (i.e., il n'y a pas de vecteur non nul a de \mathbb{R}^n tel que $Qa = 0$, ou de manière équivalente $\det(Q) \neq 0$).*

Preuve. Par le théorème précédent il existe des v.a. indépendantes Y_1, \dots, Y_n de lois $N(0, \lambda_j)$, avec $Q = A\Lambda A^*$ pour une matrice orthogonale A . Si $\det(Q) \neq 0$, on doit avoir $\lambda_j > 0$ pour tout j ($1 \leq j \leq n$), parce

que $\det(Q) = \det(\Lambda) = \prod_{j=1}^n \lambda_j$. Comme $\lambda_j > 0$ et $\mathcal{L}(Y_j) = N(0, \lambda_j)$, nous savons que le vecteur Y admet la densité

$$f_Y(y) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\lambda_j}} e^{-y_j^2/2\lambda_j},$$

et comme $X = \mu + AY$, on déduit du théorème 12.7 que X admet la densité

$$f_X(x) = \frac{1}{2\pi^{n/2} \sqrt{\det Q}} e^{-\frac{1}{2}(x-\mu) \cdot Q^{-1}(x-\mu)}. \quad (16.5)$$

Supposons à l'inverse Q dégénérée : il existe $a \in \mathbb{R}^n$ avec $a \neq 0$ et $Qa = 0$. La v.a. $Z = \langle a, X \rangle$ a la variance $\langle a, Qa \rangle = 0$, donc elle est p.s. égale à sa moyenne $\langle a, \mu \rangle$. Donc $P(X \in H) = 1$, où H est l'hyperplan affine orthogonal à a et contenant le vecteur μ , i.e. $H = \{x \in \mathbb{R}^n : \langle x - \mu, a \rangle = 0\}$. Comme la dimension de H est $n - 1$, la mesure de Lebesgue n -dimensionnelle de H vérifie $m_n(H) = 0$ (cf. exercice 1). Si X avait une densité, on aurait la propriété

$$1 = P(X \in H) = \int_{\mathbb{H}} f(x) dx = \int_{\mathbb{H}} f(x_1, \dots, x_n) dx_1 \dots dx_n, \quad (16.6)$$

Cependant

$$\int_{\mathbb{H}} f(x_1, \dots, x_n) dx_1 \dots dx_n = m_n(H) = 0$$

donc (16.6) ne peut pas être vraie et X n'admet pas de densité. ■

Commentaire. Ce corollaire montre que quand $n \geq 2$ il existe des vecteurs aléatoires gaussiens non constants, mais sans densité (si $n = 1$ une v.a. normale est soit constante, soit avec densité). Il serait tentant de définir les v.a. gaussiennes multivariées comme étant les v.a. admettant une densité de la forme (16.5), mais une telle définition serait insuffisante pour couvrir toutes les limites qu'on peut obtenir dans le théorème-limite central du chapitre 21.

Une propriété élémentaire mais importante des v.a. gaussiennes est la suivante :

Théorème 16.3. *Soit X un vecteur gaussien n -dimensionnel, et Y un vecteur gaussien m -dimensionnel indépendant de X . La v.a. $Z = (X, Y)$ à valeurs dans \mathbb{R}^{n+m} est aussi gaussienne.*

Preuve. On a à cause de l'indépendance de X et Y , et par une extension immédiate du corollaire 14.1 :

$$\varphi_Z(u) = \varphi_X(w)\varphi_Y(v), \quad u = (w, v); \quad w \in \mathbb{R}^n, \quad v \in \mathbb{R}^m.$$

Le théorème 16.1 implique, avec des notations évidentes :

$$\begin{aligned} \varphi_Z(u) &= \exp \left\{ i \langle w, \mu^X \rangle - \frac{1}{2} \langle w, Q^X w \rangle \right\} \exp \left\{ i \langle v, \mu^Y \rangle - \frac{1}{2} \langle v, Q^Y v \rangle \right\} \\ &= \exp \left\{ i \langle (w, v), (\mu^X, \mu^Y) \rangle - \frac{1}{2} \langle u, Qu \rangle \right\}, \end{aligned}$$

où

$$Q = \begin{pmatrix} Q^X & 0 \\ 0 & Q^Y \end{pmatrix}.$$

Il suffit d'appliquer de nouveau le théorème 16.1 pour obtenir le résultat. ■

On dit que deux v.a. réelles X et Y de carré intégrable sont *non corrélées* si $\text{Cov}(X, Y) = 0$. Comme

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y),$$

cela revient à dire que $E(XY) = E(X)E(Y)$. Cette propriété est vraie si X et Y sont indépendantes (théorème 12.3). La réciproque est *fausse* en général (deux v.a. peuvent être non corrélées sans être indépendantes), mais elle est *vraie* si le couple (X, Y) est gaussien. Plus généralement, on a le résultat suivant, qui est une reformulation partielle du corollaire 16.1 (pour l'obtenir, si (X_1, \dots, X_n) est un vecteur gaussien chaque couple (X_j, X_k) est aussi un vecteur gaussien bidimensionnel, auquel on peut appliquer ce corollaire) :

Théorème 16.4. *Soit X un vecteur aléatoire gaussien n -dimensionnel, et soit $j \neq k$. Les composantes X_j et X_k sont indépendantes si et seulement si elles sont non corrélées.*

Un modèle fréquemment utilisé en économie, en biologie, en ingénierie, etc., est celui de la *régression linéaire*¹, que nous allons décrire maintenant. On se donne des v.a. réelles Y_i , $1 \leq i \leq n$, de la forme

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \tag{16.7}$$

1. Le terme « linéaire » décrit la dépendance linéaire en les paramètres α et β et non en les x_i .

où les α , β et x_i sont des constantes, et où les ε_i sont des v.a. réelles. Typiquement, on peut penser qu'on mesure les quantités $\alpha + \beta x_i$ en faisant une erreur de mesure aléatoire ε_i . À cause du théorème-limite central on suppose souvent que le vecteur aléatoire $(\varepsilon_i)_{1 \leq i \leq n}$ suit une loi normale multivariée. En suivant par exemple Berger et Casella [7] on appelle x_i la *variable prédictrice* (rappelons que x_i n'est pas aléatoire), et Y_i s'appelle la « réponse »².

Supposons que $E(\varepsilon_i) = 0$ pour tout i . En prenant les espérances dans (16.7) on obtient

$$E(Y_i) = \alpha + \beta x_i. \quad (16.8)$$

Typiquement on cherche à déterminer la nature de la relation linéaire entre les Y_i et les x_i , qui serait évidente s'il n'y avait pas les perturbations induites par les ε_i . C'est-à-dire que nous cherchons à déterminer α et β en observant les Y_i , sachant que les x_i sont connus. Pour commencer, nous excluons le cas où tous les x_i sont égaux parce qu'alors nous pouvons au mieux déterminer la constante $\alpha + \beta x_1$ sans pouvoir différencier α et β . En d'autres termes, nous supposons que $\sum_{i=1}^n (x_i - \bar{x})^2 > 0$, où $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

On peut traiter ces modèles dans un cadre tout à fait général (voir par exemple le chapitre 12 de [7]), mais nous nous limiterons ici au cas le plus important pour les applications, à savoir celui où les « erreurs » sont normales.

Nous allons donc en fait supposer que les v.a. $(\varepsilon_i)_{1 \leq i \leq n}$ sont indépendantes et de même loi $N(0, \sigma^2)$. Des *estimateurs* U pour α et V pour β sont des v.a. qui dépendent des valeurs observées Y_1, \dots, Y_n et aussi si l'on veut des valeurs connues x_1, \dots, x_n et σ^2 , mais *pas* des quantités inconnues α et β . Parmi tous les estimateurs possibles, les plus simples sont les *estimateurs linéaires*, qui sont de la forme

$$U = u_0 + \sum_{i=1}^n u_i Y_i, \quad V = v_0 + \sum_{i=1}^n v_i Y_i, \quad (16.9)$$

pour des constantes u_0, \dots, u_n et v_0, \dots, v_n . Les estimateurs U et V sont dits *sans biais* s'ils vérifient $E(U) = \alpha$ et $E(V) = \beta$. Comme $E(\varepsilon_i) = 0$, c'est le cas si et seulement si

² Parfois x_i est appelée la « variable indépendante » et Y_i la « variable dépendante » ; cette terminologie, cause de confusion avec les v.a. indépendantes, ne sera pas utilisée ici.

$$\alpha = E(U) = u_0 + \alpha \left(\sum_{i=1}^n u_i \right) + \beta \left(\sum_{i=1}^n u_i x_i \right) \quad (16.10)$$

$$\beta = E(V) = v_0 + \alpha \left(\sum_{i=1}^n v_i \right) + \beta \left(\sum_{i=1}^n v_i x_i \right). \quad (16.11)$$

Ces équations doivent être satisfaites pour toutes les valeurs possibles de α et β , ce qui est vrai si et seulement si

$$u_0 = 0, \quad \sum_{i=1}^n u_i = 1 \quad \text{et} \quad \sum_{i=1}^n u_i x_i = 0. \quad (16.12)$$

$$v_0 = 0, \quad \sum_{i=1}^n v_i = 0 \quad \text{et} \quad \sum_{i=1}^n v_i x_i = 1. \quad (16.13)$$

Parmi les estimateurs U (resp. V) de la forme (16.9) et qui satisfont (16.12) (resp. (16.13)), comment en choisir des « bons » ? Une méthode standard consiste à minimiser l'erreur quadratique : i.e. U_1 et V_1 seront considérés comme « meilleurs » que U_2 et V_2 si

$$E((U_1 - \alpha)^2) \leq E((U_2 - \alpha)^2), \quad E((V_1 - \beta)^2) \leq E((V_2 - \beta)^2). \quad (16.14)$$

C'est alors un exercice classique de minimisation sous contraintes de vérifier (ce qui se trouve dans tous les cours de statistique, voir par exemple [7, p. 557-564]) que le choix suivant des coefficients u_i et v_i :

$$v_i = \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2}, \quad u_i = \frac{1}{n} - \bar{x} v_i \quad (16.15)$$

minimise l'erreur quadratique sous les contraintes de non biais (16.12) et (16.13), et donc donne les meilleurs estimateurs parmi les estimateurs linéaires sans biais. À cause de (16.15) ces estimateurs optimaux ont la forme suivante, où $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$:

$$B = \sum_{i=1}^n \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} Y_i, \quad A = \bar{Y} - \bar{x} B. \quad (16.16)$$

On peut maintenant déterminer la loi des estimateurs A et B dans le cas gaussien. C'est une propriété remarquable du cas gaussien que le couple (A, B) est alors gaussien aussi : la « gaussianité » (si l'on peut dire) est donc préservée, ce qui est extrêmement utile en statistique.

Théorème 16.5. *Supposons les v.a. $(\varepsilon_i)_{1 \leq i \leq n}$ indépendantes et de loi $N(0, \sigma^2)$, et soit*

$$Y_i = \alpha + \beta x_i + \varepsilon_i.$$

Les estimateurs A pour α et B pour β donnés par (16.16) forment une v.a. 2-dimensionnelle gaussienne, de moyennes $E(A) = \alpha$ et $E(B) = \beta$ et de matrice de covariance donnée par

$$\begin{aligned} \text{Var}(A) &= \frac{\sigma^2}{n \sum_{j=1}^n (x_j - \bar{x})^2} \sum_{j=1}^n x_j^2, \\ \text{Var}(B) &= \frac{\sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2}, \\ \text{Cov}(A, B) &= \frac{-\sigma^2 \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2}, \end{aligned}$$

où $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$.

Preuve. Comme les ε_i sont indépendantes et normales, elles forment un vecteur gaussien par l'exemple 1. Comme A et B égalent une constante plus une combinaison linéaire des ε_i , le couple (A, B) est aussi gaussien par la définition même des v.a. normales multivariées. Par construction elles sont de moyennes respectives α et β .

Pour toute v.a. U on a $\text{Var}(uU + v) = u^2 \text{Var}(U)$, de sorte que

$$\begin{aligned} \text{Var}(B) &= \text{Var} \left(\sum_{i=1}^n v_i Y_i \right) \\ &= \text{Var} \left(\sum_{i=1}^n v_i \varepsilon_i \right) \\ &= \sum_{i=1}^n v_i^2 \text{Var}(\varepsilon_i) \quad (\text{par l'indépendance des } \varepsilon_i) \\ &= \sigma^2 \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} \right)^2 \quad (\text{car } \text{Var}(\varepsilon_i) = 1) \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

De même on a

$$\begin{aligned}
\text{Var}(A) &= \sum_{i=1}^n v_i^2 \text{Var}(\varepsilon_i) \\
&= \sigma^2 \sum_{i=1}^n \left(\frac{1}{n^2} + \bar{x}^2 v_i^2 - 2 \frac{\bar{x} v_i}{n} \right) \\
&= \sigma^2 \left(\frac{1}{n} + \bar{x}^2 \sum_{i=1}^n v_i^2 \right) \\
&= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \\
\text{Cov}(A, B) &= \text{Cov} \left(\sum_{i=1}^n u_i \varepsilon_i, \sum_{i=1}^n v_i \varepsilon_i \right) \\
&= \sum_{i=1}^n u_i v_i \text{Var}(\varepsilon_i) \\
&= \sigma^2 \sum_{i=1}^n \left(\frac{v_i}{n} - \bar{x} v_i^2 \right) = -\sigma^2 \bar{x} \sum_{i=1}^n v_i^2 \\
&= -\bar{x} \text{Var}(B) \quad \blacksquare
\end{aligned}$$

Terminons ce chapitre par un exemple qui sert aussi d'avertissement : X peut être une v.a. à valeurs dans \mathbb{R}^n dont toutes les composantes sont normales, mais qui n'est pas un vecteur gaussien. Donc, *la propriété d'être un vecteur gaussien est plus forte que le fait que chaque composante soit gaussienne.*

Exemple 2. Soit Y une v.a. $N(0, 1)$, et posons pour un $a > 0$

$$Z = Y1_{\{|Y| \leq a\}} - Y1_{\{|Y| > a\}}.$$

La loi de Z est aussi $N(0, 1)$ (voir l'exercice 2), mais $Y + Z = 2Y1_{\{|Y| \leq a\}}$ n'est pas une v.a. normale puisque (par exemple) $P(Y + Z > 2a) = 0$ et $Y + Z$ n'est pas p.s. égale à une constante. Donc $X = (Y, Z)$ est une v.a. à valeurs dans \mathbb{R}^2 qui n'est pas gaussienne, bien que chacune de ses deux composantes le soient.

Cela vaut la peine de souligner encore un certain nombre de propriétés qu'ont les vecteurs gaussiens, et qui ne sont pas vraies en général pour les vecteurs aléatoires non gaussiens :

1. Les composantes sont indépendantes si et seulement si elles sont non corrélées.

2. Les composantes sont normales, donc appartiennent à la même famille de lois que la variable vectorielle.
3. Un vecteur gaussien X de loi $N(\mu, Q)$ avec Q inversible peut être transformé linéairement en un vecteur de loi $N(0, I)$ (où I est la matrice identité de même dimension que X) (cf. exercice 6) : les transformations linéaires préservent la famille de lois.
4. La densité existe si et seulement si la matrice de covariance est inversible, ce qui donne un critère simple d'existence pour la densité.
5. Les lois conditionnelles des vecteurs gaussiens sont encore des lois gaussiennes (voir l'exercice 10).

Ces six propriétés montrent une stabilité remarquable de la famille des lois gaussiennes. Il existe encore bien d'autres propriétés spécifiques de ces lois, que nous n'évoquons pas ici.

Il est toutefois intéressant de remarquer que les lois normales ne se rencontrent pas réellement dans la nature, alors qu'elles interviennent via un passage à la limite (par le théorème-limite central) et constituent donc une approximation de la réalité, souvent même une excellente approximation. Quand on dit par exemple que la taille des hommes de 20 ans en France est distribuée selon une loi normale de moyenne μ et de variance σ^2 , on veut en fait dire que les tailles sont approximativement distribuées selon cette loi normale : en effet si le modèle normal était rigoureusement exact, il y aurait une probabilité strictement positive de trouver un homme plus grand que la tour Eiffel, ou de taille négative, ce qui est évidemment absurde. Cependant ces probabilités, bien que strictement positives, sont si petites qu'elles en sont négligeables : pour tous les calculs concrets l'approximation normale est excellente, alors que la « vraie » loi est évidemment inconnue.

Exercices

1. Soit $a \in \mathbb{R}^n$ avec $a \neq 0$, et $\mu \in \mathbb{R}^n$. Soit H l'hyperplan de \mathbb{R}^n défini par

$$H = \{x \in \mathbb{R}^n : \langle x - \mu, a \rangle = 0\}.$$

Montrer que $m_n(H) = 0$ où m_n est la mesure de Lebesgue sur \mathbb{R}^n , et en déduire que

$$\int_H f(x) dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) 1_{H}(x_1, \dots, x_n) dx_1 \dots dx_n = 0$$

pour toute fonction borélienne f sur \mathbb{R}^n .

2. Soit Y de loi $N(0, 1)$ et $a > 0$. Posons

$$Z = \begin{cases} Y & \text{si } |Y| \leq a, \\ -Y & \text{si } |Y| > a. \end{cases}$$

Montrer que Z est aussi de loi $N(0, 1)$.

3. Soit X une v.a. $N(0, 1)$ et soit Z indépendante de X avec $P(Z = 1) = P(Z = -1) = \frac{1}{2}$. Soit $Y = ZX$. Montrer que Y est de loi $N(0, 1)$, mais que (X, Y) n'est pas gaussien.

4. Soit (X, Y) un vecteur gaussien bidimensionnel. On suppose que les variances de X et Y , soit σ_X^2 et σ_Y^2 , sont non nulles, et on note ρ le coefficient de corrélation de X et Y , et μ_X et μ_Y leurs moyennes. Montrer que si $-1 < \rho < 1$ la densité de (X, Y) existe et égale

$$f_{(X, Y)}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left(\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right) \right\}.$$

Montrer que si $\rho = -1$ ou $\rho = 1$, la v.a. (X, Y) n'admet pas de densité.

5. Soit $\rho \in [-1, 1]$, $\mu_1, \mu_2 \in \mathbb{R}$ et $\sigma_1 > 0$, $\sigma_2 > 0$. Construire des v.a. X_1 et X_2 normales de moyennes μ_1 et μ_2 , de variances σ_1^2 et σ_2^2 , et de coefficient de corrélation ρ . (*Indication* : Soit Y_1 et Y_2 des v.a. indépendantes $N(0, 1)$ et poser $U_1 = Y_1$ et $U_2 = \rho Y_1 + \sqrt{1-\rho^2} Y_2$, puis $X_j = \mu_j + \sigma_j U_j$ ($j = 1, 2$).)

6. Supposons X gaussien $N(\mu, Q)$ sur \mathbb{R}^n , avec $\det(Q) > 0$. Montrer qu'il existe une matrice B telle que $Y = B(X - \mu)$ soit de loi $N(0, I)$, où I est la matrice identité $n \times n$. (*Cela montre qu'un vecteur gaussien non dégénéré peut se transformer linéairement en un vecteur gaussien « standard ».*)

7. Soit $X = (X_1, \dots, X_n)$ un vecteur gaussien, et

$$Y = \sum_{j=1}^n a_j X_j,$$

Montrer que Y est $N(\mu, \sigma^2)$ avec

$$\mu = \sum_{j=1}^n a_j E(X_j)$$

et

$$\sigma^2 = \sum_{j=1}^n a_j^2 \text{Var}(X_j) + 2 \sum_{j < k} a_j a_k \text{Cov}(X_j, X_k).$$

8. Soit (X, Y) bidimensionnelle de loi $N(\mu, Q)$, où

$$Q = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

et supposons que $\det(Q) > 0$, donc (X, Y) admet une densité. Montrer que la densité conditionnelle $f_{X=r}$ est celle de la loi $N\left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(r - \mu_X), \sigma_Y^2(1 - \rho^2)\right)$. (Cf. le théorème 12.2 ou l'exercice 3 du chapitre 12.)

9. Soit X une v.a. bidimensionnelle de loi $N(\mu, Q)$ avec $\mu = (1, 1)$ et $Q = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$. Trouver la loi conditionnelle de $Y = X_1 + X_2$ sachant $Z = X_1 - X_2 = 0$.

$$\left[\text{Rép. : } f_{Z=0}(y) = \frac{1}{\sqrt{2\pi}\sqrt{\frac{20}{3}}} \exp\left\{-\frac{1}{2}\frac{(y-2)^2}{\frac{20}{3}}\right\} \right]$$

10. Soit $X = (X_1, \dots, X_n)$ de loi $N(\mu, Q)$ avec $\det(Q) > 0$. Montrer que la loi conditionnelle de (X_1, \dots, X_m) sachant $(X_{m+1} = x_{m+1}, \dots, X_n = x_n)$ (avec $1 \leq m < n$) est encore une loi gaussienne. [Cela généralise l'exercice 8.]

11. (Gut. 1995.) Soit (X, Y) une v.a. de densité sur \mathbb{P}^2 :

$$f_{(X,Y)}(x, y) = c e^{-(1+x^2)(1+y^2)},$$

où c est choisi de manière à ce que f soit effectivement une densité. Montrer que le couple (X, Y) n'est pas gaussien, mais que les densités conditionnelles $f_{X=y}$ et $f_{Y=x}$ sont des densités gaussiennes (cela montre que la réciproque de l'exercice 10 est fautive).

12. Soit (X, Y) une v.a. bidimensionnelle gaussienne de moyenne nulle, et soit ρ le coefficient de corrélation de X et Y . Montrer que si $|\rho| < 1$, la variable $Z = \frac{X}{Y}$ suit une loi de Cauchy de paramètres $\alpha = \rho \frac{\sigma_X}{\sigma_Y}$ et $\beta = \frac{\sigma_X}{\sigma_Y} \sqrt{1 - \rho^2}$. (Remarque : ce résultat a déjà été établi lorsque X et Y sont indépendantes).

13.* Soit (X, Y) une v.a. bidimensionnelle gaussienne de moyenne nulle, et soit ρ le coefficient de corrélation de X et Y . Soit β tel que

$$\cos \beta = \rho \quad (0 \leq \beta \leq \pi).$$

Montrer que

$$P(XY < 0) = \frac{\beta}{\pi}.$$

(Indication : Se rappeler que, d'après l'exercice 12, si $Z = \frac{X}{Y}$ et $z = \rho \frac{\sigma_X}{\sigma_Y}$, on a

$$F_Z(z) = \frac{1}{2} + \frac{1}{\pi} \operatorname{Arctan} \left(\frac{z\sigma_Y - \rho\sigma_X}{\sigma_X \sqrt{1 - \rho^2}} \right).$$

Soit $\alpha = \operatorname{Arcsin} \rho$ ($-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$); montrer d'abord que $P(XY < 0) = \frac{1}{2} - \frac{\alpha}{\pi}$, en utilisant que $\operatorname{Arctan} \frac{\rho}{\sqrt{1 - \rho^2}} = \operatorname{Arcsin} \rho$.)

14. Soit (X, Y) , α et ρ comme dans l'exercice 13. Montrer que

$$\begin{aligned} P(X > 0, Y > 0) &= P(X < 0, Y < 0) = \frac{1}{4} + \frac{\alpha}{2\pi}; \\ P(X > 0, Y < 0) &= P(X < 0, Y > 0) = \frac{1}{4} - \frac{\alpha}{2\pi}. \end{aligned}$$

15.* Soit (X, Y) une v.a. bidimensionnelle de densité

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma_X^2} - \frac{2xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2} \right)}.$$

Montrer que

a) $E(XY) = \rho\sigma_X\sigma_Y$

b) $E(X^2Y^2) = E(X^2)E(Y^2) + 2(E(XY))^2$

c) $E(|XY|) = \frac{2\sigma_X\sigma_Y}{\pi} (\cos \alpha + \alpha \sin \alpha)$ où $\alpha = \operatorname{Arcsin} \rho$ ($-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$) (cf. exercice 13).

16. Soit (X, Y) une v.a. gaussienne bidimensionnelle avec $\sigma_X^2 = \sigma_Y^2 > 0$, et soit ρ le coefficient de corrélation de X et Y . Montrer que X et $Y - \rho X$ sont indépendantes.

17. Soit X de loi $N(\mu, Q)$ sur \mathbb{R}^n , avec $\det(Q) > 0$. Montrer que

$$(X - \mu)^* Q^{-1} (X - \mu) \text{ suit la loi } \chi_n^2.$$

18. Soit X_1, \dots, X_n des v.a. indépendantes de loi $N(0, \sigma^2)$, et posons

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n X_j, \quad S^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{x})^2.$$

(Rappelons que d'après l'exercice 13 du chapitre 15 les v.a. \bar{x} et S^2 sont indépendantes.) Montrer que

$$\sum_{j=1}^n X_j^2 = \sum_{j=1}^n (X_j - \bar{x})^2 + n\bar{x}^2$$

et en déduire que $(n-1)S^2/\sigma^2$ admet la loi χ_{n-1}^2 et que $n\bar{x}^2/\sigma^2$ admet la loi χ_1^2 .

19. Soit $\varepsilon_1, \dots, \varepsilon_n$ des v.a. indépendantes de loi $N(0, \sigma^2)$ et soit $Y_i = \alpha + \beta x_i + \varepsilon_i$. Supposons aussi que les x_i ne sont pas tous égaux, et posons $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. On définit les *résidus de régression* par

$$\hat{\varepsilon}_i = Y_i - A - Bx_i,$$

où A et B sont donnés par (16.16).

a) Montrer que $E(\hat{\varepsilon}_i) = 0$.

b)* Montrer que

$$\text{Var}(\hat{\varepsilon}_i) = \sigma^2 \left(\frac{n-1}{n} - \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right).$$

20. On se place dans la situation de l'exercice précédent. Supposons σ inconnu, et posons

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2.$$

Montrer que $E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2$. (On a $E(\hat{\sigma}^2) \neq \sigma^2$, et $\hat{\sigma}^2$ est appelé un estimateur *biaisé* de σ^2 ; un estimateur non biaisé serait $S^2 = \frac{n}{n-2} \hat{\sigma}^2$.)

21. Toujours dans la situation des deux exercices précédents, montrer que les v.a. (A, B) et $\hat{\sigma}^2$ sont indépendantes.

Chapitre 17

Convergence des variables aléatoires

Dans les cours élémentaires de mathématiques on introduit la convergence des suites de fonctions : si $f_n : \mathbb{R} \rightarrow \mathbb{R}$, alors $\lim_{n \rightarrow \infty} f_n = f$ si $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ pour tout $x \in \mathbb{R}$, et on dit alors que la suite (f_n) converge simplement vers f . Une v.a. est bien sûr une fonction ($X : \Omega \rightarrow \mathbb{R}$ pour un espace « abstrait » Ω), et nous avons donc la même notion : une suite $X_n : \Omega \rightarrow \mathbb{R}$ converge simplement vers X si $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ pour tout $\omega \in \Omega$. Cette définition naturelle est, de manière surprenante, à peu près inutile en probabilités : l'exemple suivant explique pourquoi.

Exemple 1. Soit X_n une suite de v.a. qui sont i.i.d. (indépendantes et identiquement distribuées, i.e. de même loi), avec $P(X_n = 1) = p$ et $P(X_n = 0) = 1 - p$. Par exemple on peut imaginer qu'elles donnent les résultats successifs quand on lance (une infinité de fois) une pièce de monnaie, $X_n = 1$ (resp. $X_n = 0$) traduisant le fait que le n -ième tirage donne face (resp. pile) ; si $p \neq 1/2$ cela veut dire que la pièce est truquée. Lorsque n est grand, on s'attend à ce que la proportion de faces soit à peu près égale à p (c'est l'essence de « l'approche par les fréquences » en probabilités). Mathématiquement, on voudrait que

$$\lim_{n \rightarrow \infty} \frac{X_1(\omega) + \cdots + X_n(\omega)}{n} = p \quad \text{pour tout } \omega \in \Omega.$$

Cela est tout simplement faux : par exemple si $\omega_0 = \{f, f, f, \dots\}$ est la suite ne contenant que des faces, on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j(\omega_0) = 0.$$

Plus généralement $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j(\omega) = 0$ pour tout ω dans l'ensemble $A = \{\omega : \text{il n'y a qu'un nombre fini de faces}\}$. Plus généralement encore, on peut trouver des ω pour lesquels la fréquence converge vers n'importe quel nombre fixé dans $[0, 1]$, et d'autres pour lesquels la fréquence ne converge pas. D'ailleurs si on remplace p par un autre

nombre $q \in [0, 1]$, on change la probabilité, mais on ne change ni Ω , ni les fonctions $\frac{1}{n}(X_1 + \dots + X_n)$, alors qu'on souhaiterait que ces fonctions convergent vers p dans le premier cas, et vers q dans le second.

Bien entendu l'événement A est plutôt invraisemblable, et on peut en fait montrer (voir exercice 14) qu'il vérifie $P(A) = 0$. Ce qu'on verra plus tard (voir la « loi des grands nombres » au chapitre 20) est que

$$P \left(\left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j(\omega) = p \right\} \right) = 1.$$

Ce type de convergence, pour lequel on n'a pas convergence pour *tout* ω , mais seulement pour *presque tout* ω , est typiquement ce qui arrive pour des variables aléatoires.

Avertissement. Dans tout ce chapitre les v.a. qui interviennent sont à valeurs réelles (sauf dans la dernière remarque) et sont toutes définies sur le même espace de probabilité (Ω, \mathcal{A}, P) .

Rappelons d'abord, sous la forme d'une définition, la notion de convergence p.s. d'une suite de v.a., qui a déjà été utilisée dans le théorème 9.1 par exemple.

Définition 17.1. On dit qu'une suite $(X_n)_{n \geq 1}$ de v.a. converge presque sûrement vers une v.a. X si l'ensemble N des ω tels que la suite numérique $(X_n(\omega))$ ne converge pas vers $X(\omega)$ vérifie $P(N) = 0$. Rappelons que l'ensemble N est alors dit « négligeable ».

Noter qu'on a aussi

$$N^c = \Lambda = \left\{ \omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} \quad \text{vérifie } P(\Lambda) = 1.$$

On note la convergence presque sûre ainsi :

$$\lim_{n \rightarrow \infty} X_n = X \text{ p.s.} \quad \text{ou} \quad X_n \rightarrow X \text{ p.s.}$$

Il apparaît que même la convergence p.s. est trop forte dans certaines situations, et nous introduisons ci-dessous deux autres modes de convergence.

Définition 17.2. On dit qu'une suite $(X_n)_{n \geq 1}$ de v.a. converge dans L^p (où $p \in [1, \infty[$) vers une v.a. X si les X_n et X sont dans L^p , et si on a

$$\lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0.$$

On dit aussi que X_n converge en moyenne d'ordre p vers X , et on écrit

$$X_n \xrightarrow{L^p} X.$$

Les cas les plus importants sont $p = 1$ (convergence « en moyenne ») et $p = 2$ (convergence « en moyenne quadratique »). Quand $p = 1$, on a $|\mathbb{E}(X_n - X)| \leq \mathbb{E}(|X_n - X|)$ et $|\mathbb{E}(|X_n|) - \mathbb{E}(|X|)| \leq \mathbb{E}(|X_n - X|)$ parce que $||x| - |y|| \leq |x - y|$. Donc

$$X_n \xrightarrow{L^1} X \text{ implique } \mathbb{E}(X_n) \rightarrow \mathbb{E}(X) \text{ et } \mathbb{E}(|X_n|) \rightarrow \mathbb{E}(|X|). \quad (17.1)$$

De même, quand $X_n \xrightarrow{L^p} X$ pour $p \in]1, \infty[$, on peut montrer que $\mathbb{E}(|X_n|^p) \rightarrow \mathbb{E}(|X|^p)$: voir l'exercice 15 pour le cas $p = 2$.

Définition 17.3. On dit qu'une suite $(X_n)_{n \geq 1}$ de v.a. converge en probabilité vers une v.a. X si pour tout $\varepsilon > 0$ on a

$$\lim_{n \rightarrow \infty} \mathbb{P}(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0,$$

ou plus simplement : $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$ quand $n \rightarrow \infty$. On écrit alors

$$X_n \xrightarrow{P} X.$$

Une autre manière d'écrire les choses est la suivante : X_n converge en probabilité vers X lorsque, pour tous $\varepsilon > 0$ et $\delta > 0$ il existe $N = N(\delta, \varepsilon)$ tel que

$$n \geq N \quad \Rightarrow \quad \mathbb{P}(|X_n - X| > \varepsilon) < \delta.$$

Avant d'établir les rapports entre les différents modes de convergence on va donner un critère simple de convergence en probabilité.

Théorème 17.1. On a $X_n \xrightarrow{P} X$ si et seulement si

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{|X_n - X|}{1 + |X_n - X|} \right) = 0.$$

Preuve. On ne nuit pas à la généralité en supposant que $X = 0$, de sorte qu'il nous suffit de montrer que $X_n \xrightarrow{P} 0$ équivaut à $\lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{|X_n|}{1 + |X_n|} \right) = 0$. Supposons d'abord que $X_n \xrightarrow{P} 0$. Pour tout $\varepsilon > 0$, on a $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \varepsilon) = 0$. Mais

$$\frac{|X_n|}{1 + |X_n|} \leq \frac{|X_n|}{1 + |X_n|} 1_{\{|X_n| > \varepsilon\}} + \varepsilon 1_{\{|X_n| \leq \varepsilon\}} \leq 1_{\{|X_n| > \varepsilon\}} + \varepsilon,$$

donc

$$E\left(\frac{|X_n|}{1+|X_n|}\right) \leq E(1_{\{|X_n|>\varepsilon\}}) + \varepsilon = P(|X_n| > \varepsilon) + \varepsilon.$$

En prenant les limites on arrive à

$$\limsup_{n \rightarrow \infty} E\left(\frac{|X_n|}{1+|X_n|}\right) \leq \varepsilon;$$

comme ε est arbitrairement petit il vient $\lim_{n \rightarrow \infty} E\left(\frac{|X_n|}{1+|X_n|}\right) = 0$.

Supposons à l'inverse que $\lim_{n \rightarrow \infty} E\left(\frac{|X_n|}{1+|X_n|}\right) = 0$. La fonction $f(x) = \frac{x}{1+x}$ est strictement croissante, donc

$$\frac{\varepsilon}{1+\varepsilon} 1_{\{|X_n|>\varepsilon\}} \leq \frac{|X_n|}{1+|X_n|} 1_{\{|X_n|>\varepsilon\}} \leq \frac{|X_n|}{1+|X_n|}.$$

En prenant les espérances, puis les limites, on arrive à

$$\frac{\varepsilon}{1+\varepsilon} \lim_{n \rightarrow \infty} P(|X_n| > \varepsilon) \leq \lim_{n \rightarrow \infty} E\left(\frac{|X_n|}{1+|X_n|}\right) = 0.$$

Comme $\varepsilon > 0$ est fixé, on conclut que $\lim_{n \rightarrow \infty} P(|X_n| > \varepsilon) = 0$. ■

Remarque. Ce théorème nous dit que $X_n \xrightarrow{P} X$ si et seulement si $E(f(|X_n - X|)) \rightarrow 0$ pour la fonction $f(x) = \frac{|x|}{1+|x|}$. Un examen soigneux de sa preuve montre que la même équivalence reste vraie pour toute fonction f sur $[0, \infty[$ qui est bornée, strictement croissante, continue et nulle en 0. Par exemple $X_n \xrightarrow{P} X$ équivaut à $E(|X_n - X| \wedge 1) \rightarrow 0$, et aussi à $E(\text{Arctan } |X_n - X|) \rightarrow 0$.

Le théorème suivant nous indique que la convergence en probabilité est le mode de convergence le plus faible introduit jusqu'à présent.

Théorème 17.2. Soit $(X_n)_{n \geq 1}$ une suite de v.a.

(a) Si $X_n \xrightarrow{L^p} X$, on a $X_n \xrightarrow{P} X$.

(b) Si $X_n \rightarrow X$ p.s., on a $X_n \xrightarrow{P} X$.

Preuve. (a) En se rappelant que pour tout événement A on a $P(A) = E(1_A)$, où 1_A est la fonction indicatrice de A , on voit que

$$P(|X_n - X| > \varepsilon) = E(1_{\{|X_n - X| > \varepsilon\}}).$$

Mais $\frac{|X_n - X|^p}{\varepsilon^p} > 1$ sur l'événement $\{|X_n - X| > \varepsilon\}$, donc

$$\begin{aligned} P(|X_n - X| > \varepsilon) &\leq E\left(\frac{|X_n - X|^p}{\varepsilon^p} 1_{\{|X_n - X| > \varepsilon\}}\right) \\ &= \frac{1}{\varepsilon^p} E(|X_n - X|^p 1_{\{|X_n - X| > \varepsilon\}}), \end{aligned}$$

et comme $|X_n - X|^p \geq 0$ partout, on peut simplement omettre la fonction indicatrice, ce qui donne

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p} E(|X_n - X|^p).$$

Cette quantité tend vers 0 quand $n \rightarrow \infty$ (pour $\varepsilon > 0$ fixé), d'où le résultat.

(b) On a $\frac{|X_n - X|}{1 + |X_n - X|} \leq 1$, donc

$$\lim_{n \rightarrow \infty} E\left(\frac{|X_n - X|}{1 + |X_n - X|}\right) = E\left(\lim_{n \rightarrow \infty} \frac{|X_n - X|}{1 + |X_n - X|}\right) = E(0) = 0$$

grâce au théorème de convergence dominée de Lebesgue 9.1(f). Il nous suffit alors d'appliquer le théorème 17.1. ■

La réciproque du théorème 17.2 est fautive : il existe cependant deux « réciproques » partielles. La plus délicate concerne la convergence p.s., et s'énonce comme suit :

Théorème 17.3. *Supposons que $X_n \xrightarrow{p} X$. Il existe alors une sous-suite n_k de la suite des entiers, telle que $X_{n_k} \rightarrow X$ p.s.*

Preuve. On a $\lim_{n \rightarrow \infty} E\left(\frac{|X_n - X|}{1 + |X_n - X|}\right) = 0$ par le théorème 17.1. On peut donc trouver une suite croissante d'entiers n_k telle que $E\left(\frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|}\right) < \frac{1}{2^k}$. On a alors $\sum_{k=1}^{\infty} E\left(\frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|}\right) < \infty$, et le théorème 9.2 implique que $\sum_{k=1}^{\infty} \frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|} < \infty$ p.s.; puisque le terme général d'une série convergente tend vers 0, on en déduit que

$$\lim_{n \rightarrow \infty} |X_{n_k} - X| = 0 \quad \text{p.s.} \quad \blacksquare$$

Remarque 17.1. Ce résultat peut aussi être démontré de manière assez simple en utilisant le lemme de Borel-Cantelli (théorème 10.5).

Exemple 2. La convergence en probabilité n'implique pas la convergence presque sûre. Prenons par exemple $\Omega =]0, 1]$, avec la tribu borélienne \mathcal{A} et la probabilité uniforme P . Si A_n est un sous-intervalle de Ω de longueur a_n et si $X_n = 1_{A_n}$, on a $P(|X_n| > \varepsilon) = a_n$ pour $\varepsilon \in]0, 1[$; donc, dès que $a_n \rightarrow 0$, on a $X_n \xrightarrow{P} 0$. Soit alors $X_{n,j}$ la fonction indicatrice de l'intervalle $]\frac{j-1}{n}, \frac{j}{n}]$, pour $1 \leq j \leq n$ et $n \geq 1$. On peut réordonner la « double » suite $X_{n,j}$ selon l'ordre lexicographique (choisir d'abord n , puis j), ce qui conduit à une nouvelle suite Y_n , dont les premiers termes sont :

$$\begin{array}{cccccccc} X_{1,1} & X_{2,1} & X_{2,2} & X_{3,1} & X_{3,2} & X_{3,3} & X_{4,1} & \dots \\ Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 & Y_7 & \dots \end{array}$$

Pour tout ω et tout $n \geq 2$ il existe un j tel que $X_{n,j}(\omega) = 1$ et un autre tel que $X_{n,j}(\omega) = 0$: donc $\limsup_{m \rightarrow \infty} Y_m(\omega) = 1$ et $\liminf_{m \rightarrow \infty} Y_m(\omega) = 0$, et la suite Y_n ne converge pour aucune valeur de ω . Cependant Y_n est l'indicatrice d'un intervalle dont la longueur tend vers 0, donc $Y_n \xrightarrow{P} 0$.

Théorème 17.4. *Supposons que $X_n \xrightarrow{P} X$ et que $|X_n| \leq Y$ pour tout n , où Y est une v.a. appartenant à L^p pour un $p \in [1, \infty[$. On a alors $X \in L^p$ et $X_n \xrightarrow{L^p} X$.*

Preuve. Comme $E(|X_n|^p) \leq E(Y^p) < \infty$, on a $X_n \in L^p$. Pour tout $\varepsilon > 0$ on a aussi

$$\begin{aligned} \{|X_n| > Y + \varepsilon\} &\subset \{|X| > |X_n| + \varepsilon\} \\ &\subset \{|X| - |X_n| > \varepsilon\} \\ &\subset \{|X - X_n| > \varepsilon\}, \end{aligned}$$

donc

$$P(|X| > Y + \varepsilon) \leq P(|X - X_n| > \varepsilon),$$

et comme ceci est vrai pour chaque n , il vient

$$P(|X| > Y + \varepsilon) \leq \lim_{n \rightarrow \infty} P(|X - X_n| > \varepsilon) = 0,$$

par hypothèse. Ceci étant vrai pour tout $\varepsilon > 0$, on en déduit que

$$P(|X| > Y) \leq \lim_{m \rightarrow \infty} P\left(|X| > Y + \frac{1}{m}\right) = 0,$$

qui entraîne $|X| \leq Y$ p.s., et donc $X \in L^p$.

Supposons maintenant que X_n ne converge pas vers X dans L^p . Il existe alors une suite d'entiers strictement croissante n_k et un $\varepsilon > 0$

tels que $E(|X_{n_k} - X|^p) \geq \varepsilon$ pour tout k . La sous-suite X_{n_k} converge évidemment vers X en probabilité, donc par le théorème 17.3 elle admet une « sous-sous-suite » $X_{n_{k_j}}$ qui converge p.s. vers X . Donc les v.a. $X_{n_{k_j}} - X$ convergent p.s. vers 0, tout en restant majorées en valeur absolue par $2Y$: une application du théorème de Lebesgue donne alors que $E(|X_{n_{k_j}} - X|^p) \rightarrow 0$, en contradiction avec le fait que $E(|X_{n_k} - X|^p) \geq \varepsilon$ pour tout k . ■

Le théorème suivant est facile, mais utile.

Théorème 17.5. *Soit f une fonction continue.*

(a) *Si $X_n \rightarrow X$ p.s., alors $f(X_n) \rightarrow f(X)$ p.s.*

(b) *Si $X_n \xrightarrow{P} X$, alors $f(X_n) \xrightarrow{P} f(X)$.*

Preuve. (a) Soit N l'ensemble introduit dans la définition 17.1. On a $P(N) = 0$ par hypothèse. Si $\omega \notin N$, il vient

$$\lim_{n \rightarrow \infty} f(X_n(\omega)) = f\left(\lim_{n \rightarrow \infty} X_n(\omega)\right) = f(X(\omega)),$$

où la première égalité provient de la continuité de f : le résultat en découle.

(b) Pour chaque $k > 0$ on a :

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset \{|f(X_n) - f(X)| > \varepsilon, |X| \leq k\} \cup \{|X| > k\}. \quad (17.2)$$

Comme f est continue, elle est uniformément continue sur chaque intervalle borné. Donc, pour ε et k fixés il existe $\delta > 0$ tel que $|f(x) - f(y)| \leq \varepsilon$ si $|x - y| \leq \delta$ et $x \in [-k, k]$. Par suite

$$\begin{aligned} \{|f(X_n) - f(X)| > \varepsilon, |X| \leq k\} &\subset \{|X_n - X| > \delta, |X| \leq k\} \\ &\subset \{|X_n - X| > \delta\}. \end{aligned}$$

En combinant ceci avec (17.2) il vient

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset \{|X_n - X| > \delta\} \cup \{|X| > k\}. \quad (17.3)$$

En utilisant la sous-additivité ($P(A \cup B) \leq P(A) + P(B)$) on déduit de (17.3) que

$$P(\{|f(X_n) - f(X)| > \varepsilon\}) \leq P(\{|X_n - X| > \delta\}) + P(\{|X| > k\}).$$

Mais $\{|X| > k\}$ tend vers l'ensemble vide lorsque $k \rightarrow \infty$, donc $\lim_{k \rightarrow \infty} P(\{|X| > k\}) = 0$. Pour $\gamma > 0$ arbitraire on peut alors choisir

k de sorte que $P(|X| > k) < \gamma$. Une fois k fixé on choisit δ pour avoir (17.3), et par suite

$$\limsup_{n \rightarrow \infty} P(|f(X_n) - f(X)| > \varepsilon) \leq \lim_{n \rightarrow \infty} P(|X_n - X| > \delta) + \gamma = \gamma.$$

Comme $\gamma > 0$ est arbitraire, on a le résultat. ■

Remarque 17.2. Nous avons supposé les X_n et X réelles. Les mêmes définitions et les mêmes résultats sont valables si les X_n et X prennent leurs valeurs dans le même espace \mathbb{R}^d , à condition de comprendre la notation $|x|$ comme la norme euclidienne du vecteur $x \in \mathbb{R}^d$. De plus, dire que X_n converge vers X pour l'un des modes de convergence introduits ci-dessus revient exactement à dire que la suite des j -ièmes composantes des X_n converge vers la composante correspondante de X , pour tout $j = 1, \dots, d$.

Exercices

1. Soit les $X_{n,j}$ comme dans l'exemple 2, et $Z_{n,j} = n^{\frac{1}{p}} X_{n,j}$. On note maintenant Y_n la suite obtenue en réordonnant les $Z_{n,j}$ (au lieu des $X_{n,j}$), comme dans l'exemple 2 également. Montrer que $Y_n \rightarrow 0$ en probabilité mais pas dans L^p , bien que chaque Y_n appartienne à L^p .
2. Montrer que le théorème 17.5(h) est faux en général, lorsqu'on ne suppose pas f continue. (*Indication* : Prendre $f(x) = 1_{\{0\}}(x)$ et les X_n tendant vers 0 en probabilité.)
3. Soit X_n des v.a. indépendantes avec $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$. Montrer que

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{P} 0.$$

(*Indication* : Soit $S_n = \sum_{j=1}^n X_j$; utiliser l'inégalité de Bienaymé-Chebyshev pour $P(|S_n| > n\varepsilon)$.)

4. Soit X_n et S_n comme dans l'exercice 3. Montrer que $\frac{1}{n^2} S_n^2 \rightarrow 0$ p.s. (*Indication* : Montrer $\sum_{n=1}^{\infty} P(\frac{1}{n^2} |S_n^2| > \varepsilon) < \infty$ et utiliser le lemme de Borel-Cantelli.)
5. Si on a $|X_n| \leq Y$ p.s. pour tout n , montrer que $\sup_n |X_n| \leq Y$ p.s. également.

6. Soit $X_n \xrightarrow{P} X$. Montrer que les fonctions caractéristiques φ_{X_n} convergent simplement vers φ_X .
7. Soit X_1, \dots, X_n des v.a. indépendantes de loi de Cauchy de paramètres $\alpha = 0$ et $\beta = 1$ (donc de densité $f(x) = \frac{1}{\pi(1+x^2)}$). Montrer que $\frac{1}{n} \sum_{j=1}^n X_j$ suit aussi une loi de Cauchy. (*Indication* : Utiliser les fonctions caractéristiques ; voir l'exercice 1 du chapitre 13.)
8. Sous les mêmes hypothèses que dans l'exercice précédent, montrer qu'il n'existe aucune constante γ telle que $\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{P} \gamma$. (*Indication* : Utiliser l'exercice 7.) En déduire qu'il n'existe pas de constante γ telle que $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j = \gamma$ p.s.
9. Si les v.a. $(X_n)_{n \geq 1}$ sont de moyennes nulles et de variances $\sigma_{X_n}^2$ tendant vers 0 quand $n \rightarrow \infty$, montrer que X_n converge vers 0 dans L^2 et en probabilité.
10. Soit X_j des v.a. indépendantes, de même loi, de moyenne nulle et de variance finie. Montrer que si $S_n = \sum_{j=1}^n X_j$, alors $\frac{1}{n} S_n$ tend vers 0 dans L^2 et en probabilité.
- 11.* Supposons que $X_n \rightarrow X$ p.s., les v.a. X_n et X étant réelles. Soit $Y = \sup_n |X_n|$. Montrer que $Y < \infty$ p.s.
- 12.* Supposons que $X_n \rightarrow X$ p.s., les v.a. X_n et X étant réelles, et soit $Y = \sup_n |X_n - X|$. Montrer que $Y < \infty$ p.s. (voir l'exercice 11), et après avoir défini une nouvelle mesure de probabilité Q par

$$Q(A) = \frac{1}{c} E\left(1_A \frac{1}{1+Y}\right), \quad \text{où } c = E\left(\frac{1}{1+Y}\right),$$

montrer que X_n tend vers X dans L^1 pour la probabilité Q .

13. Soit A l'événement de l'exemple 1. Montrer que $P(A) = 0$. (*Indication* : Soit

$$A_n = \{ \text{on a Face au } n\text{-ième tirage} \}.$$

Montrer que $\sum_{n=1}^{\infty} P(A_n) = \infty$ et utiliser le lemme de Borel-Cantelli.)

14. Soit X_n et X des v.a. de L^2 , telles que X_n tende vers X dans L^2 . Montrer que $E(X_n^2) \rightarrow E(X^2)$ (*Indication* : utiliser que $|x^2 - y^2| = (x - y)^2 + 2|y||x - y|$ et l'inégalité de Cauchy-Schwarz).
- 15.* (Un autre *théorème de convergence dominée*.) Soit $(X_n)_{n \geq 1}$ des v.a. satisfaisant $X_n \xrightarrow{P} X$. Supposons que $|X_n(\omega)| \leq C$ pour une certaine constante C et tout ω . Montrer que $\lim_{n \rightarrow \infty} E(|X_n - X|) = 0$. (*Indication* : Montrer d'abord que $P(|X| \leq C) = 1$.)

Chapitre 18

Convergence en loi

Dans le chapitre 17 nous avons défini trois types de convergence pour les variables aléatoires : presque sûre, dans L^p et en probabilité. Ces trois types de convergence, bien que différant de la convergence usuelle des fonctions en analyse, sont néanmoins de même nature et peuvent être conçues comme des variantes de la convergence simple habituelle. Il existe un autre mode de convergence, de nature bien différente, mais aussi très utile en probabilité : il s'agit de la convergence « en loi », appelée aussi « convergence étroite », et parfois « convergence faible ». Comme ce dernier nom l'indique, il s'agit d'un mode faible de convergence : plus la convergence est « faible », plus il est facile pour des v.a. de converger ! Ce qui est inhabituel pour ce type de convergence, c'est que des v.a. X_n peuvent converger vers X sans que les valeurs prises par X_n se rapprochent effectivement de celles prises par X en un même point ω , et même, ces variables peuvent être définies sur des espaces de probabilité différents. En fait, il s'agit d'une convergence des lois des v.a. X_n vers la loi de X , et on a donc ainsi un mode de convergence *radicalement différent* de ce qui a été vu au chapitre 17.

Comme nous allons traiter de convergence de lois, nous commençons par examiner simplement les lois de probabilité sur \mathbb{R}^d pour un d fixé, $d \geq 1$.

Définition 18.1. Soit μ_n et μ des probabilités sur \mathbb{R}^d . On dit que la suite μ_n converge étroitement vers μ si $\int f(x)\mu_n(dx)$ converge vers $\int f(x)\mu(dx)$ pour toute fonction f réelle sur \mathbb{R}^d qui est continue et bornée.

À première vue il semble qu'il y ait une erreur typographique : on est habitué à considérer la propriété

$$\lim_{n \rightarrow \infty} \int f_n(x)\mu(dx) = \int f(x)\mu(dx);$$

mais ici la fonction f est fixée et ce sont les mesures qui varient. Noter aussi qu'on ne considère pas la convergence des intégrales pour toute fonction borélienne bornée, mais que nous imposons à f d'être en plus continue.

Définition 18.2. Soit X_n et X des v.a. à valeurs dans \mathbb{E}^d . On dit que X_n converge en loi vers X si les lois P_{X_n} convergent étroitement vers P_X , et on écrit $X_n \xrightarrow{L} X$.

Une remarque très importante est à faire tout de suite : dans cette définition, les v.a. X_n et X peuvent être définies sur des espaces distincts, puisque seules leurs lois sont en cause. Il arrive même qu'une suite X_n converge vers une limite X qui ne peut pas exister sur les espaces sur lesquels sont définies les X_n , parce que ceux-ci sont « trop petits » : par exemple X_n est une variable binomiale de taille n , convenablement normalisée, et la limite X est normale : l'espace naturel sur lequel X_n est définie contient $n + 1$ points, et sur un tel espace toutes les v.a. sont discrètes. La convergence en loi permet donc une sorte de convergence pour des v.a. pour lesquelles toute autre forme de convergence serait impossible.

Théorème 18.1. Soit X_n et X des v.a. à valeurs dans \mathbb{E}^d . On a $X_n \xrightarrow{L} X$ si et seulement si

$$\lim_{n \rightarrow \infty} E(f(X_n)) = E(f(X)),$$

pour toute fonction réelle continue et bornée sur \mathbb{R}^d .

Preuve. Il suffit de combiner les définitions 18.1 et 18.2 et d'observer que

$$E(f(X_n)) = \int f(x)P_{X_n}(dx), \quad E(f(X)) = \int f(x)P_X(dx). \quad \blacksquare$$

Pour avoir la convergence p.s. ou dans L^p ou en probabilité les v.a. X_n et X doivent être définies sur le même espace. Donc a priori la convergence en loi n'est pas comparable aux autres. On a cependant le résultat suivant :

Théorème 18.2. Soit X_n et X des v.a., toutes définies sur le même espace de probabilité (Ω, \mathcal{A}, P) . Si X_n converge vers X en probabilité, alors X_n converge aussi vers X en loi.

Preuve. Soit f une fonction réelle bornée et continue sur \mathbb{R}^d . Le théorème 17.5 implique que $f(X_n)$ converge vers $f(X)$ en probabilité également. Comme f est bornée, cette convergence a aussi lieu dans L^1 (théorème 17.4). Donc $\lim_{n \rightarrow \infty} E(f(X_n)) = E(f(X))$, et le théorème 18.1 donne le résultat. \blacksquare

Ce théorème admet une réciproque très partielle :

Théorème 18.3. Soit X_n et X des v.a. définies sur le même espace de probabilité (Ω, \mathcal{A}, P) . Si X_n converge en loi vers X et si de plus X est p.s. égale à une constante, alors X_n converge vers X en probabilité.

Preuve. Supposons que $X = a$ p.s. La fonction $f(x) = \frac{|x-a|}{1+|x-a|}$ étant bornée et continue, on a $\lim_{n \rightarrow \infty} E_f(\frac{|X_n-a|}{1+|X_n-a|}) = 0$, et donc $X_n \xrightarrow{P} X$ en vertu du théorème 17.1. ■

Il est tentant de penser que si $X_n \xrightarrow{L} X$, alors $P(X_n \in A)$ converge vers $P(X \in A)$ pour tout ensemble borélien A . Ce n'est pas vrai en général. On a $P(X_n \in A) \rightarrow P(X \in A)$ pour certains ensembles A , mais ceux-ci sont tout à fait particuliers. Cela est en rapport avec la convergence des fonctions de répartition, dans le cas de v.a. réelles : supposons que $X_n \xrightarrow{L} X$; dire que $F_n(x) = P(X_n \leq x)$ converge vers $F(x) = P(X \leq x)$ revient à dire que $F(X_n \in]-\infty, x]) \rightarrow P(X \in]-\infty, x])$, et cette propriété n'est pas toujours vraie!

Supposons donc les X_n et X à valeurs réelles, de fonctions de répartition F_n et F respectivement. Le théorème suivant est plutôt délicat et peut être omis : on remarquera que ce théorème est beaucoup plus simple dans le cas où la fonction de répartition F de la variable limite X est continue, ce qui suffit pour la plupart des applications. Nous donnons cependant la preuve dans le cas général. Pour comprendre l'énoncé, il convient de rappeler que la fonction F est croissante et continue à droite; elle admet donc partout des limites à gauche, i.e. $F(x-) = \lim_{y \rightarrow x, y < x} F(y)$ existe pour tout x (voir l'exercice 4).

Théorème 18.4. Soit X_n et X des v.a. réelles.

- (a) Si $X_n \xrightarrow{L} X$, alors $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ pour tout x dans le sous-ensemble (partout dense) de \mathbb{R} défini par $D = \{x : F(x-) = F(x)\}$. D est appelé l'ensemble des points de continuité de la fonction F .
- (b) Supposons que $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ pour tout x dans un sous-ensemble partout dense de \mathbb{R} . Alors $X_n \xrightarrow{L} X$.

Preuve de (a). Supposons que $X_n \xrightarrow{L} X$, et soit $D = \{x : F(x-) = F(x)\}$. L'ensemble D est partout dense dans \mathbb{R} puisque son complémentaire est au plus dénombrable (voir les exercices 4 et 5).

Fixons $x \in \mathbb{R}$. Pour chaque entier $p \geq 1$ on introduit les fonctions suivantes (dépendant de x), bornées et continues et vérifiant $g_p \leq 1_{]-\infty, x]} \leq f_p$:

$$f_p(y) = \begin{cases} 1 & \text{si } y \leq x \\ p(x-y) + 1 & \text{si } x < y < x + \frac{1}{p} \\ 0 & \text{si } x + \frac{1}{p} \leq y \end{cases}$$

$$g_p(y) = \begin{cases} 1 & \text{si } y \leq x - \frac{1}{p} \\ p(x-y) & \text{si } x - \frac{1}{p} < y < x \\ 0 & \text{si } x \leq y. \end{cases}$$

On a alors pour chaque p :

$$\lim_{n \rightarrow \infty} \mathbf{E}(f_p(X_n)) = \mathbf{E}(f_p(X)), \quad \lim_{n \rightarrow \infty} \mathbf{E}(g_p(X_n)) = \mathbf{E}(g_p(X)).$$

$$\mathbf{E}(q_p(X_n)) \leq F_n(x) \leq \mathbf{E}(f_p(X_n)).$$

et donc

$$\mathbf{E}(g_p(X)) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq \mathbf{E}(f_p(X)) \quad (18.1)$$

pour chaque $p \geq 1$. Mais $\lim_{p \rightarrow \infty} f_p(y) = 1_{]-\infty, x[}(y)$ et $\lim_{p \rightarrow \infty} g_p(y) = 1_{]-\infty, x[}(y)$, donc le théorème de convergence dominée de Lebesgue entraîne que

$$\begin{aligned} \lim_{p \rightarrow \infty} \mathbf{E}(f_p(X)) &= \mathbf{E}(1_{]-\infty, x[}(X)) \\ &= \mathbf{P}(X \leq x) = F(x), \end{aligned}$$

et de même $\lim_{p \rightarrow \infty} \mathbf{E}(g_p(X)) = F(x-)$ (la limite à gauche au point x). En combinant ces résultats avec (18.1) on obtient

$$F(x-) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x). \quad (18.2)$$

Si alors $x \in \mathcal{D}$ on a $F(x-) = F(x)$, donc $F_n(x) \rightarrow F(x)$.

Preuve de (b). Maintenant nous supposons que $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ pour tout $x \in \Delta$, où Δ est un sous-ensemble partout dense de \mathbb{R} . Soit f une fonction bornée et continue sur \mathbb{R} , et soit $\varepsilon > 0$. Soit aussi $r, s \in \Delta$ tels que

$$\mathbf{P}(X \notin]r, s]) = 1 - F(s) + F(r) \leq \varepsilon.$$

(On peut trouver de tels r et s car $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$, et puisque Δ est partout dense.) Comme F_n converge vers F sur Δ par hypothèse, il existe N_1 tel que pour $n \geq N_1$,

$$P(X_n \notin]r, s]) = 1 - F_n(s) + F_n(r) \leq 2\varepsilon. \quad (18.3)$$

Comme $]r, s]$ est un intervalle fermé (compact) et f est continue, elle est uniformément continue sur $]r, s]$; il existe donc un nombre fini de points $r = r_0 < r_1 < \dots < r_k = s$ appartenant à Δ (utiliser encore le fait que Δ est partout dense) et tels que

$$|f(x) - f(r_j)| \leq \varepsilon \quad \text{si} \quad r_{j-1} \leq x \leq r_j.$$

Il s'ensuit que la fonction

$$g(x) = \sum_{j=1}^k f(r_j) 1_{]r_{j-1}, r_j]}(x) \quad (18.4)$$

vérifie $|f(x) - g(x)| \leq \varepsilon$ si $x \in]r, s]$. Donc si $\alpha = \sup_x |f(x)|$, nous obtenons

$$|E(f(X_n)) - E(g(X_n))| \leq \alpha P(X_n \notin]r, s]) + \varepsilon, \quad (18.5)$$

et la même équation avec X au lieu de X_n .

D'après (18.4) on a

$$E(g(X_n)) = \sum_{j=1}^k f(r_j) \{F_n(r_j) - F_n(r_{j-1})\},$$

et

$$E(g(X)) = \sum_{j=1}^k f(r_j) \{F(r_j) - F(r_{j-1})\}.$$

Comme les r_j sont tous dans Δ on a $\lim_{n \rightarrow \infty} F_n(r_j) = F(r_j)$ pour chaque j . Les r_j étant en nombre fini, on peut trouver N_2 tel que

$$n \geq N_2 \quad \Rightarrow \quad |E(g(X_n)) - E(g(X))| \leq \varepsilon. \quad (18.6)$$

Combinons alors (18.5) pour X_n et X avec (18.6) : si $n \geq \max(N_1, N_2)$, il vient

$$\begin{aligned} & |E(f(X_n)) - E(f(X))| \\ & \leq |E(f(X_n)) - E(g(X_n))| + |E(g(X_n)) - E(g(X))| \\ & \quad + |E(g(X)) - E(f(X))| \\ & \leq (\alpha P(X_n \notin]r, s]) + \varepsilon + \varepsilon + (\alpha P(X \notin]r, s]) + \varepsilon \\ & \leq (2\alpha\varepsilon + \varepsilon) + \varepsilon + (\alpha\varepsilon + \varepsilon) \\ & \leq 3\alpha\varepsilon + 3\varepsilon. \end{aligned}$$

Comme $\varepsilon > 0$ est arbitrairement petit on a $\lim_{n \rightarrow \infty} E(f(X_n)) = E(f(X))$, d'où le résultat (voir le théorème 18.1). ■

Exemples.

1. Soit $(\mu_n)_{n \geq 1}$ une suite de probabilités sur \mathbb{R} qui sont toutes des masses de Dirac : i.e., pour chaque n il existe un point α_n tel que $\mu_n(\{\alpha_n\}) = 1$ et $\mu_n(\{\alpha_n\}^c) = \mu_n(\mathbb{R} \setminus \{\alpha_n\}) = 0$. La suite μ_n converge étroitement vers une limite μ si et seulement si α_n converge vers un réel α , et dans ce cas μ est la masse de Dirac au point α . Pour le voir, on peut bien sûr utiliser la définition 18.1 directement ; il est également instructif d'appliquer le théorème 18.4 aux fonctions de répartition $F_n(x) = 1_{[\alpha_n, \infty[}(x)$ des μ_n : la suite $F_n(x)$ converge vers une limite $F(x)$ pour tout x dans un ensemble partout dense si et seulement si α_n converge vers une limite α , et dans ce cas $F(x) = 1_{[\alpha, \infty[}(x)$. Noter que dans ce cas, l'ensemble des x pour lesquels $F_n(x) \rightarrow F(x)$ est \mathbb{R} tout entier si on a $\alpha_n \leq \alpha$ pour tout n assez grand, et est $\mathbb{R} \setminus \{\alpha\}$ sinon.
2. Soit

$$F_n(x) = \begin{cases} 0 & \text{si } x \leq -\frac{1}{n} \\ \frac{1}{2} + \frac{n}{2}x & \text{si } -\frac{1}{n} < x < \frac{1}{n} \\ 1 & \text{si } x \geq \frac{1}{n}. \end{cases}$$

Alors

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) = 1_{[0, \infty[}(x)$$

pour tout $x \neq 0$; donc l'ensemble D du théorème 18.4 est $D = \mathbb{R} \setminus \{0\}$. Ainsi des v.a. X_n de fonctions de répartition F_n vérifient $X_n \xrightarrow{\mathcal{L}} X$, où X est la v.a. constante égale à 0. On a donc montré qu'une suite de v.a. uniformément distribuées sur les intervalles $[-\frac{1}{n}, \frac{1}{n}]$ converge en loi vers 0.

3. Soit $(X_n)_{n \geq 1}$ et X des v.a. réelles admettant les densités f_n et f . La fonction de répartition

$$F(x) = \int_{-\infty}^x f(u) du$$

est continue et $F(x-) = F(x)$ pour tout x . Supposons que $f_n(x) \leq g(x)$ pour tous n et x , que $\int_{-\infty}^{\infty} g(x) dx < \infty$, et que $\lim_{n \rightarrow \infty} f_n = f$

p.p. Alors $F_n(x)$ converge vers $F(x)$ pour tout x par le théorème de convergence dominée, et $X_n \xrightarrow{L} X$.

Pour traiter cet exemple on peut aussi écrire

$$\begin{aligned} \lim_{n \rightarrow \infty} \int h(x) P_{X_n}(dx) &= \lim_{n \rightarrow \infty} \int h(x) f_n(x) dx \\ &= \int h(x) \lim_n f(x) dx \\ &= \int h(x) f(x) dx = \int h(x) P_X(dx) \end{aligned}$$

pour toute fonction bornée continue h (encore grâce au théorème de convergence dominée). Cette démonstration marche aussi lorsque les v.a. X_n et X sont à valeurs dans \mathbb{R}^d ; de plus on obtient une forme de convergence un peu plus forte que la convergence en loi, puisque la propriété $\int h dP_{X_n} \rightarrow \int h dP_X$ est vraie non seulement pour les fonctions h continues bornées, mais pour toutes les fonctions boréliennes bornées.

L'exemple précédent admet l'extension suivante, qui peut sembler un peu surprenante *a priori* : on peut échanger limite et intégrale dans un cas où les fonctions ne sont pas « dominées » par une fonction intégrable : cela est dû au fait que les densités f et f_n sont positives et d'intégrale 1.

Théorème 18.5. Soit $(X_n)_{n \geq 1}$ et X des v.a. à valeurs dans \mathbb{R}^d et admettant les densités f_n et f . Si la suite f_n converge simplement, ou même presque partout (relativement à la mesure de Lebesgue sur \mathbb{R}^d), alors $X_n \xrightarrow{L} X$.

Preuve. Soit h une fonction borélienne bornée sur \mathbb{R}^d , et soit $\alpha = \sup_x |h(x)|$. Posons $h_1(x) = h(x) + \alpha$ et $h_2(x) = \alpha - h(x)$. Ces deux fonctions sont positives, donc les fonctions $h_1 f_n$ et $h_2 f_n$ également; de plus les suites $h_i f_n$ pour $i = 1, 2$ convergent p.p. vers $h_i f$. Le lemme de Fatou (théorème 9.1) donne alors

$$\begin{aligned} E(h_i(X)) &= \int f(x) h_i(x) dx \leq \liminf_{n \rightarrow \infty} \int f_n(x) h_i(x) dx \\ &= \liminf_{n \rightarrow \infty} E(h_i(X_n)). \end{aligned} \quad (18.7)$$

Noter que $E(h(X_n)) = E(h_1(X_n)) - \alpha$ et $E(h(X_n)) = \alpha - E(h_2(X_n))$, et les mêmes égalités sont vraies avec X au lieu de X_n . Comme $\liminf(x_n) = -\limsup(-x_n)$, en appliquant (18.7) pour $i = 1$ et $i = 2$ on obtient

$$\begin{aligned} E(h(X)) &\leq \liminf_{n \rightarrow \infty} E(h(X_n)), \\ E(h(X)) &\geq \limsup_{n \rightarrow \infty} E(h(X_n)). \end{aligned}$$

Donc $E(h(X_n)) \rightarrow E(h(X))$, et le résultat est démontré. \blacksquare

Le théorème suivant est connu sous le nom de « principe de Helly ». C'est un résultat difficile, mais nous en aurons besoin pour étudier les rapports entre convergence en loi et convergence des fonctions caractéristiques. La propriété (18.8) est souvent appelée *tension de la suite de probabilités* μ_n .

Théorème 18.6. *Soit $(\mu_n)_{n \geq 1}$ une suite de probabilités sur \mathbb{R} . Supposons que*

$$\lim_{m \rightarrow \infty} \sup_n \mu_n([-m, m]^c) = 0. \quad (18.8)$$

Il existe alors une suite (n_k) d'entiers croissant vers l'infini telle que la sous-suite de probabilités $(\mu_{n_k})_{k \geq 1}$ converge étroitement.

Preuve. Soit $F_n(x) = \mu_n(]-\infty, x])$. Pour chaque x on a $0 \leq F_n(x) \leq 1$, donc la suite de réels $(F_n(x))_{n \geq 1}$ est bornée. D'après le théorème de Bolzano-Weierstrass il existe alors une suite d'entiers (n_k) (dépendant de x , bien sûr) tendant vers l'infini, telle que la suite $(F_{n_k}(x))_{k \geq 1}$ converge.

Pour montrer la convergence étroite nous devons montrer la convergence des fonctions de répartition sur un ensemble partout dense, mais qui peut être dénombrable. Nous allons donc nous restreindre à l'ensemble \mathbb{Q} des rationnels, qui est dénombrable et partout dense dans \mathbb{R} . Soit $r_1, r_2, \dots, r_j, \dots$ une énumération de tous les rationnels. Pour r_1 , il existe une sous-suite $(n_{1,k})$ de la suite de (n) de tous les entiers, telle que la limite suivante existe :

$$G(r_1) = \lim_{k \rightarrow \infty} F_{n_{1,k}}(r_1).$$

Pour r_2 , il existe une sous-suite $(n_{2,k})$ de la suite $(n_{1,k})$ telle que la limite suivante existe :

$$G(r_2) = \lim_{k \rightarrow \infty} F_{n_{2,k}}(r_2).$$

On continue de la même manière, par récurrence sur j : pour r_j on peut trouver une sous-suite $(n_{j,k})$ de la suite $(n_{j-1,k})$ telle que la limite suivante existe :

$$G(r_j) = \lim_{k \rightarrow \infty} F_{n_{j,k}}(r_j).$$

On construit alors une sous-suite unique par le « procédé diagonal », en posant $n_k = n_{k,k}$. Pour chaque j on a

$$G(r_j) = \lim_{k \rightarrow \infty} F_{n_k}(r_j),$$

puisque (n_k) est une sous-suite de la suite $(n_{j,k})$ dès que $k \geq j$.

Ensuite, on pose

$$F(x) = \inf_{\substack{y \in \mathbb{Q} \\ y > x}} G(y). \quad (18.9)$$

Comme la fonction G définie sur \mathbb{Q} est croissante, la fonction F (définie, quant à elle, sur \mathbb{R}) est aussi croissante, et elle est continue à droite par construction.

Soit $\varepsilon > 0$. Par hypothèse il existe m tel que

$$F_n([-m, m]^c) \leq \varepsilon$$

pour tous les n . Donc

$$F_n(x) \leq \varepsilon \text{ si } x < -m. \quad F_n(x) \geq 1 - \varepsilon \text{ si } x > m;$$

On a évidemment les mêmes inégalités pour $G(x)$ (pour x rationnel), donc aussi

$$\left. \begin{array}{ll} F(x) \leq \varepsilon & \text{si } x < -m \\ F(x) \geq 1 - \varepsilon & \text{si } x \geq m. \end{array} \right\} \quad (18.10)$$

Comme $0 \leq F \leq 1$ et comme F est continue à droite, la propriété (18.10) suffit à entraîner que F est la fonction de répartition d'une probabilité μ sur \mathbb{R} .

Finalement, supposons que x vérifie $F(x-) = F(x)$. Pour $\varepsilon > 0$ il existe $y, z \in \mathbb{Q}$ avec $y < x < z$

$$F(x) - \varepsilon \leq G(y) \leq F(x) \leq G(z) \leq F(x) + \varepsilon.$$

Donc pour k suffisamment grand,

$$F(x) - 2\varepsilon \leq F_{n_k}(y) \leq F_{n_k}(x) \leq F_{n_k}(z) \leq F(x) + 2\varepsilon. \quad (18.11)$$

Les inégalités (18.11) donnent

$$\begin{aligned} F(x) - 2\varepsilon &\leq F(y) \leq \liminf_{k \rightarrow \infty} F_{n_k}(x) \\ &\leq \limsup_{k \rightarrow \infty} F_{n_k}(x) \leq F(z) \leq F(x) + 2\varepsilon \end{aligned}$$

et comme $\varepsilon > 0$ est arbitrairement petit on en déduit qu'en fait $F_{n_k}(x) \rightarrow F(x)$: par suite μ_{n_k} converge étroitement vers μ par le théorème 18.4. ■

Remarque 18.1. Ce théorème admet une version multi-dimensionnelle (la preuve est analogue, en un peu plus compliqué) : si les μ_n sont des probabilités sur \mathbb{R}^d , il suffit de remplacer la condition (18.8) par

$$\lim_{m \rightarrow \infty} \sup_n \mu_n(\{x \in \mathbb{R}^d : |x| > m\}) = 0. \quad (18.12)$$

Il est parfois utile de savoir que pour vérifier que μ_n converge étroitement vers μ , il n'est pas nécessaire de montrer que $\int f d\mu_n \rightarrow \int f d\mu$ pour toutes les fonctions bornées continues f . Nous exprimons ce résultat en termes de convergence en loi de v.a. :

Théorème 18.7. Soit X_n et X des v.a. à valeurs dans \mathbb{R}^d . Pour que $X_n \xrightarrow{\mathcal{L}} X$ il faut et il suffit que $E(g(X_n)) \rightarrow E(g(X))$ pour toute fonction g bornée et lipschitzienne.

Preuve. Une fonction g est lipschitzienne s'il existe une constante k telle que $|g(x) - g(y)| \leq k\|x - y\|$ pour tous x, y . La condition nécessaire étant triviale, nous ne prouvons que la condition suffisante.

Soit f une fonction continue bornée, et $\alpha = \sup_x |f(x)|$. Supposons qu'il existe des fonctions lipschitziennes bornées g_i vérifiant $-\alpha \leq g_i \leq g_{i+1} \leq f$ et $\lim_{i \rightarrow \infty} g_i(x) = f(x)$. On a alors

$$\liminf_{n \rightarrow \infty} E(f(X_n)) \geq \lim_{i \rightarrow \infty} E(g_i(X_n)) = E(g_i(X)),$$

pour chaque i fixé. Le théorème de convergence monotone appliqué à la suite $g_i(X) + \alpha$ et à sa limite $f(X) + \alpha$ entraîne que

$$\lim_{i \rightarrow \infty} E(g_i(X)) = E(f(X)).$$

Donc

$$\liminf_{n \rightarrow \infty} E(f(X_n)) \geq E(f(X)). \quad (18.13)$$

Le même argument appliqué à la fonction $-f$ donne

$$\limsup_{n \rightarrow \infty} E(f(X_n)) \leq E(f(X)), \quad (18.14)$$

et en combinant (18.13) et (18.14) on obtient

$$\lim_{n \rightarrow \infty} E(f(X_n)) = E(f(X)),$$

ce qui montre la convergence en loi cherchée.

Il nous reste donc à construire les fonctions g_i . Pour cela, on va construire des fonctions lipschitziennes h_j telles que $\sup_k h_k(x) = f(x)$ et $h_k(x) \geq -\alpha$, et il nous suffira de poser $g_i(x) = \max\{h_1(x), \dots, h_i(x)\}$.

En remplaçant f par $\tilde{f} = f + \alpha$ si nécessaire, on peut supposer sans perte de généralité que la fonction continue et bornée f est aussi positive. Pour tout borélien A de \mathbb{R}^d définissons la fonction « distance à A » par

$$d_A(r) = \inf\{\|x - y\|; y \in A\}.$$

Pour tout rationnel $r \geq 0$ et tout entier m , posons $A_r = \{y : f(y) \leq r\}$ et

$$h_{m,r}(x) = r \wedge (m d_{A_r}(x)).$$

On a $|d_A(x) - d_A(y)| \leq \|x - y\|$, donc $|h_{m,r}(x) - h_{m,r}(y)| \leq m\|x - y\|$, et $h_{m,r}$ est lipschitzienne. Comme $0 \leq h_{m,r}(x) \leq r$ et $h_{m,r}(x) = 0$ si $f(x) \leq r$, on a $0 \leq h_{m,r}(x) \leq f(x)$.

Choisissons un point $x \in \mathbb{R}^d$ et un $\varepsilon > 0$, puis un rationnel $r > 0$ tel que $f(x) - \varepsilon < r < f(x)$. Comme f est continue, on a $f(y) > r$ pour tout y dans un voisinage de x . Donc $d_{A_r}(x) > 0$, et $h_{m,r}(x) = r > f(x) - \varepsilon$ pour m assez grand. Comme l'ensemble des rationnels est dénombrable, la famille $\{h_{m,r}; m \in \mathbb{N}, r \in \mathbb{Q}_+\}$ est dénombrable. Si $\{h_i\}_{i \geq 1}$ en constitue une énumération, on a que $\sup_i h_i(x) \geq f(x)$. Mais $h_i \leq f$ pour tout i , donc en fait $\sup_i h_i(x) = f(x)$, et la preuve est terminée. ■

Corollaire 18.1. Soit X_n et X des v.a. à valeurs dans \mathbb{R}^d . Pour que $X_n \xrightarrow{\mathcal{L}} X$ il faut et il suffit que $\mathbf{E}(g(X_n)) \rightarrow \mathbf{E}(g(X))$ pour toute fonction g bornée et uniformément continue.

Preuve. Comme toute fonction lipschitzienne est uniformément continue, il suffit d'appliquer le théorème 18.7. ■

Remarque 18.2. Dans le théorème 18.7 nous avons réduit la classe des « fonctions test » pour que des v.a. à valeurs dans \mathbb{R}^d convergent en loi. On peut se demander s'il est possible de la réduire encore. La réponse est *oui* : elle peut être réduite, par exemple, à l'ensemble des fonctions C^∞ à support compact (i.e. les fonctions indéfiniment différentiables et nulles en dehors d'un ensemble borné, ou compact) : voir les exercices 19 à 22.

Une conséquence du théorème 18.7 est :

Théorème 18.8. (Théorème de Slutsky.) Soit $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ deux suites de v.a. à valeurs dans \mathbb{R}^d . Supposons que $X_n \xrightarrow{\mathcal{L}} X$ et que $\|X_n - Y_n\| \rightarrow 0$ en probabilité. Alors $Y_n \xrightarrow{\mathcal{L}} X$.

Preuve. Grâce au théorème 18.7 il suffit de montrer que $\lim_n E(f(Y_n)) = E(f(X))$ pour toute fonction bornée lipschitzienne f . On a alors $|f(x) - f(y)| \leq k\|x - y\|$ et $|f(x)| \leq k$ pour une certaine constante k . Par suite si $\varepsilon > 0$ on a

$$\begin{aligned} |E(f(X_n) - f(Y_n))| &\leq E(|f(X_n) - f(Y_n)|) \\ &\leq k\varepsilon + E(|f(X_n) - f(Y_n)| \mathbf{1}_{\{\|X_n - Y_n\| > \varepsilon\}}) \\ &\leq k\varepsilon + 2kP(\|X_n - Y_n\| > \varepsilon). \end{aligned}$$

Mais $\lim_{n \rightarrow \infty} P(\|X_n - Y_n\| > \varepsilon) = 0$, et comme $\varepsilon > 0$ est arbitrairement petit on en déduit que $\lim_{n \rightarrow \infty} |E(f(X_n) - f(Y_n))| = 0$. Par suite

$$\lim_{n \rightarrow \infty} E(f(Y_n)) = \lim_{n \rightarrow \infty} E(f(X_n)) = E(f(X)),$$

et le théorème est prouvé. ■

Terminons ce chapitre par l'étude de la convergence en loi des v.a. discrètes (ne prenant qu'un nombre fini ou dénombrable de valeurs), comme les v.a. binomiales, de Poisson, hypergéométriques, etc. Comme l'espace d'état, disons E , est au plus dénombrable, on peut considérer que toute fonction sur E est continue : cela revient à munir E de la topologie discrète. (*Attention* : si l'espace E est contenu de manière naturelle dans \mathbb{R} , la topologie usuelle de \mathbb{R} induit sur E la topologie discrète si et seulement si le minimum de $|x - y|$ pour $x, y \in E \cap [-m, m]$ est strictement positif pour tout $m > 0$; c'est le cas par exemple si $E = \mathbb{N}$ ou $E = \mathbb{Z}$, ou si E est une suite de réels tendant vers l'infini; ce n'est pas vrai si E est une suite de réels tendant vers une limite finie.) Le théorème suivant donne une caractérisation simple de la convergence en loi dans le cas discret, et peut se comparer au théorème 18.5.

Théorème 18.9. Soit X_n et X des v.a. à valeurs dans un espace E fini ou dénombrable. Alors $X_n \xrightarrow{L} X$ si et seulement si on a

$$\lim_{n \rightarrow \infty} P(X_n = j) = P(X = j) \quad (18.15)$$

pour tout $j \in E$.

Preuve. Supposons d'abord que $X_n \xrightarrow{L} X$. On a

$$\lim_{n \rightarrow \infty} E(f(X_n)) = E(f(X))$$

pour toute fonction continue bornée f sur E . Chaque fonction indicatrice $\mathbf{1}_{\{j\}}$ étant continue et bornée, on a (18.15).

Supposons inversement (18.15). Soit f une fonction bornée et $\alpha = \sup_{j \in E} |f(j)|$. Soit $\varepsilon > 0$. Comme

$$\sum_{j \in E} \mathbb{P}(X = j) = 1$$

est une série convergente, il existe un sous-ensemble fini Λ de E tel que

$$\sum_{j \in \Lambda} \mathbb{P}(X = j) \geq 1 - \varepsilon.$$

(18.15) entraîne alors que pour n assez grand,

$$\sum_{j \in \Lambda} \mathbb{P}(X_n = j) \geq 1 - 2\varepsilon.$$

On a

$$\mathbb{E}(f(X)) = \sum_{j \in E} f(j) \mathbb{P}(X = j),$$

donc pour n assez grand il vient

$$\left. \begin{aligned} \left| \mathbb{E}(f(X)) - \sum_{j \in \Lambda} f(j) \mathbb{P}(X = j) \right| &\leq \alpha \varepsilon \\ \left| \mathbb{E}(f(X_n)) - \sum_{j \in \Lambda} f(j) \mathbb{P}(X_n = j) \right| &\leq 2\alpha \varepsilon. \end{aligned} \right\} \quad (18.16)$$

Finalement on observe que, comme Λ est fini,

$$\lim_{n \rightarrow \infty} \sum_{j \in \Lambda} f(j) \mathbb{P}(X_n = j) = \sum_{j \in \Lambda} f(j) \mathbb{P}(X = j). \quad (18.17)$$

On déduit alors de (18.16) et de (18.17) que

$$\limsup_{n \rightarrow \infty} |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq 3\alpha \varepsilon.$$

Comme $\varepsilon > 0$ est arbitrairement petit, il vient

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$$

et on en déduit que $X_n \xrightarrow{\mathcal{L}} X$ ■

Exemples (suite).

4. Si μ_λ désigne la loi de Poisson de paramètre λ , on a

$$\mu_\lambda(j) = e^{-\lambda} \frac{\lambda^j}{j!},$$

et donc si $\lambda_n \rightarrow \lambda$, on a $\mu_{\lambda_n}(j) \rightarrow \mu_\lambda(j)$ pour tout $j \in \mathbb{N}$. Par suite μ_{λ_n} converge étroitement vers μ_λ .

5. Si μ_p désigne la loi binomiale de paramètres (n, p) et si $p_k \rightarrow p$, exactement comme dans l'exemple 4 on vérifie que μ_{p_k} converge étroitement vers μ_p .
6. Soit $\mu_{n,p}$ la loi binomiale (n, p) . Considérons la suite μ_{n,p_n} où $\lim_{n \rightarrow \infty} np_n = \lambda > 0$. En reprenant l'exercice 1 du chapitre 4, on vérifie que la suite μ_{n,p_n} converge étroitement vers la loi de Poisson de paramètre λ .

Exercices

1. Montrer que si $X_n \xrightarrow{L^p} X$ ($p \geq 1$), alors $X_n \xrightarrow{\mathcal{L}} X$.
2. Soit $\alpha \in \mathbb{R}^d$ et $\varepsilon > 0$. Construire une fonction continue $f : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $0 \leq f \leq 1$, que $f(\alpha) = 0$, et que $f(x) = 1$ si $|x - \alpha| \geq \varepsilon$. (*Indication* : Résoudre d'abord l'exercice quand $d = 1$, puis généraliser au cas $d \geq 2$.)
3. Soit X une v.a. réelle de fonction de répartition F . Montrer que $F(x-) = F(x)$ si et seulement si $P(X = x) = 0$.
- 4*. Soit $g : \mathbb{R} \rightarrow [0, 1]$ croissante et continue à droite. Montrer que g a des limites à gauche en chaque point, et que l'ensemble $\Lambda = \{x : g(x-) \neq g(x)\}$ est au plus dénombrable (*Indication* : Montrer d'abord qu'il n'y a qu'un nombre fini (éventuellement nul) de points tels que $g(x) - g(x-) > \frac{1}{k}$).
- 5*. Soit F une fonction de répartition, et $D = \{x : F(x-) = F(x)\}$. Montrer que D est partout dense dans \mathbb{R} . (*Indication* : Utiliser l'exercice 4.)
6. Soit X_n une v.a. uniformément distribuée sur $[-n, n]$. Dans quel(s) sens la suite X_n converge-t-elle ? [*Rép.* : Aucun.]
7. Soit f_n des densités de probabilité sur \mathbb{R} , et supposons que $\lim_{n \rightarrow \infty} f_n(x) = e^{-x} 1_{(x>0)}$. Si les X_n sont des v.a. de densités f_n , que peut-on dire de la convergence des X_n ? [*Rép.* : $X_n \xrightarrow{\mathcal{L}} X$, où X est exponentielle de paramètre 1.]

8. Soit $(X_n)_{n \geq 1}$ des v.a. indépendantes de loi de Cauchy de paramètres $\alpha = 0$ et $\beta = 1$. Soit $Y_n = \frac{X_1 + \dots + X_n}{n}$. Montrer que Y_n converge en loi et trouver la limite. Est-ce que les Y_n convergent en probabilité ?
9. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles, telle que $\sup_n E((X^n)^2) < \infty$, et μ_n la loi de X_n . Montrer que la suite (μ_n) est tendue (i.e. vérifie (18.8)). (Indication : utiliser l'inégalité de Bienaymé-Chebyshev.)
- 10.* Soit X_n, X et Y des v.a. réelles, toutes définies sur le même espace de probabilité (Ω, \mathcal{A}, P) . Supposons que

$$\lim_{n \rightarrow \infty} E(f(X_n)g(Y)) = E(f(X)g(Y))$$

pour toutes fonctions bornées f et g , avec f continue et g borélienne. Montrer que la suite bidimensionnelle (X_n, Y) converge en loi vers (X, Y) . Si de plus $X = h(Y)$ pour une fonction borélienne h , montrer que $X_n \xrightarrow{P} X$.

11. Soit μ_{α} la loi de Pareto de paramètre α . Si $\alpha_n \rightarrow \alpha > 0$, montrer que μ_{α_n} converge étroitement vers μ_{α} .
12. Soit μ_{α} la loi géométrique de paramètre α . Si $\alpha_n \rightarrow \alpha > 0$, montrer que μ_{α_n} converge étroitement vers μ_{α} .
13. Soit $\mu_{(N,b,n)}$ une loi hypergéométrique. On fixe n et on fait tendre N et b vers l'infini, de sorte que $p = \frac{b}{N}$ reste constant. Montrer que $\mu_{(N,b,n)}$ converge étroitement vers la loi binomiale de paramètres (n, p) . [Comparer avec l'exercice 8 du chapitre 5.]
14. (Théorème de Slutsky.) Soit X_n des v.a. convergeant en loi vers X , et Y_n des v.a. convergeant en probabilité vers une constante c , toutes ces v.a. étant définies sur le même espace. Montrer (a) que $X_n Y_n \xrightarrow{L} cX$, et (b) que $\frac{X_n}{Y_n} \xrightarrow{L} \frac{X}{c}$ si $c \neq 0$.
15. Soit $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ des v.a. réelles définies sur le même espace. Si $X_n \xrightarrow{L} X$ et si $Y_n \xrightarrow{P} 0$, montrer que $X_n + Y_n \xrightarrow{L} X$.
16. Soit $(X_n)_{n \geq 1}$ des v.a. réelles de fonctions de répartition F_n , avec $X_n \xrightarrow{L} X$. Soit $p > 0$. Montrer que pour tout $N > 0$ on a

$$\int_{-N}^N |x|^p F(dx) \leq \limsup_{n \rightarrow \infty} \int_{-N}^N |x|^p F_n(dx) < \infty.$$

- 17.* Soit $(X_n)_{n \geq 1}$ et X des v.a. réelles de fonctions de répartition F_n et F vérifiant pour un $r > 0$:

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} |F_n(x) - F(x)|^r dx = 0.$$

Montrer que $X_n \xrightarrow{\mathcal{L}} X$. (*Indication* : Supposer qu'il existe un point de continuité y de F tel que $\lim_{n \rightarrow \infty} F_n(y) \neq F(y)$. Il existe alors $\varepsilon > 0$ et une sous-suite $(n_k)_{k \geq 1}$ tels que $|F_{n_k}(y) - F(y)| > \varepsilon$ pour tout k . Montrer qu'alors $|F_{n_k}(x) - F(x)| > \frac{\varepsilon}{2}$ pour $x \in]y_1, y[$ ou pour $x \in]y, y_2]$, pour des réels y_1, y_2 adéquats. Utiliser ceci pour obtenir une contradiction.)

- 18.* Soit $(F_n)_{n \geq 1}$ une suite de fonctions de répartition convergeant simplement vers une fonction de répartition F continue. Montrer que la convergence de F_n vers F est uniforme sur \mathbb{R} . (*Indication* : Commencer par montrer qu'il existe des points $x_1 < \dots < x_m$ tels que $F(x_1) < \varepsilon$, $F(x_{j+1}) - F(x_j) < \varepsilon$, et $1 - F(x_m) < \varepsilon$. Puis montrer qu'il existe N tel que si $n > N$ on ait $|F_n(x_j) - F(x_j)| < \varepsilon$, $1 \leq j \leq m$.)
19. Soit f une fonction uniformément continue sur \mathbb{R} , et X et Y deux v.a. réelles. Supposons que $|f(x) - f(y)| < \varepsilon$ si $|x - y| < \delta$. Montrer que

$$|E(f(X)) - E(f(X + Y))| \leq \varepsilon + 2 \sup_x |f(x)| P(|Y| \geq \delta).$$

- 20.* (Pollard [20].) Soit $(X_n)_{n \geq 1}$, X et Y des v.a. réelles, toutes définies sur le même espace, et supposons que $X_n + \sigma Y$ converge en loi vers $X + \sigma Y$ pour tout $\sigma > 0$. Montrer que X_n converge en loi vers X . (*Indication* : Utiliser l'exercice 19.)
21. (Pollard [20].) Soit X et Y des v.a. réelles indépendantes, avec Y de loi $N(0, 1)$. Soit f une fonction continue bornée. Montrer que

$$E(f(X + \sigma Y)) = E(f_\sigma(X))$$

où

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} f(z) e^{-\frac{1}{2}|z-x|^2/\sigma^2} dz.$$

Montrer que f_σ est bornée et \mathcal{C}^∞ .

22. Soit $(X_n)_{n \geq 1}$ et X des v.a. réelles. Montrer que X_n converge en loi vers X si et seulement si $E(f(X_n)) \rightarrow E(f(X))$ pour toute f bornée et \mathcal{C}^∞ . (*Indication* : Utiliser les exercices 20 et 21.)

Chapitre 19

Convergence en loi et fonctions caractéristiques

La convergence en loi est d'une certaine manière le cœur des probabilités et de la statistique; on va voir qu'il existe des rapports très étroits entre convergence en loi et convergence des fonctions caractéristiques, ce qui est l'une des raisons pour lesquelles les fonctions caractéristiques sont aussi utiles.

Théorème 19.1. (Théorème de continuité de P. Lévy.) Soit $(\mu_n)_{n \geq 1}$ une suite de probabilités sur \mathbb{R}^d , et $(\hat{\mu}_n)_{n \geq 1}$ la suite de leurs transformées de Fourier (ou fonctions caractéristiques).

- (a) Si μ_n converge étroitement vers une probabilité μ , alors $\hat{\mu}_n$ converge simplement vers la transformée de Fourier $\hat{\mu}$ de μ .
- (b) Si $\hat{\mu}_n(u)$ converge simplement vers une fonction f , et si de plus cette fonction est continue au point 0, alors il existe une probabilité μ sur \mathbb{R}^d dont la transformée de Fourier est $\hat{\mu} = f$, et μ_n converge étroitement vers μ .

Preuve. (a) Supposons que μ_n converge étroitement vers μ . Comme la fonction $x \mapsto e^{ux}$ est continue et bornée,

$$\hat{\mu}_n(u) = \int e^{ux} \mu_n(dx)$$

converge vers

$$\hat{\mu}(u) = \int e^{ux} \mu(dx)$$

(cette fonction est à valeurs complexes, mais on peut prendre séparément les parties réelle et imaginaire.)

(b) Bien que le théorème soit énoncé pour \mathbb{R}^d nous allons le démontrer pour \mathbb{R} seulement. Supposons que $\lim_{n \rightarrow \infty} \hat{\mu}_n(u) = f(u)$ existe pour tout u . On va montrer tout d'abord la *tension* (voir p. 164) de la suite μ_n . D'après le théorème de Fubini (théorème 10.3, et plus précisément exercice 10 du chapitre 14) on a

$$\begin{aligned} \int_{-\alpha}^{\alpha} \hat{\mu}_n(u) du &= \int_{-\alpha}^{\alpha} \left\{ \int_{-\infty}^{\infty} e^{iux} \mu_n(dx) \right\} du \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\alpha}^{\alpha} e^{iux} du \right\} \mu_n(dx); \end{aligned}$$

et comme $e^{iux} = \cos ux + i \sin ux$,

$$= \int_{-\infty}^{\infty} \left\{ \int_{-\alpha}^{\alpha} (\cos ux + i \sin ux) du \right\} \mu_n(dx).$$

$\sin ux$ étant une fonction impaire, l'intégrale de la partie imaginaire sur l'intervalle symétrique $[-\alpha, \alpha]$ est nulle, donc

$$= \int_{-\infty}^{\infty} \frac{2}{x} \sin \alpha x \mu_n(dx).$$

Comme $\int_{-\alpha}^{\alpha} 1 du = 2\alpha$, il vient

$$\begin{aligned} \frac{1}{\alpha} \int_{-\alpha}^{\alpha} (1 - \hat{\mu}_n(u)) du &= 2 - \int_{-\infty}^{\infty} \frac{2}{\alpha x} \sin \alpha x \mu_n(dx) \\ &= 2 \int_{-\infty}^{\infty} \left(1 - \frac{\sin \alpha x}{\alpha x} \right) \mu_n(dx). \end{aligned}$$

Comme $2(1 - \frac{\sin v}{v})$ est plus grand que 1 si $|v| \geq 2$ et toujours positif, l'expression ci-dessus est

$$\begin{aligned} &\geq \int_{-\infty}^{\infty} 1_{[-2, 2]^c}(\alpha x) \mu_n(dx) \\ &= \int 1_{[-2/\alpha, 2/\alpha]^c}(x) \mu_n(dx) \\ &= \mu_n \left(\left[\frac{-2}{\alpha}, \frac{2}{\alpha} \right]^c \right). \end{aligned}$$

Soit $\beta = \frac{2}{\alpha}$. On obtient alors l'estimée suivante :

$$\mu_n([- \beta, \beta]^c) \leq \frac{\beta}{2} \int_{-2/\beta}^{2/\beta} (1 - \hat{\mu}_n(u)) du. \quad (19.1)$$

Soit $\varepsilon > 0$. Comme par hypothèse f est continue en 0, il existe $\alpha > 0$ tel que $|1 - f(u)| \leq \varepsilon/4$ si $|u| \leq 2/\alpha$ (car $\hat{\mu}_n(0) = 1$ pour tout n , donc aussi $f(0) = 1$). Donc

$$\left| \frac{\alpha}{2} \int_{-2/\alpha}^{2/\alpha} (1 - f(u)) du \right| \leq \frac{\alpha}{2} \int_{-2/\alpha}^{2/\alpha} \frac{\varepsilon}{4} du = \frac{\varepsilon}{2}. \quad (19.2)$$

Comme les $\widehat{\mu}_n(u)$ sont des fonctions caractéristiques, elles vérifient $|\widehat{\mu}_n(u)| \leq 1$ et le théorème de convergence dominée implique que

$$\lim_{n \rightarrow \infty} \int_{-2/\alpha}^{2/\alpha} (1 - \widehat{\mu}_n(u)) du = \int_{-2/\alpha}^{2/\alpha} (1 - f(u)) du.$$

Il existe donc N tel que pour tout $n \geq N$ on ait

$$\left| \int_{-2/\alpha}^{2/\alpha} (1 - \widehat{\mu}_n(u)) du - \int_{-2/\alpha}^{2/\alpha} (1 - f(u)) du \right| \leq \frac{\varepsilon}{\alpha}.$$

d'où $\frac{\alpha}{2} \int_{-2/\alpha}^{2/\alpha} (1 - \widehat{\mu}_n(u)) du \leq \varepsilon$. En utilisant enfin (19.1) on arrive à $\mu_n([-a, a]^c) \leq \varepsilon$ pour tout $n \geq N$.

Il n'y a qu'un nombre fini d'entiers $n < N$, et pour chacun d'entre eux il existe un α_n tel que $\mu_n([-a_n, a_n]^c) \leq \varepsilon$. Il reste à poser $a = \max(\alpha_1, \dots, \alpha_N; \alpha)$ pour obtenir

$$\mu_n([-a, a]^c) \leq \varepsilon, \quad \text{pour tout } n.$$

Comme $\varepsilon > 0$ est arbitraire, on a donc montré que les μ_n vérifient (18.8).

On peut maintenant appliquer le théorème 18.6 de façon à obtenir une sous-suite $(n_k)_{k \geq 1}$ telle que μ_{n_k} converge étroitement vers une certaine probabilité limite μ . La partie (a) entraîne alors que

$$\lim_{k \rightarrow \infty} \widehat{\mu}_{n_k}(u) = \widehat{\mu}(u)$$

pour tout u , donc $f = \widehat{\mu}$ et f est la transformée de Fourier d'une probabilité.

Il reste à montrer que la suite $(\mu_n)_{n \geq 1}$ elle-même (et pas seulement une sous-suite) converge étroitement vers μ . Notons F_n et F les fonctions de répartition de μ_n et de μ . Soit aussi D l'ensemble des points de continuité de F , i.e.

$$D = \{x : F(x-) = F(x)\}.$$

Supposons alors que μ_n ne converge pas étroitement vers μ . D'après le théorème 18.4 il existe au moins un point $x \in D$ et une sous-suite $(n_k)_{k \geq 1}$ tels que $\beta = \lim_{k \rightarrow \infty} F_{n_k}(x)$ existe et soit différente de $F(x)$. D'après le théorème 18.6 il existe une sous-suite de la sous-suite (n_k) , soit $(n_{k_j})_{j \geq 1}$, telle que $(\mu_{n_{k_j}})_{j \geq 1}$ converge étroitement vers une limite ν . Exactement comme plus haut, on obtient que

$$\lim_{j \rightarrow \infty} \widehat{\mu}_{n_{k_j}}(u) = \widehat{\nu}(u),$$

et comme $\lim_n \hat{\mu}_n = f$ on en déduit que $\hat{\nu} = f$. Comme on a vu que $f = \hat{\mu}$, en vertu du théorème d'unicité 14.1 on a $\mu = \nu$. Mais alors le théorème 18.4 entraîne que $\beta = \lim_{j \rightarrow \infty} F_{n_k_j}(x) = F(x)$, puisque x appartient à l'ensemble des points de continuité de μ , ce qui fournit une contradiction. ■

Remarque 19.1. On pourrait en fait montrer plus : si les μ_n convergent étroitement vers μ , alors les fonctions $\hat{\mu}_n$ convergent vers $\hat{\mu}$ *uniformément sur tout compact de \mathbb{R}^d* .

Exemple. Soit $(X_n)_{n \geq 1}$ une suite de v.a. de Poisson de paramètres $\lambda_n = n$. Alors, si $Z_n = \frac{1}{\sqrt{n}}(X_n - n)$, on a

$$Z_n \xrightarrow{\mathcal{L}} Z, \quad \text{où } Z \text{ est de loi } N(0, 1).$$

Pour le voir, on remarque que

$$\begin{aligned} \mathbf{E}(e^{iuZ_n}) &= \mathbf{E}\left(e^{iu\left(\frac{1}{\sqrt{n}}(X_n - n)\right)}\right) \\ &= e^{-iu\sqrt{n}} \mathbf{E}\left(e^{i\frac{u}{\sqrt{n}}X_n}\right) \\ &= e^{-iu\sqrt{n}} e^{n(e^{i u/\sqrt{n}} - 1)} \end{aligned}$$

(cf. exemple 13.3).

En faisant un développement de Taylor pour e^z , on voit que

$$\begin{aligned} \mathbf{E}(e^{iuZ_n}) &= e^{-iu\sqrt{n}} e^{n\left(i\frac{u}{\sqrt{n}} - \frac{u^2}{2n} - \frac{i u^3}{6n^{3/2}} + \dots\right)} \\ &= e^{-iu\sqrt{n} + iu\sqrt{n} - u^2/2} e^{-\frac{h(u,n)}{\sqrt{n}}} \\ &= e^{-u^2/2} e^{-\frac{h(u,n)}{\sqrt{n}}} \end{aligned}$$

où $h(u, n)$ reste borné en n pour tout u et donc $\lim_{n \rightarrow \infty} \frac{h(u,n)}{\sqrt{n}} = 0$. Par suite

$$\lim_{n \rightarrow \infty} \varphi_{Z_n}(u) = e^{-u^2/2},$$

et comme $e^{-u^2/2}$ est la fonction caractéristique de la loi $N(0, 1)$ (exemple 13.5), on obtient notre assertion en appliquant le théorème 19.1(b).

Exercices

1. Soit $(X_n)_{n \geq 1}$ des v.a. de lois $N(\mu_n, \sigma_n^2)$. Montrer que si $\mu_n \rightarrow \mu \in \mathbb{R}$ et $\sigma_n^2 \rightarrow \sigma^2 \geq 0$, alors $X_n \xrightarrow{\mathcal{L}} X$, où X est de loi $N(\mu, \sigma^2)$.
2. Soit $(X_n)_{n \geq 1}$ des v.a. de loi $N(\mu_n, \sigma_n^2)$, et supposons que $X_n \xrightarrow{\mathcal{L}} X$. Montrer que les suites μ_n et σ_n^2 ont des limites $\mu \in \mathbb{R}$ et $\sigma^2 \geq 0$, et que X est $N(\mu, \sigma^2)$ (*Indication* : φ_{X_n} et φ_X étant les fonctions caractéristiques, écrire $\varphi_{X_n} = e^{iu\mu - \frac{u^2\sigma_n^2}{2}}$ et utiliser le théorème de Lévy pour obtenir que $\varphi_X(u) = e^{iu\mu - \frac{u^2\sigma^2}{2}}$ pour un $\mu \in \mathbb{R}$ et un $\sigma^2 \geq 0$.)
3. Soit $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ des suites de v.a. à valeurs dans \mathbb{R}^d , définies sur le même espace de probabilité. On suppose d'une part que X_n et Y_n sont indépendantes pour chaque n et que X est indépendante de Y , d'autre part que $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$. Montrer que $X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$.

Chapitre 20

La loi des grands nombres

Nous présentons dans ce chapitre l'un des résultats fondamentaux de la théorie des probabilités, celui qui en particulier « justifie » l'approche des probabilités par les fréquences, ou dans un tout autre domaine l'utilisation des méthodes de simulation de v.a. (méthodes de Monte-Carlo) pour l'intégration numérique (voir l'exemple 2).

Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles, définies sur le même espace de probabilité, et posons $S_n = \sum_{j=1}^n X_j$. Une assertion qui affirme que la suite $\frac{1}{n}S_n$, c'est-à-dire la suite des moyennes des n premières variables X_j , converge en un certain sens vers une limite s'appelle une *loi des grands nombres*. Il existe de nombreux résultats de ce type, par exemple les « théorèmes ergodiques » dans L^2 ou le théorème ergodique ponctuel, qui sont des formes de la loi des grands nombres (cf. le théorème 20.3 ci-dessous). La convergence peut être en probabilité, dans L^p , ou presque sûre ; dans ce dernier cas, on parle de *loi forte des grands nombres*.

Théorème 20.1. (Loi forte des grands nombres.) Soit $(X_n)_{n \geq 1}$ des v.a. indépendantes et de même loi (« i.i.d. »), de carré intégrable, et posons

$$\mu = E(X_j), \quad \text{et} \quad \sigma^2 = \sigma_{X_j}^2 < \infty.$$

Soit $S_n = \sum_{j=1}^n X_j$. On a alors

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j = \mu \quad \text{p.s. et dans } L^2.$$

Remarque 20.1. On écrit μ et σ^2 puisque, les v.a. X_n ayant même loi, elles ont mêmes moments. On déduit de ce qui précède la convergence $\frac{S_n}{n} \rightarrow \mu$ en probabilité, puisque les convergences dans L^2 et p.s. impliquent chacune la convergence en probabilité. En fait la convergence en probabilité est très facile à obtenir dans ce contexte (en utilisant l'inégalité de Bienaymé-Chebyshev), et cette propriété est appelée la loi faible des grands nombres. La preuve du théorème 20.1 est plus simple lorsqu'on suppose que les X_n sont dans L^3 . À l'inverse, un résultat plus fort, ne nécessitant que l'existence de la moyenne μ , est énoncé au théorème 20.2 et prouvé dans le chapitre 27.

Preuve du théorème 20.1. On peut supposer sans perdre de généralité que $\mu = E(X_j) = 0$: sinon, en effet, on peut remplacer X_j par $Z_j = X_j - \mu$ et obtenir $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n Z_j = 0$, soit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (X_j - \mu) = \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{j=1}^n X_j \right) - \mu = 0,$$

ce qui donne le résultat.

On suppose donc dans la suite que $\mu = 0$. Rappelons que $S_n = \sum_{j=1}^n X_j$ et posons $Y_n = \frac{S_n}{n}$. On a $E(Y_n) = \frac{1}{n} \sum_{j=1}^n E(X_j) = 0$. De plus $E(Y_n^2) = \frac{1}{n^2} \sum_{1 \leq j, k \leq n} E(X_j X_k)$. Cependant si $j \neq k$ on a

$$E(X_j X_k) = E(X_j)E(X_k) = 0$$

puisque X_j et X_k sont indépendants, donc

$$\begin{aligned} E(Y_n^2) &= \frac{1}{n^2} \sum_{j=1}^n E(X_j^2) & (20.1) \\ &= \frac{1}{n^2} \sum_{j=1}^n \sigma^2 = \frac{1}{n^2} (n\sigma^2) \\ &= \frac{\sigma^2}{n} \end{aligned}$$

et par suite $\lim E(Y_n^2) = 0$, ce qui montre la convergence de Y_n vers 0 dans L^2 .

La convergence dans L^2 impliquant la convergence en probabilité, on sait qu'on peut extraire de la suite (Y_n) une sous-suite qui converge p.s. vers 0. Cependant, cela ne suffit pas : nous voulons que *la suite* Y_n elle-même converge p.s. Pour le montrer nous exhibons d'abord une sous-suite particulière qui converge p.s., puis nous traitons les termes qui se trouvent entre deux éléments successifs de la sous-suite.

On a $E(Y_n^2) = \frac{\sigma^2}{n}$, donc

$$\sum_{n=1}^{\infty} E(Y_{n^2}^2) = \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} < \infty;$$

par suite, d'après le théorème 9.2 on a $\sum_{n=1}^{\infty} Y_{n^2}^2 < \infty$ p.s., donc le terme général de cette série est p.s. convergent vers 0 :

$$\lim_{n \rightarrow \infty} Y_{n^2} = 0 \quad \text{p.s.} \quad (20.2)$$

Soit ensuite $n \in \mathbb{N}$, et $p(n)$ l'unique entier tel que

$$p(n)^2 \leq n < (p(n) + 1)^2.$$

On a

$$Y_n - \frac{p(n)^2}{n} Y_{p(n)^2} = \frac{1}{n} \sum_{j=p(n)^2+1}^n X_j,$$

et comme pour (20.1) nous obtenons

$$\begin{aligned} \mathbb{E} \left(\left(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2} \right)^2 \right) &= \frac{n - p(n)^2}{n^2} \sigma^2 \\ &\leq \frac{2p(n) + 1}{n^2} \sigma^2 \\ &\leq \frac{2\sqrt{n} + 1}{n^2} \sigma^2 \leq \frac{3}{n^{\frac{3}{2}}} \sigma^2 \end{aligned}$$

en utilisant le fait que $p(n) \leq \sqrt{n}$.

Appliquons maintenant le même argument que précédemment : on a

$$\sum_{n=1}^{\infty} \mathbb{E} \left(\left(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2} \right)^2 \right) \leq \sum_{n=1}^{\infty} \frac{3\sigma^2}{n^{\frac{3}{2}}} < \infty,$$

de sorte que

$$\sum_{n=1}^{\infty} \left(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2} \right)^2 < \infty \quad \text{p.s.,}$$

et a fortiori

$$\lim_{n \rightarrow \infty} \left\{ Y_n - \frac{p(n)^2}{n} Y_{p(n)^2} \right\} = 0 \quad \text{p.s.}$$

Mais comme $\lim_{n \rightarrow \infty} Y_{p(n)^2} = 0$ p.s. par (20.2) et $\frac{p(n)^2}{n} \rightarrow 1$, on déduit que $\lim_{n \rightarrow \infty} Y_n = 0$ p.s. également, et le théorème est démontré. ■

Nous donnons ci-dessous deux autres versions de la loi des grands nombres.

Théorème 20.2. (Loi des grands nombres de Kolmogorov.) Soit (X_j) des v.a. réelles i.i.d., et soit $\mu \in \mathbb{R}$. Posons $S_n = \sum_{j=1}^n X_j$. La suite $\frac{S_n}{n}$ converge p.s. vers μ si et seulement si $\mathbb{E}(X_j) = \mu$ (et si en particulier les X_j admettent une moyenne). Dans ce cas, la convergence a aussi lieu dans L^1 .

Remarque 20.2. Ce théorème nous donne les hypothèses minimales assurant la validité de la loi des grands nombres pour les sommes de v.a. indépendantes, à savoir que les X_j soient dans L^1 . Une manière élégante de prouver le théorème 20.2 consiste à utiliser le théorème de convergence des martingales inverses (théorème 27.5).

Soit (Ω, \mathcal{A}, P) un espace de probabilité, et $T : \Omega \rightarrow \Omega$ une application injective mesurable (de (Ω, \mathcal{A}) dans lui-même). On suppose aussi que T préserve la mesure, i.e. $P(T^{-1}(A)) = P(A)$ (ou si on veut : P est sa propre image par T). On pose $T^2(\omega) = T(T(\omega))$, et on définit de manière analogue les puissances successives de T . Une partie A de Ω est dite *invariante par T* si $T^{-1}(A) = A$, ou de manière équivalente si $1_A = 1_A \circ T$.

Théorème 20.3. (Loi des grands nombres ergodique.) *Soit T une transformation injective préservant la mesure et telle que les ensembles mesurables invariants soient de probabilité 0 ou 1. Si X est une v.a. intégrable, alors*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X(T^j(\omega)) = E(X)$$

p.s. et dans L^1 .

Ce théorème est une conséquence du « théorème ergodique ponctuel », ou de Birkhoff. Il remplace l'hypothèse d'indépendance par l'hypothèse d'ergodicité (l'ergodicité est le fait que les ensembles invariants soient de probabilité 0 ou 1 : noter que les v.a. $X \circ T^j$ ne sont pas indépendantes, mais elles ont toutes même loi, puisque T préserve la mesure). Il s'appelle aussi *loi forte des grands nombres pour les suites stationnaires*.

Exemple 1. Dans l'exemple 17.1 on considérait une suite $(X_j)_{j \geq 1}$ de v.a. indépendantes de loi de Bernoulli, avec $P(X_j = 1) = p$ et $P(X_j = 0) = q = 1 - p$. Alors $S_n = \sum_{j=1}^n X_j$ est le nombre de « succès » en n tirages, et S_n/n est la fréquence des succès. La loi forte des grands nombres nous dit que

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = E(X_1) = p \quad \text{p.s.} \quad (20.3)$$

Cela fournit donc la justification souhaitée à l'approche des probabilités par les fréquences : les axiomes des probabilités, justifiés de manière intuitive par le résultat postulé (20.3), permettent en fait de montrer que ce résultat est mathématiquement correct.

Exemple 2. Nous présentons ci-dessous un exemple très simple d'une technique connue sous le nom de *méthode d'approximation de Monte-Carlo*. Soit f une fonction borélienne sur $[0, 1]$, intégrable par rapport à la mesure de Lebesgue. Il est souvent impossible de trouver une expression analytique pour l'intégrale $I = \int_0^1 f(x)dx$. Pour calculer le nombre I on peut opérer ainsi : on prend une suite de v.a. $(U_j)_{j \geq 1}$ indépendantes et uniformément distribuées sur l'intervalle $[0, 1]$, et on pose $I_n = \frac{1}{n} \sum_{j=1}^n f(U_j)$; il suit du théorème 20.2 que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n f(U_j) = E(f(U_1)) = \int_0^1 f(x)dx,$$

p.s. et dans L^1 . Si on simule sur un ordinateur la suite $(U_j)_{j \geq 1}$ en utilisant un générateur de nombres au hasard, on obtient ainsi une approximation I_n de l'intégrale I .

Bien entendu, il ne s'agit là que d'une méthode parmi beaucoup pour calculer $\int_0^1 f(x)dx$, et certainement pas la meilleure ni la plus rapide pour une intégrale sur $[0, 1]$. On verra par exemple, à l'occasion du théorème-limite central, que la « vitesse de convergence » de cette méthode est en $1/\sqrt{n}$, alors qu'une approximation de $\int_0^1 f(x)dx$ par la somme de Riemann $\frac{1}{n} \sum_{i=1}^n f(i/n)$ induit une erreur de l'ordre de $1/n$ dès que f est lipschitzienne, et que la « méthode des trapèzes » avec un pas $1/n$ conduit à une erreur de l'ordre de $1/n^2$ si f est deux fois dérivable, etc. En revanche si f est simplement borélienne la méthode de Monte-Carlo devient compétitive : et si on remplace $[0, 1]$ par, disons, un cube de l'espace \mathbb{E}^d , cette même méthode marche encore, toujours avec la vitesse $1/\sqrt{n}$, alors que les méthodes d'approximation analytiques conduisent à une vitesse dépendant de la dimension d et sont en fait impossibles à mettre en œuvre pratiquement dès que d dépasse 4 ou 5.

Exemple 3. ([10, p. 120]) Soit Ω le cercle de rayon $r = \frac{1}{2\pi}$. Soit \mathcal{A} la tribu borélienne sur ce cercle, et P la mesure de Lebesgue (de ce point de vue, le cercle n'est autre que le segment $[0, 1[$!). Soit α un nombre irrationnel et T la rotation sur Ω d'angle α . On vérifie aisément que T est injective, préserve la mesure, et que les boréliens du cercle invariants par T sont de probabilité 0 ou 1 (c'est là que l'irrationalité de α intervient). Donc le théorème 20.3 implique que si X est une fonction intégrable sur Ω ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X(x + j\alpha) = \int_0^1 X(y)dy$$

pour P-presque tout x .

Exercices

1.* (Une loi faible des grands nombres.) Soit (X_j) une suite de v.a. réelles telle que $\sup_j E(X_j^2) = c < \infty$ et $E(X_j X_k) = 0$ si $j \neq k$. Soit $S_n = \sum_{j=1}^n X_j$. Montrer que

a) $P(|\frac{1}{n}S_n| \geq \varepsilon) \leq \frac{c}{n\varepsilon^2}$ pour tout $\varepsilon > 0$;

b) $\lim_{n \rightarrow \infty} \frac{1}{n}S_n = 0$ dans L^2 et en probabilité.

(Commentaire : L'hypothèse usuelle d'indépendance a été considérablement affaiblie ici.)

2. Soit $(Y_j)_{j \geq 1}$ une suite de v.a. indépendantes binomiales de paramètres $(1, p)$, sur le même espace de probabilité. Soit $X_n = \sum_{j=1}^n Y_j$. Montrer que X_j est binomiale de paramètre j, p et que $\frac{X_j}{j}$ converge p.s. vers p .

3. Soit $(X_j)_{j \geq 1}$ des variables i.i.d. réelles intégrables, et $Y_j = e^{X_j}$. Montrer que

$$\left(\prod_{j=1}^n Y_j \right)^{\frac{1}{n}}$$

converge p.s. vers une constante α . [Rép. : $\alpha = e^{E(X_1)}$.]

4. Soit $(X_j)_{j \geq 1}$ des variables i.i.d. réelles intégrables de moyenne μ . Soit $(Y_j)_{j \geq 1}$ également i.i.d. réelles intégrables, de moyenne ν . Montrer que

$$\lim_{n \rightarrow \infty} \frac{1}{\sum_{j=1}^n Y_j} \sum_{j=1}^n X_j = \frac{\mu}{\nu} \quad \text{p.s.}$$

5. Soit $(X_j)_{j \geq 1}$ des variables i.i.d. réelles intégrables, et supposons que pour une certaine constante ν la suite $\frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \nu)$ converge en loi vers une limite Z . Montrer que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j = \nu \quad \text{p.s.}$$

(Indication : Si $Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \nu)$, montrer d'abord que $\frac{1}{\sqrt{n}}Z_n$ converge en loi vers 0.)

6. Soit $(X_j)_{j \geq 1}$ des variables i.i.d. réelles appartenant à L^p pour un $p \in [1, \infty[$. Montrer que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j^p = E(X^p) \quad \text{p.s.}$$

7. Soit $(X_j)_{j \geq 1}$ des v.a. indépendantes $N(1, 3)$. Montrer que

$$\lim_{n \rightarrow \infty} \frac{X_1 + X_2 + \cdots + X_n}{X_1^2 + X_2^2 + \cdots + X_n^2} = \frac{1}{4} \quad \text{p.s.}$$

8. Soit $(X_j)_{j \geq 1}$ des variables i.i.d. réelles intégrables de moyennes μ et de variance σ^2 . Montrer que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 = \sigma^2 \quad \text{p.s.}$$

9. Soit $(X_j)_{j \geq 1}$ des variables i.i.d. à valeurs entières et intégrables. Posons $S_n = \sum_{j=1}^n X_j$. Montrer que si $E(X_j) > 0$, alors

$$\lim_{n \rightarrow \infty} S_n = \infty \quad \text{p.s.}$$

Chapitre 21

Le théorème-limite central

Le théorème-limite central est l'un des résultats les plus impressionnants des probabilités. À partir d'hypothèses très faibles, il donne des résultats extrêmement précis. Il joue aussi un rôle de premier plan en statistique. L'idée générale est simple. Soit X_1, \dots, X_j, \dots des variables aléatoires i.i.d. de variance finie σ^2 et de moyenne notée μ , et soit $S_n = \sum_{j=1}^n X_j$. Alors, si n est grand, $\mathcal{L}(S_n) \approx N(n\mu, n\sigma^2)$. Le point clé dans ce genre de résultat est que, à part l'existence d'une variance, *absolument aucune hypothèse n'est requise*. Donc, si une certaine v.a. est en fait la somme de « beaucoup » de variables indépendantes de variances finies, alors sa loi est approximativement normale; en statistique par exemple, il est alors possible d'utiliser des procédures permettant d'estimer μ et σ^2 , et on connaît alors essentiellement tout!

Théorème 21.1. (Théorème-limite central.) Soit $(X_j)_{j \geq 1}$ une suite i.i.d. de variables aléatoires réelles avec $E(X_j) = \mu$ et $\text{Var}(X_j) = \sigma^2 \in]0, \infty[$. Soit

$$S_n = \sum_{j=1}^n X_j \quad \text{et} \quad Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Alors, les v.a. Y_n convergent en loi vers une variable $N(0, 1)$.

Noter que si $\sigma^2 = 0$, alors $X_j = \mu$ p.s. pour tout j , donc $\frac{S_n}{n} = \mu$ p.s.

Preuve. Soit φ la fonction caractéristique des v.a. $X_j - \mu$ (qui ont toutes même loi par hypothèse), et $Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Comme les X_j sont indépendantes, le théorème 15.2 entraîne que

$$\begin{aligned} \varphi_{Y_n}(u) &= \varphi_{\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (X_j - \mu)}(u) & (21.1) \\ &= \varphi_{\sum_{j=1}^n (X_j - \mu)}\left(\frac{u}{\sigma\sqrt{n}}\right) \\ &= \prod_{j=1}^n \varphi\left(\frac{u}{\sigma\sqrt{n}}\right) \\ &= \left(\varphi\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n. \end{aligned}$$

Ensuite, le fait que $E(X_j - \mu) = 0$ et $E((X_j - \mu)^2) = \sigma^2 < \infty$ entraîne par le théorème 13.2 que φ est deux fois continûment dérivable et que

$$\begin{aligned}\varphi'(u) &= iE\left((X_j - \mu)e^{iu(X_j - \mu)}\right), \\ \varphi''(u) &= -E\left((X_j - \mu)^2 e^{iu(X_j - \mu)}\right).\end{aligned}$$

Donc $\varphi'(0) = 0$ et $\varphi''(0) = -\sigma^2$. Si on fait le développement de Taylor de la fonction φ à l'ordre 2 au voisinage du point 0, on obtient

$$\varphi(u) = 1 + 0 - \frac{\sigma^2 u^2}{2} + u^2 h(u) \quad (21.2)$$

où $h(u) \rightarrow 0$ quand $u \rightarrow 0$. Rappelons que d'après (21.1),

$$\begin{aligned}\varphi_{Y_n}(u) &= \left(\varphi\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n \\ &= e^{n \log \varphi\left(\frac{u}{\sigma\sqrt{n}}\right)} \\ &= e^{n \log\left(1 - \frac{u^2}{2\sigma^2} + \frac{u^2}{n\sigma^2} h\left(\frac{u}{\sigma\sqrt{n}}\right)\right)},\end{aligned}$$

où « log » désigne la valeur principale du logarithme complexe (valant 0 au point 1 et continu dans le cercle complexe de centre 1 et de rayon 1/2). En prenant les limites quand $n \rightarrow \infty$ on voit que

$$\lim_{n \rightarrow \infty} \varphi_{Y_n}(u) = e^{-u^2/2}.$$

Le théorème de Lévy 19.1 implique alors que Y_n converge en loi vers Z , où $\varphi_Z(u) = e^{-u^2/2}$; d'après l'exemple 13.5 on sait que Z suit la loi $N(0, 1)$ (rappelons que la fonction caractéristique caractérise la loi). ■

Examinons à présent les rapports entre la loi des grands nombres et le théorème-limite central. Soit $(X_j)_{j \geq 1}$, S_n , μ et σ^2 comme dans le théorème précédent. La loi des grands nombres nous dit que

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu \quad \text{p.s. et dans } L^2, \quad (21.3)$$

La limite de S_n/n est donc μ , mais la question suivante se pose alors naturellement : à quelle vitesse cette convergence a-t-elle lieu ? Réécrivons (21.3) ainsi :

$$\lim_{n \rightarrow \infty} \left| \frac{S_n}{n} - \mu \right| = 0 \quad \text{p.s. et dans } L^2. \quad (21.4)$$

On peut alors se poser la question de savoir s'il existe un réel $\alpha > 0$ tel que, pour une certaine constante $c \neq 0$,

$$\lim_{n \rightarrow \infty} n^\alpha \left| \frac{S_n}{n} - \mu \right| = c \quad \text{p.s.}$$

Il n'existe en fait aucun α ayant cette propriété; les v.a. $n^\alpha(\frac{S_n}{n} - \mu)$ ne peuvent converger vers une constante non nulle ou vers une variable aléatoire non p.s. nulle, ni presque sûrement, ni même en probabilité. Cependant le théorème-limite central nous dit que si on prend $\alpha = \frac{1}{2}$, alors $\sqrt{n}(\frac{S_n}{n} - \mu)$ converge en loi vers une v.a. $N(0, \sigma^2)$. En ce sens, la vitesse de convergence pour la loi des grands nombres est \sqrt{n} .

On peut légèrement affaiblir les hypothèses du théorème 21.1; la même démonstration permet de démontrer le

Théorème 21.2. (Théorème-limite central.) *Soit $(X_j)_{j \geq 1}$ une suite de v.a. réelles indépendantes, mais pas nécessairement de même loi. On suppose qu'elles sont centrées : $E(X_j) = 0$, et on pose $\sigma_j^2 = \sigma_{X_j}^2$. Supposons aussi*

$$\sup_j E(|X_j|^{2+\varepsilon}) < \infty, \quad \sum_{j=1}^{\infty} \sigma_j^2 = \infty$$

pour un $\varepsilon > 0$. Si alors $S_n = X_1 + \dots + X_n$, la suite $\frac{S_n}{\sqrt{\sum_{j=1}^n \sigma_j^2}}$ converge en loi vers une v.a. $N(0, 1)$.

Le théorème 21.1 est le théorème-limite central « classique », tandis que le théorème 21.2 montre qu'il est possible d'affaiblir les hypothèses tout en obtenant le même résultat. Il existe de nombreuses versions du théorème-limite central, toutes semblables par le fait qu'elles exhibent des conditions suffisantes pour qu'une somme de v.a. convenablement normalisée converge en loi vers une variable normale. Par exemple, la théorie des martingales permet d'affaiblir notablement les hypothèses du théorème 21.2 en introduisant une certaine dépendance entre les variables X_j : voir le théorème 27.7. L'extension à d'autres formes de dépendance, connues sous le nom de « conditions de mélange » est possible mais plus difficile.

Finalement, il est important pour les applications de souligner que le théorème 21.1 admet une version d -dimensionnelle, qui est démontrée exactement de la même manière :

Théorème 21.3. (Théorème-limite central.) Soit $(X_j)_{j \geq 1}$ des variables aléatoires i.i.d. de vecteur moyen $\mu = \mathbf{E}(X_j)$ (i.e., μ_k est l'espérance de la k -ième composante X_j^k des X_j) et de matrice de covariance $Q = (q_{k,l})_{1 \leq k, l \leq d}$ (donc $q_{k,l} = \text{Cov}(X_j^k, X_j^l)$). Alors si $S_n = X_1 + \dots + X_n$, les variables aléatoires $\frac{S_n - n\mu}{\sqrt{n}}$ convergent en loi vers une variable Z (à valeurs dans \mathbb{R}^d) de loi $N(0, Q)$.

Dans ce théorème nous n'avons fait aucune hypothèse de non dégénérescence de la matrice de covariance Q : la limite Z est gaussienne, mais peut ne pas avoir de densité (rappelons que la loi $N(0, Q)$ admet une densité dans \mathbb{R}^d si et seulement si la matrice Q est inversible).

Exemples.

1. Soit $(X_j)_{j \geq 1}$ des v.a. indépendantes avec $P(X_j = 1) = p$ et $P(X_j = 0) = q = 1 - p$ (variables de Bernoulli). La v.a. $S_n = \sum_{j=1}^n X_j$ est binomiale de paramètres (n, p) . On a $\mu = \mathbf{E}(X_j) = p$ et $\sigma^2 = \sigma_{X_j}^2 = pq = p(1 - p)$. La loi des grands nombres dit que

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = p \quad \text{p.s.}$$

et le théorème-limite central dit que

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} Z, \quad \text{où } Z \text{ est } N(0, 1).$$

2. Soit $(X_j)_{j \geq 1}$ une suite i.i.d. de variables aléatoires de carré intégrable et de fonction de répartition F . On suppose que cette fonction F est inconnue, et on souhaite « l'estimer ». Nous décrivons ci-dessous une technique standard pour résoudre ce problème. Posons

$$Y_j(x) = 1_{\{X_j \leq x\}}.$$

Les v.a. $Y_j(x)$ sont aussi i.i.d. et de carré intégrable : ce sont en fait des variables de Bernoulli de paramètre $p = F(x)$. Posons ensuite

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n Y_j(x).$$

Les $F_n(x)$ sont des variables aléatoires pour tout x fixé ; pour ω fixé, la fonction $x \mapsto F_n(x) = F_n(x)(\omega)$ est clairement une fonction de répartition (celle de la probabilité $\frac{1}{n} \sum_{j=1}^n \varepsilon_{X_j(\omega)}$), appelée

fonction de répartition empirique. La loi des grands nombres entraîne la convergence p.s. et dans L^2 de $F_n(x) = \frac{1}{n} \sum_{j=1}^n Y_j(x)$ vers $E(Y_1(x)) = F(x)$:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \text{p.s. et dans } L^2.$$

Ainsi, les *fonctions de répartition empiriques* convergent vers la vraie fonction de répartition, pour chaque x fixé, au sens p.s. et L^2 . Avec un petit travail supplémentaire on peut montrer la convergence uniforme :

$$\lim_{n \rightarrow \infty} \sup_x |F_n(x) - F(x)| = 0 \quad \text{p.s. et dans } L^2.$$

Ce résultat est connu sous le nom de *théorème de Glivenko-Cantelli*. En utilisant le théorème-limite central on obtient de plus que la vitesse de convergence est \sqrt{n} : on a en effet

$$\begin{aligned} \sqrt{n}(F_n(x) - F(x)) &= \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^n Y_j(x) - E(Y_1(x)) \right) \\ &= \frac{\sum_{j=1}^n Y_j(x) - nE(Y_1(x))}{\sqrt{n}}. \end{aligned}$$

qui d'après le théorème 21.1 converge en loi vers une variable $N(0, \sigma^2(x))$, où $\sigma^2(x) = \text{Var}(Y_1(x)) = F(x)(1 - F(x))$.

Ainsi qu'il a été dit plus haut, le théorème-limite central peut être interprété comme donnant la vitesse de convergence dans la loi des grands nombres. On peut aller un cran plus loin, en se posant le problème de la « vitesse de convergence » dans le théorème-limite central lui-même : à quelle vitesse la loi de $Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ approche-t-elle la loi $N(0, 1)$? Le théorème suivant est l'une des réponses « classiques » à ce problème, lorsqu'on prend comme mesure de la « distance » entre deux lois sur \mathbb{R} le supremum de la différence des fonctions de répartition.

Théorème 21.4. (Théorème de Berry-Esseen.) *Soit $(X_j)_{j \geq 1}$ des variables aléatoires réelles i.i.d., vérifiant $E(|X_j|^3) < \infty$ et $\sigma^2 = \sigma_{X_j}^2 > 0$. Soit $G_n(x) = P(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x)$, où $\mu = E(X_j)$ et $S_n = X_1 + \dots + X_n$. Soit enfin Φ la fonction de répartition de la loi $N(0, 1)$. On a alors*

$$\sup_x |G_n(x) - \Phi(x)| \leq c \frac{E(|X_1|^3)}{\sigma^3 \sqrt{n}}$$

pour une constante c .

La démonstration de ce théorème est un peu difficile pour ce livre, et nous l'omettons. Le lecteur peut consulter, par exemple [11, p. 108], où il est prouvé qu'on peut prendre $c = 3$. La meilleure approximation de c connue à ce jour est $c = 0.7975$.

Exercices

1. Soit $(X_j)_{j \geq 1}$ des variables indépendantes telles que $P(X_j = 1) = P(X_j = 0) = \frac{1}{2}$, et $S_n = \sum_{j=1}^n X_j$. La variable $Z_n = 2S_n - n$ représente le surplus de « faces » par rapport aux « piles » après n jets d'une pièce, si « Face » correspond à $X_j = 1$. Montrer que

$$\lim_{n \rightarrow \infty} P\left(\frac{Z_n}{\sqrt{n}} < x\right) = \Phi(x)$$

où

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du.$$

2. Soit $(X_j)_{j \geq 1}$ des v.a. indépendantes de loi exponentielle double de paramètre 1 (la densité est $\frac{1}{2}e^{-|x|}$). Montrer que $\sqrt{n} \left(\frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_j^2} \right)$ converge en loi vers une variable $N(0, \frac{1}{2})$. (*Indication* : Utiliser le théorème de Slutsky (exercice 14 du chapitre 18).)
3. Construire une suite $(X_j)_{j \geq 1}$ de v.a. réelles indépendantes de carré intégrable, telle que $\lim_{j \rightarrow \infty} X_j = 1$ en probabilité et $E(X_j^2) \neq j$. Soit Y une v.a. de loi $N(0, 1)$, indépendantes de la suite (X_j) . On pose $Z_j = YX_j$. Montrer que

a) $E(Z_j) = 0$.

b) $\lim_{j \rightarrow \infty} \sigma_{Z_j}^2 = \infty$

c) Z_j converge en loi vers une v.a. $N(0, 1)$.

(*Indication* : Pour construire X_j prendre $(\Omega, \mathcal{A}, P) = ([0, 1], \mathcal{B}[0, 1], m(ds))$ où m est la mesure de Lebesgue ; poser

$$X_j(\omega) = (j+1)1_{[0, 1/j]}(\omega) + 1_{(1/j, 1]}(\omega).$$

Pour (c) utiliser le théorème de Slutsky.) [Noter que les hypothèses du théorème-limite central présenté dans ce livre ne sont pas satisfaites ici : ces hypothèses sont suffisantes, mais pas nécessaires, pour avoir une limite normale.]

4. (Durrett, [11].) Soit $(X_j)_{j \geq 1}$ des v.a. réelles i.i.d. avec $E(X_1) = 1$ et $\sigma_{X_1}^2 = \sigma^2 \in]0, \infty[$. Montrer que

$$\frac{2}{\sigma}(\sqrt{S_n} - \sqrt{n}) \xrightarrow{\mathcal{L}} Z,$$

où Z est $N(0, 1)$. (*Indication* : $\frac{S_n - n}{\sqrt{n}} = \frac{(\sqrt{S_n + \sqrt{n}})}{\sqrt{n}}(\sqrt{S_n} - \sqrt{n})$.)

5. Soit (X_j) des variables indépendantes de loi de Poisson de paramètre $\lambda = 1$ et $S_n = \sum_{j=1}^n X_j$. Montrer que $\lim_{n \rightarrow \infty} \frac{S_n - n}{\sqrt{n}} = Z$, où Z est $N(0, 1)$.
6. Soit Y^λ une v.a. de Poisson de paramètre $\lambda > 0$. Montrer que $\frac{Y^\lambda - \lambda}{\sqrt{\lambda}}$ converge en loi vers une v.a. $N(0, 1)$ quand $\lambda \rightarrow \infty$. (*Indication* : Utiliser l'exercice 5 et comparer Y^λ avec $S_{[\lambda]}$ et $S_{[\lambda]+1}$, où $[\lambda]$ désigne la « partie entière » de λ (= le plus grand entier $n \leq \lambda$)).
7. Montrer que

$$\lim_{n \rightarrow \infty} e^{-\alpha} \left(\sum_{k=0}^n \frac{n^k}{k!} \right) = \frac{1}{2}.$$

(*Indication* : utiliser l'exercice 5.)

8. Soit $(X_j)_{j \geq 1}$ des v.a. réelles i.i.d. avec $E(X_1) = 0$ et $\sigma_{X_1}^2 = \sigma^2 \in]0, \infty[$, et $S_n = \sum_{i=1}^n X_i$. Montrer que $\frac{S_n}{\sqrt{n}}$ ne converge pas en probabilité.
- 9* Soit $(X_j)_{j \geq 1}$ des v.a. réelles i.i.d. avec $E(X_1) = 0$ et $\sigma_{X_1}^2 = \sigma^2 \in]0, \infty[$, et $S_n = \sum_{i=1}^n X_i$. Montrer que

$$\lim_{n \rightarrow \infty} E \left(\frac{|S_n|}{\sqrt{n}} \right) = \sqrt{\frac{2}{\pi}} \sigma.$$

(*Indication* : Calculer $E(|Z|)$ pour une v.a. Z de loi $N(0, \sigma^2)$.)

10. (Gut, [14].) Soit $(X_j)_{j \geq 1}$ des v.a. indépendantes uniformes sur $[-1, 1]$, et

$$Y_n = \frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_j^2 + \sum_{j=1}^n X_j^3}.$$

Montrer que $\sqrt{n}Y_n$ converge. (*Rép.* : $\sqrt{n}Y_n$ converge en loi vers une v.a. $N(0, 3)$.)

11. Soit $(X_j)_{j \geq 1}$ des v.a. indépendantes, chaque X_j étant uniforme sur $[-j, j]$, et $S_n = \sum_{i=1}^n X_i$.

- a) Montrer que $\frac{S_n}{n^{3/2}}$ converge en loi vers une v.a. $N(0, \frac{1}{9})$. (*Indication* : Montrer que la fonction caractéristique de X_j est $\varphi_{X_j}(u) = \frac{\sin u_j}{u_j}$; calculer $\varphi_{S_n}(u)$, puis $\varphi_{S_n/n^{3/2}}(u)$, et prouver que la limite est $e^{-u^2/18}$ en utilisant $\sum_{j=1}^n j^2 = \frac{n(n+1)(2n+1)}{6}$).
- b) Montrer que $\frac{S_n}{\sqrt{\sum_{j=1}^n \sigma_{X_j}^2}}$ converge en loi vers une v.a. $N(0, 1)$.
 [*Commentaire* : Ceci n'est pas un cas particulier du théorème 21.2].

12* Soit $X \in L^2$ et supposons que les v.a. X et $\frac{1}{\sqrt{2}}(Y+Z)$ ont même loi, dès que Y et Z sont indépendantes et de même loi que X . Montrer que X suit une loi normale. (*Indication* : Montrer par récurrence que X a même loi que $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ si les (X_i) sont i.i.d. de même loi que X , pour tout $n = 2^m$.)

Chapitre 22

L^2 et les espaces de Hilbert

Soit un espace de probabilité (Ω, \mathcal{F}, P) . Rappelons que L^2 désigne l'ensemble des (classes d'équivalence pour la relation « égalité presque sûre » de) v.a. réelles de carré intégrable. Dans la suite on identifie de manière systématique une variable aléatoire et sa classe d'équivalence : ainsi, au lieu d'écrire $X = 0$ p.s. on écrit simplement $X = 0$.

L'espace L^2 est un espace vectoriel (cf. théorème 9.3), et on peut le munir d'un « produit scalaire » de la manière suivante : pour X et Y dans L^2 on pose

$$\langle X, Y \rangle = E(XY). \quad (22.1)$$

On a $|E(XY)| \leq E(X^2)^{\frac{1}{2}} E(Y^2)^{\frac{1}{2}} < \infty$ par l'inégalité de Cauchy-Schwarz. et la formule (22.1) a un sens. On a aussi $\langle X, Y \rangle = \langle Y, X \rangle$ (symétrie), et comme l'espérance est linéaire, il vient

$$\langle \alpha X + \beta Y, Z \rangle = \alpha \langle X, Z \rangle + \beta \langle Y, Z \rangle.$$

Finalement, on voit que

$$\langle X, X \rangle \geq 0, \quad \text{et } \langle X, X \rangle = 0 \text{ si et seulement si } X = 0$$

compte tenu de notre convention d'identifier une variable p.s. nulle avec 0. Cela conduit à poser pour $X \in L^2$:

$$\|X\| = \langle X, X \rangle^{\frac{1}{2}} = (E(X^2))^{\frac{1}{2}}. \quad (22.2)$$

Donc $\|X\| = 0$ si et seulement si $X = 0$; de plus la bilinéarité du produit scalaire, vue ci-dessus, et l'inégalité de Cauchy-Schwarz conduisent à

$$\begin{aligned} \|X + Y\|^2 &= E(X^2) + 2E(XY) + E(Y^2) \\ &\leq \|X\|^2 + 2\|X\| \|Y\| + \|Y\|^2 \\ &= (\|X\| + \|Y\|)^2, \end{aligned}$$

et on obtient l'inégalité de Minkowski :

$$\|X + Y\| \leq \|X\| + \|Y\|.$$

On a donc ainsi défini une norme, et en fait montré le :

Théorème 22.1. L^2 est un espace vectoriel normé, la norme étant associée à un produit scalaire, ce qui signifie que $\|\cdot\| = \langle \cdot, \cdot \rangle^{\frac{1}{2}}$.

La convergence d'une suite de v.a. pour cette norme est clairement la « convergence dans L^2 » définie au chapitre 17. On peut faire mieux que dans le précédent théorème, en montrant que l'espace L^2 est, de plus, *complet* : cela signifie que toute suite de Cauchy dans L^2 est une suite convergente (une suite de Cauchy (X_n) est une suite telle que $\|X_n - X_m\| \rightarrow 0$ quand m et n tendent tous deux vers l'infini ; rappelons aussi que toute suite convergente est une suite de Cauchy). Le théorème ci-dessous est parfois connu sous le nom de théorème de *Riesz-Fischer*.

Théorème 22.2. L^2 est complet pour la norme (22.2).

Preuve. Soit X_n une suite de Cauchy dans L^2 . Pour tout $\varepsilon > 0$ il existe N tel que $n, m \geq N$ entraîne $\|X_n - X_m\| \leq \varepsilon$. Choisissons une suite de epsilons de la forme $\frac{1}{2^n}$: il existe alors une sous-suite $(X_{n_k})_{k \geq 1}$ telle que $\|X_{n_k} - X_{n_{k+1}}\| \leq \frac{1}{2^k}$.

Posons

$$Y_n = \sum_{\nu=1}^n |X_{n_\nu} - X_{n_{\nu+1}}|.$$

L'inégalité triangulaire entraîne

$$E(Y_n^2) \leq \left(\sum_{p=1}^n \|X_{n_p} - X_{n_{p+1}}\| \right)^2 \leq 1.$$

La limite $Y(\omega) = \lim_{n \rightarrow \infty} Y_n(\omega)$ existe pour tout ω , puisque la suite Y_n est croissante. Comme $E(Y_n^2) \leq 1$ pour chaque n , d'après le théorème de limite monotone 9.1(d) il vient aussi $E(Y^2) \leq 1$. Donc $Y < \infty$ p.s., donc la série $X_{n_1} + \sum_{p=1}^{\infty} (X_{n_{p+1}} - X_{n_p})$ converge absolument pour presque tout ω . Comme c'est une série télescopique, on en déduit que $X_{n_p}(\omega)$ converge vers une limite $X(\omega)$ quand $p \rightarrow \infty$, et de plus $|X(\omega)| \leq |X_{n_1}(\omega)| + Y(\omega)$, tout ceci pour presque tout ω . Comme X_{n_1} et Y sont dans L^2 , on a donc aussi $X \in L^2$.

Ensuite observons que, presque sûrement,

$$X - X_{n_p} = \lim_{m \rightarrow \infty} Z_m^p, \quad \text{où } Z_m^p = \sum_{q=p}^m (X_{n_{q+1}} - X_{n_q}).$$

Comme $|Z_m^p| \leq Y$ pour tous p, m , le théorème de convergence dominée implique

$$\|X - X_{n_p}\| = \lim_{m \rightarrow \infty} \|Z_{n_p}^m\| \leq \lim_m \sum_{q=p}^m \|X_{n_{q+1}} - X_{n_q}\| \leq \frac{1}{2^{p-1}},$$

et par suite $\lim_{p \rightarrow \infty} \|X - X_{n_p}\| = 0$. Donc X_{n_p} converge vers X dans L^2 .

Finalement

$$\|X_n - X\| \leq \|X_n - X_{n_p}\| + \|X_{n_p} - X\|.$$

Donc en faisant tendre n et p vers l'infini on en déduit que $X_n \rightarrow X$ dans L^2 . ■

Définition 22.1. Un espace de Hilbert \mathcal{H} est un espace vectoriel normé complet, dont la norme provient d'un produit scalaire : $\|x\| = \langle x, x \rangle^{\frac{1}{2}}$.

On a donc établi le

Théorème 22.3. L^2 est un espace de Hilbert.

Ci-dessous nous décrivons une série de propriétés des espaces de Hilbert ; chacune s'applique évidemment à l'espace L^2 . Ci-dessous, \mathcal{H} désigne un espace de Hilbert de norme $\|\cdot\|$ et de produit scalaire $\langle \cdot, \cdot \rangle$, tandis que α et β désignent toujours des nombres réels. La convergence des vecteurs de \mathcal{H} est toujours au sens de la norme.

Définition 22.2. Deux vecteurs x et y de \mathcal{H} sont orthogonaux si $\langle x, y \rangle = 0$. Un vecteur x est orthogonal à une partie Γ de \mathcal{H} si $\langle x, y \rangle = 0$ pour tout $y \in \Gamma$.

Si $\langle x, y \rangle = 0$ on a $\|x + y\|^2 = \|x\|^2 + \|y\|^2$: c'est la version « Hilbert » du théorème de Pythagore.

Théorème 22.4. (Continuité du produit scalaire.) Si $x_n \rightarrow x$ et $y_n \rightarrow y$ dans \mathcal{H} , alors $\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$ dans \mathbb{R} (et donc aussi $\|x_n\| \rightarrow \|x\|$).

Preuve. L'inégalité de Cauchy-Schwarz, qui est valide dans tout espace de Hilbert (même démonstration que pour l'espace L^2) entraîne $|\langle x, y \rangle| \leq \|x\| \|y\|$, donc

$$\begin{aligned} |\langle x, y \rangle - \langle x_n, y_n \rangle| &= |\langle x - x_n, y_n \rangle + \langle x - x_n, y - y_n \rangle + \langle x_n, y - y_n \rangle| \\ &\leq \|x - x_n\| \|y_n\| + \|x - x_n\| \|y - y_n\| + \|x_n\| \|y - y_n\|. \end{aligned}$$

On a $\sup_n \|y_n\| < \infty$ et $\sup_n \|x_n\| < \infty$ puisque x_n et y_n convergent (par exemple, $\|x_n\| \leq \|x_n - x\| + \|x\|$ et $\|x\| < \infty$ et $\|x_n - x\| \rightarrow 0$). Donc le second membre de l'inégalité ci-dessus tend vers 0 quand $n \rightarrow \infty$. ■

Définition 22.3. Une partie \mathcal{L} de \mathcal{H} est appelée un sous-espace, ou « sous-espace vectoriel », si $x, y \in \mathcal{L}$ implique $\alpha x + \beta y \in \mathcal{L}$. On dit en outre que le sous-espace \mathcal{L} est fermé si pour toute suite (x_n) de vecteurs de \mathcal{L} convergeant vers une limite x au sens de \mathcal{H} , on a $x \in \mathcal{L}$.

Théorème 22.5. Soit Γ une partie de \mathcal{H} . L'ensemble Γ^\perp de tous les vecteurs orthogonaux à Γ est un sous-espace fermé de \mathcal{H} .

Preuve. D'abord, Γ^\perp est un sous-espace, même si Γ n'en n'est pas un : en effet si $x, y \in \Gamma^\perp$, alors $\langle x, z \rangle = 0$ et $\langle y, z \rangle = 0$ pour tout $z \in \Gamma$; donc

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle = 0,$$

et $\alpha x + \beta y \in \Gamma^\perp$. Le fait que Γ^\perp soit fermé découle du théorème 22.4. ■

Définition 22.4. Si $\Gamma \subset \mathcal{H}$ on note $d(x, \Gamma) = \inf\{\|x - y\|; y \in \Gamma\}$ la distance de x à Γ .

Si $x \in \Gamma$, on a bien sûr $d(x, \Gamma) = 0$. Inversement si $d(x, \Gamma) = 0$ et si Γ est fermé, alors $x \in \Gamma$.

Théorème 22.6. Soit \mathcal{L} un sous-espace fermé de \mathcal{H} et $x \in \mathcal{H}$. Il existe un unique vecteur $y \in \mathcal{L}$ tel que $\|x - y\| = d(x, \mathcal{L})$.

Preuve. Si $x \in \mathcal{L}$ on a évidemment $y = x$. Supposons donc que $x \notin \mathcal{L}$. On peut trouver une suite $y_n \in \mathcal{L}$ telle que $\lim_{n \rightarrow \infty} \|x - y_n\| = d(x, \mathcal{L})$. On va montrer que cette suite est de Cauchy. Observons d'abord que

$$\|y_n - y_m\|^2 = \|x - y_m\|^2 + \|x - y_n\|^2 - 2\langle x - y_m, x - y_n \rangle. \quad (22.3)$$

Si on utilise l'inégalité

$$\left\| x - \frac{y_m + y_n}{2} \right\| \leq \frac{\|x - y_m\|}{2} + \frac{\|x - y_n\|}{2}$$

on obtient que

$$\lim_{n, m \rightarrow \infty} \left\| x - \frac{y_m + y_n}{2} \right\| \leq d(x, \mathcal{L}).$$

Donc

$$\lim_{n, m \rightarrow \infty} \left\| x - \frac{y_m + y_n}{2} \right\| = d(x, \mathcal{L}),$$

puisque $\frac{y_m + y_n}{2} \in \mathcal{L}$ (car \mathcal{L} est un sous-espace). On a donc

$$\begin{aligned} d(x, \mathcal{L})^2 &= \lim_{m, n \rightarrow \infty} \left\| x - \frac{y_m + y_n}{2} \right\|^2 \\ &= \lim_{m, n \rightarrow \infty} \frac{1}{4} \{ \|x - y_m\|^2 + \|x - y_n\|^2 + 2\langle x - y_m, x - y_n \rangle \}, \end{aligned}$$

et donc

$$\lim_{n, m \rightarrow \infty} \langle x - y_m, x - y_n \rangle = d(x, \mathcal{L})^2. \quad (22.4)$$

Il reste à combiner (22.3) et (22.4) pour obtenir que la suite $(y_n)_{n \geq 1}$ est de Cauchy. Cette suite converge donc vers une limite y , qui appartient à \mathcal{L} puisque ce sous-espace est fermé. De plus $d(x, \mathcal{L}) = \|x - y\|$ à cause de la continuité (évidente) de la fonction distance.

Il reste à montrer l'unicité de y . Supposons que z soit un autre vecteur vérifiant les mêmes propriétés. La suite

$$w_{2n} = y, \quad w_{2n+1} = z,$$

est encore une suite de Cauchy par le même argument que ci-dessus, donc elle converge vers une limite *unique*, donc $z = y$. ■

Nous introduisons maintenant un concept important, celui de *projection (orthogonale)*. L'espace de Hilbert \mathcal{H} est toujours fixé, ainsi que le sous-espace fermé \mathcal{L} . La *projection d'un vecteur x sur \mathcal{L}* est le point y obtenu dans le théorème 22.6, c'est-à-dire le point de \mathcal{L} qui est « le plus proche de x ». On note Π l'opérateur de projection, i.e. $\Pi x = y$ ci-dessus. Noter que Π dépend de manière essentielle du sous-espace \mathcal{L} , et le résultat ci-dessous explique son nom de *projecteur orthogonal sur \mathcal{L}* .

Théorème 22.7. *L'opérateur de projection (ou projection orthogonale) Π de \mathcal{H} sur un sous-espace fermé \mathcal{L} vérifie les trois propriétés suivantes :*

- (i) Π est idempotent : i.e., $\Pi^2 = \Pi$;
- (ii) $\Pi x = x$ pour $x \in \mathcal{L}$; $\Pi x = 0$ pour $x \in \mathcal{L}^\perp$;
- (iii) Pour tout $x \in \mathcal{H}$, le vecteur $x - \Pi x$ est orthogonal à \mathcal{L} .

Preuve. (i) suit immédiatement du fait que $\Pi x = x$ si $x \in \mathcal{L}$.

(ii) La propriété $\Pi x = x$ si $x \in \mathcal{L}$ est évidente. Si $x \in \mathcal{L}^\perp$ on a $\|x - y\|^2 = \langle x - y, x - y \rangle = \|x\|^2 + \|y\|^2$ pour $y \in \mathcal{L}$, donc $y = 0$ minimise $y \mapsto \|x - y\|$ sur \mathcal{L} , et donc $\Pi x = 0$.

(iii) On remarque d'abord que si $y \in \mathcal{L}$:

$$\begin{aligned} \|x - \Pi x\|^2 &\leq \|x - (\Pi x + y)\|^2 \\ &= \|x - \Pi x\|^2 + \|y\|^2 - 2\langle x - \Pi x, y \rangle, \end{aligned}$$

et donc

$$2\langle x - \Pi x, y \rangle \leq \|y\|^2.$$

Comme $y \in \mathcal{L}$ est arbitraire et comme \mathcal{L} est un sous-espace, on peut remplacer y par αy pour tout $\alpha > 0$, de façon à obtenir

$$2\langle x - \Pi x, \alpha y \rangle \leq \|\alpha y\|^2,$$

et en divisant par α on obtient

$$2\langle x - \Pi x, y \rangle \leq \alpha \|y\|^2:$$

On fait tendre α vers 0 et on conclut que $\langle x - \Pi x, y \rangle \leq 0$. On montre exactement de la même manière que $\langle x - \Pi x, y \rangle \geq 0$, en considérant des $\alpha < 0$. Donc $x - \Pi x$ est orthogonal à \mathcal{L} . ■

Corollaire 22.1. *Soit Π l'opérateur de projection sur un sous-espace fermé \mathcal{L} de \mathcal{H} . Pour tout $x \in \mathcal{H}$, $x = \Pi x + (x - \Pi x)$ est l'unique représentation de x comme la somme d'un vecteur de \mathcal{L} et d'un vecteur de \mathcal{L}^\perp . De plus $x - \Pi x$ est la projection de x sur \mathcal{L}^\perp . Enfin on a $(\mathcal{L}^\perp)^\perp = \mathcal{L}$.*

Preuve. On a vu que $\Pi x \in \mathcal{L}$ et que $x - \Pi x \in \mathcal{L}^\perp$ dans le théorème 22.7(iii). Quant à l'unicité, soit $x = y + z$ une autre représentation avec $y \in \mathcal{L}$ et $z \in \mathcal{L}^\perp$. Le vecteur $y - \Pi x = z - (x - \Pi x)$ est à la fois dans \mathcal{L} et dans \mathcal{L}^\perp ; donc, étant orthogonal à lui-même, il est nul : cela montre l'unicité.

Ensuite on remarque que $\mathcal{L} \subset (\mathcal{L}^\perp)^\perp$: en effet si $x \in \mathcal{L}$ et $y \in \mathcal{L}^\perp$ on a $\langle x, y \rangle = 0$, donc $x \in (\mathcal{L}^\perp)^\perp$.

Enfin, si $x \in (\mathcal{L}^\perp)^\perp$ on a $x = y + z$ avec $y \in \mathcal{L}$ et $z \in \mathcal{L}^\perp$. On a $\langle x, z \rangle = \langle y, z \rangle + \langle z, z \rangle$, et comme $\langle y, z \rangle = 0$ (puisque $y \in \mathcal{L}$) et $\langle x, z \rangle = 0$ (puisque $z \in \mathcal{L}^\perp$ et $x \in (\mathcal{L}^\perp)^\perp$) il s'ensuit que $\langle z, z \rangle = 0$: donc $z = 0$, et $x = y$. Comme $y \in \mathcal{L}$ on a aussi $x \in \mathcal{L}$, donc $(\mathcal{L}^\perp)^\perp \subset \mathcal{L}$. ■

Corollaire 22.2. *Soit Π l'opérateur de projection sur un sous-espace fermé \mathcal{L} de \mathcal{H} . On a alors :*

$$(i) \quad \langle \Pi x, y \rangle = \langle x, \Pi y \rangle,$$

$$(ii) \quad \Pi \text{ est un opérateur linéaire : } \Pi(\alpha x + \beta y) = \alpha \Pi x + \beta \Pi y.$$

Preuve. (i) Soit $x, y \in \mathcal{H}$. En vertu du corollaire 22.1 on peut écrire les vecteurs x et y de manière unique ainsi :

$$\begin{aligned} x &= x_1 + x_2, & x_1 &\in \mathcal{L}, & x_2 &\in \mathcal{L}^\perp, \\ y &= y_1 + y_2, & y_1 &\in \mathcal{L}, & y_2 &\in \mathcal{L}^\perp. \end{aligned} \tag{22.5}$$

Par suite

$$\langle \Pi x, y \rangle = \langle x_1, y \rangle = \langle x_1, y_1 + y_2 \rangle = \langle x_1, y_1 \rangle.$$

puisque $\langle x_1, y_2 \rangle = 0$. Comme on a aussi $\langle x_2, y_1 \rangle = 0$, il vient

$$= \langle x_1 + x_2, y_1 \rangle = \langle x, y_1 \rangle = \langle x, \Pi y \rangle.$$

(ii) Avec les notations (22.5) on obtient

$$\alpha x + \beta y = (\alpha x_1 + \beta y_1) + (\alpha x_2 + \beta y_2),$$

donc

$$\Pi(\alpha x + \beta y) = \alpha x_1 + \beta y_1 = \alpha \Pi x + \beta \Pi y.$$

■

Nous terminons cette exposition par une réciproque qui dit essentiellement que si un opérateur linéaire « se comporte comme une projection », au sens des propriétés précédentes, c'est en fait une projection.

Théorème 22.8. *Soit T une application de \mathcal{H} dans un sous-espace fermé \mathcal{L} . Si pour tout x le vecteur $x - Tx$ est orthogonal à \mathcal{L} , alors T est la projection (orthogonale) sur le sous-espace \mathcal{L} .*

Preuve. On peut écrire $x = Tx + (x - Tx)$, avec $Tx \in \mathcal{L}$ et $(x - Tx) \in \mathcal{L}^\perp$. D'après le corollaire 22.1, Tx doit être la projection orthogonale de x sur \mathcal{L} .

■

Exercices

1. En utilisant la propriété $(a - b)^2 \geq 0$, montrer que $(a + b)^2 \leq 2a^2 + 2b^2$.
2. Soit $x, y \in \mathcal{H}$ (un espace de Hilbert), avec $\langle x, y \rangle = 0$. Montrer le « théorème de Pythagore » : $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.
3. Montrer que \mathbb{R}^n est un espace de Hilbert avec le produit scalaire « usuel » : si $\vec{x} = (x_1, \dots, x_n)$ et $\vec{y} = (y_1, \dots, y_n)$, alors $\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^n x_i y_i$.
4. Soit \mathcal{L} un sous-espace fermé de l'espace de Hilbert \mathcal{H} , et Π la projection sur \mathcal{L} . Montrer que Πy est l'unique vecteur de \mathcal{L} tel que $\langle \Pi y, z \rangle = \langle y, z \rangle$ pour tout $z \in \mathcal{L}$.

Chapitre 23

Espérance conditionnelle

Soit X et Y deux v.a., avec Y réelle et X prenant ses valeurs dans un ensemble fini ou dénombrable. Il arrive souvent qu'on connaisse la valeur prise par X et qu'on veuille calculer l'espérance de Y , en utilisant cette information supplémentaire, ce qui *a priori* modifie l'espérance. De manière plus formelle, cela revient à poser :

Définition 23.1. Soit X une v.a. à valeurs dans un ensemble E fini ou dénombrable, dont les points sont notés $\{x_1, x_2, \dots, x_n, \dots\}$. Soit Y une autre v.a. réelle, définie sur le même espace de probabilité (Ω, \mathcal{A}, P) . Si $P(X = x_j) > 0$ l'espérance conditionnelle de Y sachant $\{X = x_j\}$ est le nombre

$$E(Y | X = x_j) = E_{Q_j}(Y), \quad \text{espérance de } Y \text{ pour } Q_j,$$

où Q_j est la probabilité sur (Ω, \mathcal{A}) donnée par $Q(\Lambda) = P(\Lambda | X = x_j)$, pourvu que $E_{Q_j}(|Y|) < \infty$ (i.e., Y est Q_j -intégrable).

Lorsque Y est elle-même une v.a. discrète, prenant ses valeurs dans $\{y_1, y_2, \dots\}$ (une partie dénombrable de \mathbb{R}), on peut utiliser la formule (5.1) pour Y , et on obtient immédiatement le

Théorème 23.1. Dans la situation précédente, pour tout j tel que $P(X = x_j) > 0$ on a

$$E(Y | X = x_j) = \sum_{k=1}^{\infty} y_k P(Y = y_k | X = x_j),$$

pourvu que la série ci-dessus soit absolument convergente.

Au lieu de « fixer » la valeur x_j prise par X , on peut modifier légèrement notre point de vue, et chercher l'espérance conditionnelle de Y , connaissant la variable X (toujours supposée à valeurs dans l'ensemble fini ou dénombrable E), plutôt que « sachant que $X = x_j$ ». Pour cela, on pose

$$f(x_j) = \begin{cases} E(Y | X = x_j) & \text{si } P(X = x_j) > 0 \\ \text{une valeur arbitraire} & \text{si } P(X = x_j) = 0, \end{cases} \quad (23.1)$$

ce qui définit une fonction f sur E .

Définition 23.2. Soit X une v.a. à valeurs dans un espace E fini ou dénombrable, et Y une v.a. réelle définie sur le même espace de probabilité. L'espérance conditionnelle de Y sachant X est

$$E(Y | X) = f(X),$$

où f est donnée par (23.1), et pourvu que cette fonction f soit bien définie (i.e., pour tout j tel que $P(X = x_j) > 0$, la v.a. Y est intégrable par rapport à la probabilité $Q_j(\Lambda) = P(\Lambda | X = x_j)$).

Remarque 23.1. Cette définition ne définit pas réellement $E(Y | X)$ partout, mais seulement presque partout, car sa valeur est arbitraire sur les ensembles $\{X = x\}$ tels que $P(X = x) = 0$: ce sera un trait caractéristique des espérances conditionnelles dans le cas général étudié plus bas.

Exemple. Soit X une v.a. de Poisson de paramètre λ . Soit une autre v.a. S définie ainsi : quand $X = n$, on tire n fois (indépendamment) une pièce ayant la probabilité p de tomber sur Face ; S est alors le nombre total de « Faces ». On va calculer $E(S | X)$ et $E(X | S)$.

Calculons d'abord $E(S | X = n)$. Si $X = n$, la v.a. S est binomiale de paramètres n et p , donc $E(S | X = n) = pn$ et $E(S | X) = pX$.

Pour calculer $E(X | S)$ il nous faut évaluer $E(X | S = k)$, et nous commençons par chercher $P(X = n | S = k)$:

$$\begin{aligned} P(X = n | S = k) &= \frac{P(S = k | X = n)P(X = n)}{P(S = k)} \\ &= \frac{C_n^k p^k (1-p)^{n-k} \left(\frac{\lambda^n}{n!}\right) e^{-\lambda}}{\sum_{m \geq k} C_m^k p^k (1-p)^{m-k} \left(\frac{\lambda^m}{m!}\right) e^{-\lambda}} \\ &= \frac{((1-p)\lambda)^{n-k}}{(n-k)!} e^{-(1-p)\lambda} \end{aligned}$$

pour $n \geq k$, et bien sûr $P(X = n | S = k) = 0$ si $n < k$. Donc

$$E(X | S = k) = \sum_{n \geq k} n \frac{((1-p)\lambda)^{n-k}}{(n-k)!} e^{-(1-p)\lambda} = k + (1-p)\lambda.$$

et il vient

$$E(X | S) = S + (1-p)\lambda.$$

On peut vérifier directement que $E(S) = E(E(S | X))$, propriété qui découlera aussi du théorème 23.3 ci-après. Donc on a

$$E(S) = pE(X) = p\lambda.$$

Nous allons maintenant passer au *cas général* : nous voulons calculer $E(Y|X)$ quand X est une v.a. à valeurs dans un espace quelconque. L'approche précédente ne marche plus, car typiquement on a $P(X = x) = 0$ pour tout x . Cependant on a vu aussi dans le cas dénombrable que $E(Y|X) = f(X)$ pour une certaine fonction f , et c'est cette idée qui va se généraliser, grâce au théorème suivant.

Rappelons d'abord qu'on a introduit au chapitre 10 la notion de tribu engendrée par une v.a. : si par exemple X est une application de Ω dans \mathbb{R}^n , la *tribu engendrée par X* est

$$\sigma(X) = X^{-1}(\mathcal{B}^n) = \{A \subset \Omega : X^{-1}(B) = A, \text{ pour un } B \in \mathcal{B}^n\}.$$

Théorème 23.2. *Soit X une v.a. à valeurs dans \mathbb{R}^n et Y une v.a. réelle, toutes deux définies sur le même espace (Ω, \mathcal{A}) . Alors Y est mesurable par rapport à $\sigma(X)$ si et seulement s'il existe une fonction borélienne f de \mathbb{R}^n dans \mathbb{R} telle que $Y = f(X)$.*

Preuve. Si la fonction f de l'énoncé existe on a $Y^{-1}(B) = X^{-1}(f^{-1}(B))$. Si $B \in \mathcal{B}$ alors $A = f^{-1}(B) \in \mathcal{B}^n$, donc $X^{-1}(A) \in \sigma(X)$ (on peut aussi utiliser le théorème 8.2).

Supposons inversement que $Y^{-1}(B) \in \sigma(X)$ pour tout $B \in \mathcal{B}$. Supposons d'abord que Y soit « simple », i.e. $Y = \sum_{i=1}^k a_i 1_{A_i}$ pour un $k < \infty$, des a_i réels tous différents, et des A_i deux à deux disjoints. On a $A_i = Y^{-1}(\{a_i\})$, donc $A_i \in \sigma(X)$ et il existe $B_i \in \mathcal{B}^n$ avec $A_i = X^{-1}(B_i)$. Il reste à poser $f(x) = \sum_{i=1}^k a_i 1_{B_i}(x)$ pour obtenir $Y = f(X)$ avec f borélienne. Ensuite, si on suppose Y positive, on sait (cf. chapitre 9) qu'il existe une suite croissante (Y_n) de v.a. simples positives croissant vers Y et $\sigma(X)$ -mesurables. On vient de voir que $Y_n = f_n(X)$ pour des fonctions boréliennes f_n . On pose $f(x) = \limsup_{n \rightarrow \infty} f_n(x)$, qui est borélienne également, et on a

$$Y = \lim_{n \rightarrow \infty} Y_n = \lim_n f_n(X).$$

Mais

$$(\limsup_{n \rightarrow \infty} f_n)(X) = \limsup_n (f_n(X)).$$

et par suite $Y = f(X)$.

Pour Y de signe quelconque on écrit $Y = Y^+ - Y^-$, et on se ramène ainsi au cas précédent. ■

Dans ce qui suit l'espace de probabilité (Ω, \mathcal{A}, P) est fixé, et X est une v.a. sur cet espace, à valeurs dans \mathbb{R}^n . L'espace $L^2(\Omega, \mathcal{A}, P)$ est l'espace

des (classes d'équivalence pour la relation d'équivalence « égalité p.s. » de) v.a. Y vérifiant $E(Y^2) < \infty$. Comme on l'a vu au chapitre 22 c'est un espace de Hilbert avec le produit scalaire

$$\langle Y, Z \rangle = E(YZ).$$

Comme $\sigma(X) \subset \mathcal{A}$, l'espace $L^2(\Omega, \sigma(X), P)$ est aussi un espace de Hilbert, et c'est un sous-espace fermé de $L^2(\Omega, \mathcal{A}, P)$. (Noter que le produit scalaire est le même sur ces deux espaces de Hilbert.)

Définition 23.3. Soit $Y \in L^2(\Omega, \mathcal{A}, P)$. L'espérance conditionnelle de Y sachant X est l'unique élément \hat{Y} de $L^2(\Omega, \sigma(X), P)$ qui vérifie

$$E(\hat{Y}Z) = E(YZ) \quad \text{pour tout } Z \in L^2(\Omega, \sigma(X), P). \quad (23.2)$$

On écrit cette espérance conditionnelle (i.e., la variable \hat{Y}) ainsi :

$$E(Y|X)$$

L'équation (23.2) signifie simplement que \hat{Y} est la projection orthogonale de Y sur le sous-espace fermé $L^2(\Omega, \sigma(X), P)$ de $L^2(\Omega, \mathcal{A}, P)$: l'existence et l'unicité de l'espérance conditionnelle suit donc du corollaire 22.1.

Comme $E(Y|X)$ est $\sigma(X)$ -mesurable, d'après le théorème 23.2 il existe une fonction borélienne f telle que $E(Y|X) = f(X)$. Donc (23.2) équivaut au fait que

$$E(f(X)g(X)) = E(Yg(X)) \quad (23.3)$$

pour toute fonction borélienne g telle que $g(X) \in L^2$.

Dans la définition précédente la v.a. X ne joue pas de rôle réel, c'est la tribu $\sigma(X)$ qui est importante. Il est donc plutôt naturel de la remplacer par une sous-tribu arbitraire \mathcal{G} de \mathcal{A} . Là encore $L^2(\Omega, \mathcal{G}, P)$ est un sous-espace de Hilbert de $L^2(\Omega, \mathcal{A}, P)$, et on peut poser :

Définition 23.4. Soit $Y \in L^2(\Omega, \mathcal{A}, P)$ et \mathcal{G} une sous-tribu de \mathcal{A} . L'espérance conditionnelle de Y sachant \mathcal{G} est l'unique élément $E(Y|\mathcal{G})$ de $L^2(\Omega, \mathcal{G}, P)$ qui vérifie

$$E(YZ) = E(E(Y|\mathcal{G})Z) \quad (23.4)$$

pour tout $Z \in L^2(\Omega, \mathcal{G}, P)$.

Avertissement important : L'espérance conditionnelle est un élément de L^2 , c'est-à-dire une classe d'équivalence de v.a. : on peut supposer que c'est une vraie variable aléatoire, mais dans ce cas elle n'est définie qu'à un ensemble de mesure nulle près. Par suite les propriétés comme $E(Y|\mathcal{G}) \geq 0$ ou $E(Y|\mathcal{G}) = Z$, etc., doivent *toujours* être comprises au sens « presque sûr », bien qu'on omette en général de le mentionner. On dit aussi parfois : il existe une « version » de $E(Y|\mathcal{G})$ qui est positive, ou égale à Z , etc.

Théorème 23.3. Soit $Y \in L^2(\Omega, \mathcal{A}, P)$ et \mathcal{G} une sous-tribu de \mathcal{A} .

- (a) Si $Y \geq 0$ alors $E(Y|\mathcal{G}) \geq 0$.
- (b) Si $\mathcal{G} = \sigma(X)$ pour une v.a. X à valeurs dans \mathbb{R}^n , il existe une fonction borélienne f sur \mathbb{R} telle que $E(Y|\mathcal{G}) = f(X)$; .
- (c) $E(E(Y|\mathcal{G})) = E(Y)$.
- (d) L'application $Y \rightarrow E(Y|\mathcal{G})$ est linéaire.

Preuve. On a déjà vu (b). Pour (c) il suffit d'appliquer (23.4) avec $Z = 1$, et (d) découle aussi de (23.4) : en effet si U et V sont dans L^2 , on a

$$\begin{aligned} E((U + \alpha V)Z) &= E(UZ) + \alpha E(VZ) \\ &= E(E(U|\mathcal{G})Z) + \alpha E(E(V|\mathcal{G})Z) \\ &= E((E(U|\mathcal{G}) + \alpha E(V|\mathcal{G}))Z). \end{aligned}$$

et donc $E(U + \alpha V|\mathcal{G}) = E(U|\mathcal{G}) + \alpha E(V|\mathcal{G})$ à cause de l'unicité de l'espérance conditionnelle (on peut aussi dire que $E(Y|\mathcal{G})$ est la projection de Y sur le sous-espace $L^2(\Omega, \mathcal{G}, P)$, et l'opérateur projection est linéaire, cf. le corollaire 22.2).

Finalement pour (a) on utilise encore (23.4) en prenant pour Z la fonction indicatrice $1_{\{E(Y|\mathcal{G}) < 0\}}$: si $Y > 0$ p.s. on a $E(YZ) \geq 0$ puisque Y et Z sont toutes deux positives, tandis que

$$E(E(Y|\mathcal{G})Z) = E(E(Y|\mathcal{G})1_{\{E(Y|\mathcal{G}) < 0\}}) < 0 \quad \text{si} \quad P(\{E(Y|\mathcal{G}) < 0\}) > 0.$$

Cela contredit (23.3), sauf si $P(\{E(Y|\mathcal{G}) < 0\}) = 0$. ■

Remarque 23.2. Comme on l'a vu dans la démonstration du théorème 23.3, la propriété cruciale de l'espérance conditionnelle est (23.4) ; prouver l'existence et l'unicité de l'espérance conditionnelle est la seule raison pour laquelle nous avons consacré un chapitre entier aux espaces de Hilbert.

Il nous reste à étendre la notion d'espérance conditionnelle aux v.a. qui ne sont pas nécessairement de carré intégrable. Les techniques d'espace de Hilbert font alors bien sûr défaut.

Rappelons une fois encore que $L^1(\Omega, \mathcal{A}, P)$ est l'espace des (classes d'équivalences de) v.a. intégrables, et de manière analogue nous allons noter $L^+(\Omega, \mathcal{A}, P)$ l'espace des (classes d'équivalences de) v.a. positives, toujours en identifiant deux v.a. qui sont p.s. égales. Les v.a. de L^+ peuvent prendre la valeur $+\infty$.

Lemme 23.1. Soit $Y \in L^+(\Omega, \mathcal{A}, P)$ et soit \mathcal{G} une sous-tribu de \mathcal{A} . Il existe un unique élément $E(Y|\mathcal{G})$ de $L^+(\Omega, \mathcal{G}, P)$, appelé l'espérance conditionnelle de Y sachant \mathcal{G} , tel que

$$E(YX) = E(E(Y|\mathcal{G})X) \quad (23.5)$$

pour tout $X \in L^+(\Omega, \mathcal{G}, P)$, et cette espérance conditionnelle coïncide avec celle de la définition 23.4 si on a aussi $Y \in L^2(\Omega, \mathcal{A}, P)$.

Si de plus $0 \leq Y \leq Y'$, alors

$$E(Y|\mathcal{G}) \leq E(Y'|\mathcal{G}). \quad (23.6)$$

Preuve. Si Y est à la fois dans $L^2(\Omega, \mathcal{A}, P)$ et positive, on définit $E\{Y|\mathcal{G}\}$ par la définition 23.4. Si alors $X \in L^+(\Omega, \mathcal{G}, P)$, et comme les v.a. $X_n = X \wedge n$ sont de carré intégrable, il vient en appliquant deux fois le théorème de convergence monotone et (23.5) :

$$\begin{aligned} E(YX) &= \lim_n E(YX_n) \\ &= \lim_n E(E(Y|\mathcal{G})X_n) \\ &= E(E(Y|\mathcal{G})X) \end{aligned} \quad (23.7)$$

et (23.5) est satisfaite.

Soit maintenant $Y \in L^+(\Omega, \mathcal{A}, P)$. Chaque $Y_m = Y \wedge m$ est de carré intégrable, tandis que d'après le théorème 23.3 l'espérance conditionnelle est un opérateur positif, de sorte que la suite $E(Y_m|\mathcal{G})$ est croissante. Par suite on peut poser

$$E(Y|\mathcal{G}) = \lim_{m \rightarrow \infty} E(Y_m|\mathcal{G}). \quad (23.8)$$

Si $X \in L^+(\Omega, \mathcal{G}, P)$, on obtient alors en utilisant plusieurs fois le théorème de convergence monotone, ainsi que (23.8) :

$$\begin{aligned} E(YX) &= \lim_m E(Y_m X) \\ &= E\left(\lim_m E(Y_m|\mathcal{G})X\right) \\ &= E(E(Y|\mathcal{G})X), \end{aligned}$$

Si de plus on a $Y \leq Y'$, alors $Y \wedge m \leq Y' \wedge m$ pour chaque m , donc $E(Y \wedge m | \mathcal{G}) \leq E(Y' \wedge m | \mathcal{G})$ par le théorème 23.3(a), et on obtient (23.6).

Il reste à montrer l'unicité. Soit U et V deux versions de $E(Y | \mathcal{G})$, qui donc vérifient (23.8) pour toute v.a. positive \mathcal{G} -mesurable X . Soit $\Lambda_n = \{U < V \leq n\}$ et supposons que $P(\Lambda_n) > 0$. Comme on a $\Lambda_n \in \mathcal{G}$ il vient

$$E(Y1_{\Lambda_n}) = E(U1_{\Lambda_n}) = E(V1_{\Lambda_n}).$$

tandis que $0 \leq U1_{\Lambda_n} \leq V1_{\Lambda_n} \leq n$, et $P(\Lambda_n) > 0$ implique que les v.a. $V1_{\Lambda_n}$ et $U1_{\Lambda_n}$ ne sont pas p.s. égales : on en déduit que $E(U1_{\Lambda}) < E(V1_{\Lambda})$, d'où une contradiction. Il faut donc que $P(\Lambda_n) = 0$ pour tout n , et comme $\{U > V\} = \bigcup_{n \geq 1} \Lambda_n$ il vient $P(U < V) = 0$; on vérifie de même que $P(V > U) = 0$, et l'unicité en découle. ■

Théorème 23.4. *Soit $Y \in L^1(\Omega, \mathcal{A}, P)$ et soit \mathcal{G} une sous-tribu de \mathcal{A} . Il existe un unique élément $E(Y | \mathcal{G})$ de $L^1(\Omega, \mathcal{G}, P)$, appelé l'espérance conditionnelle de Y sachant \mathcal{G} , tel que*

$$E(YX) = E(E(Y | \mathcal{G})X) \quad (23.9)$$

pour tout v.a. X borné \mathcal{G} -mesurable, et cette espérance conditionnelle coïncide avec celle de la définition 23.4 (resp. du lemme 23.1) si on a aussi $Y \in L^2(\Omega, \mathcal{A}, P)$ (resp. $Y \in L^+(\Omega, \mathcal{A}, P)$). Enfin, on a aussi

(a) Si $Y \geq 0$ alors $E(Y | \mathcal{G}) \geq 0$.

(b) L'application $Y \mapsto E(Y | \mathcal{G})$ est linéaire.

Preuve. Si $Y \in L^1$, on a aussi $Y^+ \in L^1$ et $Y^- \in L^1$, où Y^+ et Y^- sont les parties positive et négative de Y . On a $Y = Y^+ - Y^-$, et on pose

$$E(Y | \mathcal{G}) = E(Y^+ | \mathcal{G}) - E(Y^- | \mathcal{G}).$$

Cette formule a bien en sens : en effet les v.a. Y^+ et Y^- , donc aussi les v.a. $E(Y^+ | \mathcal{G})$ et $E(Y^- | \mathcal{G})$ d'après le théorème 23.3(c), sont intégrables et donc p.s. finies. Il est clair que $E(Y | \mathcal{G})$ vérifie (23.9). Pour l'unicité, considérons deux versions U et V de $E(Y | \mathcal{G})$, et soit $\Lambda = \{U < V\}$. On a $\Lambda \in \mathcal{G}$, donc 1_{Λ} est borné et \mathcal{G} -mesurable, donc $E(Y1_{\Lambda}) = E(E(Y | \mathcal{G})1_{\Lambda}) = E(U1_{\Lambda}) = E(V1_{\Lambda})$. Mais si $P(\Lambda) > 0$ il vient $E(U1_{\Lambda}) < E(V1_{\Lambda})$, ce qui fournit une contradiction. Par suite $P(U < V) = 0$ et on vérifie de même que $P(V < U) = 0$.

Les dernières assertions sont des conséquences triviales de la définition ci-dessus de $E(Y | \mathcal{G})$, du lemme 23.1 et du théorème 23.3. ■

Exemple. Soit (X, Z) une v.a. bidimensionnelle admettant une densité f . Soit g une fonction borélienne bornée et $Y = g(Z)$. On cherche à calculer $E(Y | X) = E(g(Z) | X)$. On rappelle que X admet la densité f_X suivante :

$$f_X(x) = \int f(x, z) dz,$$

et qu'on a défini au chapitre 12 (cf. théorème 12.2) la *densité conditionnelle de Z sachant $X = x$* par :

$$f_{X=x}(z) = \frac{f(x, z)}{f_X(x)},$$

partout où $f_X(x) \neq 0$. Posons alors

$$h(x) = \int g(z) f_{X=x}(z) dz.$$

Pour toute fonction borélienne bornée k il vient

$$\begin{aligned} E(h(X)k(X)) &= \int h(x)k(x)f_X(x)dx \\ &= \iint g(z)f_{X=x}(z)dz k(x)f_X(x)dx \\ &= \iint g(z)\frac{f(x, z)}{f_X(x)}k(x)f_X(x)dz dx \\ &= \iint g(z)k(x)f(x, z)dz dx \\ &= E(g(Z)k(X)) = E(Yk(X)). \end{aligned}$$

Donc (23.9) implique que

$$E(Y | X) = h(X),$$

et on a ainsi décrit une méthode permettant de calculer les espérances conditionnelles lorsque les v.a. admettent des densités.

Théorème 23.5. Soit Y une v.a. positive ou intégrable sur (Ω, \mathcal{F}, P) , et \mathcal{G} une sous-tribu. Alors $E(Y | \mathcal{G}) = Y$ si et seulement si Y est \mathcal{G} -mesurable.

Preuve. Trivial à partir de la définition de l'espérance conditionnelle. ■

Théorème 23.6. *Si $Y \in L^1(\Omega, \mathcal{A}, P)$ et si X est une v.a. indépendante de Y , on a*

$$E(Y|X) = E\{Y\}.$$

Preuve. Soit g borélienne bornée. On a $E(Yg(X)) = E(Y)E(g(X))$ par l'indépendance. Le résultat découle alors immédiatement de la définition de l'espérance conditionnelle et du théorème 23.2. ■

Théorème 23.7. *Soit X, Y des v.a. réelles sur (Ω, \mathcal{A}, P) , et \mathcal{G} une sous-tribu de \mathcal{A} par rapport à laquelle X est mesurable. On a alors*

$$E(XY|\mathcal{G}) = XE(Y|\mathcal{G}).$$

dans chacun des deux cas suivants :

- (a) les v.a. X, Y et XY sont intégrables,
- (b) les v.a. X et Y sont positives.

Preuve. Supposons d'abord (b). Pour toute v.a. \mathcal{G} -mesurable et positive Z on a

$$E(XYZ) = E(XZ E(Y|\mathcal{G}))$$

par (23.5). Comme $XE(Y|\mathcal{G})$ est aussi \mathcal{G} -mesurable, on en déduit le résultat par une nouvelle application de la caractérisation (23.5).

Dans le cas (a) on observe que les v.a. X^+Y^+, X^-Y^+, X^+Y^- et X^-Y^- sont toutes intégrables et positives, donc $E(X^+Y^+|\mathcal{G}) = X^+ E(Y^+|\mathcal{G})$ d'après ce qui précède, et de même pour les trois autres produits, et toutes ces v.a. sont à valeurs finies. Il reste à appliquer la linéarité de l'espérance conditionnelle et la propriété $XY = X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+$. ■

Ensuite, nous insistons sur le fait — important — que les principaux résultats de convergence pour les espérances (ou intégrales) sont aussi valides pour les espérances conditionnelles, à condition de rajouter le qualificatif « p.s. » (qui est implicite dès qu'on parle d'espérance conditionnelle, mais que nous écrivons explicitement dans le théorème suivant).

Théorème 23.8. *Soit $(Y_n)_{n \geq 1}$ une suite de v.a. réelles sur (Ω, \mathcal{A}, P) , et \mathcal{G} une sous-tribu de \mathcal{A} .*

- (a) (Convergence monotone.) *Si les Y_n sont positives et croissent p.s. vers une limite Y , alors*

$$\lim_{n \rightarrow \infty} E(Y_n|\mathcal{G}) = E(Y|\mathcal{G}) \quad \text{p.s.};$$

(b) (*Lemme de Fatou.*) Si les Y_n sont positives, on a

$$E(\liminf_{n \rightarrow \infty} Y_n | \mathcal{G}) \leq \liminf_{n \rightarrow \infty} E(Y_n | \mathcal{G}) \quad p.s.;$$

(c) (*Théorème de convergence dominée de Lebesgue.*) Si les Y_n convergent p.s. vers une limite Y , et si on a $|Y_n| \leq Z$ pour tout n et pour une certaine v.a. $Z \in L^1(\Omega, \mathcal{A}, P)$, alors

$$\lim_{n \rightarrow \infty} E(Y_n | \mathcal{G}) = E(Y | \mathcal{G}) \quad p.s..$$

Preuve. (a) Par (23.6) on a $E(Y_{n+1} | \mathcal{G}) \geq E(Y_n | \mathcal{G})$ p.s. pour chaque n . Donc $U = \lim_{n \rightarrow \infty} E(Y_n | \mathcal{G})$ existe p.s. et pour toute v.a. positive et \mathcal{G} -mesurable X on a :

$$\begin{aligned} E(UX) &= \lim_{n \rightarrow \infty} E(E(Y_n | \mathcal{G})X) \\ &= \lim_{n \rightarrow \infty} E(Y_n X) \\ &= \lim_{n \rightarrow \infty} E(YX) \end{aligned}$$

par le théorème de convergence monotone (usuel) et (23.5). Donc $U = E(Y | \mathcal{G})$, en appliquant de nouveau (23.5).

Pour les démonstrations de (b) et (c) il suffit de reproduire les preuves des résultats similaires pour les espérances (non conditionnelles) et d'utiliser (23.5) dans le même esprit que pour (a) ci-dessus. ■

Terminons ce chapitre avec trois inégalités utiles. Les deux dernières, bien qu'ayant une version pour les espérances conditionnelles, sont données pour les espérances « ordinaires » seulement. Elles se trouvent placées ici parce qu'elles découlent naturellement de la première qui, quant à elle, est très utile dans sa version « espérances conditionnelles ».

Théorème 23.9. (Inégalité de Jensen.) Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe, et soit X une v.a. réelle, intégrable et telle que $\varphi(X)$ soit aussi intégrable. Pour toute sous-tribu \mathcal{G} on a alors

$$\varphi \circ E(X | \mathcal{G}) \leq E(\varphi(X) | \mathcal{G}).$$

Preuve. Un résultat d'analyse réelle dit que, si φ est convexe, il existe deux suites (a_n) et (b_n) de réels telles que $\varphi(x) = \sup_n (a_n x + b_n)$ pour tout x . On a alors

$$E(a_n X + b_n | \mathcal{G}) = a_n E(X | \mathcal{G}) + b_n.$$

Mais $E(a_n X + b_n | \mathcal{G}) \leq E(\varphi(X) | \mathcal{G})$, donc $a_n E(X | \mathcal{G}) + b_n \leq E(\varphi(X) | \mathcal{G})$ pour chaque n . En prenant le supremum en n , on obtient immédiatement le résultat. ■

Noter que $\varphi(x) = x^2$ est une fonction convexe. Par suite on déduit de l'inégalité de Jensen que

$$(E(X | \mathcal{G}))^2 \leq E(X^2 | \mathcal{G}).$$

Une conséquence importante de l'inégalité de Jensen est l'inégalité de Hölder pour les variables aléatoires.

Théorème 23.10. (Inégalité de Hölder.) *Soit X et Y des v.a. réelles, et $p, q \in]1, \infty[$ avec $\frac{1}{p} + \frac{1}{q} = 1$. On a alors (avec les conventions $a \times +\infty = +\infty$ si $a > 0$ et $0 \times +\infty = 0$).*

$$E(|XY|) \leq E(|X|^p)^{\frac{1}{p}} E(|Y|^q)^{\frac{1}{q}}. \quad (23.10)$$

(En particulier, si $X \in L^p$ et $Y \in L^q$ avec p, q comme ci-dessus, le produit XY appartient à L^1).

Preuve. Si le membre de droite de (23.10) est infini il n'y a rien à démontrer. S'il est nul l'un des deux termes du produit est nul, disons le premier par exemple: donc $X = 0$ p.s. et le membre de gauche de (23.10) est aussi nul. On peut donc supposer que $0 < E(X^p) < \infty$ et que $0 < E(Y^q) < \infty$, et également, sans perte de généralité, que X et Y sont positives. Posons $C = E(X^p) < \infty$, et définissons une nouvelle probabilité Q en posant pour $A \in \mathcal{A}$:

$$Q(A) = \frac{1}{C} E(1_A X^p).$$

Posons ensuite $Z = \frac{Y}{X^{p-1}} 1_{\{X>0\}}$. Comme $\varphi(x) = |x|^p$ est convexe, l'inégalité de Jensen donne

$$(E_Q(Z))^q \leq E(Z^q).$$

Donc

$$\begin{aligned} \frac{1}{C^q} E(XY)^q &= \frac{1}{C^q} E\left(\frac{Y}{X^{p-1}} X^p\right)^q \\ &= \left(E_Q\left(\frac{Y}{X^{p-1}}\right)\right)^q \\ &\leq E_Q\left(\left(\frac{Y}{X^{p-1}}\right)^q\right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{C} \mathbf{E} \left(\left(\frac{Y}{X^{p-1}} \right)^q X^p \right) \\
 &= \frac{1}{C} \mathbf{E} \left(Y^q \frac{1}{X^{(p-1)q}} X^p \right),
 \end{aligned}$$

et comme $q = \frac{p}{p-1}$ et $(p-1)q = p$,

$$\begin{aligned}
 \frac{1}{C^q} \mathbf{E}(XY)^q &= \frac{1}{C} \mathbf{E} \left(Y^q \frac{1}{X^p} X^p \right) \\
 &= \frac{1}{C} \mathbf{E}(Y^q).
 \end{aligned}$$

Par suite

$$\mathbf{E}(XY)^q \leq C^{q-1} \mathbf{E}(Y^q),$$

et en prenant la racine q -ième on obtient

$$\mathbf{E}(XY) \leq C^{\frac{q-1}{q}} \mathbf{E}(Y^q)^{\frac{1}{q}}.$$

Comme $\frac{q-1}{q} = \frac{1}{p}$ et $C = \mathbf{E}(X^p)$, on a le résultat. ■

Corollaire 23.1. (Inégalité de Minkowski.) *Soit X et Y deux v.a. réelles. On a alors*

$$\mathbf{E}(|X + Y|^p)^{\frac{1}{p}} \leq \mathbf{E}(X^p)^{\frac{1}{p}} + \mathbf{E}(Y^p)^{\frac{1}{p}}.$$

Preuve. Si $p = 1$ le résultat est évident. On suppose donc dans la suite que $1 < p < \infty$, et on prend q tel que $\frac{1}{p} + \frac{1}{q} = 1$. D'après l'inégalité de Hölder (théorème 23.10) on a

$$\begin{aligned}
 \mathbf{E}(|X + Y|^p) &= \mathbf{E}(|X| |X + Y|^{p-1}) + \mathbf{E}(|Y| |X + Y|^{p-1}) \\
 &\leq \mathbf{E}(|X|^p)^{\frac{1}{p}} \mathbf{E}(|X + Y|^{(p-1)q})^{\frac{1}{q}} + \mathbf{E}(|Y|^p)^{\frac{1}{p}} \mathbf{E}(|X + Y|^{(p-1)q})^{\frac{1}{q}}.
 \end{aligned}$$

Comme $(p-1)q = p$ et $\frac{1}{q} = 1 - \frac{1}{p}$,

$$= \left(\mathbf{E}(|X|^p)^{\frac{1}{p}} + \mathbf{E}(|Y|^p)^{\frac{1}{p}} \right) \mathbf{E}(|X + Y|^p)^{1 - \frac{1}{p}}$$

et on a le résultat. ■

L'inégalité de Minkowski permet de définir une norme (satisfaisant l'inégalité triangulaire) sur l'espace L^p des (classes d'équivalences de) v.a. vérifiant $\mathbf{E}(|X|^p) < \infty$:

Définition 23.5. Pour $X \in L^p$ on pose

$$\|X\|_p = E(|X^p|)^{\frac{1}{p}}.$$

Avec cette norme l'espace vectoriel L^p est *complet* (on dit que c'est un « espace de Banach »); mais, pour $p \neq 2$, ce n'est pas un espace de Hilbert : la norme n'est pas associée à un produit scalaire.

Exercices

Pour les exercices 1 à 6 on considère une v.a. positive Y sur un espace (Ω, \mathcal{A}, P) , et une sous-tribu \mathcal{G} de \mathcal{A} .

1. Montrer que $|E(Y | \mathcal{G})| \leq E(|Y| | \mathcal{G})$.
2. Si \mathcal{H} est une sous-tribu de \mathcal{G} , montrer que

$$E(E(Y | \mathcal{G}) | \mathcal{H}) = E(Y | \mathcal{H}).$$

3. Montrer que $E(Y | Y) = Y$ p.s.
4. Montrer que si $|Y| \leq c$ p.s., alors $|E(Y | \mathcal{G})| \leq c$ p.s. également.
5. Si $Y = \alpha$ p.s., où α est une constante, montrer que $E(Y | \mathcal{G}) = \alpha$ p.s.
6. Montrer que $\{E(Y | \mathcal{G}) = 0\} \subset \{Y = 0\}$ et $\{Y = +\infty\} \subset \{E(Y | \mathcal{G}) = +\infty\}$ p.s. (rappelons que $Y \geq 0$).
- 7* Soit X et Y des v.a. réelles indépendantes, et f une fonction borélienne telle que $f(X, Y) \in L^1(\Omega, \mathcal{A}, P)$. On pose

$$g(x) = \begin{cases} E(f(x, Y)) & \text{si } |E(f(x, Y))| < \infty, \\ 0 & \text{sinon.} \end{cases}$$

Montrer que g est borélienne et que

$$E(f(X, Y) | X) = g(X).$$

8. Soit $Y \in L^2(\Omega, \mathcal{A}, P)$ et supposons que $E(Y^2 | X) = X^2$ et $E(Y | X) = X$. Montrer que $Y = X$ p.s.
- 9* Soit Y une v.a. exponentielle de paramètre 1 (i.e., $P(Y > t) = e^{-t}$ pour $t > 0$). Calculer $E(Y | Y \wedge t)$.
10. (Inégalité de Bienaymé-Chebyshev.) Montrer que si $X \in L^2$ et $a > 0$, on a $P(|X| \geq a | \mathcal{G}) \leq \frac{E(X^2 | \mathcal{G})}{a^2}$.

11. (Cauchy-Schwarz.) Si $X, Y \in L^2$ montrer que

$$(E(XY | \mathcal{G}))^2 \leq E(X^2 | \mathcal{G})E(Y^2 | \mathcal{G}).$$

12. Si $X \in L^2$, montrer que

$$E((X - E(X | \mathcal{G}))^2) \leq E((X - E(X))^2).$$

13. Si $r \geq p \geq 1$, montrer que $L^p(\Omega, \mathcal{A}, P) \supset L^r(\Omega, \mathcal{A}, P)$.

14.* Soit Z une v.a. positive sur (Ω, \mathcal{F}, P) , avec $E(Z) = 1$. On définit une nouvelle probabilité Q en posant $Q(A) = E(1_A Z)$. Soit \mathcal{G} une sous-tribu de \mathcal{F} , et $U = E(Z | \mathcal{G})$. Montrer que $E_Q(X | \mathcal{G}) = \frac{E(XZ | \mathcal{G})}{U}$, pour toute v.a. bornée et \mathcal{F} -mesurable X . (Ici, $E_Q(X | \mathcal{G})$ désigne l'espérance conditionnelle relativement à la probabilité Q ; on utilise la convention $\frac{0}{0} = 0$.)

15. Montrer que l'espace vectoriel normé L^p est complet. (*Indication* : voir la preuve du théorème 22.2.)

16. Soit $X \in L^1(\Omega, \mathcal{F}, P)$ et soit \mathcal{G} et \mathcal{H} des sous-tribus de \mathcal{F} . Supposons de plus \mathcal{H} indépendante de $\sigma(\sigma(X), \mathcal{G})$. Montrer que

$$E(X | \sigma(\mathcal{G}, \mathcal{H})) = E(X | \mathcal{G}).$$

17. Soit $(X_n)_{n \geq 1}$ des v.a. indépendantes intégrables, et posons $S_n = \sum_{i=1}^n X_i$; et $\mathcal{G}_n = \sigma(S_n, S_{n+1}, \dots)$. Montrer que $E(X_1 | \mathcal{G}_n) = E(X_1 | S_n)$ et que $E(X_j | \mathcal{G}_n) = E(X_j | S_n)$ pour $1 \leq j \leq n$. Montrer aussi que $E(X_j | \mathcal{G}_n) = E(X_j | S_n)$ pour $1 \leq j \leq n$. (*Indication* : Utiliser l'exercice 16.)

Chapitre 24

Martingales

Rappelons la loi forte des grands nombres (théorème 20.2) : si les v.a. $(X_n)_{n \geq 1}$ sont i.i.d. avec $E(X_n) = \mu$ et $\sigma_{X_n}^2 < \infty$ (cette dernière hypothèse pouvant d'ailleurs être supprimée), et si $S_n = \sum_{j=1}^n X_j$, alors $\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu$ p.s. Comme les X_n sont indépendantes, la limite doit effectivement être p.s. constante, en vertu de la loi zéro-un (théorème 10.6). Il est évidemment intéressant d'étudier aussi des suites de v.a. qui convergent p.s. vers une limite qui est une « vraie » variable aléatoire, non constante.

La propriété qui va permettre de généraliser la loi des grands nombres est la suivante : si on considère les tribus $\mathcal{F}_n = \sigma\{S_k; k \leq n\}$, alors

$$E(S_{n+1} - (n+1)\mu | \mathcal{F}_n) = S_n - n\mu, \quad (24.1)$$

(voir l'exemple 1 ci-dessous). C'est la propriété (24.1) qui est essentielle quand on veut étudier les convergences en affaiblissant l'hypothèse d'indépendance.

Dans la suite de ce chapitre on fixe l'espace de probabilité (Ω, \mathcal{F}, P) , ainsi que la suite croissante $(\mathcal{F}_n)_{n \geq 0}$ de sous-tribus de \mathcal{F} : on a donc $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ pour tout $n \geq 0$. Une telle suite de tribus s'appelle une *filtration*. On considérera aussi des suites $(X_n)_{n \geq 0}$ de v.a. réelles indicées par \mathbb{N} : dans le contexte de la théorie des martingales, une telle suite s'appelle aussi un *processus*, en référence à la description d'un phénomène qui évolue de manière aléatoire au cours du temps (les instants sont les entiers).

Définition 24.1. Une suite $(X_n)_{n \geq 0}$ de v.a. réelles (autrement dit, un processus) est appelée une martingale si

- (i) $E(|X_n|) < \infty$ pour chaque n ;
- (ii) X_n est \mathcal{F}_n -mesurable pour chaque n ;
- (iii) $E(X_n | \mathcal{F}_m) = X_m$ p.s. pour tous $m \leq n$.

Noter que (ii) est « presque » entraîné par (iii), qui implique que X_m est p.s. égale à une variable \mathcal{F}_m -mesurable.

Exemple 1. Soit $(X_n)_{n \geq 1}$ des v.a. indépendantes intégrables centrées ; soit $\mathcal{F}_n = \sigma\{X_k; k \leq n\}$ et $S_n = \sum_{k=1}^n X_k$ (avec $\mathcal{F}_0 = \{\emptyset, \Omega\}$ = « la tribu triviale » et $S_0 = 0$). La suite $(S_n)_{n \geq 0}$ est alors une martingale, ce qui suit de :

$$\begin{aligned} E(S_n | \mathcal{F}_m) &= E(S_m + (S_n - S_m) | \mathcal{F}_m) \\ &= S_m + E(S_n - S_m | \mathcal{F}_m) \\ &= S_m + E\left(\sum_{k=m+1}^n X_k | \mathcal{F}_m\right) \\ &= S_m + \sum_{k=m+1}^n E(X_k) \\ &= S_m. \end{aligned}$$

Si les X_n ne sont pas centrées, en utilisant les v.a. $X_n - \mu$ (où $\mu = E(X_1)$) on obtient de la même manière (24.1), i.e. la suite $(S_n - n\mu)$ est une martingale.

Exemple 2. Soit Y une v.a. intégrable et $(\mathcal{F}_n)_{n \geq 0}$ une famille croissante de sous-tribus. Si on pose

$$X_n = E(Y | \mathcal{F}_n)$$

on obtient une martingale. En effet on a $E(|X_n|) \leq E(|Y|) < \infty$ et pour $m \leq n$:

$$\begin{aligned} E(X_n | \mathcal{F}_m) &= E(E(Y | \mathcal{F}_n) | \mathcal{F}_m) \\ &= E(Y | \mathcal{F}_m) \\ &= X_m \end{aligned}$$

(voir les exercices 1 et 2 du chapitre précédent).

Définition 24.2. Une martingale $X = (X_n)_{n \geq 0}$ est dite fermée par une variable aléatoire Y si cette variable est intégrable et si $X_n = E(Y | \mathcal{F}_n)$ pour tout n .

L'exemple 2 montre que toute v.a. Y intégrable donne une martingale fermée, en posant précisément $X_n = E(Y | \mathcal{F}_n)$, $n \geq 0$.

Une des propriétés importantes des martingales est que leur espérance est constante :

Théorème 24.1. Si $(X_n)_{n \geq 0}$ est une martingale, alors $n \mapsto E(X_n)$ est constant, i.e. $E(X_n) = E(X_0)$ pour tout $n \geq 0$.

Preuve. $E(X_n) = E(E(X_n | \mathcal{F}_0)) = E(X_0)$. ■

La réciproque de ce théorème est grossièrement fautive, mais il existe une réciproque partielle qui utilise les temps d'arrêt : voir le théorème 24.7.

Définition 24.3. Une variable aléatoire $T : \Omega \rightarrow \bar{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ est appelée un temps d'arrêt si $\{T \leq n\} \in \mathcal{F}_n$ pour chaque n .

Toute constante entière, ou égale à $+\infty$, est évidemment un temps d'arrêt. Les temps d'arrêt sont parfois plus utiles que les temps « fixes ». Cette terminologie suggère clairement que \mathbb{N} est un ensemble de « temps » tandis que \mathcal{F}_n représente la famille des événements qu'on peut observer avant ou à l'instant n . On peut alors imaginer un temps d'arrêt comme le premier instant où une certaine suite de v.a. X_n , avec chaque X_n observable à l'instant n (i.e. \mathcal{F}_n -mesurable), vérifie une certaine propriété, avec la convention que le temps d'arrêt vaut $+\infty$ si cela n'arrive jamais. Par exemple (X_n) est une martingale, et T est le premier temps auquel elle atteint ou dépasse 12 ; on peut écrire

$$T = \begin{cases} \inf_{n \geq 0} \{n : X_n \geq 12\} & \text{si } X_n \geq 12 \text{ pour un } n \in \mathbb{N} \\ +\infty & \text{sinon,} \end{cases}$$

ou encore,

$$T(\omega) = \inf_{n \geq 0} \{n : X_n(\omega) \geq 12\}$$

si $X_n(\omega) \geq 12$ pour au moins un entier, et $T(\omega) = +\infty$ sinon. Noter que

$$\{T \leq n\} = \bigcup_{k=0}^n \{X_k \geq 12\} \in \mathcal{F}_n$$

parce que $\{X_k \geq 12\} \in \mathcal{F}_k \subset \mathcal{F}_n$ si $k \leq n$, donc T est bien un temps d'arrêt. La terminologie « temps d'arrêt » vient de la modélisation des jeux de hasard : un joueur peut arrêter de jouer à un temps aléatoire (dépendant de ses pertes antérieures par exemple), mais uniquement sur la base des événements effectivement observables avant l'instant ou à l'instant où il s'arrête : le lecteur vérifiera que mathématiquement, cette notion équivaut à celle de la définition 24.3.

Le théorème 24.1 s'étend sans difficulté aux temps d'arrêts T qui sont bornés (i.e., il existe une constante c telle que $T \leq c$ identiquement). Si T est un temps d'arrêt fini, on notera X_T la variable $X_T(\omega) = X_{T(\omega)}(\omega) = \sum_{n=0}^{\infty} X_n(\omega) 1_{\{T(\omega)=n\}}$.

Théorème 24.2. Soit T un temps d'arrêt borné par c , et $(X_n)_{n \geq 0}$ une martingale. On a alors $E(X_T) = E(X_0)$.

Preuve. En supposant (sans perte de généralité) que c est un entier, on a

$$\begin{aligned} E(X_T) &= E\left(\sum_{n=0}^{\infty} X_n 1_{\{T=n\}}\right) \\ &= E\left(\sum_{n=0}^c X_n 1_{\{T=n\}}\right) \\ &= \sum_{n=0}^c E(X_n 1_{\{T=n\}}). \end{aligned}$$

Comme $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$ on voit que $\{T = n\} \in \mathcal{F}_n$, et par suite

$$\begin{aligned} &= \sum_{n=0}^c E(E(X_c | \mathcal{F}_n) 1_{\{T=n\}}) \\ &= \sum_{n=0}^c E(X_c 1_{\{T=n\}}) \\ &= E\left(X_c \sum_{n=0}^c 1_{\{T=n\}}\right) \\ &= E(X_c) = E(X_0), \end{aligned}$$

où la dernière égalité vient du théorème 24.1. ■

La tribu \mathcal{F}_n représente la classe des événements observables jusqu'à l'instant n ; on veut généraliser cette notion aux temps d'arrêt.

Définition 24.4. Soit T un temps d'arrêt. La tribu antérieure à T est la tribu

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n \text{ pour tout } n\}.$$

Pour que cette définition ait un sens, nous avons besoin du résultat suivant :

Théorème 24.3. Si T est un temps d'arrêt, la classe \mathcal{F}_T définie ci-dessus est une tribu.

Preuve. Que Ω soit dans \mathcal{F}_T est évident. Si $A \in \mathcal{F}_T$ on a

$$A^c \cap \{T \leq n\} = \{T \leq n\} \setminus (A \cap \{T \leq n\}),$$

donc $A^c \in \mathcal{F}_T$. Enfin si les $(A_i)_{i \geq 1}$ sont dans \mathcal{F}_T , alors

$$\left(\bigcup_{i=1}^{\infty} A_i \right) \cap \{T \leq n\} = \bigcup_{i=1}^{\infty} (A_i \cap \{T \leq n\}) \in \mathcal{F}_n.$$

Donc \mathcal{F}_T , étant stable par complémentation et par réunions dénombrables, est une tribu. ■

Théorème 24.4. *Si S et T sont des temps d'arrêt vérifiant $S \leq T$, on a $\mathcal{F}_S \subset \mathcal{F}_T$.*

Preuve. Comme $S \leq T$ on a $\{T \leq n\} \subset \{S \leq n\}$. Donc si $A \in \mathcal{F}_S$ il vient

$$A \cap \{T \leq n\} = A \cap \{S \leq n\} \cap \{T \leq n\};$$

mais $A \cap \{S \leq n\} \in \mathcal{F}_n$ et $\{T \leq n\} \in \mathcal{F}_n$, donc $A \cap \{T \leq n\} \in \mathcal{F}_n$, donc $A \in \mathcal{F}_T$. ■

Dans le théorème suivant on suppose donnée une suite $(X_n)_{n \geq 0}$ de v.a. réelles, telle que chaque X_n soit \mathcal{F}_n -mesurable. Soit T un temps d'arrêt fini, de sorte que X_T soit bien défini.

Théorème 24.5. X_T est \mathcal{F}_T -mesurable.

Preuve. Soit $\Lambda \in \mathcal{B}$. On veut montrer que $\{X_T \in \Lambda\} \in \mathcal{F}_T$, ce qui signifie que pour tout n on a $\{X_T \in \Lambda\} \cap \{T \leq n\} \in \mathcal{F}_n$. Mais

$$\begin{aligned} \{X_T \in \Lambda\} \cap \{T \leq n\} &= \bigcup_{k=1}^n \{X_T \in \Lambda\} \cap \{T = k\} \\ &= \bigcup_{k=1}^n \{X_k \in \Lambda\} \cap \{T = k\}, \end{aligned}$$

et $\{X_k \in \Lambda\} \cap \{T = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ pour $k \leq n$. ■

Les deux théorèmes suivants montrent que la « propriété de martingale » est valide pour les temps d'arrêt aussi bien que pour les temps fixes. Il s'agit de résultats simples, mais très puissants.

Théorème 24.6. (Théorème d'arrêt de Doob.) *Soit $(X_n)_{n \geq 0}$ une martingale. Soit S et T deux temps d'arrêt bornés par une constante c , avec $S \leq T$ p.s. On a alors*

$$E(X_T | \mathcal{F}_S) = X_S \quad \text{p.s.}$$

Preuve. D'abord, $|X_T| \leq \sum_{n=0}^c |X_n|$ est intégrable (on a supposé, sans perte de généralité, que c est un entier). De plus X_S est \mathcal{F}_S -mesurable en vertu du théorème précédent. Il reste donc à prouver que $E(X_T Z) = E(X_S Z)$ pour toute v.a. Z bornée et \mathcal{F}_S -mesurable. Grâce à un argument standard il suffit même de montrer que pour tout $A \in \mathcal{F}_S$, alors

$$E(X_T 1_A) = E(X_S 1_A)$$

(ceci implique par linéarité qu'on a $E(X_T Z) = E(X_S Z)$ pour les v.a. Z qui sont « simples », puis par limite croissante que c'est vrai pour Z positive, et enfin par différence que c'est vrai pour Z bornée).

Soit donc $A \in \mathcal{F}_S$. Définissons un nouveau « temps aléatoire » R par

$$R(\omega) = S(\omega)1_A(\omega) + T(\omega)1_{A^c}(\omega).$$

R est aussi un temps d'arrêt : en effet, on a

$$\{R \leq n\} = (A \cap \{S \leq n\}) \cup (A^c \cap \{T \leq n\}),$$

et $A \cap \{S \leq n\} \in \mathcal{F}_n$ car $A \in \mathcal{F}_S$. Comme $A \in \mathcal{F}_S$ on a $A^c \in \mathcal{F}_S$, donc $A^c \in \mathcal{F}_T$ par le théorème 24.4. Donc $A^c \cap \{T \leq n\} \in \mathcal{F}_n$, d'où on déduit que $\{R \leq n\} \in \mathcal{F}_n$. Donc $E(X_R) = E(X_T) = E(X_0)$ par le théorème 24.2. Mais

$$E(X_R) = E(X_S 1_A + X_T 1_{A^c}),$$

$$E(X_T) = E(X_T 1_A + X_T 1_{A^c})$$

et en soustrayant on obtient

$$E(X_S 1_A) - E(X_T 1_A) = 0. \quad \blacksquare$$

On peut maintenant établir une réciproque partielle au théorème 24.1.

Théorème 24.7. Soit $(X_n)_{n \geq 0}$ une suite de v.a. intégrables, telle que chaque X_n soit \mathcal{F}_n -mesurable. Si $E(X_T) = E(X_0)$ pour tout temps d'arrêt T borné, alors (X_n) est une martingale.

Preuve. Soit $0 \leq m < n < \infty$ et $\Lambda \in \mathcal{F}_m$. Posons

$$T(\omega) = \begin{cases} m & \text{si } \omega \in \Lambda^c, \\ n & \text{si } \omega \in \Lambda \end{cases}$$

T est un temps d'arrêt (vérification immédiate), donc

$$E(X_0) = E(X_T) = E(X_m 1_{\Lambda^c} + X_n 1_{\Lambda}).$$

Cependant on a aussi $E(X_0) = E(X_m 1_{\Lambda^c} + X_m 1_{\Lambda})$ (appliquer l'hypothèse avec $T \equiv m$). En soustrayant, on obtient $E(X_n 1_{\Lambda}) = E(X_m 1_{\Lambda})$, et comme Λ est arbitraire dans \mathcal{F}_m on obtient $E(X_n | \mathcal{F}_m) = X_m$ p.s. ■

Exercices

Dans les exercices 1 à 11 on considère deux temps d'arrêt S et T relatifs à une suite croissante donnée $(\mathcal{F}_n)_{n \geq 0}$ de sous-tribus.

1. Si $T \equiv n$ montrer que $\mathcal{F}_T = \mathcal{F}_n$.
2. Montrer que $S \wedge T = \min(S, T)$ est un temps d'arrêt.
3. Montrer que $S \vee T = \max(S, T)$ est un temps d'arrêt.
4. Montrer que $S + T$ est un temps d'arrêt.
5. Montrer que aT est un temps d'arrêt pour a entier supérieur ou égal à 1.
6. Montrer que $\mathcal{F}_{S \wedge T} \subset \mathcal{F}_T \subset \mathcal{F}_{S \vee T}$.
7. Montrer que T est un temps d'arrêt si et seulement si $\{T = n\} \in \mathcal{F}_n$ pour tout $n \geq 0$.
8. Soit $\Lambda \in \mathcal{F}_T$ et posons

$$T_{\Lambda}(\omega) = \begin{cases} T(\omega) & \text{si } \omega \in \Lambda, \\ \infty & \text{si } \omega \notin \Lambda. \end{cases}$$

Montrer que T_{Λ} est aussi un temps d'arrêt.

9. Montrer que T est \mathcal{F}_T -mesurable.
10. Montrer que $\{S < T\}$, $\{S \leq T\}$, et $\{S = T\}$ sont tous dans $\mathcal{F}_S \cap \mathcal{F}_T$.
- 11.* Montrer que si Y est une v.a. intégrable,

$$E(E(Y | \mathcal{F}_T) | \mathcal{F}_S) = E(E(Y_{\text{id}} \mathcal{F}_S) | \mathcal{F}_T) = E(Y | \mathcal{F}_{S \wedge T}).$$

12. Soit $(M_n)_{n \geq 0}$ une martingale avec $M_n \in L^2$ pour tout n . Soit S et T des temps d'arrêt bornés avec $S \leq T$. Montrer que M_S et M_T sont dans L^2 , que

$$E((M_T - M_S)^2 | \mathcal{F}_S) = E(M_T^2 - M_S^2 | \mathcal{F}_S),$$

et que

$$E((M_T - M_S)^2) = E(M_T^2) - E(M_S^2).$$

13. Soit φ une fonction convexe et soit $(M_n)_{n \geq 0}$ une martingale. Montrer que $n \rightarrow E(\varphi(M_n))$ est une fonction croissante. (*Indication* : Utiliser l'inégalité de Jensen [théorème 23.9].)
14. Soit X_n une suite de v.a. réelles intégrables avec $E(X_n | \mathcal{F}_{n-1}) = 0$ et X_n \mathcal{F}_n -mesurable pour chaque n . Soit $S_n = \sum_{k=0}^n X_k$. Montrer que $(S_n)_{n \geq 0}$ est une martingale.

Chapitre 25

Surmartingales et sous-martingales

Au chapitre 24 nous avons défini les martingales via une égalité pour certaines espérances conditionnelles. Si on remplace cette égalité par une inégalité, on obtient les surmartingales et les sous-martingales. Nous supposons de nouveau donné un espace de probabilité (Ω, \mathcal{F}, P) muni d'une suite croissante $(\mathcal{F}_n)_{n \geq 0}$ de sous-tribus de \mathcal{F} .

Définition 25.1. Une suite de v.a. réelles $(X_n)_{n \geq 0}$ (un processus) est appelée une sous-martingale (resp. une surmartingale) si

- (i) $E(|X_n|) < \infty$ pour chaque n ;
- (ii) X_n est \mathcal{F}_n -mesurable pour chaque n ;
- (iii) $E(X_n | \mathcal{F}_m) \geq X_m$ p.s. (resp. $\leq X_m$ p.s.) pour tous $m \leq n$.

La suite $(X_n)_{n \geq 0}$ est une martingale si et seulement si c'est à la fois une sous-martingale et une surmartingale.

Théorème 25.1. Si $(M_n)_{n \geq 0}$ est une martingale et si φ est une fonction convexe telle que chaque $\varphi(M_n)$ soit intégrable, alors $(\varphi(M_n))_{n \geq 0}$ est une sous-martingale.

Preuve. Soit $m \leq n$. On a $E(M_n | \mathcal{F}_m) = M_m$ p.s., donc $\varphi(E(M_n | \mathcal{F}_m)) = \varphi(M_m)$ p.s., et comme φ est convexe l'inégalité de Jensen (théorème 23.9) donne

$$E(\varphi(M_n) | \mathcal{F}_m) \geq \varphi(E(M_n | \mathcal{F}_m)) = \varphi(M_m).$$

Corollaire 25.1. Si $(M_n)_{n \geq 0}$ est une martingale, alors $(|M_n|)_{n \geq 0}$ est une sous-martingale.

Preuve. Appliquer le théorème 25.1 avec $\varphi(x) = |x|$.

Théorème 25.2. Soit T un temps d'arrêt borné par $C \in \mathbb{N}$, et $(X_n)_{n \geq 0}$ une sous-martingale. On a alors $E(X_T) \leq E(X_C)$.

Preuve. La démonstration, analogue à celle du théorème 24.2, est omise.

Le résultat suivant explique un autre type de rapports entre sous-martingales et martingales.

Théorème 25.3. (Décomposition de Doob.) Soit $X = (X_n)_{n \geq 0}$ une sous-martingale. Il existe une martingale $M = (M_n)_{n \geq 0}$ avec $M_0 = 0$ et une suite de v.a. $A = (A_n)_{n \geq 0}$ avec $A_{n+1} \geq A_n$ p.s. et $A_0 = 0$ (on dit : un processus croissant), chaque variable A_{n+1} étant en outre \mathcal{F}_n -mesurable, tels que

$$X_n = X_0 + M_n + A_n.$$

De plus, cette décomposition est p.s. unique.

Preuve. Posons $A_0 = 0$ et

$$A_n = \sum_{k=1}^n E(X_k - X_{k-1} | \mathcal{F}_{k-1}) \quad \text{pour } n \geq 1.$$

X étant une sous-martingale, on a $E\{X_k - X_{k-1} | \mathcal{F}_{k-1}\} \geq 0$ p.s., donc $A_{k+1} \geq A_k$ p.s., tandis que par construction A_{k+1} est \mathcal{F}_k -mesurable. Par ailleurs on a

$$E(X_n | \mathcal{F}_{n-1}) - X_{n-1} = E(X_n - X_{n-1} | \mathcal{F}_{n-1}) = A_n - A_{n-1},$$

et donc

$$E(X_n | \mathcal{F}_{n-1}) - A_n = X_{n-1} - A_{n-1};$$

mais $A_n \in \mathcal{F}_{n-1}$, donc

$$E(X_n - A_n | \mathcal{F}_{n-1}) = X_{n-1} - A_{n-1}. \quad (25.1)$$

Si on pose $M_n = X_n - A_n - X_0$ il suit de (25.1) que $M = (M_n)_{n \geq 0}$ est une martingale avec $M_0 = 0$, et on a la décomposition souhaitée.

Quant à l'unicité, supposons que

$$X_n = X_0 + M_n + A_n, \quad n \geq 0.$$

$$X_n = X_0 + L_n + C_n, \quad n \geq 0,$$

soient deux décompositions. En soustrayant l'une de l'autre, on arrive à

$$L_n - M_n = A_n - C_n. \quad (25.2)$$

Comme A_n et C_n sont \mathcal{F}_{n-1} -mesurables, il en est de même de $L_n - M_n$, et donc

$$L_n - M_n = E(L_n - M_n | \mathcal{F}_{n-1}) = L_{n-1} - M_{n-1} = A_{n-1} - C_{n-1} \quad \text{p.s.}$$

Par une récurrence évidente on obtient alors que $L_n - M_n = L_0 - M_0 = 0$ p.s. (car $L_0 = M_0 = 0$). Par suite $L_n = M_n$ p.s., donc aussi $A_n = C_n$ p.s., et l'unicité est démontrée. ■

Corollaire 25.2. Soit $X = (X_n)_{n \geq 0}$ une surmartingale. Il existe une (p.s.) unique décomposition

$$X_n = X_0 + M_n - A_n, \quad n \geq 0$$

avec $(M_n)_{n \geq 0}$ et $(A_n)_{n \geq 0}$ comme dans le théorème 25.3.

Preuve. Soit $Y_n = -X_n$. La suite $(Y_n)_{n \geq 0}$ est une sous-martingale, qui admet donc une décomposition de Doob p.s. unique

$$Y_n = Y_0 + L_n + C_n,$$

et il suffit de poser $M_n = -L_n$ et $A_n = C_n$. ■

Exercices

1. Montrer que $X = (X_n)_{n \geq 0}$ est une sous-martingale si et seulement si $-X = (-X_n)_{n \geq 0}$ est une surmartingale.
2. Montrer que si $X = (X_n)_{n \geq 0}$ est à la fois une sous-martingale et une surmartingale, c'est une martingale.
3. Soit $(X_n)_{n \geq 0}$ une sous-martingale de décomposition de Doob $X_n = X_0 + M_n + A_n$. Montrer que $E(A_n) < \infty$ pour tout n .
4. Soit $M = (M_n)_{n \geq 0}$ une martingale avec $M_0 = 0$, telle que $E(M_n^2) < \infty$ pour tout n , et soit $X_n = M_n^2$. Montrer que $X = (X_n)_{n \geq 0}$ est une sous-martingale, et que sa décomposition de Doob $X_n = L_n + A_n$ vérifie $E(M_n^2) = E(A_n)$.
5. Soit M et A comme dans l'exercice 4. Montrer que $A_n - A_{n-1} = E((M_n - M_{n-1})^2 | \mathcal{F}_{n-1})$.
6. Soit $(X_n)_{n \geq 0}$ une sous-martingale. Montrer que si φ est une fonction convexe croissante sur \mathbb{R} telle que chaque v.a. $\varphi(X_n)$ soit intégrable, alors $(\varphi(X_n))_{n \geq 0}$ est aussi une sous-martingale.
7. Soit $(X_n)_{n \geq 0}$ une suite croissante de v.a. intégrables, chaque X_n étant \mathcal{F}_n -mesurable. Montrer que cette suite est une sous-martingale.

Chapitre 26

Les inégalités de martingales

L'une des raisons qui font que les martingales jouent un rôle central en probabilité est que leur structure engendre plusieurs inégalités très puissantes. Nous suivons ci-dessous la présentation de Bass [1].

L'espace de probabilité (Ω, \mathcal{F}, P) et la suite croissante $(\mathcal{F}_n)_{n \geq 0}$ de sous-tribus de \mathcal{F} sont encore fixés. Soit $M = (M_n)_{n \geq 0}$ une suite de v.a. intégrables, chaque M_n étant \mathcal{F}_n -mesurable. On pose $M_n^* = \sup_{j \leq n} |M_j|$. Remarquer que $(M_n^*)_{n \geq 0}$ est un « processus croissant », et donc une sous-martingale (voir l'exercice 7 du chapitre 25), puisque

$$E(M_n^*) \leq E\left(\sum_{j=1}^n |M_j|\right) < \infty.$$

Grâce à l'inégalité de Markov (corollaire 5.1), on a

$$P(M_n^* \geq \alpha) = E(I_{\{M_n^* \geq \alpha\}}) \leq \frac{E(M_n^*)}{\alpha}.$$

Notre première inégalité indique que dans le cas d'une martingale on peut remplacer M_n^* par $|M_n|$ dans le membre de droite : on a donc un contrôle du supremum pour les temps $0, 1, \dots, n$ par la variable « terminale » $|M_n|$.

Théorème 26.1. (Inégalité maximale de Doob.) *Soit $M = (M_n)_{n \geq 0}$ une martingale, ou une sous-martingale positive. On a alors*

$$P(M_n^* \geq \alpha) \leq \frac{E(|M_n|)}{\alpha}.$$

Preuve. Soit $T = \min\{j : |M_j| \geq \alpha\}$ (rappelons notre convention selon laquelle l'infimum d'un ensemble vide d'entiers vaut $+\infty$). Comme la fonction $\varphi(x) = |x|$ est convexe et croissante sur \mathbb{R}_+ , le processus $(|M_n|)$ est une sous-martingale (théorème 25.1 si M est une martingale, exercice 24.6 si M est une sous-martingale positive). Les ensembles $\{T \leq n, |M_T| \geq \alpha\}$ et $\{M_n^* \geq \alpha\}$ sont égaux, donc

$$P(M_n^* \geq \alpha) = P(T \leq n, |M_T| \geq \alpha) \leq E\left(\frac{|M_T|}{\alpha} I_{\{T \leq n\}}\right).$$

et comme $M_T = M_{T \wedge n}$ sur $\{T \leq n\}$, on a

$$P(M_n^* \geq \alpha) \leq \frac{1}{\alpha} E(|M_{T \wedge n}| 1_{\{T \leq n\}}) \leq \frac{E(|M_{T \wedge n}|)}{\alpha} \leq \frac{E(|M_n|)}{\alpha}$$

par le théorème 25.2. ■

Avant de prouver notre seconde inégalité, nous avons besoin d'un lemme intéressant en lui-même.

Lemme 26.1. *Soit X une v.a. positive et $p > 0$. On a alors*

$$E(X^p) = \int_0^\infty p\lambda^{p-1} P(X > \lambda) d\lambda.$$

(Les deux membres de l'égalité ci-dessus peuvent être égaux à $+\infty$.)

Preuve. On a

$$\int_0^\infty p\lambda^{p-1} P(X > \lambda) d\lambda = \int_0^\infty p\lambda^{p-1} E(1_{\{X > \lambda\}}) d\lambda,$$

et le théorème de Fubini (voir l'exercice 15 du chapitre 10) donne

$$= E\left(\int_0^\infty p\lambda^{p-1} 1_{\{X > \lambda\}} d\lambda\right) = E\left(\int_0^X p\lambda^{p-1} d\lambda\right) = E(X^p). \quad \blacksquare$$

Théorème 26.2. (Inégalité de Doob.) *Soit $M = (M_n)_{n \geq 0}$ une sous-martingale positive ou une martingale (de signe quelconque). Soit $1 < p < \infty$. Il existe une constante c_p dépendant seulement de p (mais pas de M) telle que*

$$E((M_n^*)^p) \leq c_p E(|M_n|^p).$$

Preuve. Le processus $(|M_n|)$ est une sous-martingale : c'est évident si M est une sous-martingale positive, et lorsque M est une martingale c'est le corollaire 25.1. Soit $X_n = M_n 1_{\{|M_n| > \frac{\alpha}{2}\}}$. Pour n fixé, posons

$$Z_j = E(X_n | \mathcal{F}_j), \quad 0 \leq j \leq n.$$

Noter que $(Z_j)_{0 \leq j \leq n}$ est une martingale, et aussi que $M_n^* \leq Z_n^* + \frac{\alpha}{2}$, puisque

$$\begin{aligned} |M_j| &\leq E(|M_n| | \mathcal{F}_j) \\ &= E(|M_n| 1_{\{|M_n| > \frac{\alpha}{2}\}}) + |M_n| 1_{\{|M_n| \leq \frac{\alpha}{2}\}} | \mathcal{F}_j) \\ &= E(X_n + |M_n| 1_{\{|M_n| \leq \frac{\alpha}{2}\}} | \mathcal{F}_j) \\ &\leq E(X_n | \mathcal{F}_j) + \frac{\alpha}{2} \\ &= Z_j + \frac{\alpha}{2}. \end{aligned}$$

Le théorème 26.1 implique alors que

$$\begin{aligned} \mathbb{P}(M_n^* > \alpha) &\leq \mathbb{P}\left(Z_n^* > \frac{\alpha}{2}\right) \\ &\leq \frac{2}{\alpha} \mathbb{E}(Z_n) = \frac{2}{\alpha} \mathbb{E}(X_n) \\ &= \frac{2}{\alpha} \mathbb{E}(|M_n| 1_{\{|M_n| > \frac{\alpha}{2}\}}). \end{aligned}$$

D'après le lemme 26.1 il vient

$$\begin{aligned} \mathbb{E}((M_n^*)^p) &= \int_0^\infty p\lambda^{p-1} \mathbb{P}(M_n^* > \lambda) d\lambda \\ &\leq \int_0^\infty 2p\lambda^{p-2} \mathbb{E}(|M_n| 1_{\{|M_n| > \frac{\lambda}{2}\}}) d\lambda \end{aligned}$$

et le théorème de Fubini (voir l'exercice 15 du chapitre 10) entraîne

$$\begin{aligned} \mathbb{E}((M_n^*)^p) &= \mathbb{E}\left(\int_0^{2|M_n|} 2p\lambda^{p-2} d\lambda \mid M_n\right) \\ &= \frac{2^p p}{p-1} \mathbb{E}(|M_n|^p). \end{aligned}$$

Remarquer que cette démonstration donne la valeur $\frac{2^p p}{p-1}$ à la constante c_p . Avec un peu plus de travail, on peut montrer que $c_p = (\frac{p}{p-1})^p$ (c'est la valeur optimale de c_p), de sorte qu'on peut reformuler le théorème 26.2 ainsi :

Théorème 26.3. (Inégalité de Doob.) *Soit $M = (M_n)_{n \geq 0}$ une sous-martingale positive ou une martingale. Soit $1 < p < \infty$. On a alors*

$$\mathbb{E}((M_n^*)^p)^{\frac{1}{p}} \leq \frac{p}{p-1} \mathbb{E}(|M_n|^p)^{\frac{1}{p}},$$

ou en utilisant la notation de la norme de L^p :

$$\|M_n^*\|_p \leq \frac{p}{p-1} \|M_n\|_p.$$

La dernière inégalité de ce chapitre sera utilisée pour démontrer le théorème de convergence des martingales au chapitre 27. Introduisons d'abord la notion de « montées » : soit $X = (X_n)_{n \geq 0}$ un processus, et soit $a < b$. Le nombre de montées de X de a à b est le nombre de fois où

le processus X « monte » d'un point en dessous de a jusqu'à un point au dessus de b à un instant ultérieur. On peut exprimer ceci d'une manière plus mathématique en utilisant les temps d'arrêt. Posons

$$T_0 = 0,$$

et définissons par récurrence sur $j \geq 0$ les S_{j+1} et T_{j+1} :

$$S_{j+1} = \min\{k \geq T_j : X_k \leq a\}, \quad T_{j+1} = \min\{k > S_{j+1} : X_k \geq b\}, \quad (26.1)$$

avec la convention usuelle que l'infimum de l'ensemble vide vaut $+\infty$; avec la convention symétrique que le supremum de l'ensemble vide vaut 0, on peut enfin poser

$$U_n = \max\{j : T_j \leq n\} \quad (26.2)$$

et U_n est exactement le nombre de montées de a à b pour X , entre les instants 0 et n .

Théorème 26.4. (Inégalité des montées de Doob.) Soit $X = (X_n)_{n \geq 0}$ une sous-martingale. Soit $a < b$ et notons U_n le nombre de montées de a à b pour X , entre les temps 0 et n (cf. (26.2)). On a alors

$$E(U_n) \leq \frac{1}{b-a} E((X_n - a)^+)$$

où $(X_n - a)^+ = \max(X_n - a, 0)$.

Preuve. Soit $Y_n = (X_n - a)^+$. La fonction $\varphi(x) = (x - a)^+$ étant convexe et croissante, d'après l'exercice 6 du chapitre 25, $Y = (Y_n)_{n \geq 0}$ est une sous-martingale positive. Comme à l'évidence $S_{n+1} > n$, on a l'égalité suivante :

$$Y_n = Y_{S_1 \wedge n} + \sum_{i=1}^n (Y_{T_i \wedge n} - Y_{S_i \wedge n}) + \sum_{i=1}^n (Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}). \quad (26.3)$$

Par ailleurs chaque montée de X entre les instants 0 et n correspond à l'existence d'un entier i tel que $S_i < T_i \leq n$, avec $Y_{S_i} = 0$ et $Y_{T_i} = Y_{T_i \wedge n} \geq b - a$, tandis qu'on a par construction $Y_{T_i \wedge n} - Y_{S_i \wedge n} \geq 0$ pour tout i . Par suite

$$\sum_{i=1}^n (Y_{T_i \wedge n} - Y_{S_i \wedge n}) \geq (b - a)U_n.$$

Étant donné (26.3), et comme $Y_{S_i \wedge n} \geq 0$, il vient alors

$$(b-a)U_n \leq Y_n - \sum_{i=1}^n (Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}).$$

On prend alors l'espérance des deux membres. Comme Y est une sous-martingale, comme les temps d'arrêt $T_i \wedge n$ et $S_{i+1} \wedge n$ sont bornés (par n) et comme $T_i \wedge n \leq S_{i+1} \wedge n$, le théorème d'arrêt 25.2 implique que l'espérance de chacun des termes de la somme à droite de (26.3) est positive, de sorte qu'on obtient

$$(b-a)E(U_n) \leq E(Y_n). \quad \blacksquare$$

Exercices

1. Soit $Y_n \in L^2$ des variables aléatoires telles que $\lim_{n \rightarrow \infty} E(Y_n^2) = 0$. Soit $(\mathcal{F}_k)_{k \geq 0}$ une suite croissante de sous-tribus et posons $X_k^n = E(Y_n | \mathcal{F}_k)$. Montrer que $\lim_{n \rightarrow \infty} E(\sup_k (X_k^n)^2) = 0$.
2. Soit X et Y des v.a. positives vérifiant

$$\alpha P(X \geq \alpha) \leq E(Y 1_{\{X \geq \alpha\}}),$$

pour tout $\alpha > 0$. Montrer que si $p > 1$ et $\frac{1}{p} + \frac{1}{q} = 1$, on a $E(X^p) \leq E(qX^{p-1}Y)$.

3. Soit X et Y comme dans l'exercice 2. Supposons que $\|X\|_p < \infty$ et que $\|Y\|_p < \infty$. Montrer que $\|X\|_p \leq q\|Y\|_p$. (*Indication* : Utiliser l'exercice 2 et l'inégalité de Hölder.)
4. Établir le résultat de l'exercice 3 sans l'hypothèse $\|X\|_p < \infty$.
- 5* Utiliser l'exercice 4 pour démontrer le théorème 26.3.

Chapitre 27

Les théorèmes de convergence de martingales

Dans le chapitre 17 nous avons étudié des théorèmes de convergence du type suivant : une certaine forme de convergence, plus éventuellement telle ou telle condition supplémentaire, entraîne une autre forme de convergence. Pour les martingales, on a un phénomène un peu surprenant : aucune forme de convergence n'est *a priori* supposée – seulement une certaine structure – et on obtient la convergence. Cela rend ce genre de théorème singulier en analyse, la seule situation un peu analogue ne se rencontrant qu'en théorie ergodique.

Théorème 27.1. (Théorème de convergence des sous-martingales.) *Soit $(X_n)_{n \geq 0}$ une sous-martingale telle que $\sup_n E(X_n^+) < \infty$. Alors la suite X_n converge p.s. vers une limite X à valeurs réelles. De plus X est dans L^1 . Attention : nous n'affirmons pas que X_n converge vers X dans L^1 ; ceci est faux en général.*

Preuve. Pour $a < b$ notons $U_n(a, b)$ le nombre de montées de la suite (X_k) de a à b , entre les instants 0 et n , ainsi que nous l'avons défini en (26.2). La suite $U_n(a, b)$ est évidemment croissante en n , et elle converge donc (pour tout ω) vers une limite $U(a, b)$, à valeurs dans $\mathbb{N} \cup \{+\infty\}$. En utilisant le théorème de convergence monotone, le théorème 26.4 et le fait que $(x - a)^+ \leq x^+ + |a|$ pour tous réels a, x , on obtient

$$\begin{aligned} E(U(a, b)) &= \lim_{n \rightarrow \infty} E(U_n(a, b)) \\ &\leq \frac{1}{b-a} \sup_n E((X_n - a)^+) \\ &\leq \frac{1}{b-a} \left(\sup_n E(X_n^+) + |a| \right) \leq \frac{c}{b-a} < \infty \end{aligned}$$

pour une certaine constante c finie. Comme $E(U(a, b)) < \infty$, on a aussi $P(U(a, b) < \infty) = 1$. C'est-à-dire que la suite X_n « monte » de a à b presque sûrement un nombre fini de fois le long de *tous les entiers*. Il s'ensuit que si

$$\Lambda_{a,b} = \{ \limsup_{n \rightarrow \infty} X_n \geq b ; \liminf_{n \rightarrow \infty} X_n \leq a \},$$

alors $P(\Lambda_{a,b}) = 0$. Si $\Lambda = \bigcup_{\substack{a < b \\ a, b \in \mathbb{Q}}} \Lambda_{a,b}$ (où \mathbb{Q} désigne l'ensemble des rationnels), et comme la famille des couples de rationnels est dénombrable, on a aussi $P(\Lambda) = 0$. Mais

$$\Lambda = \{ \limsup_n X_n > \liminf_n X_n \}.$$

et on en conclut que X_n converge p.s. vers une limite X .

Noter qu'il est pour le moment possible *a priori* que X prenne les valeurs $+\infty$ et $-\infty$, et nous devons montrer que ces deux éventualités sont en fait impossibles. Comme (X_n) est une sous-martingale on a $E(X_n) \geq E(X_0)$, et donc

$$\begin{aligned} E(|X_n|) &= E(X_n^+) + E(X_n^-) \\ &= 2E(X_n^+) - E(X_n) \\ &\leq 2E(X_n^+) - E(X_0), \end{aligned} \tag{27.1}$$

donc

$$E(|X|) = E(\lim_n |X_n|) \leq \liminf_{n \rightarrow \infty} E(|X_n|) \leq 2 \sup_n E(X_n^+) - E(X_0) < \infty,$$

par le lemme de Fatou et (27.1) combiné à l'hypothèse que $\sup_n E(X_n^+) < \infty$. Donc X est intégrable (dans L^1), et donc *a fortiori* p.s. fini. ■

Corollaire 27.1. *Si (X_n) est une surmartingale positive, ou une martingale bornée inférieurement, ou bornée supérieurement (par une constante), la suite (X_n) converge p.s. vers une limite X appartenant à L^1 .*

Preuve. Si (X_n) est une surmartingale positive, alors $(-X_n)$ est une sous-martingale négative et on peut appliquer le théorème 27.1.

Si (X_n) est une martingale bornée inférieurement, disons par a , alors $(X_n - a)$ est une martingale positive et il suffit d'appliquer ce qui précède. Si enfin (X_n) est une martingale bornée supérieurement, il suffit de considérer $(-X_n)$. ■

Le théorème 27.1 donne la convergence presque sûre vers une v.a. X qui est dans L^1 . Mais il ne donne pas la convergence dans L^1 . Pour l'obtenir il nous faut une hypothèse légèrement plus forte, pour laquelle nous avons besoin du concept d'intégrabilité uniforme.

Définition 27.1. Un sous-ensemble \mathcal{H} de L^1 est dit uniformément intégrable si

$$\lim_{c \rightarrow \infty} \sup_{X \in \mathcal{H}} E(1_{\{|X| \geq c\}} |X|) = 0.$$

Voici deux conditions assurant l'uniforme intégrabilité.

Théorème 27.2. Soit \mathcal{H} une famille de v.a. réelles.

- (a) Si $\sup_{X \in \mathcal{H}} E(|X|^p) < \infty$ pour un $p > 1$, la famille \mathcal{H} est uniformément intégrable.
- (b) S'il existe une v.a. Y intégrable telle que $|X| \leq Y$ p.s. pour tout $X \in \mathcal{H}$, la famille \mathcal{H} est uniformément intégrable.

Preuve. (a) Soit $k = \sup_{X \in \mathcal{H}_n} E(|X|^p)$. Si $x \geq c > 0$, on a $x^{1-p} \leq c^{1-p}$, et en multipliant par x^p on obtient $x \leq c^{1-p} x^p$. Par suite

$$E(|X|1_{\{|X| > c\}}) \leq c^{1-p} E(|X|^p 1_{\{|X| > c\}}) \leq \frac{k}{c^{p-1}},$$

donc $\lim_{c \rightarrow \infty} \sup_{X \in \mathcal{H}} E(|X|1_{\{|X| > c\}}) \leq \lim_{c \rightarrow \infty} \frac{k}{c^{p-1}} = 0$.

(b) Comme $|X| \leq Y$ a.s. pour tout $X \in \mathcal{H}$, on a

$$|X|1_{\{|X| > c\}} \leq Y1_{\{Y > c\}}.$$

Mais $\lim_{c \rightarrow \infty} Y1_{\{Y > c\}} = 0$ p.s.; donc le théorème de convergence dominée de Lebesgue entraîne que

$$\begin{aligned} \lim_{c \rightarrow \infty} \sup_{X \in \mathcal{H}} E(|X|1_{\{|X| > c\}}) &\leq \lim_{c \rightarrow \infty} E(Y1_{\{Y > c\}}) \\ &= E(\lim_{c \rightarrow \infty} Y1_{\{Y > c\}}) = 0. \end{aligned}$$

Pour des résultats complémentaires sur l'intégrabilité uniforme, nous recommandons [18, p. 16-21] ou [19].

Théorème 27.3. (Théorème de convergence des martingales.)

- (a) Soit $(X_n)_{n \geq 1}$ une martingale. Si la suite de v.a. $(X_n)_{n \geq 1}$ est uniformément intégrable, elle converge p.s. et dans L^1 vers une v.a. X intégrable, et de plus on a $X_n = E(X | \mathcal{F}_n)$ pour tout n .
- (b) Inversement, soit $Y \in L^1$ et considérons la martingale $X_n = E(Y | \mathcal{F}_n)$. La suite $(X_n)_{n \geq 1}$ est alors uniformément intégrable.

En utilisant la terminologie de la définition 24.2, nous avons donc que (X_n) est une martingale fermée si et seulement si elle est uniformément intégrable.

Preuve. (a) Comme $(X_n)_{n \geq 1}$ est uniformément intégrable, pour tout $\varepsilon > 0$ il existe c tel que $\sup_n E(|X_n|1_{\{|X_n|>c\}}) \leq \varepsilon$. Donc

$$E(|X_n|) = E(|X_n|1_{\{|X_n|>c\}}) + E(|X_n|1_{\{|X_n|\leq c\}}) \leq \varepsilon + c.$$

Par suite $(X_n)_{n \geq 1}$ est « bornée dans L^1 ». En particulier $\sup_n E(X_n^+) < \infty$ et le théorème 27.1 implique

$$\lim_{n \rightarrow \infty} X_n = X \text{ existe p.s. et } X \text{ est dans } L^1.$$

Pour montrer la convergence $X_n \rightarrow X$ dans L^1 , posons

$$f_c(x) = \begin{cases} c & \text{si } x > c, \\ x & \text{si } |x| \leq c, \\ -c & \text{si } x < -c. \end{cases}$$

La fonction f est lipschitzienne bornée. À cause de l'uniforme intégrabilité, pour tout $\varepsilon > 0$ il existe c tel que

$$E(|f_c(X_n) - X_n|) < \frac{\varepsilon}{3} \text{ pour tout } n; \quad (27.2)$$

$$E(|f_c(X) - X|) < \frac{\varepsilon}{3}. \quad (27.3)$$

Comme $\lim X_n \rightarrow X$ p.s. on a $f_c(X_n) \rightarrow f_c(X)$ p.s., et le théorème de convergence dominée de Lebesgue implique que pour tout $n \geq N$, avec N assez grand,

$$E(|f_c(X_n) - f_c(X)|) < \frac{\varepsilon}{3}. \quad (27.4)$$

Donc en utilisant (27.2), (27.3) et (27.4), on obtient

$$E(|X_n - X|) < \varepsilon \text{ pour } n \geq N.$$

Donc $X_n \rightarrow X$ dans L^1 . Il reste à montrer que $E(X | \mathcal{F}_n) = X_n$. Soit $\Lambda \in \mathcal{F}_n$ et $m \geq n$. Alors

$$E(X_n 1_\Lambda) = E(X_m 1_\Lambda)$$

par la propriété de martingale. Mais

$$\begin{aligned} |E(X_m 1_\Lambda) - E(X 1_\Lambda)| &\leq E(|X_m - X| 1_\Lambda) \\ &\leq E(|X_m - X|) \end{aligned}$$

qui tend vers 0 quand $m \rightarrow \infty$. Donc $E(X_n 1_\Lambda) = E(X 1_\Lambda)$ et on en conclut que $E(X | \mathcal{F}_n) = X_n$ p.s.

(b) On sait déjà que $(X_n)_{n \geq 0}$ est une martingale. Si $c > 0$ on a

$$X_n 1_{\{|X_n| > c\}} = E(Y 1_{\{|X_n| > c\}} | \mathcal{F}_n),$$

puisque $\{|X_n| > c\} \in \mathcal{F}_n$. Donc pour tout $d > 0$ on a

$$\begin{aligned} E(|X_n| 1_{\{|X_n| > c\}}) &\leq E(|Y| 1_{\{|X_n| > c\}}) \\ &\leq E(|Y| 1_{\{|Y| > d\}}) + d P(|X_n| > c) \\ &\leq E(|Y| 1_{\{|Y| > d\}}) + \frac{d}{c} E(|X_n|). \end{aligned} \quad (27.5)$$

Prenons $\varepsilon > 0$. Choisissons d tel que le premier terme dans (27.5) soit plus petit que $\varepsilon/2$, puis c tel que le second terme soit plus petit que $\varepsilon/2$: on a alors $E(|X_n| 1_{\{|X_n| > c\}}) \leq \varepsilon$ pour tout n , et nous avons terminé. ■

La propriété de martingale est $E(X_m | \mathcal{F}_n) = X_n$ p.s. pour $m \geq n$. Jusqu'à présent nous avons supposé que m et n étaient des entiers positifs, mais on peut aussi prendre des entiers négatifs, i.e. considérer que l'ensemble d'indices est $-\mathbb{N}$, l'ensemble des entiers négatifs. Dans ce cas si $|m| > |n|$, mais m et n sont des entiers négatifs, on a $m < n$. De manière à minimiser les risques de confusion, on supposera toujours que m et n sont des entiers positifs et on écrira X_{-n} . Ainsi, une *martingale inverse* sera un processus $X = (X_{-n})_{n \in \mathbb{N}}$ vérifiant $X_{-n} \in L^1$ et X_{-n} est \mathcal{F}_{-n} -mesurable pour tout n , et

$$E(X_{-n} | \mathcal{F}_{-m}) = X_{-m} \quad \text{p.s.}, \quad (27.6)$$

si $0 \leq n < m$. Ci-dessus les sous-tribus vérifient $\mathcal{F}_{-m} \subset \mathcal{F}_{-n}$ si $0 \leq n < m$.

Théorème 27.4. (Théorème de convergence des martingales inverses.)

Soit une martingale inverse $(X_{-n}, \mathcal{F}_{-n})_{n \geq 0}$, et soit $\mathcal{F}_{-\infty} = \bigcap_{n=0}^{\infty} \mathcal{F}_{-n}$. Alors la suite (X_{-n}) converge p.s. et dans L^1 quand $n \rightarrow \infty$ vers une limite X qui est finie et intégrable.

Preuve. Si $a < b$ notons $U_{-n}(a, b)$ le nombre de montées de a à b du processus $(X_{-n})_{n \geq 0}$ entre les instants $-n$ et 0 . La suite $U_{-n}(a, b)$ croît vers une limite $U(a, b)$ à valeurs dans $\mathbb{N} \cup \{+\infty\}$ quand $n \rightarrow \infty$. Exactement comme dans la preuve du théorème 27.1 on peut écrire

$$\begin{aligned} E(U(a, b)) &= \lim_{n \rightarrow \infty} E(U_{-n}) \\ &\leq \frac{1}{b-a} E((X_0 - a)^+) < \infty, \end{aligned}$$

donc $P(U(a, b) < \infty) = 1$. À nouveau le même argument que dans le théorème 27.1 montre que X_{-n} converge p.s. vers une limite X , *a priori* à valeurs dans $[-\infty, +\infty]$.

La fonction $\varphi(x) = x^+ = x \vee 0$ est convexe, donc l'inégalité de Jensen (théorème 23.9) et (27.6) impliquent que $X_{-n}^+ \leq E(X_0^+ | \mathcal{F}_{-n})$, donc $E(X_{-n}^+) \leq E(X_0^+)$. En appliquant le lemme de Fatou et le fait que $X_{-n}^+ \geq 0$ et $X_{-n}^+ \rightarrow X^+$ p.s., on arrive à

$$E(X^+) \leq \liminf_n E(X_{-n}^+) \leq E(X_0^+) < \infty,$$

de sorte que $X^+ \in L^1$. Le même argument appliqué à la martingale $(-X_{-n})$ montre que $X^- \in L^1$, donc $X \in L^1$.

Il reste à montrer la convergence dans L^1 . Pour cela, on remarque d'abord que dans la preuve du théorème 27.3 nous avons en particulier montré que si $X_{-n} \rightarrow X$ p.s., si $X \in L^1$, et si la suite (X_{-n}) est uniformément intégrable, alors on a aussi $X_{-n} \rightarrow X$ dans L^1 . La partie (b) de ce théorème montre aussi que la famille des v.a. $E(X_0 | \mathcal{G})$, lorsque \mathcal{G} décrit la classe de toutes les sous-tribus de \mathcal{F} , est uniformément intégrable. Comme $X_{-n} = E(X_0 | \mathcal{F}_{-n})$, on déduit de tous ces rappels la convergence dans L^1 souhaitée. ■

Comme application du théorème 27.4 nous prouvons la *loi des grands nombres de Kolmogorov*.

Théorème 27.5. (Loi des grands nombres.) Soit $(X_n)_{n \geq 1}$ une suite i.i.d. de variables aléatoires réelles intégrables. On a alors

$$\frac{X_1 + \cdots + X_n}{n} \rightarrow E(X_1) \quad \text{p.s. et dans } L^1.$$

Preuve. Soit $S_n = X_1 + \cdots + X_n$ et $\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2}, \dots)$. On a $\mathcal{F}_{-n} \subset \mathcal{F}_{-m}$ si $n \geq m$, et le processus

$$M_{-n} = E(X_1 | \mathcal{F}_{-n})$$

est une martingale inverse. Noter que $E(M_{-n}) = E(X_1)$ pour chaque n . En utilisant le fait que les X_n sont i.i.d., un argument de symétrie permet de vérifier aisément que pour $1 \leq j \leq n$ on a

$$E(X_1 | \mathcal{F}_{-n}) = E(X_j | \mathcal{F}_{-n}) \quad \text{p.s.} \quad (27.7)$$

(voir l'exercice 17 du chapitre 23). Donc

$$M_{-n} = E(X_1 | \mathcal{F}_{-n}) = E(X_2 | \mathcal{F}_{-n}) = \cdots = E(X_n | \mathcal{F}_{-n}),$$

et par suite

$$M_{-n} = \frac{1}{n} \sum_{j=1}^n E(X_j | \mathcal{F}_{-n}) = E\left(\frac{S_n}{n} | \mathcal{F}_{-n}\right) = \frac{S_n}{n} \quad \text{p.s.}$$

Le théorème 27.4 implique alors que $\frac{S_n}{n}$ converge p.s. et dans L^1 vers une limite intégrable X , qui vérifie nécessairement $E(X) = E(X_1)$. De plus X est mesurable par rapport à la tribu asymptotique, donc la loi zéro-un de Kolmogorov (théorème 10.6) entraîne que X est p.s. égale à une constante, qui est donc nécessairement son espérance. ■

Le théorème 27.5 a été publié pour la première fois en 1933 [17], sans l'aide de la théorie des martingales; cette théorie a été développée par J. L. Doob bien des années plus tard.

Voici une autre application du théorème de convergence des martingales, due aussi à Kolmogorov.

Théorème 27.6. (Kolmogorov.) *Soit $(Y_n)_{n \geq 1}$ des v.a. réelles indépendantes, centrées et de carré intégrable. Supposons que $\sum_{n=1}^{\infty} E(Y_n^2) < \infty$, et soit $S_n = \sum_{j=1}^n Y_j$. La suite S_n converge alors p.s. vers une limite finie, notée $\sum_{j=1}^{\infty} Y_j$.*

Preuve. Soit $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$. On a vu au chapitre 24 que $(S_n)_{n \geq 1}$ est une \mathcal{F}_n -martingale. De plus, $\sup_n E(S_n^+) \leq \sup_n (E(S_n^2) + 1) \leq \sum_{n=1}^{\infty} E(Y_n^2) + 1 < \infty$. Le résultat découle alors du théorème 27.1. ■

Nous venons de voir ci-dessus un exemple de martingale, à savoir la suite des sommes partielles (S_n) de v.a. indépendantes et centrées. Si ces v.a. indépendantes sont de plus de même loi et de variance finie, disons σ^2 , le théorème-limite central nous dit que la suite $\frac{S_n}{\sqrt{n}}$ converge en loi vers une v.a. $N(0, \sigma^2)$. La théorie des martingales nous permet d'affaiblir à la fois l'hypothèse d'indépendance, et l'hypothèse que ces v.a. ont même loi. En ce sens, elle permet de mieux cerner ce qui est réellement indispensable pour avoir un théorème-limite central. Comme dans tout ce chapitre, nous supposons donnée la suite croissante $(\mathcal{F}_n)_{n \geq 0}$ de sous-tribus de \mathcal{F} , sur l'espace (Ω, \mathcal{F}, P) .

Théorème 27.7. (Théorème-limite central pour les martingales.) *Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles telles que pour tout $n \geq 1$:*

- (i) $E(X_n | \mathcal{F}_{n-1}) = 0$
- (ii) $E(X_n^2 | \mathcal{F}_{n-1}) = 1$
- (iii) $E(|X_n|^3 | \mathcal{F}_{n-1}) \leq K < \infty$.

Soit $S_n = \sum_{i=1}^n X_i$ pour $n \geq 1$. La suite $\frac{1}{\sqrt{n}}S_n$ converge alors en loi vers une v.a. $N(0, 1)$.

Preuve. Nous allons utiliser les fonctions caractéristiques pour prouver ce théorème. Pour $u \in \mathbb{R}$, rappelons que la fonction caractéristique d'une v.a. réelle X est $\varphi_X(u) = E(e^{iuX})$. On définit la « fonction caractéristique conditionnelle » de la v.a. X_j/\sqrt{n} comme étant la v.a.

$$\varphi_{n,j}(u) = E\left(e^{iu\frac{1}{\sqrt{n}}X_j} \mid \mathcal{F}_{j-1}\right).$$

On peut faire un développement de Taylor au voisinage de 0 :

$$e^{iu\frac{1}{\sqrt{n}}X_j} = 1 + iu\frac{1}{\sqrt{n}}X_j - \frac{u^2}{2n}X_j^2 - \frac{iu^3}{6n^{\frac{3}{2}}}\bar{X}_j^3 \quad (27.8)$$

où \bar{X}_j est un nombre (aléatoire) situé entre 0 et X_j . En prenant l'espérance conditionnelle des deux membres de (27.8), on obtient

$$\varphi_{n,j}(u) = 1 + iu\frac{1}{\sqrt{n}}E(X_j \mid \mathcal{F}_{j-1}) - \frac{u^2}{2n}E(X_j^2 \mid \mathcal{F}_{j-1}) - \frac{iu^3}{6n^{\frac{3}{2}}}E(\bar{X}_j^3 \mid \mathcal{F}_{j-1})$$

et les hypothèses (i) et (ii) donnent

$$\varphi_{n,j}(u) - 1 - \frac{u^2}{2n} = \frac{u^3}{6n^{\frac{3}{2}}}E(\bar{X}_j^3 \mid \mathcal{F}_{j-1}). \quad (27.9)$$

Rappelons que $S_p = \sum_{j=1}^p X_j$, donc pour $1 \leq p \leq n$ on a

$$\begin{aligned} E\left(e^{iu\frac{1}{\sqrt{n}}S_p}\right) &= E\left(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}e^{iu\frac{1}{\sqrt{n}}X_p}\right) \\ &= E\left(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}E\left(e^{iu\frac{1}{\sqrt{n}}X_p} \mid \mathcal{F}_{p-1}\right)\right) \\ &= E\left(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}\varphi_{n,p}(u)\right). \end{aligned} \quad (27.10)$$

En utilisant (27.10) et (27.9) on arrive à

$$E\left(e^{iu\frac{1}{\sqrt{n}}S_p}\right) = E\left(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}\left(1 - \frac{u^2}{2n} - \frac{iu^3}{6n^{\frac{3}{2}}}\bar{X}_j^3\right)\right)$$

et donc

$$E\left(e^{iu\frac{1}{\sqrt{n}}S_p} - \left(1 - \frac{u^2}{2n}\right)e^{iu\frac{1}{\sqrt{n}}S_{p-1}}\right) = E\left(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}\frac{iu^3}{6n^{\frac{3}{2}}}\bar{X}_j^3\right). \quad (27.11)$$

En prenant le module dans chaque membre de (27.11) et en utilisant le fait que $|\overline{X}_j| \leq |X_j|$ et l'hypothèse (iii), il vient

$$\begin{aligned} \left| \mathbb{E} \left(e^{iu \frac{1}{\sqrt{n}} S_p} - \left(1 - \frac{u^2}{2n}\right) e^{iu \frac{1}{\sqrt{n}} S_{p-1}} \right) \right| \\ \leq \mathbb{E} \left(\left| e^{iu \frac{1}{\sqrt{n}} S_{p-1}} \right| \frac{|u|^3}{6n^{\frac{3}{2}}} \mathbb{E}(|X_j|^3 | \mathcal{F}_{j-1}) \right) \\ \leq K \frac{|u|^3}{6n^{\frac{3}{2}}}. \end{aligned} \quad (27.12)$$

Fixons $u \in \mathbb{R}$. Lorsque $n \rightarrow \infty$, pour n assez grand on a $n \geq \frac{u^2}{2}$ et donc $0 \leq 1 - \frac{u^2}{2n} \leq 1$. En multipliant (27.12) par $(1 - \frac{u^2}{2n})^{n-p}$ on arrive, pour n assez grand, à

$$\left| \left(1 - \frac{u^2}{2n}\right)^{n-p} \mathbb{E} \left(e^{iu \frac{1}{\sqrt{n}} S_p} \right) - \left(1 - \frac{u^2}{2n}\right)^{n-p+1} \mathbb{E} \left(e^{iu \frac{1}{\sqrt{n}} S_{p-1}} \right) \right| \leq K \frac{|u|^3}{6n^{\frac{3}{2}}}. \quad (27.13)$$

Noter que (27.13) est vraie également pour chaque n fixé, pour un certain K_n : en prenant K assez grand on peut donc supposer (27.13) vraie pour tout n .

Finalement en additionnant et en utilisant le caractère télescopique des différents termes, on voit que

$$\begin{aligned} \mathbb{E} \left(e^{iu \frac{1}{\sqrt{n}} S_n} \right) - \left(1 - \frac{u^2}{2n}\right)^n = \sum_{p=1}^n \left\{ \left(1 - \frac{u^2}{2n}\right)^{n-p} \mathbb{E} \left(e^{iu \frac{1}{\sqrt{n}} S_p} \right) \right. \\ \left. - \left(1 - \frac{u^2}{2n}\right)^{n-(p-1)} \mathbb{E} \left(e^{iu \frac{1}{\sqrt{n}} S_{p-1}} \right) \right\} \end{aligned}$$

de sorte que par l'inégalité triangulaire et (27.13) on voit que

$$\left| \mathbb{E} \left(e^{iu \frac{1}{\sqrt{n}} S_n} \right) - \left(1 - \frac{u^2}{2n}\right)^n \right| \leq n \frac{K|u|^3}{6n^{\frac{3}{2}}} = K \frac{|u|^3}{6\sqrt{n}}. \quad (27.14)$$

Comme le membre de droite de (27.14) tend vers 0 et que

$$\lim_{n \rightarrow \infty} \left(1 - \frac{u^2}{2n}\right)^n = e^{-\frac{u^2}{2}},$$

on obtient

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(e^{iu \frac{S_n}{\sqrt{n}} \right) = e^{-\frac{u^2}{2}}.$$

Le théorème de Lévy 19.1 donne alors le résultat. ■

Remarque. Si S_n est la martingale du théorème 22.4, on sait que le théorème de convergence p.s. ne peut pas être vérifié : en effet, si on avait $\lim_{n \rightarrow \infty} S_n = S$ p.s. avec $S \in L^1$, on aurait aussi $\lim_{n \rightarrow \infty} \frac{S_n}{\sqrt{n}} = 0$ p.s., et la convergence en loi de $\frac{S_n}{\sqrt{n}}$ vers une v.a. normale non dégénérée serait impossible. Ce qui rend la convergence p.s. de S_n impossible est le comportement des variances conditionnelles des accroissements X_n , i.e. l'hypothèse (ii) du théorème 22.4.

Nous terminons notre exposé sur les martingales par un exemple tiré de l'analyse. Cet exemple montre la souplesse d'utilisation et la variété des applications de la théorie des martingales : nous utilisons le résultat « probabiliste » de convergence des martingales pour montrer de manière simple un résultat « analytique » de convergence de fonctions.

Exemple ([13]). Soit f une fonction dans $L^p[0, 1]$ pour la mesure de Lebesgue sur $[0, 1]$. Définissons les *fonctions de Rademacher* sur $[0, 1]$ comme suit. Posons $R_0(x) = 1$ pour $0 \leq x \leq 1$. Si $n \geq 1$, nous posons pour $0 \leq x \leq 1$:

$$R_n(x) = \begin{cases} 1 & \text{si } \frac{2j-1}{2^n} \leq x < \frac{2j}{2^n}, \quad \text{pour un } j \text{ dans } \{1, \dots, 2^n\} \\ -1 & \text{sinon.} \end{cases}$$

Notons P la probabilité sur $[0, 1]$ égale à la mesure de Lebesgue. \mathcal{F} désigne la tribu borélienne. On a alors

$$E(R_n) = \int_0^1 R_n(x) dx = 0$$

et

$$\text{Var}(R_n) = E(R_n^2) = \int_0^1 R_n(x)^2 dx = 1.$$

Finalement on observe que R_n et R_m (qui peuvent être ici considérées comme des variables aléatoires) sont indépendantes si $n \neq m$. (voir l'exercice 8.)

Ensuite, nous définissons les *fonctions de Haar* comme suit :

$$H_0(x) = R_0(x),$$

$$H_1(x) = R_1(x);$$

Pour $n \geq 2$, écrivons $n = 1 + 2 + \dots + 2^{r-2} + \lambda = 2^{r-1} - 1 + \lambda$, avec $r \geq 2$ et $1 \leq \lambda \leq 2^{r-1}$; posons alors

$$H_n(x) = \begin{cases} \sqrt{2^{r-1}} R_n(x) & \text{si } \frac{2\lambda-2}{2^r} \leq x < \frac{2\lambda}{2^r}, \\ 0 & \text{sinon.} \end{cases}$$

Ensuite, on note $\mathcal{F}_n = \sigma(H_0, H_1, \dots, H_n)$ la plus petite tribu rendant les fonctions H_0, \dots, H_n toutes mesurables. On vérifie aisément que

$$\int_{\Lambda} H_{n+1}(x) dx = 0 \quad \text{si } \Lambda \in \mathcal{F}_n; \quad (27.15)$$

(voir l'exercice 9). On a de plus

$$\begin{aligned} \int_0^1 H_n(x) dx &= 0, \\ \int_0^1 H_n(x)^2 dx &= 1. \end{aligned}$$

On a alors le théorème suivant :

Théorème 27.8. Soit (H_n) les fonctions de Haar sur $[0, 1]$, et soit $f \in L^p[0, 1]$ pour un $p \geq 1$. Posons

$$\begin{aligned} \alpha_r &= \int_0^1 H_r(x) f(x) dx, \\ S_n(x) &= \sum_{r=0}^n \alpha_r H_r(x). \end{aligned} \quad (27.16)$$

Alors S_n converge vers f presque partout (pour la mesure de Lebesgue). De plus si $S^*(x) = \sup_n |S_n(x)|$, on a

$$\int_0^1 S^*(x)^p dx \leq \left(\frac{p}{p-1} \right)^p \int_0^1 |f(x)|^p dx.$$

Preuve. Montrons d'abord que la suite (S_n) est une martingale. On a

$$\begin{aligned} \mathbf{E}(S_{n+1} | \mathcal{F}_n) &= S_n + \mathbf{E}(\alpha_{n+1} H_{n+1} | \mathcal{F}_n) \\ &= S_n + \alpha_{n+1} \mathbf{E}(H_{n+1} | \mathcal{F}_n) \\ &= S_n, \end{aligned}$$

où on a utilisé (27.15). On a même :

$$S_n = E(f | \mathcal{F}_n), \quad (27.17)$$

qui est le résultat clé. C'est l'endroit où nous avons besoin de la forme (27.16) pour les coefficients α_r (voir l'exercice 10.)

Nous montrons ensuite que S_n vérifie $\sup_n E(S_n^+) < \infty$, pour $p > 1$ (c'est l'hypothèse pour pouvoir appliquer le théorème de convergence des martingales 27.1). Nous montrons en fait un peu plus, grâce à l'inégalité de Jensen : comme $\varphi(u) = |u|^p$ est convexe pour $p > 1$, il vient

$$\begin{aligned} \int_0^1 |S_n(x)|^p dx &= E(|E(f | \mathcal{F}_n)|^p) \\ &\leq E(E(|f|^p | \mathcal{F}_n)) \\ &= E(|f|^p) \\ &= \int_0^1 |f(x)|^p dx < \infty. \end{aligned}$$

Comme $S_n^+ \leq |S_n| \leq |S_n|^p + 1$, on a alors

$$\begin{aligned} \sup_n E(S_n^+) &\leq \sup_n (E(|S_n|^p) + 1) \\ &\leq E(|f|^p) + 1 < \infty. \end{aligned}$$

Le théorème 27.1 montre alors que

$$\lim_{n \rightarrow \infty} S_n = f \quad \text{p.s.,}$$

tandis que l'inégalité de Doob (théorème 26.2) donne

$$\begin{aligned} E(S^{*p}) &\leq \left(\frac{p}{p-1}\right)^p E(|S_n|^p) \\ &\leq \left(\frac{p}{p-1}\right)^p E(|f|^p), \end{aligned}$$

ou de manière équivalente :

$$\int_0^1 (S^*(x))^p dx \leq \left(\frac{p}{p-1}\right)^p \int_0^1 |f(x)|^p dx. \quad \blacksquare$$

Remarquons enfin que des résultats analogues au théorème 27.8 sont vrais pour les séries de Fourier classiques, mais c'est un peu plus difficile à montrer.

Exercices

1. (Une preuve de la loi zéro-un de Kolmogorov à l'aide des martingales.) Soit (X_n) une suite de v.a. indépendantes, et \mathcal{C}_∞ la tribu asymptotique correspondante (voir le théorème 10.6). Soit $C \in \mathcal{C}_\infty$. Montrer que $E(1_C | \mathcal{F}_n) = P(C)$ pour tout n , où $\mathcal{F}_n = \sigma(X_j; 0 \leq j \leq n)$. Montrer ensuite que $\lim_{n \rightarrow \infty} E(1_C | \mathcal{F}_n) = 1_C$ p.s., et en déduire que $P(C) = 0$ ou $P(C) = 1$.
2. Une martingale (X_n) est dite *bornée dans L^2* si $\sup_n E(X_n^2) < \infty$. Soit (X_n) une martingale telle que chaque X_n soit dans L^2 . Montrer que (X_n) est bornée dans L^2 si et seulement si

$$\sum_{n=1}^{\infty} E((X_n - X_{n-1})^2) < \infty.$$

(Indication : cf. l'exercice 12 du chapitre 24.)

3. Soit (X_n) une martingale bornée dans L^2 . Montrer que

$$\sup_n E(|X_n|) < \infty$$

et en conclure que X_n converge p.s. et dans L^1 vers une limite X .

- 4* Sous les hypothèses de l'exercice 3, montrer que X_n converge vers X dans L^2 également.
5. (Signes aléatoires.) Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes avec $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$. Soit $(\alpha_n)_{n \geq 1}$ une suite de réels. Montrer que la série $\sum_{n=1}^{\infty} \alpha_n X_n$ est p.s. convergente dès que $\sum_{n=1}^{\infty} \alpha_n^2 < \infty$.
6. Soit $(X_n)_{n \geq 1}$ une suite i.i.d. de variables positives vérifiant $E\{X_1\} = 1$. Soit $R_n = \prod_{i=1}^n X_i$. Montrer que (R_n) est une martingale.
7. Montrer que si $n \neq m$ les fonctions de Rademacher R_n et R_m sont indépendantes pour $P = \lambda$ (mesure de Lebesgue sur $[0, 1]$).
8. Soit (H_n) les fonctions de Haar, et soit $\Lambda \in \mathcal{F}_n = \sigma(H_0, H_1, \dots, H_n)$. Montrer que si $n \geq 1$,

$$\int_{\Lambda} H_n(x) dx = 0.$$

9. Soit $f \in L^p[0, 1]$, et S_n défini en (27.16). Montrer que $E(f | \mathcal{F}_n) = S_n$. (Indication : Montrer que

$$\int_{\Lambda} f(x) dx = \int_{\Lambda} S_n(x) dx \quad \text{si } \Lambda \in \mathcal{F}_n$$

en utilisant que les fonctions de Haar forment un système orthonormal, i.e.

$$\int_0^1 H_n(x)H_m(x)dx = 0 \quad \text{si } n \neq m \text{ et } \int_0^1 H_n(x)^2 dx = 1.$$

10. Utiliser le théorème de convergence des martingales pour montrer la « loi zéro-un » suivante : soit (\mathcal{F}_n) une suite croissante de sous-tribus, et (\mathcal{G}_n) une autre suite, décroissante, de sous-tribus, avec $\mathcal{G}_1 \subset \sigma(\cup_{n=1}^{\infty} \mathcal{F}_n)$. Supposons que \mathcal{F}_n et \mathcal{G}_n soient indépendantes pour chaque n . Montrer que si $\Lambda \in \cap_{n=1}^{\infty} \mathcal{G}_n$, alors $P(\Lambda) = 0$ ou $P(\Lambda) = 1$.
11. Soit \mathcal{H} un sous-ensemble de L^1 . Soit G une fonction positive croissante sur $[0, \infty[$, telle que

$$\lim_{t \rightarrow \infty} \frac{G(t)}{t} = \infty.$$

Supposons en outre que $\sup_{X \in \mathcal{H}} E(G(X)) < \infty$. Montrer que \mathcal{H} est uniformément intégrable. (Cela étend le théorème 27.2(a).)

Chapitre 28

Le théorème de Radon-Nikodym

Soit (Ω, \mathcal{F}, P) un espace de probabilité. Considérons une v.a. positive X d'espérance égale à 1. Pour tout $\Lambda \in \mathcal{F}$ on pose

$$Q(\Lambda) = E(1_{\Lambda}X). \quad (28.1)$$

On obtient ainsi une nouvelle probabilité Q : en effet on a d'abord

$$Q(\Omega) = E(1_{\Omega}X) = E(X) = 1,$$

et ensuite si A_1, A_2, A_3, \dots sont dans \mathcal{F} et deux à deux disjoints il vient

$$\begin{aligned} Q\left(\bigcup_{i=1}^{\infty} A_i\right) &= E(1_{\{\cup_{i=1}^{\infty} A_i\}}X) \\ &= E\left(\sum_{i=1}^{\infty} 1_{A_i}X\right) \\ &= \sum_{i=1}^{\infty} E(1_{A_i}X) \\ &= \sum_{i=1}^{\infty} Q(A_i) \end{aligned}$$

et on a la σ -additivité ; ci-dessus, on a pu échanger la somme et l'espérance grâce au théorème de convergence monotone (théorème 9.1(d)).

La probabilité Q jouit de deux propriétés remarquables :

- (i) Si $P(\Lambda) = 0$ alors $Q(\Lambda) = 0$ (car $Q(\Lambda) = E(1_{\Lambda}X)$ et $1_{\Lambda} = 0$ p.s., donc aussi $1_{\Lambda}X = 0$ p.s.)
- (ii) Pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que si $\Lambda \in \mathcal{F}$ et $P(\Lambda) < \delta$, alors $Q(\Lambda) < \varepsilon$.

En fait, (ii) ci-dessus découle de manière générale de (i), comme nous allons l'énoncer formellement, et le montrer.

Théorème 28.1. *Soit P, Q deux probabilités telles que $P(\Lambda) = 0$ implique $Q(\Lambda) = 0$. Alors, pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que si $\Lambda \in \mathcal{F}$ et $P(\Lambda) < \delta$, on ait $Q(\Lambda) < \varepsilon$.*

Preuve. Supposons le résultat faux. Il existerait alors une suite $\Lambda_n \in \mathcal{F}$ avec $P(\Lambda_n) < \frac{1}{2^n}$ et $Q(\Lambda_n) \geq \varepsilon$ pour tout n et pour un certain $\varepsilon > 0$. Posons $\Lambda = \limsup_{n \rightarrow \infty} \Lambda_n$. Le lemme de Borel-Cantelli (théorème 10.5) entraîne que $P(\Lambda) = 0$. Puis, en remarquant que le lemme de Fatou admet une version symétrique pour les lim sup (établie en passant dans la preuve du théorème 9.1(f)), on obtient

$$Q(\Lambda) \geq \limsup_{n \rightarrow \infty} Q(\Lambda_n) \geq \varepsilon,$$

et on arrive à une contradiction. ■

Cela vaut la peine de remarquer que les conditions (i) et (ii) sont en fait *équivalentes*. En effet, on vient de montrer que (i) implique (ii), et la réciproque est presque évidente : si $P(\Lambda) = 0$, pour tout $\varepsilon > 0$ on a $P(\Lambda) < \delta$ pour le δ associé au ε dans (ii), donc $Q(\Lambda) < \varepsilon$; comme ε est arbitrairement petit on a donc $Q(\Lambda) = 0$.

Définition 28.1. Soit P une probabilité et Q une mesure finie. On dit que Q est absolument continue par rapport à P si, chaque fois qu'on a $P(\Lambda) = 0$ pour un $\Lambda \in \mathcal{F}$, on a aussi $Q(\Lambda) = 0$. On note cette propriété ainsi : $Q \ll P$.

Exemple. On a vu que la formule (28.1) donne une probabilité Q vérifiant $Q \ll P$.

Un tel exemple se présentant naturellement est celui de la probabilité conditionnelle $Q(\Lambda) = P(\Lambda | A)$, lorsque $P(A) > 0$. L'absolue continuité est évidente à vérifier, mais c'est aussi un cas particulier de l'exemple précédent avec la v.a. $X = \frac{1}{P(A)} \mathbf{1}_A$.

Le théorème de Radon-Nikodym caractérise en fait toutes les probabilités absolument continues par rapport à P : ce sont exactement celles qui se mettent sous la forme (28.1) : notre exemple initial couvre donc toutes les situations possibles. Nous montrons une version un peu simplifiée, pour les tribus dites « séparables », et notre démonstration suit celle de P. A. Meyer [18].

Définition 28.2. Une tribu \mathcal{F} est dite *séparable* si $\mathcal{F} = \sigma(A_1, \dots, A_n, \dots)$: en d'autres termes, \mathcal{F} est engendrée par une suite dénombrable d'événements.

Théorème 28.2. (Radon-Nikodym.) Soit (Ω, \mathcal{F}, P) un espace de probabilité avec une tribu \mathcal{F} qui est séparable. Si Q est une mesure finie sur \mathcal{F} telle que $Q \ll P$, il existe alors une v.a. positive intégrable X telle que pour tout $\Lambda \in \mathcal{F}$ on ait

$$Q(\Lambda) = E(\mathbf{1}_\Lambda X).$$

De plus X est p.s. unique (i.e. si X' vérifie les mêmes propriétés, alors $X' = X$ P-p.s.), et on l'écrit souvent $X = \frac{dQ}{dP}$.

Preuve. Si $Q(\Omega) = 0$, le résultat est évident avec $X \equiv 0$. On suppose donc que $Q(\Omega) > 0$, et on peut alors « normaliser » Q en posant $\tilde{Q} = \frac{1}{Q(\Omega)} Q$, ce qui revient à supposer que Q est une probabilité. Soit $(A_n)_{n \geq 0}$ une suite engendrant la tribu \mathcal{F} , et définissons la famille croissante de sous-tribus $(\mathcal{F}_n)_{n \geq 1}$ ainsi :

$$\mathcal{F}_n = \sigma(A_0, \dots, A_n).$$

Il existe alors pour chaque n une partition finie $A_{n,1}, A_{n,2}, \dots, A_{n,k_n}$ de Ω , telle que les éléments de la tribu \mathcal{F}_n soient exactement les réunions (nécessairement finies !) d'éléments de cette partition. Les $A(n, i)$ sont appelés les *atomes* de \mathcal{F}_n . Posons alors

$$X_n(\omega) = \sum_{i=1}^{k_n} \frac{Q(A_{n,i})}{P(A_{n,i})} 1_{A_{n,i}}(\omega), \quad (28.2)$$

avec la convention $\frac{0}{0} = 0$ (comme $Q \ll P$ le numérateur est nul chaque fois que le dénominateur est nul). On va montrer que la suite $(X_n)_{n \geq 0}$ est en fait une martingale.

D'abord, le fait que X_n soit \mathcal{F}_n -mesurable est évident. Ensuite, soit $m \leq n$. Exactement comme dans la preuve du théorème 24.6, pour montrer que $E(X_n | \mathcal{F}_m) = X_m$ il suffit de montrer que pour tout $\Lambda \in \mathcal{F}_m$ on a

$$\int_{\Lambda} X_n dP = \int_{\Lambda} X_m dP. \quad (28.3)$$

On peut écrire

$$\begin{aligned} \int_{\Lambda} X_n dP &= \int_{\Lambda} \sum_{i=1}^{k_n} \frac{Q(A_{n,i})}{P(A_{n,i})} 1_{A_{n,i}} dP \\ &= \int \sum_{i=1}^{k_n} \frac{Q(A_{n,i})}{P(A_{n,i})} 1_{A_{n,i} \cap \Lambda} dP \\ &= \sum_{i=1}^{k_n} \frac{Q(A_{n,i})}{P(A_{n,i})} P(A_{n,i} \cap \Lambda). \end{aligned}$$

Mais comme $\Lambda \in \mathcal{F}_m$, l'ensemble Λ s'écrit comme la réunion (disjointe) des $A(m, i)$ pour i dans une partie I de $\{1, \dots, k_m\}$. Donc $\Lambda \cap A_{n,i} = A_{n,i}$ si $i \in I$ et $\Lambda \cap A_{n,i} = \emptyset$ sinon, de sorte que

$$\begin{aligned} \int_{\Lambda} X_n dP &= \sum_{i \in I} \frac{Q(A_{n,i})}{P(A_{n,i})} P(A_{n,i} \cap \Lambda) \\ &= \sum_{i \in I} Q(A_{n,i}) = Q(\Lambda) \end{aligned}$$

où on a encore une fois utilisé le fait que $Q(A_{n,i}) = 0$ si $P(A_{n,i}) = 0$. Comme $\Lambda \in \mathcal{F}_n$ on a de même $\int_{\Lambda} X_m dP = Q(\Lambda)$. Donc (28.3) est vérifié, et si on prend de plus $\Lambda = \Omega$ on obtient $\int X_n dP = Q(\Omega) = 1 < \infty$, donc X_n est P -intégrable. Donc $(X_n)_{n \geq 1}$ est bien une martingale.

En fait, cette martingale (X_n) est uniformément intégrable. En effet on a

$$\int_{\{X_n > c\}} X_n dP = Q(X_n > c);$$

et par l'inégalité de Markov il vient

$$P(X_n > c) \leq \frac{E(X_n)}{c} = \frac{1}{c}.$$

Soit alors $\varepsilon > 0$ et le $\delta > 0$ qui lui est associé comme dans le théorème 28.1. Si $c > 1/\delta$ on a $P(X_n > c) < \delta$, donc $Q(X_n > c) \leq \varepsilon$, donc $\int_{\{X_n > c\}} X_n dP \leq \varepsilon$: en conséquence, la suite (X_n) est uniformément intégrable. On peut alors appliquer le second théorème de convergence des martingales 27.3, de façon à obtenir que X_n converge p.s. et dans L^1 vers une limite X , et de plus

$$E(X | \mathcal{F}_n) = X_n.$$

Si alors on pose $R(\Lambda) = E(1_{\Lambda} X)$ pour tout $\Lambda \in \mathcal{F}$, on définit une nouvelle probabilité R , qui coïncide avec Q sur chaque tribu \mathcal{F}_n (puisque si $\Lambda \in \mathcal{F}_n$ on a $R(\Lambda) = E(1_{\Lambda} X) = E(1_{\Lambda} X_n) = Q(\Lambda)$). Le théorème des classes monotones 6.3 entraîne alors que $R = Q$, puisque $\mathcal{F} = \sigma(\mathcal{F}_n; n \geq 1)$. ■

Remarque. On peut utiliser le théorème précédent pour montrer une version plus générale, sans l'hypothèse de séparabilité. Pour une démonstration du théorème 28.3 ci-dessous, nous référons à [27, p. 147-149] par exemple.

Théorème 28.3. (Radon-Nikodym.) *Soit (Ω, \mathcal{F}, P) un espace de probabilité. Si Q est une mesure finie sur \mathcal{F} telle que $Q \ll P$, il existe alors une v.a. positive intégrable X telle que pour tout $\Lambda \in \mathcal{F}$ on ait*

$$Q(\Lambda) = E(1_{\Lambda} X).$$

De plus X est p.s. unique (i.e. si X' vérifie les mêmes propriétés, alors $X' = X$ P-p.s.).

Le théorème de Radon-Nikodym est directement en rapport avec les espérances conditionnelles : supposons donnés l'espace de probabilité (Ω, \mathcal{F}, P) et la sous-tribu \mathcal{G} de \mathcal{F} . Pour toute v.a. positive intégrable X , la formule $Q(\Lambda) = E(X1_\Lambda)$ pour $\Lambda \in \mathcal{G}$ définit une mesure finie Q sur (Ω, \mathcal{G}) , et $P(\Lambda) = 0$ implique $Q(\Lambda) = 0$. Donc $Y = \frac{dQ}{dP}$ existe sur l'espace (Ω, \mathcal{G}) , ce qui implique en particulier que Y est \mathcal{G} -mesurable. Par ailleurs on a pour tout $\Lambda \in \mathcal{G}$:

$$E(Y1_\Lambda) = Q(\Lambda) = E(X1_\Lambda).$$

Donc Y est une version de $E(X|\mathcal{G})$. En fait il est possible de démontrer le théorème de Radon-Nikodym à l'aide seulement de la théorie de la mesure (sans martingales et donc sans espérances conditionnelles) ; puis on peut *construire* l'espérance conditionnelle de la manière suggérée ci-dessus : cela évite de recourir à la théorie des espaces de Hilbert (mais la preuve « directe » du théorème de Radon-Nikodym n'est pas particulièrement aisée).

Signalons enfin que si P est une probabilité sur \mathbb{R} admettant la densité f , alors comme $P(A) = \int_A f(x)dx$ on a que la mesure P est absolument continu par rapport à la mesure de Lebesgue m (ici, m est une mesure σ -finie, mais le théorème de Radon-Nikodym « marche » encore dans ce cas), et on peut écrire $f = \frac{dP}{dm}$.

Exercices

1. Supposons que $Q \ll P$ et que $P \ll Q$: on dit alors que P et Q sont équivalentes et on écrit $Q \sim P$. Montrer que $X = \frac{dQ}{dP}$ vérifie $X > 0$ P-p.s.
2. Supposons que $Q \sim P$, et soit $X = \frac{dQ}{dP}$. Montrer que $\frac{1}{X} = \frac{dP}{dQ}$.
3. Soit μ une mesure telle que $\mu = \sum_{n=1}^{\infty} \alpha_n P_n$, pour une suite (P_n) de probabilités et une suite (α_n) de nombres strictement positifs vérifiant $\sum_n \alpha_n = 1$. Montrer que μ est une probabilité. Soit une autre suite (Q_n) de probabilités et une autre suite (β_n) de réels positifs avec $\sum_n \beta_n < \infty$. Si on a $Q_n \ll P_n$ pour chaque n , montrer que la mesure finie $\nu = \sum_{n=1}^{\infty} \beta_n Q_n$ vérifie $\nu \ll \mu$.
4. Soit P et Q deux probabilités et $R = \frac{P+Q}{2}$. Montrer que $P \ll R$.

Bibliographie

- [1] R. Bass (1995), *Probabilistic Techniques in Analysis*; Springer-Verlag; New York.
- [2] H. Bauer (1996), *Probability Theory*; Walter de Gruyter; Berlin.
- [3] J. Bernoulli (1713), *Ars Conjectandi*; Thurnisiorum; Bâle.
- [4] M. Briane, G. Pagès (1998), *Théorie de l'intégration*; Vuibert; Paris.
- [5] G. Cardano (1961), *Liber de ludo aleae*; *The Book on Games of Chance*; Sidney Gould (Traducteur); Holt, Rinehart and Winston.
- [6] H. Cartan (1967); *Calcul différentiel*; Hermann; Paris.
- [7] G. Casella and R. L. Berger (1990), *Statistical Inference*; Wadsworth; Belmont, USA.
- [8] G. Choquet (1969); *Topologie*; Masson; Paris.
- [9] A. De Moivre (1718), *The Doctrine of Chances; or, A Method of Calculating the Probability of Events in Play*; W. Pearson; Londres. Et, en 1756, *The Doctrine of Chances (Third Edition)*, réimpression en 1967 : Chelsea, New York.
- [10] J. Doob (1994), *Measure Theory*; Springer-Verlag; New York.
- [11] R. Durrett (1991), *Probability : Theory and Examples*; Wadsworth and Brooks-Cole; Belmont, USA.
- [12] W. Feller (1971), *An Introduction to Probability Theory and Its Applications* (Volume II); John Wiley; New York.
- [13] A. Garsia (1970), *Topics in Almost Everywhere Convergence*; Markham; Chicago.
- [14] A. Gut (1995). *An Intermediate Course in Probability*; Springer-Verlag; New York.
- [15] N: B. Haaser and J. A. Sullivan (1991), *Real Analysis*; Dover; New York.
- [16] C. Huygens (1657); *Voir (Euvres Complètes de Christiaan Huygens* (traduction française en 1920), Den Haag : Nijhoff.
- [17] A. N. Kolmogorov (1933), *Grundbegriffe der Wahrscheinlichkeitrechnung*. Traduction anglaise : *Foundations of the Theory of Probability*, (1950), Nathan Morrison (traducteur); Chelsea; New York.
- [18] P. A. Meyer (1966), *Probabilités et Potentiel*; Hermann; Paris.
- [19] J. Neveu (1964), *Bases mathématiques du calcul des probabilités*. Paris.

- [20] D. Pollard (1984), *Convergence of Stochastic Processes*; Springer-Verlag; New York.
- [21] M. H. Protter and P. Protter (1988), *Calculus with Analytic Geometry* (Fourth Edition); Jones and Bartlett; Boston.
- [22] M. Sharpe (1988), *General Theory of Markov Processes*; Academic Press; New York.
- [23] G. F. Simmons (1963), *Introduction to Topology and Modern Analysis*; McGraw-Hill; New York.
- [24] S. M. Stigler (1986), *The History of Statistics : The measurement of uncertainty before 1900*; Harvard University Press; Cambridge, USA.
- [25] D. W. Stroock (1990), *A Concise Introduction to the Theory of Integration*; World Scientific; Singapour.
- [26] S. J. Taylor (1973), *Introduction to Measure and Integration*; Cambridge University Press; Cambridge, Grande-Bretagne.
- [27] D. Williams (1991), *Probability with Martingales*; Cambridge; Grande-Bretagne.

Index des notations

- 1_A fonction indicatrice, 11, 53
 2^Ω classe de toutes les parties de Ω , 4, 7
 A^* transposée de A , 98
 $A_n \rightarrow A$ convergence des ensembles A_n vers A , 11
 $E(X)$ espérance de X , 30, 55, 56
 $E(X|\mathcal{G})$ espérance conditionnelle de X si \mathcal{G} , 206
 $E_Q(X|\mathcal{G})$ espérance conditionnelle de X si \mathcal{G} sous Q , 216
 $f * g$ produit de convolution des fonctions f et g , 129
 H_n fonction de Haar, 244
 J_g matrice jacobienne, 98
 $L^1 := \mathcal{L}^1$ modulo l'égalité p.s., 58
 $L^p := \mathcal{L}^p$ modulo l'égalité p.s., 58
 m mesure de Lebesgue sur \mathbb{R} , 83
 m_n mesure de Lebesgue sur \mathbb{R}^n , 93
 $N(\mu, Q)$, 133
 $N(\mu, \sigma^2)$ loi normale de moyenne μ et de variance σ^2 , 46, 131
 $P(A|B)$, 15, 16
 $P \otimes Q$, 74
 P_X loi de la variable X , 5, 29, 54
 $P_X \otimes P_Y$ produit des lois P_X et P_Y , 75
 $P_{(X,Y)}$ loi du couple $\langle X, Y \rangle$, 75
 $Q \ll P$ Q absolument continu par rapport à P , 250
 R_n fonction de Rademacher, 244
 $X \vee Y$ maximum de X et Y , 53
 $X \wedge Y$ minimum de X et Y , 53
 X^+ maximum de X et 0, 56
 X^- maximum de $-X$ et 0, 56
 $X^{-1}(\mathcal{F})$ image inverse par X de la tribu \mathcal{F} , 51
 $X_n \xrightarrow{L^p} X$ convergence dans L^p de X_n vers X , 149
 $X_n \xrightarrow{P} X$ convergence en probabilité de X_n vers X , 149
 $X_n \xrightarrow{\mathcal{L}} X$ convergence en loi de X_n vers X , 158
 χ^2 loi du chi-deux, 89
 χ_n^2 loi du chi-deux à n degrés de liberté, 90
 \emptyset ensemble vide, 4, 7
 $\frac{dQ}{dP}$ dérivée de Radon-Nykodim de Q par rapport à P , 251
 $\Gamma(\alpha)$ fonction gamma, 46
 Γ^\perp ensemble des vecteurs orthogonaux à Γ , 198
 $\hat{\mu}$ transformée de Fourier de μ , 110
 \langle, \rangle produit scalaire, 110, 195
 $\text{Cov}(X, Y)$ covariance de X et Y , 80, 97
 $\mu_X * \mu_Y$ produit de convolution, 123
 $\otimes_{n=1}^\infty A_n$, 76
 Π opérateur de projection, 199
 $\liminf_n A_n$ (limite inférieure des ensembles A_n), 11
 $\limsup_n A_n$ (limite supérieure des ensembles A_n), 11, 78
 $\sigma(C)$ tribu engendrée par C , 7
 σ^2, σ_X^2 variance, 31, 63
 $\{\omega\}$ singleton, 23
 $\text{Var}(X)$ variance de X , 63, voir aussi σ^2, σ_X^2
 $\|X\|$ norme de X , 195
 $\|X\|_p$ norme dans L^p , 215
 \mathbb{N} ensemble des entiers naturels, 5
 \mathbb{Q} ensemble des rationnels, 8
 \mathbb{R} ensemble des réels, 2
 \mathbb{Z} ensemble des entiers relatifs, 84
 \mathcal{B} tribu borélienne de \mathbb{R} , 8, 41
 \mathcal{B}^n tribu borélienne de \mathbb{R}^n , 93
 C^∞ ensemble des fonctions indéfiniment différentiables, 167
 C_∞ tribu asymptotique, 79
 $\mathcal{E} \otimes \mathcal{F}$, 73
 \mathcal{F}_T tribu antérieure au temps d'arrêt T , 220
 \mathcal{L}^1 ensemble des variables intégrables, 30, 56
 $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ espace \mathcal{L}^1 sur (Ω, \mathcal{A}, P) , 56
 \mathcal{L}^p ensemble des variables de puissance p ème intégrable, 58
 \mathcal{N} classe des ensembles négligeables, 39

Index

A

additivité, 9
 σ -additivité, 9, 37
algèbre, 7, 37

B

Bayes, 18
Bernoulli
 fonction caractéristique, 113
 loi de variable de, 32, 125, 182, 190
Bernoulli, Jacob, 1, 131
Berry-Esseen, 191
bêta (fonction, loi), 67
Bienaymé-Chebyshev (inégalité de), 32, 63, 215

binomiale
 fonction caractéristique, 113
 loi, 25, 27, 33, 126, 170, 190, 204
binomiale négative (loi), 33
Bolzano-Weierstrass (théorème de), 164
Bonferroni (inégalités de), 14
Borel (tribu de), 8
 sur \mathbb{R} , 8, 41
 sur \mathbb{R}^n , 93
Borel-Cantelli (lemme de), 77
borélien (ensemble), 8
Box-Muller (simulation de), 107

C

Cauchy (loi de), 47, 65, 90, 105, 155
 et loi normale, 105
Cauchy (suite de), 196
Cauchy-Schwarz (inégalité de), 62, 216
changement de variable (formule du), 99
chi-deux (loi du), 89, 90, 103, 127
Cobb-Douglas (loi de), 46
coefficient de corrélation, 97
complètement convergent, 81
convergence (de variables aléatoires)
 dans L^p , 148
 en loi, 158
 en moyenne d'ordre p , 149
 en probabilité, 149
 presque sûre, 148
 simple, 147

convergence étroite (de mesures), 157
convolution de fonctions (produit de), 129
convolution de mesures de probabilité (produit de), 123
covariance, 80, 97
 et indépendance, 97
 matrice de, 98

D

densité
 sur \mathbb{R} , 45, 84
 sur \mathbb{R}^n , 94
densité conditionnelle, 95, 210
deux à deux indépendants, 16
Dirac (masse de, mesure de), 45, 162
dominée (théorème de convergence), 57, 212
Doob
 décomposition de, 226
 inégalité de, 230, 231
 inégalité des montées de, 232
 inégalité maximale de, 229
 théorème d'arrêt de, 221

E

ergodique (loi forte des grands nombres), 182
espérance, 30, 55, 56
espérance conditionnelle
 définie dans L^1 , 209
 définie dans L^2 , 206
estimateur, 123, 138
événement, 4, 8
exponentielle
 fonction caractéristique, 115
 loi, 45, 65, 102
exponentielle double (loi), 47

F

Fatou (lemme de), 57, 212
fonction de répartition, 41, 54
 empirique, 191
Fourier (transformée de), 110
Fubini (théorème de), 74, 82

G

- Galton-McAlister (loi de), 46
 gamma (loi), 46, 68, 90, 127
 et χ^2 , 90, 103, 127
 fonction caractéristique, 116
 Gauss (loi de), voir loi normale
 Gauss C.F., 131
 géométrique (loi), 26, 33
 Glivenko-Cantelli (théorème de), 191
 Gosset, W., 103

H

- Haar (fonctions de, systèmes de), 244
 Helly (principe de), 164
 Hölder (inégalité de), 213
 hypergéométrique (loi), 24, 28

I

- i.i.d. (indépendantes de même loi),
 131
 image de P par X, 54
 indépendant(e)s
 deux à deux, 16
 événements, 15
 tribus, 71
 variables aléatoires, 71
 indicatrice (fonction), 53
 intensité (d'une variable aléatoire),
 69

J

- Jacobien(ne) : matrice, déterminant,
 98
 Jensen (inégalité de), 47, 212

K

- Kolmogorov, 181, 240, 241, 247
 Kolmogorov (loi forte des grands
 nombres de), 181, 240

L

- Laplace, 131
 Laplace (loi de), voir exponentielle
 double
 Lebesgue (mesure de)
 sur \mathbb{R} , 83
 sur \mathbb{R}^n , 93

- Lebesgue (théorème de convergence
 dominée), 57, 212
 Lévy (théorème de), 173
 Lévy, P., 132
 linéaire (estimateur), 138
 linéaire (régression), 137
 logistique (loi), 69
 lognormale (loi), 46, 47, 68, 91, 121
 loi d'une variable aléatoire, 5, 29, 54
 loi forte des grands nombres, 179,
 181, 240
 loi zéro-un (0-1), 79

M

- marginale (densité), 95
 Markov (inégalité de), 31
 martingale, 217
 fermée, 218, 237
 inverse, 239
 théorème de convergence, 235, 237
 théorème-limite central, 241
 Mellin (transformée de), 121
 mesurable (fonction), 51
 Minkowski (inégalité de), 214
 Moivre, A. de, 131
 monotone (théorème de
 convergence), 57, 211
 monotones (théorème des classes),
 38, 39
 Monte Carlo, 183
 montées (nombre de), 232
 multivariée (loi normale ou
 gaussienne), 132

N

- négligeable (ensemble), 39
 non corrélées (variables aléatoires),
 137
 normale (loi), 46, 65, 89, 100, 101,
 103, 126, 127, 131
 fonction caractéristique, 114, 115,
 126, 132
 multivariée, 132
 simulation de la, 107
 normé (espace vectoriel), 195
 normé complet (espace vectoriel),
 196

O

orthogonale (matrice), 135
orthogonaux (vecteurs), 197

P

p.p., presque partout, 95
p.s., presque sûrement, 39, 57
Pareto (loi de), 33
Pascal (loi de), 33
Poincaré (identité de), 13
Poisson (loi de), 26, 27, 32, 126, 170.
201
fonction caractéristique, 114
probabilité, 4, 9
probabilité conditionnelle, 16
projection (opérateur de), 199
Pythagore (théorème de), 197

R

Rademacher (fonction de), 244
Radon-Nikodym (théorème de,
dérivée de), 250, 252
Rayleigh (loi de), 105, 106
régression, 137
régression (résidus de), 146
Riemann (fonction zêta de), 34
Riesz-Fischer (théorème de), 196

S

semi-définie positive (matrice), 98
signes aléatoires, 247
simple (variable aléatoire), 55
singleton, 23
Slutsky (théorème de), 167
sous-additivité, 13
sous-espace (de Hilbert), 198
sous-martingale, 225
stable
par différences, 38
par intersections finies, 38
par limites croissantes, 38
statistique d'ordre, 107
Stone-Weierstrass (théorème de), 118
Student (loi de), 103
surmartingale, 225
symétrique
densité, 91
variable aléatoire, 91

T

temps d'arrêt, 219
borné, 219
tension, tendue, 164
théorème-limite central, 187, 189
Tonelli-Fubini (théorème de), 74
topologique (espace), 52
transformation préservant la mesure,
182
triangulaire (loi), 52
tribu, 7
engendrée par \mathcal{C} , 7
triviale, 8
tribu asymptotique, 79

U

uniforme (loi)
(fonction caractéristique), 114
continue, 45
discrète, 24, 34
uniforme intégrabilité, 237
unimodale (loi), 48

V

variable aléatoire, 5, 29, 51
variable aléatoire simple, 42, 55
variance, 31, 63

W

Weibull (loi de), 46

Z

zéro-un (loi), 79
Zêta (loi), 33

En 28 courts chapitres, cet ouvrage expose, dans un style simple et dépouillé, les notions fondamentales de la théorie des probabilités. Il conduit le lecteur des premiers rudiments aux principaux théorèmes-limites et à la notion d'espérance conditionnelle, aboutissement traditionnel des cours de licence ou de première année de master.

Les derniers chapitres sont consacrés à un aperçu de la théorie des martingales. Ils constituent une initiation aux processus stochastiques, en même temps que l'exposé d'une théorie qui est à la base de la plupart des applications actuelles des probabilités.

La lecture de ce livre ne présuppose aucune connaissance en probabilités. Les connaissances nécessaires en analyse sont celles des deux premières années d'université. Les notions plus avancées, comme l'intégrale de Lebesgue, les espaces de Hilbert, sont introduites et étudiées quand elles deviennent nécessaires à la progression de l'exposé, ce qui rend l'ouvrage accessible aux lecteurs non universitaires intéressés notamment par les applications.

Le cours est complété par 331 énoncés d'exercices.

Jean Jacod est professeur à l'Université Paris VI-Pierre et Marie Curie. Philip Protter est professeur à Cornell University (Ithaca, N. Y., États-Unis).

Collection enseignement des mathématiques



PUSCULES

Pile ou Face

**Une introduction
aux théorèmes limites du
Calcul des Probabilités**

Emmanuel LESIGNE



PILE ou FACE

**Une introduction aux théorèmes
limites du Calcul des Probabilités**

Emmanuel Lesigne



Opuscules

Collection dirigée par Pierre Damphousse

N° 1 - *Découvrir l'arithmétique*, par Pierre Damphousse, 2000, 128 pages.

N° 2 - *PILE ou FACE, une introduction aux théorèmes limites du Calcul des Probabilités*, par Emmanuel Lesigne, 2001, 128 pages.

Emmanuel Lesigne
Département de mathématiques
Faculté des Sciences et Techniques
Université François Rabelais
Parc de Grandmont
37200 Tours, France
`lesigne@univ-tours.fr`

ISBN 2-7298-0679-2

AVANT-PROPOS

Si je lance un grand nombre de fois une pièce de monnaie équilibrée, je m'attends à observer environ autant de fois chacun des côtés de la pièce : le côté «pile» et le côté «face». L'objet de cet opuscule est de donner un sens précis à la phrase précédente. Quand le nombre de lancers devient très grand, la fréquence du nombre de piles doit tendre vers $1/2$; en quel sens ? à quelle vitesse ? quelles sont les «déviations» normales ?

Pile ou Face est une introduction au Calcul des Probabilités, et plus particulièrement à l'étude des propriétés de convergence sur les grandes séries d'observations. L'objectif est de présenter un domaine scientifique qui allie utilité et esthétique.

Le Calcul des Probabilités est la branche des mathématiques consacrée à l'étude des phénomènes aléatoires. Un «phénomène aléatoire» est une expérience dont on considère que le résultat dépend du hasard, soit parce que les conditions exactes de sa réalisation nous sont inconnues, soit parce que le hasard «existe» réellement. Nous n'entrerons pas dans cette discussion et partirons d'un modèle mathématique postulé.

Pile ou Face propose les premières étapes de la modélisation mathématique de ces phénomènes et la déduction rigoureuse de lois que l'on observe quand ils se répètent indépendamment.

Un triple parti pris a été adopté dans la rédaction de cet ouvrage.

- Le niveau de connaissance mathématique requis pour en aborder la lecture est celui du cours d'analyse des deux premières années universitaires, ou des classes préparatoires scientifiques. En particulier, la théorie de la mesure ne sera pas utilisée. Ainsi cet opuscule s'adresse aux étudiants dès leur entrée en licence ou dans une formation d'ingénieur, mais aussi aux enseignants et à toute personne munie d'une formation scientifique de base.

- Nous nous imposons le niveau de rigueur usuel d'un traité de mathématiques. Les définitions et les énoncés sont précis et les démonstrations complètes.

- Nous nous limitons pour l'essentiel à l'étude du jeu de pile ou face, avec une pièce de monnaie éventuellement faussée : nous décrivons les lois mathématiques que l'on doit observer quand on répète des expériences identiques, indépendantes et à deux issues (le succès et l'échec). Ce choix pourrait paraître trop restrictif mais c'est un fait remarquable que le simple jeu de pile ou face recelle une grande part de la complexité du Calcul des Probabilités. C'était l'avis de BOREL quand il écrivait : «*Le jeu de pile ou face, dont le principe est si simple, possède un très grand caractère de généralité et conduit, lorsqu'on l'étudie en détail, aux Mathématiques les plus élevées*» (Émile Borel, *Principes et formules classiques du Calcul des Probabilités*, Chapitre V : *Jeu de pile ou face* ; 1924).

Cet ouvrage est une invitation à la découverte du Calcul des Probabilités. Les lois que nous y décrivons sont de difficulté variable, car «élémentaire» n'est pas synonyme de «simple» ; elles sont diverses et précises. Que le lecteur ne s'attende pas à trouver ici des recettes pour gagner au Loto ou au casino. Au contraire les mathématiques nous apprennent qu'il n'existe pas de meilleure stratégie que l'abstention dans ces jeux de hasard.

Respectant l'excellente règle proposée par Pierre Damphousse, initiateur et animateur de cette collection d'opuscules, je me suis efforcé de donner des références historiques et des notices biographiques précises. Une bibliographie succincte est proposée dans le dernier chapitre.

Le contenu de *Pile ou Face* est décrit dans la table des matières et dans le premier chapitre, qui serviront de guide au lecteur.

Pour conclure cet avant-propos, je veux remercier l'ensemble des personnes qui m'ont aidé à progresser dans la connaissance scientifique : la liste des collègues et des étudiants qu'il me faudrait citer à cette occasion serait trop longue. Je tiens toutefois à remercier nominativement Jean Blanchard et Jean-Pierre Conze qui sont directement à l'origine de mon intérêt pour les mathématiques, et mes amis et collègues Pierre Damphousse, Marc Peigné et Elisabeth Rouy qui m'ont aidé dans la rédaction de ce petit livre.

Emmanuel Lesigne
Tours, février 2001

Table des matières

0	Prérequis et présentation générale	1
1	Modélisation d'une expérience aléatoire	2
	1.1. Expérience élémentaire	2
	1.2. Suite d'expériences élémentaires	4
2	Variables aléatoires	5
3	Indépendance	7
4	La distribution binomiale	11
5	La loi faible des grands nombres	13
6	Estimation des grands écarts	16
7	Le théorème limite central	20
	7.1. Énoncé	20
	7.2. Commentaires	21
	7.3. Exemples d'applications	24
	7.4. Démonstration du théorème limite central	27
8	Estimation des écarts modérés	35
9	Une loi d'Arcsinus	41
	9.1. Introduction	41
	9.2. Énoncés	42
	9.3. Le principe de réflexion	43
	9.4. Démonstration de la loi d'Arcsinus	48
	9.5. Démonstration de la loi des retours en zéro	54

10	La loi forte des grands nombres	57
	10.1. Événements presque sûrs, événements indépendants	57
	10.2. La loi forte des grands nombres de BOREL	61
	10.3. Extension au cas des suites aléatoires pouvant prendre plus de deux valeurs	64
	10.4. Nombres normaux	65
	10.5. Les lemmes de BOREL-CANTELLI	68
11	La loi du logarithme itéré	75
	11.1. Introduction	75
	11.2. L'estimation de HAUSDORFF	77
	11.3. L'estimation de HARDY et LITTLEWOOD	78
	11.4. La loi du logarithme itéré de KHINTCHINE	79
12	Récurrence des marches aléatoires	85
	12.1. Introduction et formalisme	85
	12.2. La marche aléatoire aux plus proches voisins sur \mathbb{Z}	87
	12.3. Généralités	87
	12.4. Récurrence des marches aléatoires sur \mathbb{Z}^N	96
13	Épilogue	106
—————		
	<i>Biographies</i>	107
	<i>Éléments de bibliographie</i>	115
	<i>Index</i>	117

0. Prérequis et présentation générale

Tout au long de ce texte nous noterons \mathbb{N} l'ensemble des nombres entiers naturels et \mathbb{R} l'ensemble des nombres réels.

Les notions dont la connaissance est requise pour aborder la lecture de l'opuscule se trouvent dans tout traité de premier cycle universitaire, et elles sont faciles à énumérer :

- théorie des ensembles élémentaire : ensembles, ensembles produits, applications ;
- analyse combinatoire : dénombrement, combinaisons ;
- nombres réels ; suites : limites, suites équivalentes, comparaison de suites (le sens des notations \sim , o et O est rappelé page 18) ;
- fonctions réelles d'une variable réelle : limites et continuité, fonctions classiques ; intégrale d'une fonction continue sur un intervalle de \mathbb{R} , sommes de RIEMANN.

Dans le Calcul des Probabilités, on a pris l'habitude d'appeler *théorème limite* tout résultat de convergence obtenu par observation d'une suite d'expériences aléatoires. Chacun des chapitres de 5 à 12 est centré sur un type de théorème limite.

Pile ou Face peut être partagé en trois parties. La première, contenant les chapitres 1 à 4, donne les bases du modèle mathématique décrivant une *expérience aléatoire finie*, c'est-à-dire une expérience aléatoire ayant un nombre fini d'issues possibles ; cette partie est prolongée dans les paragraphes 1 et 3 du chapitre 10, où l'on aborde la description des suites infinies d'expériences aléatoires. La seconde partie est constituée des chapitres 5 à 9 ; on y traite de théorèmes qui s'énoncent uniquement avec les probabilités sur des expériences finies : les piliers fondamentaux sont la loi faible des grands nombres et le théorème limite central, qui sont précisés par les estimations d'écarts, grands ou modérés ; les lois d'Arcsinus entrent aussi dans cette catégorie. La troisième partie est constituée des chapitres 10 à 13 ; on y aborde la modélisation des expériences aléatoires non finies, ce qui permet de donner diverses formes de la loi forte des grands nombres, une preuve complète de la loi du logarithme itéré, et les résultats de base sur la récurrence des marches aléatoires.

À partir du cinquième, chaque chapitre débute par un texte de présentation. Un résumé du contenu de l'opuscule peut être obtenu en rassemblant ces introductions.

Un traité de niveau élémentaire sur le calcul des probabilités pourrait être constitué par les chapitres 1 à 7 de ce livre, auxquels il conviendrait d'ajouter des éléments de la théorie du dénombrement, la présentation des lois discrètes et continues les plus courantes, ainsi qu'un exposé sur les probabilités conditionnelles.

1. Modélisation d'une expérience aléatoire

1.1. Expérience élémentaire. — Nous commençons par la présentation du modèle mathématique décrivant une expérience aléatoire ayant un nombre fini d d'issues possibles. Chaque issue est représentée par un

ensemble
des issues
ensemble
des
éventualités

$$\Omega := \{\omega^1, \omega^2, \dots, \omega^d\}.$$

probabilité
d'une issue

À chaque issue (ou éventualité) ω^i est associée une *probabilité* p_i ; chaque probabilité p_i est un nombre réel positif ou nul et $\sum_{i=1}^d p_i = 1$.

Il est important de noter que, dans ce modèle, les probabilités des éventualités sont données *a priori* ; le travail qui consiste à déterminer ces probabilités à partir d'observations relève de la Statistique qui est une branche mathématique sœur, mais distincte, du Calcul des Probabilités. La Statistique mathématique utilise les outils et résultats qui sont présentés dans cet opuscule.

événement
probabilité
d'un
événement

Revenons à notre modèle pour introduire un peu de vocabulaire. Un sous-ensemble de Ω est appelé un *événement* et on définit la *probabilité d'un événement* comme la somme des probabilités des issues appartenant à cet événement : si $A \subset \Omega$ est un événement, sa probabilité $P(A)$ est définie par

$$P(A) := \sum_{\omega_i \in A} p_i.$$

On a en particulier $P(\{\omega_i\}) = p_i$, que l'on écrit simplement $P(\omega_i) = p_i$.

fonction
indicatrice

On note $\mathbf{1}_A$ la *fonction indicatrice* de A , c'est-à-dire l'application de Ω dans $\{0, 1\}$ qui prend la valeur 1 sur A et la valeur 0 sur son

complémentaire A^c . Ainsi, on a

$$P(A) = \sum_{i=1}^d p_i \mathbf{1}_A(\omega_i).$$

En résumé, le modèle mathématique est donné par un couple (Ω, P) où Ω est un ensemble fini et P est une application de l'ensemble des sous-ensembles de Ω dans l'intervalle $[0, 1]$ vérifiant les conditions suivantes : $P(\Omega) = 1$ et, si A et B sont deux sous-ensembles de Ω tels que $A \cap B = \emptyset$, alors $P(A \cup B) = P(A) + P(B)$. Un tel couple (Ω, P) est appelé un *espace probabilisé fini* et l'application P est appelée une *probabilité*. Les règles de calcul suivantes sont de vérification immédiate : $P(\emptyset) = 0$; si $A \subset \Omega$ alors $P(A^c) = 1 - P(A)$; si $A, B \subset \Omega$ alors $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Dans le cas particulier où toutes les issues ont même probabilité, on dit que l'espace Ω est muni de la *probabilité uniforme*. La probabilité d'un événement se calcule alors facilement : c'est le quotient de son nombre d'éléments par d , le nombre d'éléments de Ω . C'est cette situation qui est décrite dans l'expression classique : « la probabilité d'un événement est le rapport du nombre de cas favorables au nombre de cas possibles ».

Voici quelques exemples :

1. Le lancer d'une pièce de monnaie équilibrée est décrit par un ensemble Ω à deux éléments et d'une probabilité attribuant la même valeur aux deux issues. En convenant de noter 1 l'issue « face » et 0 l'issue « pile », on a : $\Omega = \{0, 1\}$ et $P(0) = P(1) = \frac{1}{2}$. On dit alors que l'espace $\{0, 1\}$ est muni de la probabilité uniforme $(\frac{1}{2}, \frac{1}{2})$.

2. Plus généralement, le modèle décrivant une expérience aléatoire à deux issues, conventionnellement baptisées « succès » et « échec », est déterminé par un paramètre p , nombre réel entre 0 et 1 qui représente la probabilité de succès. En notant 1 pour le succès et 0 pour l'échec, on a : $\Omega = \{0, 1\}$; $P(0) = 1 - p$; $P(1) = p$. On dit alors que l'espace $\{0, 1\}$ est muni de la probabilité $(1 - p, p)$.

3. Le tirage d'un numéro au Loto est modélisé par le couple (Ω, P) où $\Omega = \{1, 2, 3, \dots, 49\}$ et P est la probabilité uniforme sur Ω . Ici, pour chaque $\omega \in \Omega$, on a $P(\omega) = \frac{1}{49}$. Si on ne s'intéresse qu'à la parité du numéro, on considère un espace Ω à deux éléments muni de la probabilité $(\frac{24}{49}, \frac{25}{49})$.

4. Des expériences simples peuvent conduire rapidement à des espaces probabilisés énormes. Par exemple, pour décrire le tirage d'une « main de bridge », c'est-à-dire un choix de 13 cartes à jouer parmi 52, l'espace Ω aura six cent trente cinq milliards treize millions cinq cent cinquante neuf mille

six cents éléments (il s'agit du nombre de combinaisons C_{52}^{13} , cf. Chapitre 4). Si l'on suppose que le jeu est correctement battu avant d'être distribué, on munira cet espace de la probabilité uniforme.

Comme toute les branches des mathématiques, le Calcul des Probabilités a quelques notations qui lui sont propres. Si X est une application de Ω dans un ensemble E et si F est un sous-ensemble de E il est d'usage de noter $(X \in F)$ l'image réciproque de F par X . Ainsi,

$$(X \in F) := \{\omega \in \Omega : X(\omega) \in F\} .$$

Ici X s'interprète comme un élément aléatoire de l'ensemble E et $X(\omega)$ est la valeur de cet élément sous l'aléa ω . La probabilité de l'événement $(X \in F)$ est simplement notée $P(X \in F)$.

1.2. Suite d'expériences élémentaires. — L'exemple qui nous intéresse plus particulièrement dans ce texte est celui d'une répétition d'expériences identiques et indépendantes. Dans la première partie de l'opuscule, nous ne considérons que des suites finies d'expériences. L'étude des suites infinies débute au chapitre 10.

On considère donc une expérience aléatoire «composée» qui consiste à effectuer n fois une expérience aléatoire élémentaire, que l'on suppose à deux issues possibles, le succès, désigné par le chiffre 1, ou l'échec, désigné par le chiffre 0. Notre modèle devra traduire le fait que les expériences élémentaires sont identiques et indépendantes. On note p la probabilité de succès et $q = 1 - p$ la probabilité d'échec. Une issue de l'expérience composée est notée par une suite finie de n zéros ou uns. L'espace Ω , noté ici Ω_n , est l'ensemble des n -uplets de 0 et de 1. Autrement dit, $\Omega_n = \{0, 1\}^n$. Les éléments de Ω_n sont notés $\omega = (\omega_1, \omega_2, \dots, \omega_n)$, avec $\omega_i = 0$ ou 1. La probabilité sur Ω_n est notée P_n . Nous traduisons le fait que les expériences élémentaires sont identiques par la règle suivante :

$$\text{pour chaque } i \text{ entre 1 et } n, \quad P_n(\omega_i = 0) = q \quad \text{et} \quad P_n(\omega_i = 1) = p .$$

Nous traduisons le fait que le résultat de la $(i + 1)$ -ème expérience élémentaire est indépendant des résultats des i expériences élémentaires qui l'ont précédée par la règle suivante :

$$\text{pour tout } (e_1, e_2, \dots, e_i) \in \{0, 1\}^i,$$

$$P_n(\omega_{i+1} = 1 \text{ et } (\omega_1, \omega_2, \dots, \omega_i) = (e_1, e_2, \dots, e_i))$$

$$= P_n(\omega_{i+1} = 1) \times P_n((\omega_1, \omega_2, \dots, \omega_i) = (e_1, e_2, \dots, e_i)) .$$

Par récurrence sur n , on constate que ces deux règles définissent de façon unique la probabilité P_n ; si pour chaque issue ω de l'expérience composée, on note $S_n(\omega)$ le nombre de succès :

$$S_n(\omega) = S_n((\omega_1, \omega_2, \dots, \omega_n)) := \omega_1 + \omega_2 + \dots + \omega_n ,$$

alors la probabilité P_n est donnée par

$$P_n(\omega) = p^{S_n(\omega)} q^{n-S_n(\omega)} .$$

On dit que l'espace $\Omega_n = \{0, 1\}^n$ est muni de la *probabilité produit* $P_n = (q, p)^{\otimes n}$. probabilité produit

Dans le cas où les probabilités de succès et d'échec sont égales, l'espace Ω_n est muni de la probabilité uniforme. Ceci est conforme à l'intuition : dans l'expérience qui consiste à lancer n fois une pièce équilibrée et à noter les résultats successifs, toutes les issues sont équiprobables. La probabilité d'un événement est alors simplement égale à son cardinal divisé par 2^n .

2. Variables aléatoires

Soit (Ω, P) un espace probabilisé fini. Une application définie sur Ω et à valeurs réelles est appelée une *variable aléatoire* . Traditionnellement, les variables aléatoires sont désignées par des lettres majuscules comme par exemple variable aléatoire

$$X : \Omega \rightarrow \mathbb{R} \\ \omega \mapsto X(\omega) .$$

La *loi de la variable aléatoire* X est la donnée des probabilités des événements déterminés par les valeurs de X . Si la variable aléatoire X prend les valeurs x_1, x_2, \dots, x_k , alors les événements $(X = x_i)$, $1 \leq i \leq k$, forment une partition de Ω et la loi de la variable aléatoire est la donnée des $(x_i, P(X = x_i))$, $1 \leq i \leq k$. Dans ces expressions le mot «loi» peut être remplacé par son synonyme «distribution». loi d'une variable aléatoire

On associe à la variable aléatoire X son *espérance* (ou *moyenne*) notée $E[X]$ et définie par la formule espérance d'une variable aléatoire

$$E[X] = \sum_{i=1}^k x_i P(X = x_i) .$$

Cette notion d'espérance et son nom (*expectatio*, en latin) ont été introduits par Christiaan HUYGENS dans une analyse des justes mises pour les jeux de hasard^{2.1}.

¹ C. Huygens, *De ratiociniis in aleae ludo*, 1657.

Ce nombre $E\{X\}$ représente la moyenne, sous la probabilité P , des valeurs prises par la variable aléatoire X . On déduit directement de cette définition les propriétés suivantes :

- si X est constante, alors $E[X] = X$; en particulier, $E[E\{X\}] = E\{X\}$;
- si $X \geq 0$, alors $E[X] \geq 0$;
- $|E\{X\}| \leq E\{|X|\}$;
- l'application E agit linéairement sur l'espace vectoriel réel des variables aléatoires sur Ω (i.e. si X, X' sont deux variables aléatoires et λ un nombre réel, alors $E[X + X'] = E[X] + E[X']$ et $E[\lambda X] = \lambda E[X]$).

Pour vérifier cette dernière affirmation, on peut remarquer qu'une variable aléatoire X peut s'écrire de multiples façons comme combinaison linéaire de fonctions indicatrices d'événements et que, si $X = \sum_i y_i \mathbf{1}_{A_i}$ est une telle écriture, alors on a toujours $E[X] = \sum_i y_i P(A_i)$.

Cette dernière remarque permet aussi de démontrer la formule permettant de calculer l'espérance d'une fonction d'une variable aléatoire : si X est une variable aléatoire, et si f est une fonction réelle définie sur l'ensemble E des valeurs prises par X , alors $f(X) := f \circ X$ est une variable aléatoire et on a

$$E[f(X)] = \sum_{x \in E} f(x) P(X = x). \quad (2.1)$$

Les deux énoncés suivants donnent des inégalités plus remarquables par leur utilité que par leur difficulté.

inégalité de Markov **PROPOSITION 2.1** (Inégalité de MARKOV). — *Soit X est une variable aléatoire positive. Pour tout $a > 0$,*

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Démonstration. — Notons x_1, x_2, \dots, x_k les valeurs prises par X . Ce sont des nombres réels positifs (ou nuls) et on a

$$P(X \geq a) = \sum_{i: x_i \geq a} P(X = x_i) \leq \sum_{i: x_i \geq a} \frac{x_i}{a} P(X = x_i) \leq \frac{1}{a} E[X].$$

— QED (Proposition 2.1)

Soit X une variable aléatoire. Pour tout $a > 0$,

$$P(|X - E[X]| \geq a) \leq \frac{1}{a^2} E[(X - E[X])^2].$$

Pour justifier ce corollaire, il suffit d'appliquer l'inégalité de Markov à la variable aléatoire $(X - E[X])^2$.

La quantité $E[(X - E[X])^2]$ est appelée la *variance* de la variable aléatoire X et notée $\text{var}(X)$. En développant le carré et en utilisant la linéarité de l'espérance on obtient :

variance
d'une
variable
aléatoire

$$\text{var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

La racine carrée de la variance de X est appelée l'*écart type* de X . L'écart type représente une valeur moyenne de l'écart entre les valeurs de la variable aléatoire et son espérance.

écart type
d'une
variable
aléatoire

Terminons par une remarque sur les notations : si la probabilité est notée P_n , l'espérance sera naturellement notée E_n .

3. Indépendance

Nous allons introduire à présent la notion d'indépendance pour une famille d'événements, puis pour une famille de variables aléatoires. Nous considérons comme auparavant un espace probabilisé fini (Ω, P) .

Intuitivement, on dit que deux événements sont indépendants si la réalisation de l'un ne modifie pas la probabilité de l'autre. Cela se traduit par la définition suivante. Deux sous-ensembles A et B de Ω sont des *événements indépendants* si $P(A \cap B) = P(A) \times P(B)$. En dehors du cas trivial où $P(B) = 0$, cela s'écrit encore $P(A \cap B)/P(B) = P(A)/P(\Omega)$.

événements
indépendants

EXEMPLE. — Le modèle que nous avons adopté précédemment pour décrire la réalisation de deux expériences à deux issues identiques et indépendantes, avec probabilité de succès p et probabilité d'échec $q = 1 - p$, est le

² M. Bienaymé, *Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés*, Journal de Mathématiques pures et appliquées, vol. 12, p. 158-176, 1867.

¹ P.L. Tchebychev, *Des valeurs moyennes*, Journal de Mathématiques pures et appliquées, vol. 12, p. 177-184, 1867.

suivant :

$$\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\} = \{(\omega_1, \omega_2) : \omega_i = 0 \text{ ou } 1\} ,$$

$$P(0, 0) = q^2 , P(0, 1) = P(1, 0) = pq , P(1, 1) = p^2 .$$

Le modèle associé à la réalisation de deux expériences (toujours à deux issues) indépendantes mais pas nécessairement identiques, dont les probabilités de succès sont notées p_1 et p_2 , est donné par le même espace Ω muni de la probabilité :

$$P(0, 0) = (1 - p_1)(1 - p_2) , P(0, 1) = (1 - p_1)p_2 ,$$

$$P(1, 0) = p_1(1 - p_2) , P(1, 1) = p_1p_2 .$$

Le lecteur vérifiera le fait élémentaire que la probabilité P est entièrement déterminée par les valeurs de $P(\omega_i = 1)$, $i = 1, 2$, et la condition «les événements $(\omega_1 = 1)$ et $(\omega_2 = 1)$ sont indépendants». —◇

Une famille d'événements $\{A_1, A_2, \dots, A_k\}$ est appelée une *famille d'événements indépendants* si, pour tous nombres entiers i_1, i_2, \dots, i_ℓ tels que $1 \leq i_1 < i_2 < \dots < i_\ell \leq k$, on a

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_\ell}) = P(A_{i_1}) \times P(A_{i_2}) \times \dots \times P(A_{i_\ell}) .$$

Ici, quelques mises en garde s'imposent.

- Il ne suffit pas que $P(\bigcap_{i=1}^k A_i) = \prod_{i=1}^k P(A_i)$ pour que les événements soient indépendants (cela est particulièrement évident si un des événements est vide).
- Il ne suffit pas que les événements soient deux à deux indépendants pour qu'ils soient indépendants «dans leur ensemble». L'exemple le plus simple pour illustrer ce fait est dû à BERNSTEIN ; le voici : reprenons le modèle de deux lancers d'une pièce équilibrée, en décidant arbitrairement de qualifier «succès» l'apparition de «face» ; c'est le premier des deux exemples précédents avec $p = \frac{1}{2}$: considérons les trois événements suivants : $A_1 := (\omega_1 = 1)$, *i.e.* le premier lancer donne face, $A_2 := (\omega_2 = 1)$, *i.e.* le second lancer donne face, et $A_3 := (\omega_1 = \omega_2)$, *i.e.* les deux lancers donnent des résultats identiques ; chacune des trois paires formées par deux de ces événements est une paire d'événements indépendants, mais la famille $\{A_1, A_2, A_3\}$ n'est pas une famille d'événements indépendants ; en effet, on a

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \quad \text{et} \quad P(A_1) = P(A_2) = P(A_3) = \frac{1}{2} .$$

variables
aléatoires
indépendantes

La notion d'indépendance de variables aléatoires se déduit naturellement de celle d'indépendance d'événements. Soit $\{X_1, X_2, \dots, X_k\}$ une famille de variables aléatoires définies sur (Ω, \mathcal{P}) . Ces variables aléatoires sont dites *indépendantes* si, pour tous x_1, x_2, \dots, x_k réels, les événements $(X_1 = x_1), (X_2 = x_2), \dots, (X_k = x_k)$ sont indépendants. A titre de petit exercice d'entraînement, le lecteur pourra vérifier les propriétés suivantes :

- les variables aléatoires X_1, X_2, \dots, X_k sont indépendantes si et seulement si, pour tous B_1, B_2, \dots, B_k sous-ensembles de \mathbb{R} , les événements $(X_1 \in B_1), (X_2 \in B_2), \dots, (X_k \in B_k)$ sont indépendants ;
- la propriété d'indépendance n'est évidemment pas liée à l'ordre dans lequel sont rangées les variables, et elle passe aux sous-familles ;
- les variables aléatoires X_1, X_2, \dots, X_k sont indépendantes si et seulement si, pour tout j entre 2 et k , pour tous x_1, x_2, \dots, x_j , les deux événements $(X_1 = x_1 \text{ et } X_2 = x_2 \text{ et } \dots \text{ et } X_{j-1} = x_{j-1})$ et $(X_j = x_j)$ sont indépendants (on dit que la variable aléatoire X_j est indépendante du *vecteur aléatoire* $(X_1, X_2, \dots, X_{j-1})$) ;
- des sous-ensembles de Ω forment une famille d'événements indépendants si et seulement si leurs indicatrices forment une famille de variables aléatoires indépendantes.

Remarque sur les notations : dans l'avant-dernière des affirmations précédentes, nous avons écrit l'événement $(X_1 = x_1) \cap (X_2 = x_2)$ sous la forme $(X_1 = x_1 \text{ et } X_2 = x_2)$. Cette utilisation de la conjonction de coordination *et* pour l'intersection ensembliste est naturelle et nous l'adoptons dans la suite.

Les deux propositions suivantes nous seront très utiles.

PROPOSITION 3.1. — Soit $(X_i)_{i \in I}$ une famille finie de variables aléatoires indépendantes et J, K deux sous-ensembles de I . Si $J \cap K = \emptyset$, si Y est une fonction réelle de $(X_i)_{i \in J}$ et si Z est une fonction réelle de $(X_i)_{i \in K}$, alors Y et Z sont deux variables aléatoires indépendantes.

Démonstration. — Soit f , respectivement g , une application de \mathbb{R}^J , respectivement \mathbb{R}^K , dans \mathbb{R} . Supposons que $Y = f((X_j)_{j \in J})$ et $Z = g((X_k)_{k \in K})$. Soient $a, b \in \mathbb{R}$. Notons A l'ensemble fini des $(x_j)_{j \in J}$ tels que $f((x_j)_{j \in J}) = a$ et il existe $\omega \in \Omega$ tel que $(X_j(\omega))_{j \in J} = (x_j)_{j \in J}$. Notons B l'ensemble fini des $(x_k)_{k \in K}$ tels que $g((x_k)_{k \in K}) = b$ et il existe $\omega \in \Omega$ tel que $(X_k(\omega))_{k \in K} = (x_k)_{k \in K}$. On a

$$(Y = a) = \bigcup_{(x_j) \in A} \bigcap_{j \in J} (X_j = x_j), \quad (Z = b) = \bigcup_{(x_k) \in B} \bigcap_{k \in K} (X_k = x_k)$$

et

$$(Y = a \text{ et } Z = b) = \bigcup_{(x_j) \in A, (x_k) \in B} \bigcap_{i \in J \cup K} (X_i = x_i)$$

Cette expression est une réunion d'événements deux à deux disjoints ; on a donc

$$P(Y = a \text{ et } Z = b) = \sum_{(x_j) \in A, (x_k) \in B} P \left(\bigcap_{i \in J \cup K} (X_i = x_i) \right).$$

En utilisant l'indépendance des variables aléatoires X_i on obtient

$$\begin{aligned} P(Y = a \text{ et } Z = b) &= \sum_{(x_j) \in A, (x_k) \in B} \prod_{i \in J \cup K} P(X_i = x_i) \\ &= \left(\sum_{(x_j) \in A} \prod_{j \in J} P(X_j = x_j) \right) \left(\sum_{(x_k) \in B} \prod_{k \in K} P(X_k = x_k) \right) \\ &= P(Y = a) P(Z = b). \end{aligned}$$

Cela prouve que les variables aléatoires Y et Z sont indépendantes.

— QED (Proposition 3.1)

PROPOSITION 3.2. — *Si X et Y sont deux variables aléatoires indépendantes, alors*

$$E[XY] = E[X] \times E[Y] \quad \text{et} \quad \text{var}(X + Y) = \text{var}(X) + \text{var}(Y).$$

Démonstration. — Si on note A l'ensemble (fini) des valeurs prises par X , et B l'ensemble des valeurs prises par Y , on a

$$E[XY] = \sum_{(a,b) \in A \times B} ab P(X = a \text{ et } Y = b)$$

et la propriété d'indépendance nous donne immédiatement

$$\begin{aligned} E[XY] &= \sum_{a \in A, b \in B} ab P(X = a) P(Y = b) \\ &= \sum_{a \in A} a P(X = a) \sum_{b \in B} b P(Y = b) = E[X] E[Y]. \end{aligned}$$

Cette formule est utilisée dans le calcul suivant :

$$\begin{aligned} \text{var}(X + Y) &= E \{ (X + Y)^2 \} - (E[X + Y])^2 \\ &= (E[X^2] + E[Y^2] + 2E[XY]) - (E[X]^2 + E[Y]^2 + 2E[X]E[Y]) \end{aligned}$$

$$= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \text{var}(X) + \text{var}(Y) .$$

— QED (Proposition 3.2)

4. La distribution binomiale

Dans la suite de cet opuscule, nous nous concentrons sur le problème suivant : notons S_n le nombre de succès observés au cours d'une suite de n expériences aléatoires identiques et indépendantes, à deux issues ; que peut-on dire du comportement de la suite (S_n) quand n tend vers l'infini ?

Nous notons p la probabilité de succès pour chacune des expériences élémentaires et nous travaillons dans l'espace probabilisé (Ω_n, \mathbb{P}_n) décrit dans le paragraphe 1.2. Sur l'espace $\Omega_n = \{0, 1\}^n$, l'application S_n est la variable aléatoire définie par $S_n(\omega) = \sum_{i=1}^n \omega_i$.

Rappelons la définition des *coefficients binomiaux* $\binom{n}{k}$. Si k et n sont deux nombres entiers tels que $0 \leq k \leq n$, on désigne par $\binom{n}{k}$ le nombre de sous-ensembles à k éléments dans un ensemble à n éléments. (Dans la littérature mathématique française classique ces coefficients sont notés C_n^k .) Ces nombres vérifient la relation de récurrence du triangle de PASCAL

coefficients
binomiaux

$$\binom{n}{0} = \binom{n}{n} = 1 \quad \text{et, si } 0 \leq k < n, \text{ alors} \quad \binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1} .$$

Ils sont donnés par la formule

$$\binom{n}{k} = \frac{n!}{k! \times (n-k)!} ,$$

et ils interviennent dans la formule du binôme de NEWTON :

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} . \quad (4.1)$$

PROPOSITION 4.1. — *La variable aléatoire S_n ne prend que des valeurs entières entre 0 et n et, pour tout k entre 0 et n , on a*

$$\mathbb{P}_n(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} .$$

On dit que la variable aléatoire S_n suit la loi binomiale de paramètres n et p .

loi
binomiale

Démonstration. — On note $q = 1 - p$. Pour chaque $\omega \in \Omega_n$ on a $P_n(\omega) = p^{S_n(\omega)} q^{n-S_n(\omega)}$. L'événement $(S_n = k)$ est formé d'issues ω ayant toutes mêmes probabilités égales à $p^k q^{n-k}$. La probabilité de cet événement est donc égale à $p^k q^{n-k}$ multiplié par son cardinal, c'est-à-dire par le nombre de façons d'obtenir k succès parmi n expériences : ce nombre est $\binom{n}{k}$. — QED (Proposition 4.1)

loi de Bernoulli Une variable aléatoire X suit la *loi de BERNOULLI* de paramètre p si elle prend les valeurs 0 et 1, avec $P(X = 1) = p$ et $P(X = 0) = 1 - p$. C'est la loi binomiale de paramètres 1 et p . Une autre façon d'introduire la loi binomiale de paramètres n et p est donnée par la proposition suivante, qui se déduit aisément des considérations du précédent chapitre.

PROPOSITION 4.2. — Si X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes, toutes de loi de BERNOULLI de paramètre p , alors leur somme $X_1 + X_2 + \dots + X_n$ est une variable aléatoire qui suit la loi binomiale de paramètres n et p .

PROPOSITION 4.3. — On a $E_n[S_n] = np$ et $\text{var}(S_n) = np(1 - p)$.

Démonstration (Proposition 4.3) — De la proposition 4.1, on déduit que

$$E[S_n] = \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k}$$

et, en utilisant la formule (2.1),

$$\text{var}(S_n) = \sum_{k=0}^n (k - E[S_n])^2 \binom{n}{k} p^k (1 - p)^{n-k}.$$

Les formules annoncées s'obtiennent alors facilement à partir de la formule du binôme de NEWTON. (On peut utiliser l'astuce suivante : à partir de la formule du binôme (4.1), utiliser les formules obtenues en dérivant deux fois chacun des membres par rapport à la variable a .)

Remarquons que la proposition 4.3 se déduit aussi immédiatement des propositions 4.2 et 3.2, puisque si X est une variable aléatoire de loi de BERNOULLI de paramètre p , alors on a $E[X] = p$ et $\text{var}(X) = p(1 - p)$. — QED (Proposition 4.3)

5. La loi faible des grands nombres

Le sens communément donné à la probabilité de succès d'une expérience aléatoire est la fréquence des succès dans une suite de répétitions (identiques et indépendantes) de cette expérience. Or, dans notre modèle mathématique, nous avons choisi une autre voie puisque la probabilité de succès est donnée *a priori*. La loi des grands nombres va réconcilier ces deux points de vue ; ce théorème établit qu'au cours d'un grand nombre de réalisations de l'expérience, la probabilité pour que la fréquence des succès soit proche de la probabilité théorique de succès est grande, c'est-à-dire peu éloignée de 1. C'est une première justification du modèle mathématique du Calcul des Probabilités.

loi des
grands
nombres

Nous reprenons le cadre présenté dans le paragraphe précédent : $\Omega_n = \{0, 1\}^n$ est l'espace des issues pour n réalisations de l'expérience ; nous supposons ces expériences identiques et indépendantes, ce qui nous conduit à munir l'espace Ω_n de la probabilité produit $P_n = (1-p, p)^{\otimes n}$ où p est un paramètre entre 0 et 1, représentant la probabilité théorique de succès (cf. §1.2.) ; on note encore S_n la variable aléatoire qui compte le nombre de succès :

$$S_n(\omega) = \omega_1 + \omega_2 + \dots + \omega_n \quad \text{si } \omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Omega_n .$$

On sait que la variable aléatoire S_n suit la loi binomiale de paramètres n et p .

La fréquence empirique des succès est $\frac{S_n}{n}$ et, quand n est grand, on s'attend à ce que cette fréquence soit proche de p . Le théorème suivant nous dit que cela est vrai avec une forte probabilité. Ce résultat apparaît pour la première fois dans une œuvre posthume de Jacques BERNOULLI^{5.1}.

THÉORÈME 5.1 (Loi faible des grands nombres). —

Pour chaque $\epsilon > 0$,

$$P_n \left(\left| \frac{S_n}{n} - p \right| > \epsilon \right) \xrightarrow{n \rightarrow \infty} 0$$

loi faible
des
grands
nombres

et cette convergence est uniforme en p .

La dénomination de *loi des grands nombres* nous vient de Siméon

1. Bernoulli, *Ars conjectandi*, 1713.

Denis POISSON ^{5.2}, qui a généralisé le résultat de BERNOULLI au cas où les probabilités de succès peuvent varier entre les expériences. (Nous rencontrerons ce résultat plus tard : affirmation (10.4), Théorème 10.12.)

Démonstration (Loi faible des grands nombres) —

Nous savons que la variance de la variable aléatoire S_n est $\text{var}(S_n) = np(1-p)$. Grâce à l'inégalité de BIENAYMÉ-TCHEBYCHEV (Corollaire 2.2), on peut écrire

$$P_n(|S_n - np| > n\epsilon) \leq \frac{1}{(n\epsilon)^2} \text{var}(S_n) = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2},$$

ce qui prouve le théorème. — QED (Loi faible des grands nombres)

Dans le cas des lancers d'une pièce de monnaie équilibrée ($p = 1/2$), la loi faible des grands nombres nous dit que la proportion de suites de n «0» ou «1» dans lesquelles la fréquence des «1» diffère de $1/2$ de moins de ϵ tend vers 1 quand n tend vers l'infini.

La loi faible des grands nombres de BERNOULLI possède une jolie application au problème de l'approximation uniforme d'une fonction continue sur un intervalle de \mathbb{R} par un polynôme. Le théorème de WEIERSTRASS nous dit que, pour toute fonction réelle f définie et continue sur un intervalle fermé et borné $[a, b]$ de \mathbb{R} et pour tout $\epsilon > 0$, il existe une fonction polynôme réelle g telle que

$$\sup_{a \leq x \leq b} |f(x) - g(x)| < \epsilon.$$

polynômes de Bernstein Serge BERNSTEIN^{5.3} a remarqué que ce résultat peut s'obtenir comme corollaire de la loi des grands nombres, et que cette méthode a l'avantage de donner une formule explicite pour les polynômes d'approximation.

PROPOSITION 5.2 (Polynômes d'approximation de BERNSTEIN). — Soit f une fonction réelle définie et continue sur l'intervalle $[0, 1]$. On a :

$$\sup_{0 \leq x \leq 1} \left| f(x) - \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \right| \xrightarrow{n \rightarrow \infty} 0$$

^{5.2} D.S. Poisson, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*, Paris, 1837.

^{5.3} S. Bernstein, *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités*, Soobschs. Charkovskovo Mat. Obsch, vol. 13, p. 1-2, 1912.

On déduit bien sûr de ce théorème une formule pour des polynômes d'approximation sur un intervalle $[a, b]$ quelconque : ils sont donnés par

$$x \mapsto \sum_{k=0}^n \binom{n}{k} f\left(a + (b-a)\frac{k}{n}\right) \left(\frac{x-a}{b-a}\right)^k \left(\frac{b-x}{b-a}\right)^{n-k}.$$

Démonstration (Proposition 5.2) — Soit $\epsilon > 0$. La fonction f étant uniformément continue, il existe $\eta > 0$ tel que

$$0 \leq x, y \leq 1 \text{ et } |x - y| < \eta \Rightarrow |f(x) - f(y)| < \epsilon.$$

Considérons l'espace probabilisé (Ω_n, \mathbb{P}_n) introduit précédemment et la variable aléatoire $f\left(\frac{S_n}{n}\right)$. Les polynômes d'approximation de BERNSTEIN apparaissent comme l'espérance de cette variable aléatoire ; on a en effet

$$\mathbb{E}_n \left[f\left(\frac{S_n}{n}\right) \right] = \sum_{k=0}^n f\left(\frac{k}{n}\right) \mathbb{P}_n(S_n = k) = \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) p^k (1-p)^{n-k}.$$

La loi des grands nombres nous dit qu'il existe un entier n_0 , indépendant du paramètre p , tel que, pour tout $n \geq n_0$, on ait

$$\mathbb{P}_n \left(\left| \frac{S_n}{n} - p \right| > \eta \right) < \epsilon.$$

On a :

$$\left| \mathbb{E}_n \left[f\left(\frac{S_n}{n}\right) \right] - f(p) \right| = \left| \sum_{k=0}^n \left(f\left(\frac{k}{n}\right) - f(p) \right) \mathbb{P}_n(S_n = k) \right|.$$

Grâce à l'inégalité triangulaire cette expression est majorée par :

$$\begin{aligned} & \sum_{\left| \frac{k}{n} - p \right| \leq \eta} \left| f\left(\frac{k}{n}\right) - f(p) \right| \mathbb{P}_n(S_n = k) \\ & + \sum_{\left| \frac{k}{n} - p \right| > \eta} \left(\left| f\left(\frac{k}{n}\right) \right| + |f(p)| \right) \mathbb{P}_n(S_n = k) \\ & \leq \sum_{\left| \frac{k}{n} - p \right| \leq \eta} \epsilon \mathbb{P}_n(S_n = k) + \sum_{\left| \frac{k}{n} - p \right| > \eta} 2 \sup_{0 \leq x \leq 1} |f(x)| \mathbb{P}_n(S_n = k) \\ & = \epsilon \mathbb{P}_n \left(\left| \frac{S_n}{n} - p \right| \leq \eta \right) + 2 \sup_{0 \leq x \leq 1} |f(x)| \mathbb{P}_n \left(\left| \frac{S_n}{n} - p \right| > \eta \right). \end{aligned}$$

Pour tout $n \geq n_0$, on a donc

$$\left| \mathbb{E}_n \left[f \left(\frac{S_n}{n} \right) \right] - f(p) \right| \leq \epsilon + 2\epsilon \sup_{0 \leq x \leq 1} |f(x)|.$$

Cela prouve bien que $\left| \mathbb{E}_n \left[f \left(\frac{S_n}{n} \right) \right] - f(p) \right|$ est arbitrairement petit, uniformément en p , quand n est assez grand. — QED (Proposition 5.2)

6. Estimation des grands écarts

Plusieurs des résultats que nous exposons dans la suite de cet ouvrage sont des raffinements de la loi faible des grands nombres. Nous présentons à présent le premier d'entre eux, qui établit que la vitesse de convergence dans cette loi des grands nombres est exponentielle.

En utilisant l'inégalité de BIENAYMÉ-TCHEBYCHEV dans la démonstration de la loi faible des grands nombres, nous avons obtenu l'estimation suivante :

$$\mathbb{P}_n \left(\left| \frac{S_n}{n} - p \right| \geq \epsilon \right) \leq \frac{p(1-p)}{n\epsilon^2}.$$

Comme l'a remarqué Serge BERNSTEIN^{6.1}, cette majoration peut être considérablement améliorée, au moins pour les grandes valeurs de n .

Pour tout $\epsilon \in]0, 1-p[$, posons

$$h_+(\epsilon) := (p + \epsilon) \ln \frac{p + \epsilon}{p} + (1 - p - \epsilon) \ln \frac{1 - p - \epsilon}{1 - p}.$$

THÉORÈME 6.1 (Estimation des grands écarts, ou estimation des grandes déviations). — On a $h_+(\epsilon) > 0$ et, pour tout $n \geq 1$,

grands écarts
ou grandes
déviations

$$\mathbb{P}_n \left(\frac{S_n}{n} \geq p + \epsilon \right) \leq e^{-nh_+(\epsilon)}.$$

Voici deux conséquences directes de ce théorème :

1. Si $0 < \epsilon < p$, alors $h_-(\epsilon) := h_+(-\epsilon)$ est bien défini et > 0 et on a

$$\mathbb{P}_n \left(\frac{S_n}{n} \leq p - \epsilon \right) \leq e^{-nh_-(\epsilon)}.$$

^{6.1} S. Bernstein, *Sur une modification de l'inégalité de Tchebychev*, Ann. Sc. Instit. Sav. Ukraine, Sect. Math. I, 1924.

(Ces affirmations se justifient en appliquant le théorème après interversion des notions de succès et d'échec ; ainsi S_n est changé en $n - S_n$ et p est changé en $1 - p$.)

2. Si $0 < \epsilon < \min(p, 1 - p)$, alors

$$P_n \left(\left| \frac{S_n}{n} - p \right| \geq \epsilon \right) \leq e^{-nh_+(\epsilon)} + e^{-nh_-(\epsilon)},$$

qui tend vers zéro à vitesse exponentielle quand n tend vers l'infini.

Démonstration (Estimation des grands écarts) — Fixons $t > 0$. On a :

$$P_n \left(\frac{S_n}{n} \geq p + \epsilon \right) = P_n \left(e^{t(S_n - np - n\epsilon)} \geq 1 \right).$$

En utilisant l'inégalité de MARKOV, on obtient :

$$\begin{aligned} P_n \left(\frac{S_n}{n} \geq p + \epsilon \right) &\leq E_n \left[e^{t(S_n - np - n\epsilon)} \right] \\ &= e^{-nt(p+\epsilon)} E_n \left[e^{tS_n} \right] = e^{-nt(p+\epsilon)} \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

La formule du binôme de NEWTON nous donne alors

$$P_n \left(\frac{S_n}{n} \geq p + \epsilon \right) \leq e^{-nt(p+\epsilon)} (1 - p + pe^t)^n = e^{-n(t(p+\epsilon) - \ln(1 - p + pe^t))}.$$

Cela étant vrai pour tout $t > 0$, on obtient finalement

$$P_n \left(\frac{S_n}{n} \geq p + \epsilon \right) \leq e^{-nh}$$

où $h = \sup_{t>0} (t(p + \epsilon) - \ln(1 - p + pe^t))$. Il ne nous reste donc plus qu'à démontrer que

$$h = (p + \epsilon) \ln \frac{p + \epsilon}{p} + (1 - p - \epsilon) \ln \frac{1 - p - \epsilon}{1 - p} > 0.$$

Cela se déduit de l'étude des variations de la fonction $g : t \mapsto t(p + \epsilon) - \ln(1 - p + pe^t)$ sur $[0, +\infty[$. Résumons cette étude en laissant les calculs au lecteur : on a $g(0) = 0$ et $g'(0) = \epsilon > 0$, ce qui prouve que la borne supérieure de g est strictement positive ; la dérivée de g s'annule au point

$t = \ln \frac{(p + \epsilon)(1 - p)}{p(1 - p - \epsilon)}$; en ce point la fonction g est maximale et sa valeur est bien $h_+(\epsilon)$. — QED (Estimation des grands écarts)

Il est bien sûr facile d'illustrer l'estimation des grands écarts par des applications numériques. En voici quelques unes :

1. Si je lance 100 fois une pièce équilibrée, la probabilité pour que le nombre de « piles » soit supérieur ou égal à 60 est inférieure à 0,14. On a ici $(p; n; \epsilon) = (0,5; 100; 0,1)$.

2. Si je lance 1000 fois une pièce équilibrée, la probabilité pour que le nombre de « piles » soit supérieur ou égal à 600 est inférieure à $2 \cdot 10^{-9}$. On a ici $(p; n; \epsilon) = (0,5; 1000; 0,1)$.

3. Si je lance 1000 fois une pièce équilibrée, la probabilité pour que le nombre de « piles » soit supérieur ou égal à 540 est inférieure à 0,05. On a ici $(p; n; \epsilon) = (0,5; 1000; 0,04)$.

Il est utile de connaître le comportement de l'exposant $h_+(\epsilon)$ quand ϵ tend vers zéro. Un simple développement limité du logarithme à l'ordre 2 donne le résultat suivant.

PROPOSITION 6.2. — *Quand ϵ tend vers 0, on a*

$$h_+(\epsilon) = \frac{\epsilon^2}{2p(1-p)} + O(\epsilon^3).$$

La proposition suivante, qui conclut ce chapitre, nous dit que l'estimation donnée par le théorème des grands écarts est optimale, en ce sens que l'exposant de l'exponentielle ne peut pas être amélioré.

PROPOSITION 6.3. — *Pour tout $\epsilon \in]0, 1-p[$, on a*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \left(P_n \left(\frac{S_n}{n} \geq p + \epsilon \right) \right) = -h_+(\epsilon). \quad (6.1)$$

Pour démontrer cette proposition, nous utiliserons la célèbre formule de STIRLING : la suite $(n^{-(n+1/2)} e^n n!)$ _{$n \geq 1$} converge vers une limite c strictement positive. Cette formule est démontrée sous une forme plus précise dans le chapitre suivant (Proposition 7.2), où elle joue un rôle fondamental.

Rappelons ici le sens des notations usuelles de comparaison de suites, telles qu'on les utilise dans toute la suite de l'opuscule. Soient (u_n) et (v_n) deux suites réelles. Nous notons $u_n = O(v_n)$ si u_n est au plus de l'ordre de grandeur de v_n , c'est-à-dire si il existe une constante $k > 0$ telle que, pour tout n , on ait $|u_n| \leq kv_n$. Nous notons $u_n = o(v_n)$ si u_n est au négligeable devant v_n , c'est-à-dire si, pour tout $\epsilon > 0$, il existe $n_0 \geq 1$ tel que, pour

tout $n \geq n_0$, on ait $|u_n| \leq \epsilon v_n$. Enfin, nous notons $u_n \sim v_n$ si les suites sont équivalentes, c'est-à-dire si $u_n - v_n = o(|v_n|)$ (pour les suites qui ne s'annulent pas cela s'exprime aussi par le fait que le rapport u_n/v_n tend vers 1 quand n tend vers l'infini). À titre d'exemple, la formule de STIRLING peut s'écrire : $n! \sim cn^{n+1/2}e^{-n}$.

Démonstration (Proposition 6.3) — L'estimation des grands écarts s'écrit, pour tout $n \geq 1$,

$$\frac{1}{n} \ln (\mathbb{P}_n(S_n \geq n(p + \epsilon))) \leq -h_+(\epsilon). \quad (6.2)$$

Nous devons rechercher une minoration. Notons $k_n := 1 + [n(p + \epsilon)]$ le nombre entier immédiatement supérieur à $n(p + \epsilon)$. On a bien sûr

$$\mathbb{P}_n(S_n \geq n(p + \epsilon)) \geq \mathbb{P}_n(S_n = k_n)$$

et nous allons montrer que

$$\lim \frac{1}{n} \ln (\mathbb{P}_n(S_n = k_n)) = -h_+(\epsilon). \quad (6.3)$$

Cela assurera que

$$\liminf \frac{1}{n} \ln (\mathbb{P}_n(S_n \geq n(p + \epsilon))) \geq -h_+(\epsilon)$$

ce qui, associé à (6.2), prouvera le résultat annoncé.

Les suites (k_n) et $(n - k_n)$ tendent vers l'infini avec n et on a :

$$\mathbb{P}_n(S_n = k_n) = \frac{n!}{k_n!(n - k_n)!} p^{k_n} (1 - p)^{n - k_n}.$$

En appliquant la formule de STIRLING aux trois suites de factorielles, on obtient

$$\mathbb{P}_n(S_n = k_n) \sim \frac{1}{c} \sqrt{\frac{n}{k_n(n - k_n)}} \left(\frac{np}{k_n}\right)^{k_n} \left(\frac{n(1 - p)}{n - k_n}\right)^{n - k_n}. \quad (6.4)$$

Cette équivalence entre deux suites qui tendent vers zéro s'étend à leurs logarithmes. Étudions le comportement du logarithme de la dernière expression. On a $k_n \sim n(p + \epsilon)$, $n - k_n \sim n(1 - p - \epsilon)$ et donc, à une constante multiplicative près,

$$\sqrt{\frac{n}{k_n(n - k_n)}} \sim \frac{1}{\sqrt{n}}.$$

D'où

$$\lim \frac{1}{n} \ln \left(\frac{1}{c} \sqrt{\frac{n}{k_n(n - k_n)}} \right) = 0 .$$

On a

$$k_n \ln \left(\frac{np}{k_n} \right) = n(p + \epsilon) \ln \left(\frac{p}{p + \epsilon} \right) \\ + n(p + \epsilon) \ln \left(\frac{n(p + \epsilon)}{k_n} \right) + (k_n - n(p + \epsilon)) \ln \left(\frac{np}{k_n} \right)$$

et, en utilisant simplement le fait que $k_n - n(p + \epsilon)$ reste borné, on en déduit

$$\lim \frac{1}{n} \ln \left(\left(\frac{np}{k_n} \right)^{k_n} \right) = (p + \epsilon) \ln \left(\frac{p}{p + \epsilon} \right) .$$

De façon similaire, on montre que

$$\lim \frac{1}{n} \ln \left(\left(\frac{n(1 - p)}{n - k_n} \right)^{n - k_n} \right) = (1 - p - \epsilon) \ln \left(\frac{1 - p}{1 - p - \epsilon} \right) .$$

Des trois derniers calculs de limites et de (6.4), on déduit (6.3), ce qui achève la démonstration. — QED (Proposition 6.3)

7. Le théorème limite central

7.1. Énoncé. — Comme l'indique son nom étrange, le *théorème limite central* occupe la place centrale (Leo BREIMAN parle du «trône») du Calcul des Probabilités. Ce théorème, connu aussi sous le nom de *théorème de convergence vers la loi de GAUSS*, a été obtenu grâce aux travaux successifs d'Abraham DE MOIVRE^{7.1}, Pierre Simon LAPLACE^{7.2} et Carl Friedrich GAUSS^{7.3}. Son champ d'application est extrêmement vaste et c'est ce qui lui donne son caractère fascinant ; il établit le rôle universel de la loi de GAUSS, la fameuse «courbe en cloche». Conformément à notre objectif, nous nous limiterons ici à l'étude des suites de variables aléatoires indépendantes, identiquement distribuées et ne prenant que deux valeurs.

^{7.1} A. De Moivre, *The Doctrine of Chances*, Londres, 1718.

^{7.2} P.S. Laplace, *Théorie analytique des probabilités*, Paris, 1812.

^{7.3} C.F. Gauss, *Theoria combinationis observationum erroribus minimus obnoxiae*, 1821.

Revenons donc aux notations précédemment introduites. L'espace $\Omega_n = \{0, 1\}^n$ est muni de la probabilité produit $P_n = (p, 1 - p)^{\otimes n}$ et on note $S_n(\omega) = \omega_1 + \omega_2 + \dots + \omega_n$ si $\omega = (\omega_1, \omega_2, \dots, \omega_n)$. On supposera dorénavant que $0 < p < 1$, ce qui élimine deux cas où notre analyse est triviale.

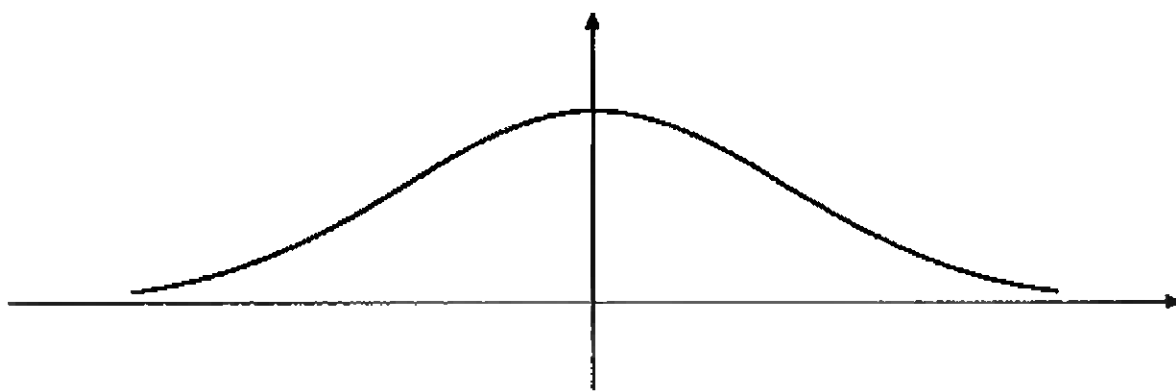
La loi faible des grands nombres donne un premier contrôle des fluctuations de S_n autour de sa valeur moyenne np : si n est assez grand, des fluctuations de l'ordre de grandeur de n sont très improbables. D'un autre côté, on sait que l'espérance de $(S_n - np)^2$ est égale à $np(1 - p)$. Cela suggère que les fluctuations probables de S_n autour de sa valeur moyenne sont de l'ordre de \sqrt{n} . Le théorème limite central va confirmer cette intuition et nous donner un contrôle quantitatif (mais asymptotique) de ces fluctuations. En voici l'énoncé précis.

THÉORÈME 7.1 (Théorème limite central). — Soient a et b deux éléments de $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ tels que $a < b$. On a théorème limite central

$$P_n \left(a \leq \frac{S_n - np}{\sqrt{p(1-p)}\sqrt{n}} \leq b \right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-x^2/2) dx \quad (7.1)$$

La suite de ce chapitre est consacrée d'abord à quelques commentaires autour de ce résultat, ensuite à la présentation de certaines de ses applications et enfin à sa démonstration.

7.2. Commentaires. — Le graphe de la fonction $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ est la courbe de GAUSS . courbe de Gauss



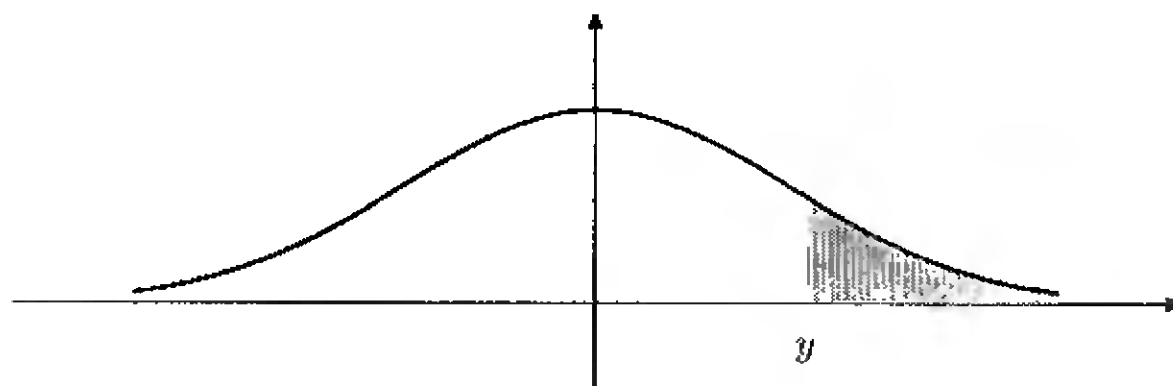
Au cours de la démonstration du théorème, on utilisera le fait que que

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = 1 .$$

(Cela signifie que la fonction $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ est une densité de probabilité sur \mathbb{R} , mais ce point de vue ne sera pas développé ici.)

La fonction $x \mapsto e^{-x^2/2}$ ne possède pas de primitive qui puisse s'exprimer avec les fonctions classiques, et donc les intégrales du type $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$ ne possèdent pas d'expression plus simple. Toutefois leurs valeurs peuvent être calculées numériquement avec une grande précision. Ces valeurs sont tabulées dans beaucoup d'ouvrages car elles jouent un grand rôle dans les nombreuses applications du théorème limite central en statistiques et probabilités.

La fonction $\Phi : y \mapsto \frac{1}{\sqrt{2\pi}} \int_y^{+\infty} e^{-x^2/2} dx$ qui mesure l'aire de « la queue de la courbe de GAUSS » décroît très vite quand y croît dans $]0, +\infty[$.



On a par exemple

$$\Phi(1) \approx 0,1587, \quad \Phi(2) \approx 0,0228 \quad \text{et} \quad \Phi(4) \approx 3,2 \cdot 10^{-5}.$$

D'autre part, pour tout $y > 0$,

$$\frac{1}{\sqrt{2\pi}(y+1)} \left(e^{-y^2/2} - e^{-(y+1)^2/2} \right) \leq \Phi(y) \leq \frac{1}{\sqrt{2\pi}y} e^{-y^2/2}$$

et, quand y tend vers $+\infty$,

$$\Phi(y) \sim \frac{1}{\sqrt{2\pi}y} e^{-y^2/2}.$$

La vérification de ces affirmations est laissée au lecteur en guise de petit exercice d'analyse.

Le théorème limite central peut être renforcé par un résultat de convergence uniforme : la convergence (7.1) est uniforme en a et b . Cette affirmation se justifie en utilisant un résultat classique d'analyse élémentaire : toute suite de fonctions monotones sur \mathbb{R} qui converge simplement vers une fonction continue et bornée est uniformément convergente.

En utilisant l'expression de la loi binomiale, le théorème limite central peut aussi s'écrire sous la forme suivante : si $y \in \mathbb{R}$ et si $k(n)$ désigne la partie entière de $np + y\sqrt{np(1-p)}$, alors

$$\sum_{j=0}^{k(n)} \binom{n}{j} p^j (1-p)^{n-j} \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-x^2/2} dx .$$

En effet, on a :

$$\sum_{j=0}^{k(n)} \binom{n}{j} p^j (1-p)^{n-j} = \mathbf{P}_n \left(S_n \leq np + y\sqrt{np(1-p)} \right) .$$

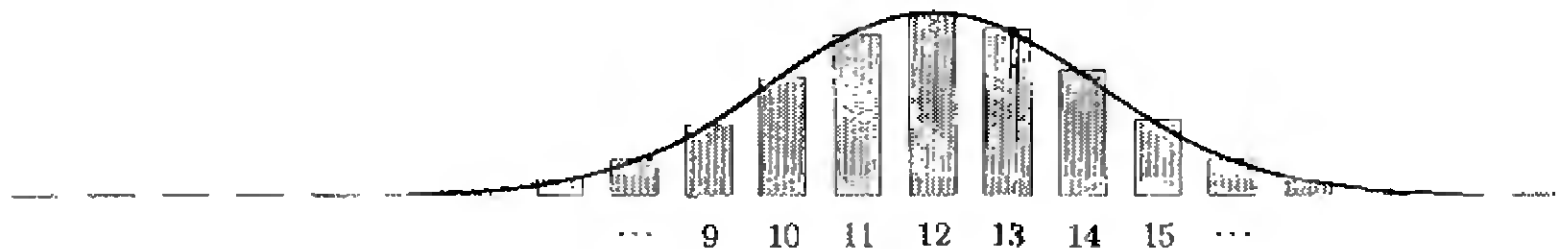
On pourrait aussi écrire

$$\sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j} \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{k-np}{\sqrt{np(1-p)}}} e^{-x^2/2} dx ,$$

c'est-à-dire, en effectuant le changement de variable $t = np + x\sqrt{np(1-p)}$,

$$\sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j} \approx \frac{1}{\sqrt{2\pi np(1-p)}} \int_{-\infty}^k \exp \left(-\frac{(t-np)^2}{2np(1-p)} \right) dt .$$

Cela peut être illustré par l'ajustement d'une courbe en cloche et de l'histogramme représentant la distribution binomiale : il suffit de tracer sur une même figure le graphe de la fonction prenant la valeur $\binom{n}{j} p^j (1-p)^{n-j}$ en chaque point j entier entre 1 et n et le graphe de la fonction d'une variable réelle $t \mapsto \frac{1}{\sqrt{2\pi np(1-p)}} \exp \left(-\frac{(t-np)^2}{2np(1-p)} \right)$. Sur la figure suivante, on a représenté ces deux fonctions pour les valeurs des paramètres $n = 20$ et $p = 0,6$.



On observe sur ce graphique que

$$\binom{n}{j} p^j (1-p)^{n-j} \approx \frac{1}{\sqrt{2\pi np(1-p)}} \exp \left(-\frac{(j-np)^2}{2np(1-p)} \right) .$$

Cela sera confirmé par le théorème de MOIVRE-LAPLACE, qui est énoncé et démontré dans la suite de ce chapitre, comme étape dans la démonstration du théorème limite central.

7.3. Exemples d'applications. — Nous donnons à présent quelques exemples d'applications du théorème limite central : d'abord une application « théorique » à la loi des grands nombres, ensuite des exemples plus pratiques de calcul de probabilités, de calcul d'intervalle de confiance et de test statistique.

La loi faible des grands nombres est une conséquence immédiate du théorème limite central. Soient en effet $\epsilon > 0$ et $\delta > 0$ fixés. Il existe $a > 0$ tel que $\Phi(a) < \delta$ et, pour tout entier n assez grand, on a $\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} \geq a$.

Pour un tel entier n , on a

$$\begin{aligned} \mathbf{P}_n \left(\left| \frac{S_n}{n} - p \right| \geq \epsilon \right) &= \mathbf{P}_n \left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq -\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} \right) \\ &\quad + \mathbf{P}_n \left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} \right), \end{aligned}$$

donc

$$\begin{aligned} \mathbf{P}_n \left(\left| \frac{S_n}{n} - p \right| \geq \epsilon \right) \\ \leq \mathbf{P}_n \left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq -a \right) + \mathbf{P}_n \left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq a \right). \end{aligned}$$

Le théorème limite central nous dit que, pour tout entier n assez grand,

$$\left| \mathbf{P}_n \left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq -a \right) - \Phi(a) \right| < \delta$$

et

$$\left| \mathbf{P}_n \left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq a \right) - \Phi(a) \right| < \delta.$$

Grâce au choix de a et aux trois inégalités précédentes, on peut conclure que, pour tout n assez grand,

$$\mathbf{P}_n \left(\left| \frac{S_n}{n} - p \right| \geq \epsilon \right) \leq 4\delta.$$

Nous avons ainsi retrouvé le fait que

$$\mathbf{P}_n \left(\left| \frac{S_n}{n} - p \right| \geq \epsilon \right) \longrightarrow 0.$$

L'examen de cet argument nous montre que nous avons en fait beaucoup mieux : pour toute suite réelle (u_n) telle que $\lim_{n \rightarrow \infty} \frac{u_n}{\sqrt{n}} = 0$, on a

$$P_n \left(u_n \left| \frac{S_n}{n} - p \right| \geq \epsilon \right) \rightarrow 0.$$

D'autres conséquences du théorème limite central sont présentées dans la suite de cet ouvrage. En particulier, dans le chapitre 10, nous donnons un sens à l'affirmation *la suite (S_n) est presque sûrement non bornée* et nous la démontrons.

Passons à présent à quelques applications plus pratiques du théorème limite central. Ces applications sont basées sur le fait que l'on estime une probabilité du type

$$P_n (np + \sqrt{na} \leq S_n \leq np + \sqrt{nb})$$

par sa valeur approchée

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{p(1-p)}}}^{\frac{b}{\sqrt{p(1-p)}}} e^{-x^2/2} dx.$$

Mais tel qu'il est énoncé, le théorème ne nous dit rien sur la qualité de cette approximation. Il existe des résultats théoriques donnant une majoration explicite de l'erreur d'approximation, en particulier l'inégalité de BERRY-ÈSSEEN. Nous n'aborderons pas ce sujet, qui est traité dans les ouvrages plus avancés de Calcul des Probabilités. Il semble admis dans la littérature que l'approximation donnée par le théorème limite central est acceptable dès que $np(1-p) > 18$. Ce point mériterait une étude plus détaillée, qui est certainement possible avec les moyens de calcul modernes (on pourra aussi consulter la discussion apparaissant dans le livre de William FELLER^{7.4}).

Au cours du chapitre précédent, nous avons déterminé une majoration de la probabilité pour que, après 100 lancers d'une pièce équilibrée, le nombre de « piles » soit supérieur ou égal à 60. Le théorème limite central nous donne une estimation de cette probabilité : nous avons ici $p = 0,5$ et

$$P_{100}(S_{100} \geq 60) = P_{100} \left(\frac{S_{100} - 50}{5} \geq 2 \right) \approx \Phi(2) \approx 2/100.$$

⁷ W. Feller, *Introduction to Probability Theory and its Applications*, volume 1, chap VII.3.

On peut estimer de la même façon la probabilité pour que après 1000 lancers, le nombre de « piles » soit supérieur ou égal à 540 :

$$P_{1000}(S_{1000} \geq 540) = P_{1000} \left(\frac{S_{1000} - 500}{5\sqrt{10}} \geq \frac{8}{\sqrt{10}} \right) \\ \approx \Phi \left(\frac{8}{\sqrt{10}} \right) \approx 6/1000.$$

Voici un autre exemple dans lequel nous approchons de deux concepts des statistiques : intervalle de confiance et test d'hypothèse (nous ne donnons pas dans ce livre de définition précise de ces notions).

Entre 1871 et 1900, on a relevé en Suisse 1 359 670 naissances de garçons et 1 285 086 naissances de filles. Ces données sont elles compatibles avec l'hypothèse que les sexes des nouveaux nés sont des caractères aléatoires indépendants et que les deux sexes apparaissent avec égale probabilité ?

Posons $n = 1\,359\,670 + 1\,285\,086 = 2\,644\,756$ et considérons que n expériences ont été réalisées. Sans craindre l'accusation de sexisme, baptisons succès la naissance d'un garçon et échec la naissance d'une fille. Supposons un moment que les expériences sont indépendantes et que la probabilité de succès est $1/2$, et déterminons un nombre C tel que la probabilité pour que nombre de succès dépasse C soit inférieure à 10^{-5} .

Les tables numériques nous indiquent que $\Phi(4,5) < 10^{-5}$ et le théorème limite central nous dit que

$$P_n(S_n \geq C) \approx \Phi \left(\frac{2C - n}{\sqrt{n}} \right).$$

On peut donc choisir $C \geq \frac{1}{2}(n + 4,5\sqrt{n})$, ce qui donne ici $C = 1\,326\,037$.

On constate alors que le nombre de succès effectivement observés est largement supérieure à ce seuil. Si notre hypothèse sur la loi des naissances était exacte, on observerait un phénomène qui se produit avec une probabilité extrêmement faible. Il est donc raisonnable de rejeter l'hypothèse. On conclut que les données démographiques suisses contredisent le fait que les sexes des nouveaux nés soient des caractères aléatoires indépendants et également distribués.

REMARQUE. — On peut être gêné par l'utilisation dans ce calcul du théorème limite central qui n'est qu'un résultat asymptotique ; c'est pourtant ainsi qu'il est utilisé dans la pratique. Notons toutefois que, sur le dernier

exemple traité, l'estimation des grands écarts donnait un résultat voisin : cette estimation s'écrit en effet :

$$P_n(S_n \geq C) \leq \exp\left(-nh_+\left(\frac{C}{n} - \frac{1}{2}\right)\right),$$

ce qui nous donne, avec $n = 2\,644\,756$ et $C = 1\,326\,037$,

$$P_n(S_n \geq C) \leq 5 \cdot 10^{-5}.$$

—◇

7.4. Démonstration du théorème limite central. — La démonstration se fait en plusieurs étapes. Son point de départ, qui en constitue l'ingrédient essentiel, est la formule de STIRLING. Cette formule donne, sous une forme précise, un équivalent de la suite $(n!)$, et donc une estimation des coefficients binomiaux $\binom{n}{k}$ quand n , k et $n - k$ sont grands. Nous recherchons une estimation de la probabilité que S_n se trouve dans un certain intervalle $[np + a\sqrt{p(1-p)}\sqrt{n}, np + b\sqrt{p(1-p)}\sqrt{n}]$. Pour ce faire, nous allons donner une estimation de $P_n(S_n = k)$ quand n est grand et que l'entier k est dans l'intervalle considéré. C'est l'objet du théorème de MOIVRE-LAPLACE (dans un cadre plus général que le modèle de « pile ou face », ce résultat porte le nom de *théorème limite local*). Le théorème limite central s'en déduit par l'apparition de sommes de RIEMANN.

Cette démonstration n'utilise que des outils élémentaires de l'analyse réelle. Une autre méthode de démonstration utilise la transformée de FOURIER des mesures, que les probabilistes appellent la *fonction caractéristique*. Ce point de vue, qui est bien adapté à une situation beaucoup plus générale que le simple jeu de pile ou face, ne sera pas développé ici.

PROPOSITION 7.2 (Formule de STIRLING). — *Posons, pour tout* formule de Stirling
entier $n > 0$,

$$n! = \sqrt{2\pi n} n^{n+1/2} e^{-n} (1 + \epsilon_n).$$

Il existe une constante réelle A *telle que* $|\epsilon_n| < \frac{A}{n}$.

Démonstration. — Dans une première étape montrons qu'il existe $c_1 \in \mathbb{R}$ tel que

$$\ln(n!) = c_1 + \left(n + \frac{1}{2}\right) \ln n - n + O\left(\frac{1}{n}\right).$$

Cette estimation est basée sur la comparaison de la série de terme général $\ln n$ et de l'intégrale de la fonction logarithme. On écrit :

$$\ln(n!) = \int_{1/2}^{n+1/2} \ln t \, dt + \sum_{k=1}^n \left(\ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right). \quad (7.2)$$

On a d'une part :

$$\int_{1/2}^{n+1/2} \ln t \, dt = [t \ln t - t]_{1/2}^{n+1/2} = \left(n + \frac{1}{2}\right) \ln \left(n + \frac{1}{2}\right) - n + c_2,$$

pour une certaine constante c_2 ; en utilisant le développement limité du logarithme au voisinage de 1, on écrit

$$\ln \left(n + \frac{1}{2}\right) = \ln(n) + \ln \left(1 + \frac{1}{2n}\right) = \ln(n) + \frac{1}{2n} + O\left(\frac{1}{n^2}\right),$$

ce qui nous donne

$$\int_{1/2}^{n+1/2} \ln t \, dt = \left(n + \frac{1}{2}\right) \ln(n) - n + c_3 + O\left(\frac{1}{n}\right), \quad (7.3)$$

pour une certaine constante c_3 .

D'autre part :

$$\begin{aligned} \ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt &= \ln k - [t \ln t - t]_{k-1/2}^{k+1/2} \\ &= \ln k - \left(k + \frac{1}{2}\right) \ln \left(k + \frac{1}{2}\right) + \left(k - \frac{1}{2}\right) \ln \left(k - \frac{1}{2}\right) + 1 \\ &= -\frac{1}{2} \ln \left(1 - \frac{1}{4k^2}\right) - k \left(\ln \left(1 + \frac{1}{2k}\right) - \ln \left(1 - \frac{1}{2k}\right) \right) + 1; \end{aligned}$$

grâce au développement limité du logarithme cela s'écrit

$$\begin{aligned} \ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt &= O\left(\frac{1}{k^2}\right) \\ &\quad - k \left[\left(\frac{1}{2k} - \frac{1}{8k^2} + O\left(\frac{1}{k^3}\right) \right) - \left(-\frac{1}{2k} - \frac{1}{8k^2} + O\left(\frac{1}{k^3}\right) \right) \right] + 1, \end{aligned}$$

ce qui prouve qu'il existe $c_4 \in \mathbb{R}$ tel que

$$\left| \ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right| \leq \frac{c_4}{k^2};$$

ainsi $\left(\ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right)$ est le terme général d'une série absolument convergente ; notons

$$c_5 := \sum_{k=1}^{+\infty} \left(\ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right) .$$

Remarquons que

$$\sum_{k=n+1}^{+\infty} \left| \ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right| \leq c_4 \sum_{k=n+1}^{+\infty} \frac{1}{k^2} < c_4 \sum_{k=n+1}^{+\infty} \frac{1}{k(k-1)} = \frac{c_4}{n} ;$$

ce qui prouve que

$$\sum_{k=1}^n \left(\ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right) = c_5 + O\left(\frac{1}{n}\right) . \quad (7.4)$$

En réunissant (7.2), (7.3) et (7.4), on obtient

$$\ln(n!) = \left(n + \frac{1}{2}\right) \ln(n) - n + c_3 + c_5 + O\left(\frac{1}{n}\right) ,$$

ce qui conclut notre première étape. En notant $d := e^{c_3+c_5}$, nous avons obtenu le résultat suivant :

$$n! = dn^{n+1/2}e^{-n}(1 + \epsilon_n)$$

où $\epsilon_n = O\left(\frac{1}{n}\right)$.

Pour achever la démonstration de la proposition nous devons à présent montrer que $d = \sqrt{2\pi}$. Nous allons le faire en utilisant les intégrales de WALLIS définies pour $n \in \mathbb{N}$ par

intégrales
de Wallis.

$$I_n := \int_0^{\pi/2} \sin^n t \, dt .$$

En intégrant par parties, on a, pour $n \geq 2$,

$$I_n = \int_0^{\pi/2} \sin^{n-1} t \cdot \sin t \, dt = \int_0^{\pi/2} (n-1) \sin^{n-2} t \cos t \cdot \cos t \, dt ,$$

d'où

$$I_n = (n-1) \int_0^{\pi/2} \sin^{n-2} t (1 - \sin^2 t) \, dt$$

et la formule de récurrence

$$nI_n = (n-1)I_{n-2}.$$

On a de plus $I_0 = \frac{\pi}{2}$ et $I_1 = 1$, ce qui nous permet d'écrire, pour $n \geq 1$,

$$I_{2n} = \frac{(2n-1)(2n-3)\dots 1}{(2n)(2n-2)\dots 2} I_0 = \frac{(2n)!}{2^{2n}(n!)^2} \frac{\pi}{2}$$

et

$$I_{2n+1} = \frac{(2n)(2n-2)\dots 2}{(2n+1)(2n-1)\dots 3} I_1 = \frac{2^{2n}(n!)^2}{(2n+1)!}.$$

Remarquons que la formule de récurrence entraîne

$$\lim_{n \rightarrow \infty} \frac{I_n}{I_{n-2}} = 1.$$

Comme de plus $I_{n-2} \geq I_{n-1} \geq I_n$ on a aussi

$$\lim_{n \rightarrow \infty} \frac{I_n}{I_{n-1}} = 1,$$

ce qui peut s'écrire

$$\lim_{n \rightarrow \infty} \frac{((2n)!)^2(2n+1)\frac{\pi}{2}}{2^{4n}(n!)^4} = 1.$$

D'après la première étape de notre démonstration, nous savons que, quand $n \rightarrow \infty$, on a $n! \sim dn^{n+1/2}e^{-n}$. En remplaçant, dans la limite précédente, les factorielles par leurs équivalents, on conclut directement que $d^2 = 2\pi$, ce qui achève la démonstration. — QED (Formule de STIRLING)

théorème de
Moivre-
Laplace

PROPOSITION 7.3 (Théorème de MOIVRE-LAPLACE). — Si l'on note, pour $0 \leq k \leq n$,

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) (1 + \delta_n(k)),$$

alors, pour tout $a > 0$,

$$\lim_{n \rightarrow \infty} \max_{k: |k-np| < a\sqrt{n}} |\delta_n(k)| = 0.$$

Démonstration. — Le nombre réel a est fixé. Notons I_n l'ensemble des nombres entiers k compris entre $np - a\sqrt{n}$ et $np + a\sqrt{n}$. Soient $(s_{n,k})_{n>0, k \in I_n}$ et $(t_n)_{n>0}$ deux familles de nombres réels; nous écrivons $s_{n,k} = O_u(t_n)$ si il existe une constante réelle C telle que, pour tout $n > 0$ et tout $k \in I_n$, $|s_{n,k}| < Ct_n$ (l'indice u affecté au O est là pour rappeler que l'estimation est uniforme en k).

D'après la formule de STIRLING, on a

$$\left. \begin{aligned} \binom{n}{k} p^k (1-p)^{n-k} &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{n}{k} p\right)^k \left(\frac{n}{n-k} (1-p)\right)^{n-k} \frac{1 + \epsilon_n}{(1 + \epsilon_k)(1 + \epsilon_{n-k})} \end{aligned} \right\} (7.5)$$

Si $k \in I_n$, on a

$$\left. \begin{aligned} \frac{n}{(np + a\sqrt{n})(n(1-p) + a\sqrt{n})} &\leq \frac{n}{k(n-k)} \\ &\leq \frac{n}{(np - a\sqrt{n})(n(1-p) - a\sqrt{n})} \end{aligned} \right\} (7.6)$$

On déduit de cet encadrement que

$$\frac{n}{k(n-k)} = \frac{1}{np(1-p)} \left(1 + O_u\left(n^{-1/2}\right)\right) \quad (7.7)$$

puis que

$$\sqrt{\frac{n}{k(n-k)}} = \frac{1}{\sqrt{np(1-p)}} \left(1 + O_u\left(n^{-1/2}\right)\right). \quad (7.8)$$

D'autre part, en utilisant le développement limité $\ln(1+t) = t - \frac{t^2}{2} + O(t^3)$,

le fait que $\frac{k - np}{k} = O_u(n^{-1/2})$ et $\frac{k - np}{n - k} = O_u(n^{-1/2})$, et de nouveau (7.7), on peut écrire

$$\begin{aligned} &\ln \left(\left(\frac{n}{k} p\right)^k \left(\frac{n}{n-k} (1-p)\right)^{n-k} \right) \\ &= k \ln \left(1 - \frac{k - np}{k} \right) + (n - k) \ln \left(1 + \frac{k - np}{n - k} \right) \\ &= -\frac{1}{2}(k - np)^2 \left(\frac{1}{k} + \frac{1}{n - k} \right) + k O_u\left(n^{-3/2}\right) + (n - k) O_u\left(n^{-3/2}\right) \\ &= -\frac{1}{2}(k - np)^2 \frac{1}{np(1-p)} + O_u\left(n^{-1/2}\right). \end{aligned}$$

On a donc

$$\left. \begin{aligned} & \left(\frac{n}{k}p\right)^k \left(\frac{n}{n-k}(1-p)\right)^{n-k} \\ & = \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) \left(1 + O_u\left(n^{-1/2}\right)\right) . \end{aligned} \right\} \quad (7.9)$$

Enfin, puisque

$$\begin{aligned} \epsilon_n &< \frac{A}{n}, \quad \epsilon_k < \frac{A}{k}, \quad \epsilon_{n-k} < \frac{A}{n-k} \\ \frac{1}{k} &= O_u\left(\frac{1}{n}\right), \quad \frac{1}{n-k} = O_u\left(\frac{1}{n}\right), \end{aligned}$$

on a

$$\frac{1 + \epsilon_n}{(1 + \epsilon_k)(1 + \epsilon_{n-k})} = \left(1 + O_u\left(\frac{1}{n}\right)\right) . \quad (7.10)$$

Rassemblant les estimations (7.8), (7.9) et (7.10) dans l'identité (7.5), on obtient

$$\begin{aligned} & \binom{n}{k} p^k (1-p)^{n-k} \\ & = \frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) \left(1 + O_u\left(n^{-1/2}\right)\right) . \end{aligned}$$

— QED (Théorème de MOIVRE-LAPLACE)

Pour déduire le théorème limite central du théorème de MOIVRE-LAPLACE, nous utilisons les lemmes suivants.

LEMME 7.4. — Soit $[a, b]$ un segment de la droite réelle \mathbb{R} et f une fonction définie sur \mathbb{R} , nulle en dehors de $[a, b]$ et continue sur $[a, b]$. On a

$$\lim_{h \rightarrow 0, h > 0} h \sum_{k=-\infty}^{+\infty} f(t + kh) = \int_a^b f(x) dx ,$$

uniformément en $t \in \mathbb{R}$

Démonstration. — L'expression considérée est, aux effets de bord près, une somme de RIEMANN. En utilisant l'uniforme continuité de la fonction f sur le segment, on peut aisément vérifier le résultat : soit $\epsilon > 0$ fixé ; si h est choisi assez proche de zéro pour que

$$x, y \in [a, b] \text{ et } |x - y| < h \Rightarrow |f(x) - f(y)| < \epsilon ,$$

et si

$$\{k \in \mathbb{Z} \mid a \leq t + kh \leq b\} = \{i, i+1, i+2, \dots, j\}$$

alors, en notant $M := \sup_{a \leq x \leq b} |f(x)|$, on a

$$\begin{aligned} & \left| h \sum_{k: t+kh \in [a, b]} f(t+kh) - \int_a^b f(x) dx \right| \\ & \leq hM + \sum_{k=i}^{j-1} \left| hf(t+kh) - \int_{t+kh}^{t+(k+1)h} f(x) dx \right| + 2hM \\ & \leq 3hM + (j-i)h\epsilon \leq 3hM + (b-a)\epsilon. \end{aligned}$$

— QED (Lemme 7.4)

Le lemme suivant est un résultat classique laissé en exercice aux lecteurs qui ne l'auraient pas encore rencontré.

LEMME 7.5. — On a

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = 1.$$

Démonstration (Théorème limite central) — Nous commençons par traiter le cas où a et b sont deux nombres réels. Notons J_n l'intervalle $J_n := [a\sqrt{np(1-p)}, b\sqrt{np(1-p)}]$. On a

$$\begin{aligned} \mathbb{P}_n(S_n - np \in J_n) &= \sum_{k=0}^n \mathbf{1}_{J_n}(k - np) \cdot \mathbb{P}_n(S_n = k) \\ &= \frac{1}{\sqrt{2\pi p(1-p)}n} \sum_{k=0}^n \mathbf{1}_{J_n}(k - np) \cdot \exp\left(-\frac{(k - np)^2}{2np(1-p)}\right) \cdot (1 + \delta_n(k)) \end{aligned}$$

où $\lim_{n \rightarrow \infty} \max_k |\delta_n(k)| = 0$, d'après la proposition 7.3, puisque $k - np$ reste dans J_n . On en déduit que

$$\left. \begin{aligned} & \mathbb{P}_n(S_n - np \in J_n) \\ &= \frac{1 + \delta_n}{\sqrt{2\pi p(1-p)}n} \sum_{k=0}^n \mathbf{1}_{J_n}(k - np) \cdot \exp\left(-\frac{(k - np)^2}{2np(1-p)}\right) \end{aligned} \right\} (7.11)$$

où $\lim_{n \rightarrow \infty} \delta_n = 0$. Quand n est assez grand, l'expression (7.11) est égale à l'expression obtenue en remplaçant la somme $\sum_{k=0}^n$ par la somme \sum_k .

Ainsi, cette expression est équivalente à

$$\left. \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{p(1-p)n}} \times \sum_{k \in \mathbb{Z}} \left[\mathbf{1}_{[a,b]} \left(\frac{k}{\sqrt{np(1-p)}} - \sqrt{\frac{np}{1-p}} \right) \times \exp \left(-\frac{1}{2} \left(\frac{k}{\sqrt{np(1-p)}} - \sqrt{\frac{np}{1-p}} \right)^2 \right) \right] \right\} (7.12)$$

On reconnaît ici une expression du type de celle étudiée dans le lemme 7.4, avec $h = (np(1-p))^{-1/2}$ et $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Quand n tend vers l'infini,

l'expression (7.12) tend vers $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$.

Le théorème est démontré quand a et b sont réels. Pour traiter les cas où $a = -\infty$ ou $b = +\infty$ il nous suffit, grâce au lemme 7.5 de démontrer par exemple que, pour tout b réel,

$$\mathbf{P}_n \left(S_n - np \leq b \sqrt{np(1-p)} \right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-x^2/2} dx. \quad (7.13)$$

Soient $b \in \mathbb{R}$ et $\epsilon > 0$. Fixons $c > \max(0, b)$ tel que $\frac{1}{\sqrt{2\pi}} \int_c^{+\infty} e^{-x^2/2} dx < \epsilon$.

On a alors $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-c} e^{-x^2/2} dx < \epsilon$ et, grâce au lemme 7.5,

$$\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-x^2/2} dx > 1 - 2\epsilon.$$

On peut écrire

$$\left| \mathbf{P}_n \left(S_n - np \leq b \sqrt{np(1-p)} \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-x^2/2} dx \right| \leq A_n + B_n + C$$

où

$$A_n := \mathbf{P}_n \left(S_n - np < -c \sqrt{np(1-p)} \right),$$

$$B_n := \left| \mathbf{P}_n \left(-c \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right) - \frac{1}{\sqrt{2\pi}} \int_{-c}^b e^{-x^2/2} dx \right|,$$

$$C := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-c} e^{-x^2/2} dx.$$

On a

$$0 \leq A_n \leq 1 - P_n \left(-c\sqrt{np(1-p)} \leq S_n - np \leq c\sqrt{np(1-p)} \right)$$

et, d'après la première partie de la démonstration,

$$\begin{aligned} \lim_{n \rightarrow \infty} P_n \left(-c\sqrt{np(1-p)} \leq S_n - np \leq c\sqrt{np(1-p)} \right) \\ = \frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-x^2/2} dx > 1 - 2\epsilon. \end{aligned}$$

Cela montre que, pour tout n assez grand, on a $A_n < 2\epsilon$.

Toujours d'après la première partie de la démonstration, $\lim_{n \rightarrow \infty} B_n = 0$, et, par choix de c , on a $C < \epsilon$. Nous avons démontré (7.13).

—— QED (Théorème limite central)

8. Estimation des écarts modérés

Nous conservons ici les notations des chapitres précédents. Ainsi S_n désigne le nombre de succès observés dans une suite de n réalisations indépendantes d'une expérience dont la probabilité de succès est p . La fréquence empirique des succès est S_n/n . L'estimation des grands écarts nous donne l'ordre de grandeur de la probabilité que la fréquence empirique s'écarte de p quand n tend vers l'infini. D'un autre côté, le théorème limite central nous dit en particulier que si (a_n) est une suite réelle qui tend vers l'infini, alors la probabilité que la fréquence empirique s'écarte de p d'un ordre de grandeur comparable à a_n/\sqrt{n} tend vers zéro quand n tend vers l'infini. L'estimation des écarts modérés va nous donner, sous certaines hypothèses sur la suite (a_n) , des précisions sur cette convergence. Pour établir cette estimation, nous utiliserons une généralisation du théorème limite central, en considérant des « bornes mobiles ». L'estimation des écarts modérés nous servira pour établir, dans le chapitre 11, la célèbre loi du logarithme itéré.

Quelles estimations peut-on espérer ? Le théorème limite central nous dit que, quand n tend vers l'infini,

$$P_n \left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)} \frac{a}{\sqrt{n}} \right) \sim \Phi(a).$$

L'estimation des écarts modérés affirme que cet équivalent est conservé quand le nombre a est autorisé à tendre vers l'infini avec n , mais pas trop

vite. Rappelons que, si $\lim_{n \rightarrow \infty} a_n = +\infty$, alors

$$\Phi(a_n) = \frac{1}{\sqrt{2\pi}} \int_{a_n}^{+\infty} e^{-x^2/2} dx \sim \frac{1}{a_n \sqrt{2\pi}} \exp\left(-\frac{a_n^2}{2}\right).$$

Le théorème suivant apparaît dans un article de Harald CRAMER^{8.1}.

THÉORÈME 8.1 (Estimation des écarts modérés). —

Si $\lim_{n \rightarrow \infty} a_n = +\infty$ et $\lim_{n \rightarrow \infty} \frac{a_n}{n^{1/6}} = 0$, alors

$$\mathbb{P}_n \left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)} \frac{a_n}{\sqrt{n}} \right) \sim \frac{1}{a_n \sqrt{2\pi}} \exp\left(-\frac{a_n^2}{2}\right).$$

Remarquons que la conclusion du théorème est fautive si la suite (a_n) tend trop vite vers l'infini. Par exemple (considérer la suite $a_n = \sqrt{n}$) les suites $(\Phi(\sqrt{n}))$ et $(\mathbb{P}_n(\frac{S_n}{n} - p \geq \sqrt{p(1-p)}))$ ne sont pas équivalentes. En effet, on a

$$\lim \frac{1}{n} \ln \Phi(\sqrt{n}) = -\frac{1}{2}$$

mais, d'après la proposition 6.3, pour $p < \frac{1}{2}$, on a $\sqrt{p(1-p)} < 1-p$ et

$$\lim \frac{1}{n} \ln \mathbb{P}_n \left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)} \right) = -h_+ \left(\sqrt{p(1-p)} \right).$$

Le théorème 8.1 va être établi en utilisant un raffinement des techniques utilisées dans la démonstration du théorème limite central. La première étape consiste à « optimiser » la preuve du théorème de MOIVRE-LAPLACE (Proposition 7.3). Ce théorème donne un équivalent, quand n tend vers l'infini, de $\mathbb{P}_n(S_n = k_n)$, avec $k_n = np + O(\sqrt{n})$. Une lecture attentive de sa démonstration convainc que cet équivalent s'étend au cas où $k_n = np + o(n^{2/3})$. C'est l'objet de l'énoncé suivant.

PROPOSITION 8.2. — Si l'on note, pour $0 \leq k \leq n$,

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) \cdot (1 + \delta_n(k)),$$

alors, pour toute suite réelle positive (c_n) tendant vers zéro, on a

$$\lim_{n \rightarrow \infty} \max_{k: |k-np| < c_n n^{2/3}} |\delta_n(k)| = 0.$$

^{8.1} H. Cramer, *Sur un nouveau théorème-limite de la théorie des probabilités*, Actualités Scientifiques et Industrielles, vol. 736, p. 5-23, Hermann, Paris, 1938.

Démonstration. — Notons I'_n l'ensemble des nombres entiers k tels que $|k - np| \leq c_n n^{2/3}$. Nous partons de la formule (7.5). La formule (7.6) devient, pour tout $k \in I'_n$,

$$\frac{n}{(np + c_n n^{2/3})(n(1-p) + c_n n^{2/3})} \leq \frac{n}{k(n-k)} \leq \frac{n}{(np - c_n n^{2/3})(n(1-p) - c_n n^{2/3})}$$

On déduit de cet encadrement que

$$\frac{n}{k(n-k)} = \frac{1}{np(1-p)} \left(1 + O_u(c_n n^{-1/3})\right)$$

puis que

$$\sqrt{\frac{n}{k(n-k)}} = \frac{1}{\sqrt{np(1-p)}} \left(1 + O_u(c_n n^{-1/3})\right). \quad (8.1)$$

D'autre part on a $\frac{k - np}{k} = O_u(c_n n^{-1/3})$ et $\frac{k - np}{n - k} = O_u(c_n n^{-1/3})$. Le développement limité du logarithme nous donne

$$\begin{aligned} \ln \left(\left(\frac{n}{k}p\right)^k \left(\frac{n}{n-k}(1-p)\right)^{n-k} \right) \\ = -\frac{1}{2}(k - np)^2 \left(\frac{1}{k} + \frac{1}{n-k}\right) + kO_u(c_n^3 n^{-1}) + (n-k)O_u(c_n^3 n^{-1}) \\ = -\frac{1}{2}(k - np)^2 \frac{1}{np(1-p)} + O_u(c_n^3). \end{aligned}$$

Autrement dit

$$\left. \begin{aligned} \left(\frac{n}{k}p\right)^k \left(\frac{n}{n-k}(1-p)\right)^{n-k} \\ = \exp\left(-\frac{(k - np)^2}{2np(1-p)}\right) \left(1 + O_u(c_n^3)\right). \end{aligned} \right\} \quad (8.2)$$

Enfin, comme dans la démonstration de la proposition 7.3, on a

$$\frac{1 + \epsilon_n}{(1 + \epsilon_k)(1 + \epsilon_{n-k})} = \left(1 + O_u\left(\frac{1}{n}\right)\right). \quad (8.3)$$

Rassemblant les estimations (8.1), (8.2) et (8.3) dans l'identité (7.5), on obtient

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{(k - np)^2}{2np(1-p)}\right) \cdot \left(1 + O_u(c'_n)\right),$$

où $c'_n := \max(c_n n^{-1/3}, c_n^3, n^{-1})$.

— QED (Proposition 8.2)

En suivant la démarche qui nous a conduit du théorème de MOIVRE-LAPLACE au théorème limite central, nous allons déduire de la proposition 8.2 une version du théorème limite central avec bornes mobiles.

PROPOSITION 8.3. — Soient (k_n) et (ℓ_n) deux suites de nombres entiers telles que $k_n < \ell_n$ et telles que $k_n, \ell_n = np + o(n^{2/3})$ quand n tend vers l'infini. En notant $a_n := \frac{k_n - np}{\sqrt{np(1-p)}}$ et $b_n := \frac{\ell_n - np}{\sqrt{np(1-p)}}$, on a, quand n tend vers l'infini,

$$P_n(k_n \leq S_n < \ell_n) \sim \frac{1}{\sqrt{2\pi}} \int_{a_n}^{b_n} e^{-x^2/2} dx .$$

Si les suites (a_n) et (b_n) convergent vers des limites respectives a et b telles que $a < b$, on retrouve ici l'énoncé du théorème limite central.

Démonstration. — Notons $h(n) := \frac{1}{\sqrt{np(1-p)}}$. Nous ne considérons que des entiers n assez grand pour que $0 \leq k_n < \ell_n \leq n$. On écrit

$$P_n(S_n = j) = \frac{h(n)}{\sqrt{2\pi}} \exp\left(-\frac{(j - np)^2}{2np(1-p)}\right) \cdot (1 + \delta_n(j))$$

et

$$P_n(k_n \leq S_n < \ell_n) = \frac{h(n)}{\sqrt{2\pi}} \sum_{j=k_n}^{\ell_n-1} \exp\left(-\frac{(j - np)^2}{2np(1-p)}\right) \cdot (1 + \delta_n(j)) .$$

Les hypothèses sur les suites (k_n) , (ℓ_n) et la proposition 8.2 nous permettent d'affirmer que la suite $(\delta_n(j))_{n \geq 1}$ tend vers zéro uniformément quand $k_n \leq j \leq \ell_n$. Il nous suffit donc de démontrer que

$$h(n) \sum_{j=k_n}^{\ell_n-1} \exp\left(-\frac{(j - np)^2}{2np(1-p)}\right) \sim \int_{a_n}^{b_n} e^{-x^2/2} dx . \quad (8.4)$$

C'est bien sûr une propriété de sommes de RIEMANN. Posons

$$x(j) := \frac{j - np}{\sqrt{np(1-p)}} .$$

Ainsi, $a_n = x(k_n)$ et $b_n = x(\ell_n)$. La formule à démontrer (8.4) s'écrit :

$$h(n) \sum_{j=k_n}^{\ell_n-1} \exp\left(-\frac{x(j)^2}{2}\right) - \int_{a_n}^{b_n} e^{-x^2/2} dx = o\left(\int_{a_n}^{b_n} e^{-x^2/2} dx\right) . \quad (8.5)$$

Supposons un moment que $a_n \geq 0$. En remarquant que, pour $k_n \leq j < \ell_n$, on a

$$h(n) \exp\left(-\frac{x(j+1)^2}{2}\right) < \int_{x(j)}^{x(j+1)} e^{-x^2/2} dx < h(n) \exp\left(-\frac{x(j)^2}{2}\right),$$

on obtient

$$0 \leq h(n) \sum_{j=k_n}^{\ell_n-1} e^{-x(j)^2/2} - \int_{a_n}^{b_n} e^{-x^2/2} dx \leq h(n) \left(e^{-a_n^2/2} - e^{-b_n^2/2} \right).$$

Mais on a

$$\int_{a_n}^{b_n} e^{-x^2/2} dx \geq \frac{1}{b_n} \int_{a_n}^{b_n} x e^{-x^2/2} dx = \frac{1}{b_n} \left(e^{-a_n^2/2} - e^{-b_n^2/2} \right).$$

Or on sait que $b_n = o(n^{1/6})$, et donc $h(n) = o(b_n^{-1})$. Les deux dernières assertions entraînent bien (8.5).

Dans le cas où $a_n \geq 0$, la proposition est démontrée. Le cas où $b_n \leq 0$ est bien sûr identique, par symétrie. Dans le cas où $a_n < 0 < b_n$, on peut utiliser les mêmes arguments, en séparant la comparaison de l'intégrale et de la somme discrète en trois parties : une partie sur chacun des intervalles \mathbb{R}^+ et \mathbb{R}^- et un terme éventuel autour de zéro. Les détails sont ici laissés au lecteur pointilleux. — QED (Proposition 8.3)

Démonstration (Théorème 8.1) — La suite (a_n) tend vers l'infini, moins vite que la suite $(n^{1/6})$. Considérons une suite réelle (d_n) telle que $d_n = o(a_n)$ et $d_n \geq \sqrt{2} \ln a_n$ (on peut choisir par exemple $d_n = \ln a_n$ ou $d_n = \sqrt{a_n}$). Pour chaque n , notons k_n le plus petit nombre entier supérieur ou égal à $np + \sqrt{np(1-p)}a_n$, et ℓ_n le plus petit nombre entier supérieur ou égal à $np + \sqrt{np(1-p)}(a_n + d_n)$. On a

$$\mathbf{P}_n \left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)} \frac{a_n}{\sqrt{n}} \right) = \mathbf{P}_n(S_n \geq k_n),$$

que l'on écrit

$$\mathbf{P}_n \left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)} \frac{a_n}{\sqrt{n}} \right) = \mathbf{P}_n(k_n \leq S_n < \ell_n) + \mathbf{P}_n(S_n \geq \ell_n).$$

Du fait que $a_n = o(n^{1/6})$, on déduit que $k_n, \ell_n = np + o(n^{2/3})$ et la proposition 8.3 s'applique. En notant $a'_n := \frac{k_n - np}{\sqrt{np(1-p)}}$ et $b_n := \frac{\ell_n - np}{\sqrt{np(1-p)}}$,

on a

$$P_n(k_n \leq S_n < l_n) \sim \frac{1}{\sqrt{2\pi}} \int_{a'_n}^{b_n} e^{-x^2/2} dx,$$

que l'on écrit

$$P_n(k_n \leq S_n < l_n) \sim \frac{1}{\sqrt{2\pi}} \int_{a_n}^{b_n} e^{-x^2/2} dx - \frac{1}{\sqrt{2\pi}} \int_{a_n}^{a'_n} e^{-x^2/2} dx.$$

Pour conclure, il nous suffit de montrer que

$$\int_{a_n}^{b_n} e^{-x^2/2} dx \sim \frac{1}{a_n} \exp\left(-\frac{a_n^2}{2}\right), \quad (8.6)$$

$$\int_{a_n}^{a'_n} e^{-x^2/2} dx = o\left(\frac{1}{a_n} \exp\left(-\frac{a_n^2}{2}\right)\right), \quad (8.7)$$

et

$$P_n(S_n \geq l_n) = o\left(\frac{1}{a_n} \exp\left(-\frac{a_n^2}{2}\right)\right). \quad (8.8)$$

On a

$$\int_{a_n}^{b_n} e^{-x^2/2} dx \leq \frac{1}{a_n} \int_{a_n}^{+\infty} x e^{-x^2/2} dx = \frac{1}{a_n} \exp\left(-\frac{a_n^2}{2}\right),$$

et, puisque $b_n \geq a_n + d_n$,

$$\begin{aligned} \int_{a_n}^{b_n} e^{-x^2/2} dx &\geq \int_{a_n}^{a_n+d_n} e^{-x^2/2} dx \geq \frac{1}{a_n+d_n} \int_{a_n}^{a_n+d_n} x e^{-x^2/2} dx \\ &= \frac{1}{a_n+d_n} \left(\exp\left(-\frac{a_n^2}{2}\right) - \exp\left(-\frac{(a_n+d_n)^2}{2}\right) \right) \sim \frac{1}{a_n} \exp\left(-\frac{a_n^2}{2}\right), \end{aligned}$$

ce qui prouve (8.6).

Du fait que $0 \leq a'_n - a_n \leq (np(1-p))^{-1/2}$, on déduit directement

$$\int_{a_n}^{a'_n} e^{-x^2/2} dx \leq \frac{1}{\sqrt{np(1-p)}} \exp\left(-\frac{a_n^2}{2}\right),$$

et (8.7) en découle, puisque $a_n = o(\sqrt{n})$.

La dernière estimation se déduit de celle des grands écarts (Théorème 6.1). On a

$$P_n(S_n \geq l_n) \leq \exp\left(-nh_+ \left(\sqrt{p(1-p)} \frac{b_n}{\sqrt{n}}\right)\right).$$

Or au voisinage de zéro, $h_+(\epsilon) = \frac{\epsilon^2}{2p(1-p)} + O(\epsilon^3)$, donc

$$P_n(S_n \geq \ell_n) \leq \exp\left(-\frac{b_n^2}{2} + O\left(\frac{b_n^3}{\sqrt{n}}\right)\right) \sim \exp\left(-\frac{b_n^2}{2}\right),$$

puisque $b_n = o(n^{1/6})$. Enfin, on a

$$\exp\left(-\frac{b_n^2}{2}\right) \leq \exp\left(-\frac{(a_n + d_n)^2}{2}\right) = o\left(\exp\left(-\frac{d_n^2}{2}\right) \exp\left(-\frac{a_n^2}{2}\right)\right).$$

Or, par choix de d_n , on a $\exp\left(-\frac{d_n^2}{2}\right) \leq \frac{1}{a_n}$. Cela prouve (8.8).

—— QED (Théorème 8.1)

9. Une loi d'Arcsinus

9.1. Introduction. — Jules et Jim jouent à pile ou face. Le jeu consiste simplement à effectuer des lancers successifs d'une pièce de monnaie équilibrée et à attribuer un point à Jim quand la pièce affiche «pile», un point à Jules quand la pièce affiche «face». Si le jeu dure assez longtemps, et puisqu'il est équitable, on pourrait s'attendre à ce que Jules soit en tête à peu près la moitié du temps, Jim en tête à peu près l'autre moitié, les instants d'égalité revenant régulièrement. Cela est tout à fait faux : les instants d'égalité sont de plus en plus rares quand la partie s'allonge et, très probablement, un des deux joueurs sera en tête la plus grande partie du temps !

Nous montrerons par exemple que, avec une probabilité supérieure à $2/3$, un des deux joueurs sera en tête pendant plus des $3/4$ du temps. Ou encore que, avec une probabilité supérieure à $1/2$, un des deux joueurs sera en tête pendant plus de 85% du temps. Le fait qu'un même joueur soit en tête plus de 97% du temps n'a rien d'anormal : cela se produit avec une probabilité supérieure à $1/5$.

Ces résultats se déduisent du théorème 9.1, appelé *loi d'Arcsinus*, qui ^{loi} donne une estimation asymptotique précise de la probabilité pour que Jules ^{d'Arcsinus} soit en tête une proportion fixée du temps de jeu. Cette loi a été découverte par Paul LÉVY^{9.1}.

^{9.1} P. Lévy, *Sur certains processus stochastiques homogènes*, Compositio Mathematica, vol.27, p. 283-339, 1939.

loi du
nombre de
retours
en zéro

Voici un autre paradoxe, relié au précédent : au cours d'une partie de n lancers, Jules et Jim seront à égalité un nombre Z_n de fois. Dans une approche naïve, on pourrait penser que la variable aléatoire Z_n croît proportionnellement à n . En fait, nous verrons que Z_n croît beaucoup moins vite, à une vitesse de l'ordre de \sqrt{n} . C'est l'objet du théorème 9.2, appelé *loi du nombre de retours en zéro*.

9.2. Énoncés. — Notre modèle mathématique reste le même que dans les chapitres précédents, avec ici l'hypothèse $p = 1/2$, qui traduit le fait que la pièce est équilibrée. Ainsi, l'espace $\Omega_n = \{0, 1\}^n$ est muni de la probabilité uniforme P_n et les calculs de probabilités se réduisent à des questions de dénombrement. Nous notons $S_n(\omega) = \omega_1 + \omega_2 + \dots + \omega_n$ si $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ et

$$M_n := 2S_n - n$$

(avec la convention $M_0 = S_0 = 0$).

marche
aléatoire
simple

Si S_n est le nombre de « faces » apparues au cours de n lancers, alors M_n est l'avance de Jules sur Jim au bout de n lancers. Dans la littérature probabiliste, la suite (M_n) est appelée *la marche aléatoire simple*. La loi d'Arcsinus décrit, pour n grand, le comportement de la variable aléatoire

$$T_n := \text{card}\{k : 0 \leq k \leq n \text{ et } M_k > 0\}$$

(qui est le nombre d'instant où Jules est en tête)

et la loi des retours en zéro décrit le comportement de la variable aléatoire

$$U_n := \text{card}\{k : 0 < k \leq n \text{ et } M_k = 0\}$$

(qui est le nombre d'instant où les joueurs sont à égalité).

THÉORÈME 9.1 (Loi d'Arcsinus). — Pour tout nombre réel α entre 0 et 1,

$$P_n(T_n < n\alpha) \xrightarrow{n \rightarrow \infty} \frac{1}{\pi} \int_0^\alpha \frac{1}{\sqrt{x(1-x)}} dx = \frac{2}{\pi} \text{Arcsin } \sqrt{\alpha}.$$

Voici une illustration numérique de cette loi : notons p_i la probabilité pour que Jules soit en tête durant une fraction du temps de jeu T_n/n comprise entre $\frac{i}{10}$ et $\frac{i+1}{10}$; nous avons, au moins si le jeu dure assez longtemps, $p_0 = p_9 \approx 0,20 > p_1 = p_8 \approx 0,090 > p_2 = p_7 \approx 0,074 > p_3 = p_6 \approx 0,067 > p_4 = p_5 \approx 0,064$.

THÉORÈME 9.2 (Loi des retours en zéro). — Pour tout $\alpha > 0$,

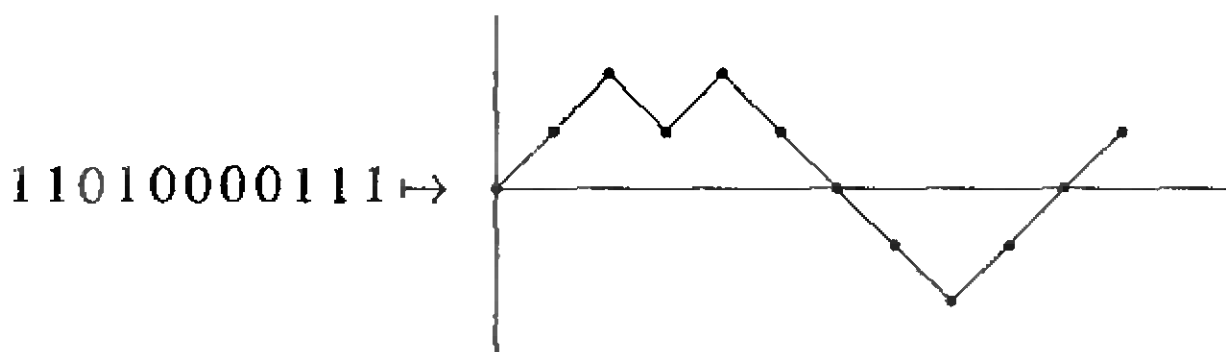
$$\mathbb{P}_n(U_n < \alpha\sqrt{n}) \xrightarrow{n \rightarrow \infty} \sqrt{\frac{2}{\pi}} \int_0^\alpha e^{-x^2/2} dx$$

Ce théorème nous dit par exemple que si Jules et Jim lancent une pièce 10 000 fois alors, avec une probabilité supérieure à $1/2$, ils se retrouveront à égalité moins de 68 fois. Et s'ils lancent la pièce cent fois plus longtemps, 68 doit être remplacé par 680.

Une conséquence immédiate de la loi des retours en zéro est que, pour tout $\epsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}_n(U_n > n\epsilon) = 0$. Cela entraîne que la variable aléatoire $T_n + U_n = \text{card}\{k : 0 < k \leq n \text{ et } M_k \geq 0\}$ vérifie la même loi d'Arcsinus que la variable aléatoire T_n .

Comme le théorème limite central, les deux théorèmes précédents seront démontrés en trois étapes : calculs explicites de probabilités, approximations utilisant la formule de STIRLING et utilisation de sommes de RIEMANN. Notons enfin que de nombreux autres résultats de nature similaire se trouvent dans la littérature.

9.3. Le principe de réflexion. — Nous allons utiliser une représentation graphique de l'espace des issues $\Omega_n = \{0, 1\}^n$. Convenons d'appeler chemin dans le plan \mathbb{R}^2 toute courbe C continue affine par morceaux réunion finie de segments du type $[(i, j), (i+1, j+1)]$ ou $[(i, j), (i+1, j-1)]$, avec i et j entiers. Un tel chemin possède une origine (a, b) et une extrémité (c, d) , points à coordonnées entières sur la courbe C tels que, pour tout $(i, j) \in C$, on ait $a \leq i \leq c$. On appelle longueur du chemin C le nombre $c - a$ (la longueur euclidienne de C est $(c - a)\sqrt{2}$). À tout élément ω de Ω_n , on associe le chemin $\cup_{i=0}^{n-1} [(i, M_i(\omega)), (i+1, M_{i+1}(\omega))]$, d'origine $(0, 0)$ et d'extrémité $(n, M_n(\omega))$. Cette représentation de Ω_n est bien sûr injective. Elle décrit exactement l'évolution de l'avance de Jules sur Jim au cours du jeu. Notons que tout chemin d'origine $(0, 0)$ et d'extrémité d'abscisse n représente un élément de Ω_n .



Si c et d sont deux nombres entiers tels que $0 \leq |d| \leq c$, le nombre de chemins d'origine $(0, 0)$ et d'extrémité (c, d) est nul si $c + d$ est impair et égal à $2^c P_c(M_c = d) = 2^c P_c(S_c = (c + d)/2) = \binom{c}{(c+d)/2}$ si $c + d$ est pair. Plus généralement, si a, b, c et d sont des nombres entiers tels que $0 \leq |d - b| \leq c - a$ et tels que $c - a + d - b$ soit pair, alors le nombre de chemins d'origine (a, b) et d'extrémité (c, d) est égal à $\binom{c-a}{(c-a+d-b)/2}$.

La proposition suivante, qui est une réécriture d'un résultat dû à Désiré ANDRÉ^{9.2} joue un rôle essentiel dans notre analyse.

principe de réflexion $n > 0$ des nombres entiers. On a

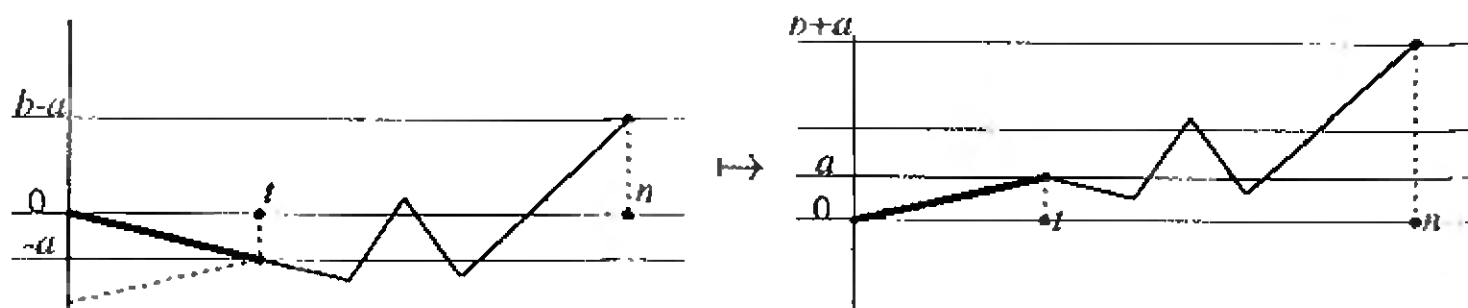
PROPOSITION 9.3 (Principe de réflexion). — Soient $a, b \geq 0$ et $n > 0$ des nombres entiers. On a

$$P_n(M_n = b - a \text{ et } M_k = -a \text{ pour un } k \text{ entre } 0 \text{ et } n) = P_n(M_n = b + a) .$$

Démonstration. — Nous supposons $a > 0$, le cas $a = 0$ étant trivial. Il s'agit de démontrer que deux sous-ensembles de l'espace Ω_n ont le même nombre d'éléments. En utilisant notre représentation graphique, il faut montrer que le nombre de chemins d'origine $(0, 0)$ et d'extrémité $(n, b - a)$ rencontrant la droite horizontale d'ordonnée $-a$ est égal au nombre de chemins d'origine $(0, 0)$ et d'extrémité $(n, a + b)$. Par translation verticale, il faut montrer que le nombre de chemins d'origine $(0, a)$ et d'extrémité (n, b) rencontrant l'axe horizontal « $y = 0$ » (appelés «chemins du premier type») est égal au nombre de chemins d'origine $(0, -a)$ et d'extrémité (n, b) (appelés «chemins du second type»). Si C est un chemin du premier type, on note $t(C)$ le plus petit $i > 0$ tel que $(i, 0) \in C$. Le chemin C est réunion d'un chemin C_1 d'origine $(0, 0)$ et d'extrémité $(t(C), 0)$ et d'un chemin C_2 d'origine $(t(C), 0)$ et d'extrémité (n, b) . On associe à C le chemin C' obtenu par réunion de C_2 et du symétrique de C_1 par rapport à l'axe horizontal. Le chemin C' est du second type et la correspondance $C \leftrightarrow C'$ établit une

^{9.2} D. André, *Solution directe du problème résolu par M. Bertrand*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 105, p. 436-437, 1887. Il s'agit de la résolution du *problème du scrutin* : lors du décompte des voix dans une élection opposant deux candidats, quelle est la probabilité pour que le candidat gagnant reste en tête durant tout le dépouillement ?

bijection entre les ensembles de chemins de chacun des types.



— QED (Principe de réflexion)

Nous connaissons la probabilité pour qu'au bout de $2n$ lancers les deux joueurs soient à égalité :

$$P_{2n}(M_{2n} = 0) = P_{2n}(S_{2n} = n) = 2^{-2n} \binom{2n}{n}.$$

Nous allons voir successivement, dans les deux corollaires suivants, que cette probabilité est la même que celle de l'événement «un même joueur reste en tête pendant toute la partie de $2n$ lancers» et que celle de l'événement «Jim n'est en tête à aucun moment de la partie».

COROLLAIRE 9.4. —

$$P_{2n}(M_1 \neq 0, M_2 \neq 0, \dots \text{ et } M_{2n} \neq 0) \\ = 2 P_{2n}(M_1 > 0, M_2 > 0, \dots \text{ et } M_{2n} > 0);$$

$$P_{2n}(M_1 > 0, M_2 > 0, \dots \text{ et } M_{2n} > 0) = 2^{-(2n+1)} \binom{2n}{n}.$$

Démonstration. — Par symétrie, la première identité est évidente.

Pour démontrer la seconde, reprenons notre représentation graphique. Nous devons compter les chemins d'origine $(0, 0)$ et de longueur $2n$ qui sont contenus dans le demi-plan supérieur (et qui ne rencontrent pas l'axe horizontal, sauf en leur origine). Nous devons donc sommer, entre $k = 1$ et $k = n$, le nombre de chemins d'origine $(1, 1)$ et d'extrémité $(2n, 2k)$ ne rencontrant pas l'axe horizontal. Il n'y a qu'un chemin reliant les points $(1, 1)$ et $(2n, 2n)$, et ce chemin ne rencontre pas l'axe horizontal. Si $1 \leq k < n$, le nombre de chemins reliant le point $(1, 1)$ au point $(2n, 2k)$ et ne rencontrant pas l'axe horizontal est égal au nombre de chemins reliant le point $(1, 1)$ au point $(2n, 2k)$ moins le nombre de chemins reliant le point $(1, 1)$ au point $(2n, 2k)$ et rencontrant l'axe horizontal. Le nombre de chemins reliant le point $(1, 1)$ au point $(2n, 2k)$ est égal à $\binom{2n-1}{n+k-1}$. D'après le principe de

réflexion, le nombre de chemins reliant le point $(1, 1)$ au point $(2n, 2k)$ et rencontrant l'axe horizontal est égal au nombre de chemins reliant le point $(1, -1)$ au point $(2n, 2k)$, qui est égal à $\binom{2n-1}{n+k}$.

En conclusion, le nombre de chemins d'origine $(0, 0)$ et de longueur $2n$ qui sont contenus dans le demi-plan supérieur est égal à

$$1 + \sum_{k=1}^{n-1} \left(\binom{2n-1}{n+k-1} - \binom{2n-1}{n+k} \right).$$

Cette expression se simplifie et ce nombre est égal à $\binom{2n-1}{n}$; on vérifie enfin sans peine que $\binom{2n-1}{n} = \frac{1}{2} \binom{2n}{n}$. Cela achève la démonstration du corollaire.

— QED (Corollaire 9.4)

Exprimons à présent la probabilité pour qu'à aucun moment de la partie Jim ne soit en tête.

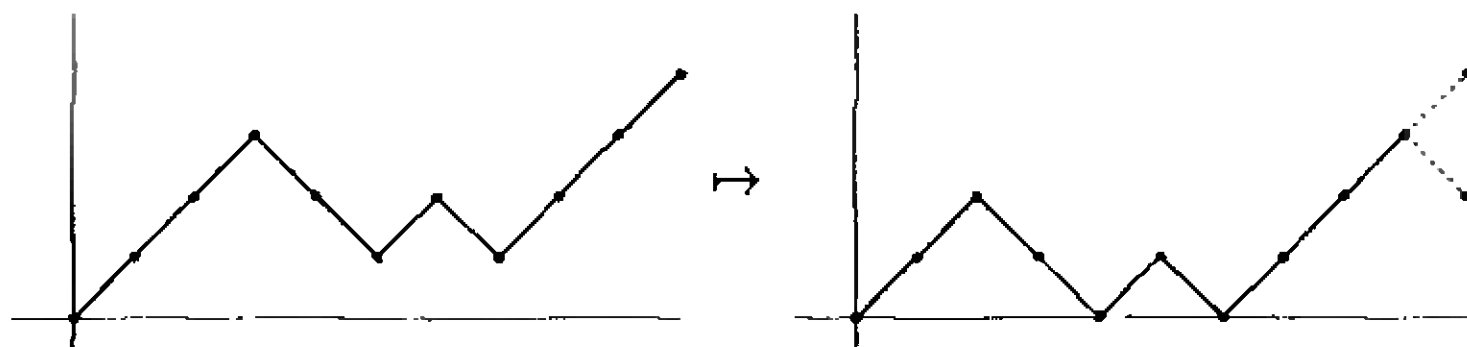
COROLLAIRE 9.5. —

$$\mathbb{P}_{2n}(M_1 \geq 0, M_2 \geq 0, \dots \text{ et } M_{2n} \geq 0) = 2^{-2n} \binom{2n}{n}.$$

Démonstration. — Nous voulons montrer que

$$\begin{aligned} \mathbb{P}_{2n}(M_1 \geq 0, M_2 \geq 0, \dots \text{ et } M_{2n} \geq 0) \\ = 2\mathbb{P}_{2n}(M_1 > 0, M_2 > 0, \dots \text{ et } M_{2n} > 0). \end{aligned}$$

Le nombre de chemins d'origine $(0, 0)$, de longueur $2n$, contenus dans le demi-plan supérieur et ne rencontrant pas l'axe horizontal est égal au nombre de chemins d'origine $(0, 0)$, de longueur $2n - 1$, contenus dans le demi-plan supérieur (y compris l'axe horizontal). En effet, ces deux ensembles de chemins sont en bijection : à tout chemin du premier type, on associe un chemin du second type en ôtant le segment initial et en effectuant une translation de 1 vers la droite et de 1 vers le bas.



Remarquons que M_{2n-1} n'est jamais nul. À tout chemin d'origine $(0, 0)$, de longueur $2n - 1$, contenu dans le demi-plan supérieur (y compris l'axe horizontal) on peut donc associer, en ajoutant un segment terminal, exactement deux chemins d'origine $(0, 0)$, de longueur $2n$, contenus dans le demi-plan supérieur (y compris l'axe horizontal). Le cardinal de l'événement $(M_1 \geq 0, M_2 \geq 0, \dots \text{ et } M_{2n} \geq 0)$ est donc double du cardinal de l'événement $(M_1 > 0, M_2 > 0, \dots \text{ et } M_{2n} > 0)$.

— QED (Corollaire 9.5)

Le principe de réflexion permet aussi de calculer la probabilité pour que Jules soit resté en tête tout le long de la partie et que, à la fin, Jules et Jim soient à égalité.

COROLLAIRE 9.6. —

$$P_{2n}(M_1 > 0, M_2 > 0, \dots, M_{2n-1} > 0 \text{ et } M_{2n} = 0) = \frac{1}{n2^{2n}} \binom{2n-2}{n-1}.$$

Démonstration. — Il s'agit de compter les chemins d'origine $(0, 0)$, d'extrémité $(2n, 0)$ et entièrement contenus entre ces deux points dans le demi-plan supérieur ouvert. Autrement dit, il s'agit de compter les chemins d'origine $(1, 1)$, d'extrémité $(2n-1, 1)$ et entièrement contenus dans le demi-plan supérieur ouvert. Leur nombre est égal au nombre de chemins d'origine $(1, 1)$ et d'extrémité $(2n-1, 1)$, moins le nombre de chemins d'origine $(1, 1)$ et d'extrémité $(2n-1, 1)$ rencontrant l'axe horizontal. Le nombre de chemins d'origine $(1, 1)$ et d'extrémité $(2n-1, 1)$ est égal à $\binom{2n-2}{n-1}$. Grâce au principe de réflexion, on sait que le nombre de chemins d'origine $(1, 1)$ et d'extrémité $(2n-1, 1)$ rencontrant l'axe horizontal est égal au nombre de chemins d'origine $(1, -1)$ et d'extrémité $(2n-1, 1)$; ce nombre est égal à $\binom{2n-2}{n-2}$. Finalement, la vérification de $\binom{2n-2}{n-1} - \binom{2n-2}{n-2} = \frac{1}{n} \binom{2n-2}{n-1}$ est immédiate.

— QED (Corollaire 9.6)

Du corollaire 9.6 on déduit une formule combinatoire fort utile.

LEMME 9.7. — Si n et k sont des nombres entiers tels que $0 \leq k < n$, alors

$$\sum_{j=1}^{n-k} \frac{1}{j} \binom{2j-2}{j-1} \binom{2(n-j)-k}{n-j} = \binom{2n-k-1}{n}.$$

$$2 \sum_{j=1}^n \frac{1}{j} \binom{2j-2}{j-1} \binom{2(n-j)}{n-j} = \binom{2n}{n}.$$

Démonstration. — Nous allons compter de deux façons les chemins d'origine $(0, 0)$ et d'extrémité $(2n - k, k)$. D'une part ce nombre est égal à $\binom{2n-k}{n}$. D'autre part, il est égal au nombre de chemins d'origine $(1, 1)$ et d'extrémité $(2n - k, k)$ additionné au nombre de chemins d'origine $(1, -1)$ et d'extrémité $(2n - k, k)$. Le nombre de chemins d'origine $(1, 1)$ et d'extrémité $(2n - k, k)$ est égal à $\binom{2n-k-1}{n-1}$. Pour tout chemin d'origine $(1, -1)$ et d'extrémité $(2n - k, k)$, il existe un entier $j \geq 1$ minimal tel que le chemin passe par le point $(2j, 0)$; on a nécessairement $j \leq n - k$. Le nombre de chemins d'origine $(1, -1)$ et d'extrémité $(2n - k, k)$ est donc égal à la somme, pour j variant de 1 à $n - k$, du produit du nombre de chemins d'origine $(0, 0)$, d'extrémité $(2j, 0)$ et entièrement contenus entre ces deux points dans le demi-plan ouvert inférieur, par le nombre de chemins d'origine $(2j, 0)$ et d'extrémité $(2n - k, k)$. Grâce au corollaire 9.6, cette phrase complexe signifie que le nombre de chemins d'origine $(1, -1)$ et d'extrémité $(2n - k, k)$ est égal à

$$\sum_{j=1}^{n-k} \left(\frac{1}{j} \binom{2j-2}{j-1} \right) \binom{2(n-j)-k}{n-j}.$$

Cette expression est égale à $\binom{2n-k}{n} - \binom{2n-k-1}{n-1}$, c'est-à-dire à $\binom{2n-k-1}{n}$.

La seconde formule du lemme 9.7 s'obtient en considérant le cas particulier $k = 0$ et en remarquant que $\binom{2n}{n} = 2 \binom{2n-1}{n}$.

— QED (Lemme 9.7)

9.4. Démonstration de la loi d'Arcsinus. — Nous allons définir une nouvelle suite de variables aléatoires (T'_n) dont on peut expliciter les lois et qui seront proches des variables T_n qui nous intéressent. Nous notons T'_n le nombre d'entiers k entre 1 et n tels que, dans les k premiers ou dans les $k - 1$ premiers lancers, il est apparu strictement plus de « faces » que de « piles ». Avec notre interprétation graphique, T'_n associe, à chaque chemin de longueur n et d'origine $(0, 0)$, le nombre de segments élémentaires situés dans le demi-plan supérieur. Plus formellement, on a

$$T'_n := \text{card}\{k : 0 < k \leq n \text{ et } (M_k > 0 \text{ ou } M_{k-1} > 0)\}.$$

Tout segment élémentaire est de la forme

$$\left[(2k-1, M_{2k-1}), (2k, M_{2k-1} \pm 1) \right] \\ \text{ou} \quad \left[(2k-2, M_{2k-1} \pm 1), (2k-1, M_{2k-1}) \right]$$

où M_{2k-1} est un nombre entier non nul. Ce segment est dans le demi-plan supérieur si et seulement si $M_{2k-1} > 0$. Ainsi, on a

$$T'_{2n} = 2 \text{ card} \{ k : 0 < k \leq n \text{ et } M_{2k-1} > 0 \} .$$

PROPOSITION 9.8. — Pour tout $n > 0$, pour tout k entre 0 et n , on a

$$\mathbb{P}_{2n}(T'_{2n} = 2k) = 2^{-2n} \binom{2k}{k} \binom{2(n-k)}{n-k} . \quad (9.1)$$

Démonstration. — On a

$$(T'_{2n} = 2n) = (\text{pour tout } k \text{ entre } 1 \text{ et } 2n, M_k \geq 0)$$

et la probabilité de cet événement est donnée par le corollaire 9.5. La formule générale (9.1) sera démontrée par récurrence sur n . Pour $n = 1$, la vérification est immédiate. Fixons $N > 1$ et supposons la formule (9.1) vraie pour tout n entre 1 et $N - 1$, et tout k entre 0 et n .

La probabilité de l'événement

$$(T'_{2N} = 0) = (\text{pour tout } k \text{ entre } 1 \text{ et } 2N, M_k \leq 0)$$

est donnée par le corollaire 9.5. Si $0 < T'_{2N} < 2N$, alors il existe j entre 1 et N tel que $M_{2j} = 0$. Pour chaque $\omega \in \Omega_{2N}$ tel que $0 < T'_{2N}(\omega) < 2N$, on note

$$t(\omega) := \min \{ j > 0 : M_{2j}(\omega) = 0 \} .$$

Soit k fixé tel que $0 < k < N$.

$$\mathbb{P}_{2N}(T'_{2N} = 2k) = \sum_{j=1}^N \mathbb{P}_{2N}(T'_{2N} = 2k \text{ et } t = 2j \text{ et } M_1 > 0) \\ + \sum_{j=1}^N \mathbb{P}_{2N}(T'_{2N} = 2k \text{ et } t = 2j \text{ et } M_1 < 0) .$$

Si $j > k$, l'événement $(T'_{2N} = 2k \text{ et } t = 2j \text{ et } M_1 > 0)$ est vide. Si $j \leq k$, le cardinal de cet événement est égal au nombre de chemins d'origine $(0, 0)$, d'extrémité $(2j, 0)$ et contenus entre ces deux points dans le demi-plan

supérieur ouvert multiplié par le nombre de chemins d'origine $(2j, 0)$, de longueur $2(N - j)$ et dont $2(k - j)$ segments élémentaires sont dans le demi-plan supérieur. Le nombre de chemins d'origine $(0, 0)$, d'extrémité $(2j, 0)$ et contenus entre ces deux points dans le demi-plan supérieur ouvert est donné par le corollaire 9.6 ; il est égal à $\frac{1}{j} \binom{2j-2}{j-1}$. Le nombre de chemins d'origine $(2j, 0)$, de longueur $2(N - j)$ et dont $2(k - j)$ segments élémentaires sont dans le demi-plan supérieur est donné par l'hypothèse de récurrence ; il est égal à $\binom{2(k-j)}{k-j} \binom{2(N-k)}{N-k}$. Ainsi, on a

$$P_{2N} (T'_{2N} = 2k \text{ et } t = 2j \text{ et } M_1 > 0) = \frac{1}{j2^{2N}} \binom{2j-2}{j-1} \binom{2(k-j)}{k-j} \binom{2(N-k)}{N-k}.$$

Si $j > N - k$, l'événement $(T'_{2N} = 2k \text{ et } t = 2j \text{ et } M_1 < 0)$ est vide. Si $j \leq N - k$, le cardinal de cet événement est égal au nombre de chemins d'origine $(0, 0)$, d'extrémité $(2j, 0)$ et contenus entre ces deux points dans le demi-plan inférieur ouvert multiplié par le nombre de chemins d'origine $(2j, 0)$, de longueur $2(N - j)$ et dont $2k$ segments élémentaires sont dans le demi-plan supérieur. Le nombre de chemins d'origine $(0, 0)$, d'extrémité $(2j, 0)$ et contenus entre ces deux points dans le demi-plan inférieur ouvert est donné par le corollaire 9.6 ; il est égal à $\frac{1}{j} \binom{2j-2}{j-1}$. Le nombre de chemins d'origine $(2j, 0)$, de longueur $2(N - j)$ et dont $2k$ segments élémentaires sont dans le demi-plan supérieur est donné par l'hypothèse de récurrence ; il est égal à $\binom{2k}{k} \binom{2(N-j-k)}{N-j-k}$. Ainsi, on a

$$P_{2N} (T'_{2N} = 2k \text{ et } t = 2j \text{ et } M_1 < 0) = \frac{1}{j2^{2N}} \binom{2j-2}{j-1} \binom{2k}{k} \binom{2(N-j-k)}{N-j-k}.$$

Nous pouvons conclure en utilisant la seconde formule du lemme 9.7 :

$$\begin{aligned} P_{2N} (T'_{2N} = 2k) &= \sum_{j=1}^k \frac{1}{j2^{2N}} \binom{2j-2}{j-1} \binom{2(k-j)}{k-j} \binom{2(N-k)}{N-k} + \sum_{j=1}^{N-k} \frac{1}{j2^{2N}} \binom{2j-2}{j-1} \binom{2k}{k} \binom{2(N-j-k)}{N-j-k} \\ &= \left(\frac{1}{2^{2N}} \binom{2(N-k)}{N-k} \right) \left(\frac{1}{2} \binom{2k}{k} \right) + \left(\frac{1}{2^{2N}} \binom{2k}{k} \right) \left(\frac{1}{2} \binom{2(N-k)}{N-k} \right) \\ &= \frac{1}{2^{2N}} \binom{2(N-k)}{N-k} \binom{2k}{k}. \end{aligned}$$

Cela achève le raisonnement par récurrence.

— QED (Proposition 9.8)

Montrons à présent que les variables aléatoires T'_{2n} vérifient la loi d'Arcsinus.

PROPOSITION 9.9. — Pour tous nombres réels a et b tels que $0 \leq a \leq b \leq 1$, on a

$$\lim_{n \rightarrow \infty} P_{2n}(2na \leq T'_{2n} \leq 2nb) = \frac{1}{\pi} \int_a^b \frac{1}{\sqrt{x(1-x)}} dx.$$

Démonstration. — Dans une première étape, nous allons étudier le cas $0 < a < b < 1$. À partir de la proposition 9.8, une application directe de la formule de STIRLING (Proposition 7.2) nous donne, pour $0 < k < n$,

$$P_{2n}(T'_{2n} = 2k) = \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}} (1 + \epsilon(k))(1 + \epsilon(n-k)),$$

avec $\lim_{k \rightarrow \infty} \epsilon(k) = 0$. Ainsi,

$$P_{2n}(T'_{2n} = 2k) = \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}} (1 + \epsilon(n, k)),$$

avec $\lim_{n \rightarrow \infty} \epsilon(n, k) = 0$, uniformément quand k varie dans l'intervalle d'entiers $[na, nb] \cap \mathbb{Z}$. Cela entraîne que

$$\begin{aligned} P_{2n}(2na \leq T'_{2n} \leq 2nb) &= \sum_{na \leq k \leq nb} P_{2n}(T'_{2n} = 2k) \sim \sum_{na \leq k \leq nb} \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}}. \end{aligned}$$

Le terme de droite ci-dessus s'écrit

$$\frac{1}{n\pi} \sum_{k=0}^n \mathbf{1}_{[a,b]}\left(\frac{k}{n}\right) \frac{1}{\sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right)}}$$

et on reconnaît une somme de Riemann qui tend vers $\frac{1}{\pi} \int_a^b \frac{1}{\sqrt{x(1-x)}} dx$, ce qui conclut la première étape de la démonstration.

Soit $\epsilon > 0$. Puisque l'intégrale $\int_0^1 \frac{1}{\sqrt{x(1-x)}} dx$ est convergente, pour tout $a > 0$ assez proche de zéro, on a

$$\frac{1}{\pi} \int_0^a \frac{1}{\sqrt{x(1-x)}} dx < \epsilon \quad \text{et} \quad \frac{1}{\pi} \int_{1-a}^1 \frac{1}{\sqrt{x(1-x)}} dx < \epsilon.$$

Fixons un tel nombre a . D'après la première étape de la démonstration, on a, pour tout n assez grand,

$$\left| \frac{1}{\pi} \int_a^{1-a} \frac{1}{\sqrt{x(1-x)}} dx - P_{2n}(2na \leq T'_{2n} < 2n(1-a)) \right| < \epsilon.$$

Or

$$\mathbb{P}_{2n} (T'_{2n} < 2na) + \mathbb{P}_{2n} (2na \leq T'_{2n} < 2n(1-a)) + \mathbb{P}_{2n} (T'_{2n} \geq 2n(1-a)) = 1$$

et $\frac{1}{\pi} \int_0^1 \frac{1}{\sqrt{x(1-x)}} dx = 1$. On en déduit que, pour tout n assez grand,

$$\left| \frac{1}{\pi} \int_0^a \frac{1}{\sqrt{x(1-x)}} dx + \frac{1}{\pi} \int_{1-a}^1 \frac{1}{\sqrt{x(1-x)}} dx - \mathbb{P}_{2n} (T'_{2n} < 2na) - \mathbb{P}_{2n} (T'_{2n} \geq 2n(1-a)) \right| < \epsilon,$$

donc

$$\mathbb{P}_{2n} (T'_{2n} < 2na) + \mathbb{P}_{2n} (T'_{2n} \geq 2n(1-a)) < 3\epsilon.$$

Nous avons démontré qu'il existe $a > 0$ tel que, pour tout n assez grand,

$$\mathbb{P}_{2n} (T'_{2n} < 2na) < 3\epsilon.$$

Puisque $\mathbb{P}_{2n} (T'_{2n} < 2na)$ est une fonction croissante de a , qui, pour chaque n fixé, tend vers 0 quand a tend vers 0, nous pouvons en déduire que $\lim_{a \rightarrow 0} \mathbb{P}_{2n} (T'_{2n} < 2na) = 0$ uniformément en n .

Grâce à nouveau à la première étape de la démonstration, on a alors, pour tout $b \in]0, 1[$,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{2n} (T'_{2n} \leq 2nb) = \frac{1}{\pi} \int_0^b \frac{1}{\sqrt{x(1-x)}} dx.$$

Les mêmes arguments permettent de montrer que, si $0 < a < b < 1$,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{2n} (2na < T'_{2n} < 2nb) = \frac{1}{\pi} \int_a^b \frac{1}{\sqrt{x(1-x)}} dx.$$

et

$$\lim_{n \rightarrow \infty} \mathbb{P}_{2n} (T'_{2n} < 2na) = \frac{1}{\pi} \int_0^a \frac{1}{\sqrt{x(1-x)}} dx.$$

et le résultat annoncé s'en déduit immédiatement.

— QED (Proposition 9.9)

Pour achever la démonstration du théorème 9.1 il ne nous reste plus qu'à comparer les variables aléatoires T_n et T'_n . On a

$$T_{2n} = T'_{2n} - \text{card} \{k : 1 \leq k \leq n, M_{2k-1} > 0 \text{ et } M_{2k} = 0\} ,$$

et donc

$$|T_{2n} - T'_{2n}| \leq \text{card} \{k : 1 \leq k \leq n \text{ et } M_{2k} = 0\} = U_{2n} .$$

Nous démontrons dans le prochain paragraphe la loi des temps de retours en zéro, qui entraîne en particulier que

$$\text{pour tout } \epsilon > 0, \quad \lim_{n \rightarrow \infty} P_{2n}(U_{2n} > 2n\epsilon) = 0 ;$$

mais on peut donner de ceci une justification directe. On a en effet

$$E_{2n}[U_{2n}] = E_{2n} \left[\sum_{k=1}^n \mathbf{1}_{(M_{2k}=0)} \right] = \sum_{k=1}^n P_{2n}(M_{2k} = 0) = \sum_{k=1}^n 2^{-2k} \binom{2k}{k} ;$$

donc, d'après l'inégalité de Markov, pour tout $\epsilon > 0$,

$$P_{2n}(U_{2n} > 2n\epsilon) \leq \frac{1}{2n\epsilon} \sum_{k=1}^n 2^{-2k} \binom{2k}{k} .$$

Or il est facile de vérifier, à l'aide de la formule de STIRLING, que

$$\lim_{k \rightarrow \infty} 2^{-2k} \binom{2k}{k} = 0 ,$$

et, grâce au principe de CESÀRO, on peut conclure que

$$\lim_{n \rightarrow \infty} P_{2n}(U_{2n} > 2n\epsilon) = 0 .$$

L'événement $(T_{2n} < 2n\alpha)$ étant inclus dans la réunion

$$(|T_{2n} - T'_{2n}| > 2n\epsilon) \cup (T'_{2n} \leq 2n(\alpha + \epsilon)) ,$$

on a

$$P_{2n}(T_{2n} < 2n\alpha) \leq P_{2n}(|T_{2n} - T'_{2n}| > 2n\epsilon) + P_{2n}(T'_{2n} \leq 2n(\alpha + \epsilon)) .$$

D'une part, quand $n \rightarrow \infty$,

$$P_{2n}(|T_{2n} - T'_{2n}| > 2n\epsilon) \leq P_{2n}(U_{2n} > 2n\epsilon) \longrightarrow 0 .$$

d'autre part, d'après la proposition 9.9,

$$\lim_{n \rightarrow \infty} \mathbf{P}_{2n} (T'_{2n} \leq 2n(\alpha + \epsilon)) = \frac{1}{\pi} \int_0^{\alpha + \epsilon} \frac{1}{\sqrt{x(1-x)}} dx$$

et

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \int_0^{\alpha + \epsilon} \frac{1}{\sqrt{x(1-x)}} dx = \frac{1}{\pi} \int_0^{\alpha} \frac{1}{\sqrt{x(1-x)}} dx ;$$

donc

$$\limsup_{n \rightarrow \infty} \mathbf{P}_{2n} (T_{2n} < 2n\alpha) \leq \frac{1}{\pi} \int_0^{\alpha} \frac{1}{\sqrt{x(1-x)}} dx .$$

Puisque $T_{2n} \leq T'_{2n}$, on déduit directement de la proposition 9.9 que

$$\liminf_{n \rightarrow \infty} \mathbf{P}_{2n} (T_{2n} < 2n\alpha) \geq \frac{1}{\pi} \int_0^{\alpha} \frac{1}{\sqrt{x(1-x)}} dx .$$

Et on peut conclure que

$$\lim_{n \rightarrow \infty} \mathbf{P}_{2n} (T_{2n} < 2n\alpha) = \frac{1}{\pi} \int_0^{\alpha} \frac{1}{\sqrt{x(1-x)}} dx .$$

Il ne nous reste plus qu'à vérifier que

$$\lim_{n \rightarrow \infty} \mathbf{P}_{2n+1} (T_{2n+1} < (2n+1)\alpha) = \frac{1}{\pi} \int_0^{\alpha} \frac{1}{\sqrt{x(1-x)}} dx ,$$

et cela s'obtient aisément, par exemple à partir des encadrements suivants

$$\mathbf{P}_{2n+1} (T_{2n+1} < (2n+1)\alpha) \leq \mathbf{P}_{2n} (T_{2n} < (2n+1)\alpha) ,$$

$$\mathbf{P}_{2n+1} (T_{2n+1} < (2n+1)\alpha) \geq \mathbf{P}_{2n+2} (T_{2n+2} < (2n+1)\alpha) .$$

— QED (Loi d'Arcsinus)

9.5. Démonstration de la loi des retours en zéro. — Nous allons déterminer explicitement la loi de la variable aléatoire U_{2n} . Il est clair que cette variable aléatoire ne peut prendre que des valeurs entières entre 0 et n . Notons que, pour chaque nombre entier k entre 0 et n , la probabilité $\mathbf{P}_{2n} (U_{2n} = k)$ est égale à 2^{-2n} multiplié par le nombre de chemins de longueur $2n$ issus de $(0, 0)$ rencontrant l'axe horizontal en exactement $k + 1$ points.

PROPOSITION 9.10. — *Pour tout k entre 0 et n , on a*

$$\mathbf{P}_{2n} (U_{2n} = k) = \frac{1}{2^{2n-k}} \binom{2n-k}{n} .$$

On remarque que $\mathbf{P}_{2n} (U_{2n} = k)$ est une fonction décroissante de k .

Démonstration. — La formule $\mathbb{P}_{2n}(U_{2n} = 0) = 2^{-2n} \binom{2n}{n}$ a déjà été démontrée ; c'est l'objet du corollaire 9.4. Pour démontrer la formule générale, nous raisonnons par récurrence sur n . Si $U_{2n}(\omega) > 0$, alors il existe un $j \geq 1$ minimal tel que $M_{2j}(\omega) = 0$. Cet entier j est noté $t(\omega)$. Soit $k > 0$. On a

$$\mathbb{P}_{2n}(U_{2n} = k) = \sum_{j=1}^n \mathbb{P}_{2n}(t = j \text{ et } U_{2n} = k) .$$

Si $j > n - k + 1$, alors l'événement $(t = j \text{ et } U_{2n} = k)$ est vide. Si $j \leq n - k + 1$ alors le cardinal de cet événement est égal au nombre de chemins d'origine $(0, 0)$, d'extrémité $(2j, 0)$ et ne rencontrant pas l'axe horizontal entre ces deux points, multiplié par le nombre de chemins d'origine $(2j, 0)$, de longueur $2n - 2j$ et rencontrant l'axe horizontal en exactement k points. Ces nombres sont donnés respectivement par le corollaire 9.6 et par une hypothèse de récurrence. On obtient ainsi :

$$\text{card}(t = j \text{ et } U_{2n} = k) = \frac{2}{j} \binom{2j-2}{j-1} 2^{k-1} \binom{2n-2j-k+1}{n-j} .$$

On en déduit que :

$$\text{card}(U_{2n} = k) = 2^k \sum_{j=1}^{n-k+1} \frac{1}{j} \binom{2j-2}{j-1} \binom{2n-2j-k+1}{n-j} ,$$

et le lemme 9.7 permet de conclure que

$$\text{card}(U_{2n} = k) = 2^k \binom{2n-k}{n} .$$

— QED (Proposition 9.10)

Démonstration (Théorème 9.2) — Le nombre $\alpha > 0$ est fixé. Nous emploierons une notation déjà utilisée dans la démonstration du théorème limite central : si $(s_{n,k})_{n>0, 0 \leq k < \alpha\sqrt{2n}}$ et $(t_n)_{n>0}$ sont deux familles de nombres réels, nous écrirons $s_{n,k} = O_u(t_n)$ si la valeur absolue de $s_{n,k}$ est majorée, à constante multiplicative près, par t_n , uniformément en k .

La proposition 9.10 et la formule de STIRLING nous permettent d'estimer $\mathbb{P}_{2n}(U_{2n} = k)$ quand $0 \leq k < \alpha\sqrt{2n}$. On a ainsi :

$$\mathbb{P}_{2n}(U_{2n} = k) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2n-k}{n(n-k)}} \left(1 - \frac{k}{2n}\right)^n \left(1 + \frac{k}{2(n-k)}\right)^{n-k} \left(1 + O_u\left(\frac{1}{n}\right)\right) .$$

D'où :

$$P_{2n}(U_{2n} = k) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2}{n}} \left(1 + O_u\left(\frac{1}{\sqrt{n}}\right)\right) \times \\ \exp\left[n\left(-\frac{k}{2n} - \frac{k^2}{8n^2}\right) + (n-k)\left(\frac{k}{2(n-k)} - \frac{k^2}{8(n-k)^2}\right) + O_u\left(\frac{1}{\sqrt{n}}\right)\right],$$

et finalement

$$P_{2n}(U_{2n} = k) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2}{n}} \exp\left(-\frac{k^2}{4n}\right) \left(1 + O_u\left(\frac{1}{\sqrt{n}}\right)\right).$$

Cette estimation uniforme en k permet d'écrire

$$P_{2n}(U_{2n} < \alpha\sqrt{2n}) \\ = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2n}} \sum_{0 \leq k < \alpha\sqrt{2n}} \exp\left(-\frac{1}{2} \left(\frac{k}{\sqrt{2n}}\right)^2\right) \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right).$$

On reconnaît dans cette expression une somme de RIEMANN (cf. Lemme 7.4) et on conclut que

$$\lim_{n \rightarrow \infty} P_{2n}(U_{2n} < \alpha\sqrt{2n}) = \sqrt{\frac{2}{\pi}} \int_0^\alpha \exp\left(-\frac{1}{2}x^2\right) dx.$$

On remarque finalement que, pour tout n , $U_{2n+1} = U_{2n}$, car M_{2n+1} n'est jamais nul. Donc

$$P_{2n+1}(U_{2n+1} < \alpha\sqrt{2n+1}) = P_{2n}(U_{2n} < \alpha\sqrt{2n+1}),$$

et on vérifie sans peine, grâce à ce qui précède, que

$$\lim_{n \rightarrow \infty} \left(P_{2n}(U_{2n} < \alpha\sqrt{2n+1}) - P_{2n}(U_{2n} < \alpha\sqrt{2n})\right) = 0.$$

Nous avons démontré que

$$\lim_{n \rightarrow \infty} P_n(U_n < \alpha\sqrt{n}) = \sqrt{\frac{2}{\pi}} \int_0^\alpha e^{-x^2/2} dx.$$

— QED (Loi des retours en zéro)

10. La loi forte des grands nombres

Nous allons à présent décrire une nouvelle classe de propriétés asymptotiques du jeu de pile ou face ; ces propriétés concernent les suites infinies de lancers d'une pièce et mettent en jeu la notion d'événement presque sûr.

La loi faible des grands nombres nous dit que, si l'on joue à pile ou face assez longtemps, la probabilité que la fréquence de succès observée soit proche de la probabilité théorique de succès est grande. La loi forte des grands nombres nous dira que si l'on joue à pile ou face sans s'arrêter, alors la fréquence de succès observée S_n/n tend vers la probabilité de succès p quand le nombre de lancers n tend vers l'infini. Cet énoncé a besoin d'être précisé. En effet, une suite infinie de « piles » ou « faces » telle que la fréquence d'apparition des « piles » ne converge pas (ou converge vers une valeur arbitraire) n'est théoriquement pas exclue. Mais elle l'est du point de vue du Calcul des Probabilités. Ici intervient la notion d'événement presque sûr. La loi forte des grands nombres nous dira que, presque sûrement, la suite (S_n/n) converge vers p . Ce résultat fondamental est dû à Émile BOREL^{10.1}. En suivant encore les idées de BOREL, nous illustrerons la loi forte des grands nombres par la notion de nombre normal.

loi forte
des grands
nombres

Ce chapitre est organisé de la façon suivante : dans le premier paragraphe, la notion d'événement presque sûr est définie ; la loi forte des grands nombres est établie dans le second paragraphe, comme conséquence de l'estimation des grands écarts ; les troisième et quatrième paragraphes, qui traitent des nombres normaux, sont essentiellement indépendants du cinquième où sont présentés les lemmes de BOREL-CANTELLI et d'autres approches de la loi forte des grands nombres.

La loi du logarithme itéré, qui fait l'objet du chapitre suivant, donne une estimation de la vitesse de convergence dans la loi forte des grands nombres.

10.1. Événements presque sûrs, événements indépendants. —

L'expérience que nous modélisons à présent est une suite infinie d'expériences élémentaires identiques, indépendantes et à deux issues, appelées succès et échec. Nous noterons p la probabilité de succès ; c'est un paramètre fixé entre 0 et 1. Nous noterons $q = 1 - p$ la probabilité d'échec. Considérons

^{10.1} E. Borel, *Sur les probabilités dénombrables et leurs applications arithmétiques*, Rendiconti del Circolo Math. di Palermo, vol. 26, p. 247-271, 1909.

l'ensemble Ω des suites infinies de 0 et de 1 :

$$\Omega = \{ \omega = (\omega_n)_{n \geq 1} : \text{pour tout } n \geq 1, \omega_n = 0 \text{ ou } 1 \} .$$

Chaque ω représente une issue possible de l'expérience : la n -ième coordonnée ω_n est égale à 1 si la n -ième expérience élémentaire est un succès. On notera $S_n(\omega) = \omega_1 + \omega_2 + \dots + \omega_n$ le nombre de succès observés après n expériences.

événement
de type fini

On appellera *événement de type fini* tout sous-ensemble de Ω défini par une condition ne portant que sur un nombre fini de coordonnées, c'est-à-dire un événement dont on peut contrôler la réalisation au bout d'un nombre fini (et fixé) d'expériences élémentaires. Avec les notations utilisées dans les chapitres précédents, on a les définitions formelles suivantes.

Un sous-ensemble A de Ω est un *événement de type fini* si il existe un nombre entier $n = n(A) \geq 1$ et un sous-ensemble A' de Ω_n tels que

$$A = \{ \omega \in \Omega : \omega^{(n)} \in A' \} \text{ où } \omega^{(n)} := (\omega_1, \omega_2, \dots, \omega_n) .$$

probabilité
d'un
événement
de type fini

On appelle *probabilité de l'événement de type fini* A le nombre

$$P(A) := P_{n(A)}(A') .$$

Ainsi, on a

$$P(A) = \sum_{\omega^{(n)} \in A'} p^{S_n(\omega)} q^{n-S_n(\omega)} .$$

Dans la définition d'un événement de type fini A , le nombre $n(A)$ n'est pas déterminé de façon unique par la donnée de A (si un nombre convient, tout nombre plus grand convient encore) ; pour que la définition de la probabilité de A ait un sens, il est donc nécessaire que la valeur donnée par l'expression précédente ne dépende pas du choix de $n(A)$. La vérification de ce fait est aisée et laissée au lecteur.

algèbre
de Boole

clan

L'ensemble \mathcal{E} des événements de type fini contient Ω et \emptyset , il est stable par passage au complémentaire et par réunion ou intersection finie. Un tel ensemble de sous-ensembles de Ω est appelé une *algèbre de BOOLE* ou *clan* de sous-ensembles de Ω . La probabilité P est une application de \mathcal{E} dans l'intervalle $[0,1]$; elle vérifie $P(\Omega) = 1$ et, pour tous A, B dans \mathcal{E} , si $A \cap B = \emptyset$, alors $P(A \cup B) = P(A) + P(B)$.

Nous pouvons à présent donner la principale définition de ce paragraphe.

événement
négligeable

Un sous-ensemble N de Ω est un *événement négligeable* si, pour tout $\epsilon > 0$, il existe une famille dénombrable $\{A_k : k \geq 1\}$ d'événements de

type fini telle que

$$N \subset \bigcup_{k \geq 1} A_k \quad \text{et} \quad \sum_{k \geq 1} P(A_k) < \epsilon.$$

Un sous-ensemble de Ω est un *événement presque sûr* si son complémentaire est négligeable. Si A est un événement presque sûr, on dira «*presque sûrement, ω appartient à A* ». événement presque sûr

Nota bene : nous ne considérerons dans cet ouvrage que des événements de type fini, des événements presque sûrs ou des événements négligeables ; nous omettons, en général le qualificatif «*de type fini*».

Nous utiliserons la propriété d'*invariance par décalage* de la probabilité P : si A est un événement et si k est un entier positif, alors l'événement $\{\omega \in \Omega : (\omega_k, \omega_{k+1}, \omega_{k+2}, \dots) \in A\}$ a même probabilité que l'événement A .

PROPOSITION 10.1. — *Tout sous-ensemble de Ω contenu dans un événement négligeable est un événement négligeable. Toute réunion dénombrable d'événements négligeables est négligeable. Si p n'est égal ni à 0, ni à 1, alors tout sous-ensemble dénombrable de Ω est négligeable.*

Démonstration. — La première affirmation est évidente à partir de la définition.

Soit $\{N_n : n \geq 1\}$ une famille dénombrable d'événements négligeables. Notons $N = \bigcup_{n \geq 1} N_n$ et fixons $\epsilon > 0$. Pour chaque n , il existe une famille dénombrable $\{A_{n,k} : k \geq 1\}$ d'événements de type fini telle que

$$N_n \subset \bigcup_{k \geq 1} A_{n,k} \quad \text{et} \quad \sum_{k \geq 1} P(A_{n,k}) < \epsilon 2^{-n}.$$

La famille d'événements $\{A_{n,k} : n, k \geq 1\}$ recouvre N et la somme des probabilités de ces événements est inférieure à $\sum_n \epsilon 2^{-n} = \epsilon$. Cela prouve que N est un événement négligeable.

On suppose que p est différent de 0 et 1. Tout singleton de Ω est un événement négligeable. En effet, si ω est un élément fixé de Ω , alors, pour tout $n \geq 1$, le singleton $\{\omega\}$ est contenu dans l'événement $\{\omega' \in \Omega : \omega'^{(n)} = \omega^{(n)}\}$ dont la probabilité est majorée par $(\max(p, q))^n$, qui peut être rendu arbitrairement petit en choisissant n assez grand. Enfin, tout ensemble dénombrable est réunion dénombrable de singletons.

— QED (Proposition 10.1)

Soit $(A_i)_{i \in I}$ une famille d'événements. On dira que les événements A_i sont *indépendants* si, pour toute famille finie d'indices deux à deux distincts $i_1, i_2, \dots, i_n \in I$, on a

$$P \left(\bigcap_{k=1}^n A_{i_k} \right) = \prod_{k=1}^n P(A_{i_k}) .$$

Le lecteur vérifiera, à titre d'exercice, la propriété élémentaire suivante : si les événements A_i sont indépendants, alors les événements complémentaires A_i^c sont aussi indépendants.

PROPOSITION 10.2. — *Des événements qui dépendent d'ensembles de coordonnées disjoints sont indépendants.*

Cet énoncé peut être récrit de la façon suivante. Soit $(A_i)_{i \in I}$ une famille d'événements. On suppose qu'à chaque $i \in I$ sont associés un sous-ensemble fini E_i de \mathbb{N} et un sous-ensemble A'_i de $\{0, 1\}^{E_i}$ tels que

$$E_i \cap E_j = \emptyset \quad \text{si} \quad i \neq j$$

et

$$A_i = \{ \omega \in \Omega : (\omega_n)_{n \in E_i} \in A'_i \} .$$

Alors les événements $(A_i)_{i \in I}$ sont indépendants.

La vérification de cette proposition est aisée à partir de la définition de la probabilité. Nous laissons le lecteur s'en convaincre par lui-même.

Donnons à présent quelques exemples d'événements négligeables. On suppose dans ces exemples que le paramètre p n'est égal ni à 0, ni à 1.

EXEMPLE. — L'ensemble des suites périodiques à partir d'un certain rang est négligeable. (Cet exemple est un cas particulier du suivant.) —◇

EXEMPLE. — Soit b un mot de l'alphabet $\{0, 1\}$, c'est-à-dire une suite finie formée de 0 et de 1. L'ensemble des suites infinies de 0 et de 1 dans lesquelles le mot b n'apparaît pas est négligeable.

Puisque l'ensemble des mots est dénombrable, on en déduit, en utilisant la proposition 10.1, que *presque sûrement, dans la suite ω tous les mots possibles apparaissent.*

Vérifions que si b est un mot fixé, de longueur $j > 0$, alors l'ensemble des ω dans lesquels ce mot n'apparaît pas est négligeable. Pour chaque $m \geq 0$, on note A_m l'événement formé des ω tels que

$$(\omega_{mj+1}, \omega_{mj+2}, \dots, \omega_{(m+1)j}) \neq b .$$

On a $P(A_0) < 1$ et la propriété d'invariance par décalage entraîne que tous les événements A_m ont même probabilité. En utilisant la proposition 10.2 on constate de plus que les événements A_m sont indépendants. On a donc $P(\bigcap_{k \leq m} A_k) = (P(A_0))^{m+1}$. L'ensemble des suites ω telles que, pour tout $n \geq 0$,

$$(\omega_{n+1}, \omega_{n+2}, \dots, \omega_{n+j}) \neq b$$

est contenu dans $\bigcap_{k \leq m} A_k$, dont la probabilité peut être choisie arbitrairement petite, en prenant m assez grand. Cet ensemble est donc négligeable. —◇

EXEMPLE. — En utilisant le théorème limite central, on peut démontrer que si $(a_n)_{n \geq 1}$ est une suite non majorée de nombres réels positifs, alors, presque sûrement,

$$\limsup_{n \rightarrow \infty} a_n \sqrt{n} \left| \frac{S_n}{n} - p \right| = +\infty .$$

Un résultat beaucoup plus précis sera démontré dans le chapitre 11, mais nous donnons ci-dessous quelques indications pour une preuve de cette affirmation, qui constituera un bon entraînement pour le lecteur.

Soient m un nombre > 0 et (n_k) une suite strictement croissante d'entiers. L'ensemble des ω tels que $\limsup_n a_n \sqrt{n} \left| \frac{S_n(\omega)}{n} - p \right| < m$ est contenu dans la réunion sur k des événements de type fini $\left(a_{n_k} \sqrt{n_k} \left| \frac{S_{n_k}}{n_k} - p \right| < m \right)$. Or, d'après le théorème limite central, ces événements sont de probabilité arbitrairement petite pourvu que n_k et a_{n_k} soient assez grands. Pour tout $\epsilon > 0$, on peut choisir la suite (n_k) de sorte que $\sum_k P \left(a_{n_k} \sqrt{n_k} \left| \frac{S_{n_k}}{n_k} - p \right| < m \right) < \epsilon$. On en déduit que, presque sûrement, $\limsup_n a_n \sqrt{n} \left| \frac{S_n}{n} - p \right| \geq m$. On peut conclure en utilisant la proposition 10.1. —◇

10.2. La loi forte des grands nombres de BOREL. —

THÉORÈME 10.3 (Loi forte des grands nombres de BOREL). —

Presque sûrement, on a

$$\lim_{n \rightarrow +\infty} \frac{1}{n} S_n(\omega) = p . \tag{10.1}$$

Démonstration. — Notons $R_n = \frac{1}{n}S_n - p$. La suite $(R_n(\omega))_{n \geq 1}$ ne tend pas vers zéro si et seulement si il existe $m \geq 1$ tel que, pour tout $n \geq 1$, il existe $k \geq n$ tel que $|R_k(\omega)| > \frac{1}{m}$. Autrement dit, l'ensemble des ω ne vérifiant pas (10.1) est

$$\bigcup_{m \geq 1} \bigcap_{n \geq 1} \bigcup_{k \geq n} \left\{ \omega \in \Omega : |R_k(\omega)| > \frac{1}{m} \right\}.$$

Nous voulons démontrer que cet ensemble est un événement négligeable. Grâce à la Proposition 10.1, il nous suffit de démontrer que, pour chaque $m \geq 1$,

$$N_m := \bigcap_{n \geq 1} \bigcup_{k \geq n} \left\{ \omega \in \Omega : |R_k(\omega)| > \frac{1}{m} \right\}$$

est un événement négligeable.

Pour chaque $k \geq 1$ notons $A_{m,k} := \left\{ \omega \in \Omega : |R_k(\omega)| > \frac{1}{m} \right\}$. D'après l'estimation des grands écarts, il existe une constante $c = c(p, m) > 0$ telle que $P(A_{m,k}) \leq e^{-ck}$. La série $\sum_{k \geq 1} e^{-ck}$ est convergente. En conséquence, pour tout $\epsilon > 0$, il existe $n \geq 1$ tel que $\sum_{k \geq n} P(A_{m,k}) < \epsilon$. Or $N_m \subset \bigcup_{k \geq n} A_{m,k}$. Cela prouve que N_m est un événement négligeable.

— QED (Loi forte des grands nombres de BOREL)

Après avoir introduit la notion de variable aléatoire, nous énoncerons (Théorème 10.5) une généralisation de la loi forte des grands nombres de BOREL. Montrons toutefois que celle-ci possède déjà un large champ d'application. Cette loi des grands nombres nous dit que dans une suite infinie de lancers d'une pièce équilibrée, les fréquences asymptotiques de «pile» et de «face» sont égales à $1/2$. Mais elle nous dit beaucoup d'autres choses : si l'on compte les apparitions de n'importe quel événement dans la suite, on observe une fréquence asymptotique égale à sa probabilité. Cette affirmation est rendue précise par le corollaire 10.5 ci-dessous.

PROPOSITION 10.4. — *Soit $(A_n)_{n \geq 1}$ une suite d'événements indépendants équiprobables de probabilité notée $P(A)$. Presque sûrement, la fréquence asymptotique des réalisations de ces événements est égale à leur probabilité. Autrement dit, presque sûrement,*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \text{card} \{k : 1 \leq k \leq n \text{ et } \omega \in A_k\} = P(A).$$

Cette proposition contient le théorème 10.3 (il suffit de poser $A_n := (\omega_n = 1)$ pour s'en assurer). En fait sa démonstration va nous montrer qu'elle n'est qu'une réécriture de ce théorème.

Démonstration. — À chaque ω , on associe une suite $\rho = (\rho_n)_{n \geq 1}$ de 0 et de 1 en posant $\rho_n = 1$ si $\omega \in A_n$, et $\rho_n = 0$ sinon. Il s'agit alors de démontrer que,

$$\text{presque sûrement,} \quad \lim \frac{1}{n} \sum_{k=1}^n \rho_k = P(\rho_i = 1) = P(A). \quad (10.2)$$

Pour chaque n et chaque $(\epsilon_1, \epsilon_2, \dots, \epsilon_n) \in \{0, 1\}^n$, si on note $s := \sum_k \epsilon_k$, on a

$$P(\rho_k = \epsilon_k, 1 \leq k \leq n) = P(A)^s (1 - P(A))^{n-s}.$$

Les estimations et les calculs utilisés pour démontrer que, presque sûrement, $\lim \frac{1}{n} \sum_{k=1}^n \omega_k = p$ s'appliquent à l'identique pour démontrer 10.2.

—— QED (Proposition 10.4)

COROLLAIRE 10.5. — *Soit A un événement de type fini. Pour chaque entier $n \geq 1$ et chaque $\omega \in \Omega$, notons $S(A, n, \omega)$ le nombre d'entiers k entre 1 et n tels que $(\omega_k, \omega_{k+1}, \omega_{k+2}, \dots) \in A$.*

$$\text{Presque sûrement,} \quad \lim_{n \rightarrow \infty} \frac{1}{n} S(A, n, \omega) = P(A).$$

Démonstration. — Il existe un nombre entier $m > 0$ tel que l'événement A ne dépende que des composantes d'indice inférieur ou égal à m . Autrement dit, il existe $A' \subset \Omega_m$ tel que

$$A = \{\omega : (\omega_1, \omega_2, \dots, \omega_m) \in A'\}.$$

Pour chaque nombre entier j entre 1 et m on considère la suite $(A_{j,n})_{n \geq 0}$ d'événements définis par

$$\begin{aligned} A_{j,n} &:= \{\omega : (\omega_{j+nm}, \omega_{j+nm+1}, \omega_{j+nm+2}, \dots) \in A\} \\ &= \{\omega : (\omega_{j+nm}, \omega_{j+nm+1}, \dots, \omega_{j+(n+1)m-1}) \in A'\}. \end{aligned}$$

Quand j est fixé et n varie, les événements $A_{j,n}$ sont indépendants et ont tous même probabilité égale à $P(A)$.

Notons $S(A, j, n, \omega)$ le nombre d'entiers k entre 0 et $n - 1$ tels que $\omega \in A_{j,k}$. En utilisant la proposition 10.4, on obtient :

$$\text{presque sûrement,} \quad \lim_{n \rightarrow \infty} \frac{1}{n} S(A, j, n, \omega) = P(A).$$

Mais on a

$$\frac{1}{nm} S(A, nm, \omega) = \frac{1}{m} \sum_{j=1}^m \frac{1}{n} S(A, j, n, \omega)$$

et la conclusion est alors directe grâce à l'encadrement :

$$\frac{1}{(n+1)m} S(A, nm, \omega) \leq \frac{1}{nm+k} S(A, nm+k, \omega) \leq \frac{1}{nm} S(A, (n+1)m, \omega)$$

pour tout k entre 0 et m .

— QED (Corollaire 10.5)

Si $b = (b_1, b_2, \dots, b_j)$ est un mot de l'alphabet $\{0, 1\}$, et si on note $s := \sum_{i=1}^j b_i$, nous appelons $p^s q^{j-s}$ la *probabilité du mot* b .

Le résultat suivant est un cas particulier du corollaire 10.5.

COROLLAIRE 10.6. — *Presque sûrement, dans la suite ω , tout mot b apparaît avec une fréquence asymptotique égale à sa probabilité.*

10.3. Extension au cas des suites aléatoires pouvant prendre plus de deux valeurs. — En vue de présenter des illustrations et applications de la loi des grands nombres, nous allons légèrement élargir le cadre de notre étude, en considérant le modèle d'une suite d'expériences aléatoires élémentaires ayant plusieurs issues possibles. Cette extension est purement formelle et n'apporte aucune difficulté supplémentaire, c'est pourquoi nous la présentons rapidement. Considérons donc une expérience aléatoire ayant d issues, numérotées de 1 à d . À chaque issue i est associée un nombre p_i qui représente sa probabilité. Les p_i sont des nombres positifs et leur somme est égale à 1. L'espace probabilisé naturellement associé à la réalisation d'une expérience est $\Omega_1 = \{1, 2, \dots, d\}$ et la probabilité d'un sous-ensemble A de Ω_1 est $P(A) = \sum_{i \in A} p_i$. L'espace des réalisations d'une suite finie de n expériences élémentaires indépendantes est l'ensemble produit $\Omega_n = \Omega_1^n$ qui sera muni de la probabilité P_n caractérisée par

$$P_n(\omega^{(n)}) = \prod_{k=1}^n p_{\omega_k} = \prod_{i=1}^d p_i^{\text{card}\{k : 1 \leq k \leq n, \omega_k=i\}},$$

si $\omega^{(n)} = (\omega_1, \omega_2, \dots, \omega_n) \in \Omega_n$.

L'espace des réalisations d'une suite infinie d'expériences élémentaires indépendantes est l'ensemble $\Omega = \Omega_1^{\mathbb{N}^*}$ des suites d'éléments de Ω_1 . On peut alors définir comme dans le cas $d = 2$ les notions d'événement de type fini,

de probabilité d'un événement de type fini, d'événements indépendants, et d'événement négligeable ou presque sûr.

Si l'on observe la réalisation d'une suite d'événements, on est ramené au cas des suites d'expériences à deux issues. La loi forte des grands nombres nous dit en particulier que, presque sûrement, pour chaque i ,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \text{card} \{k : 1 \leq k \leq n \text{ et } \omega_k = i\} = p_i .$$

Et le corollaire 10.5 peut être repris mot pour mot.

10.4. Nombres normaux. — Nous allons à présent relier les propriétés du jeu de pile ou face aux propriétés statistiques des développements décimaux de nombres réels. Nous ne considérons pas seulement le développement décimal mais le développement dans une base arbitraire, déterminée par un nombre entier strictement positif b . Nous commençons par rappeler cette notion de *développement b -adique* puis nous définirons la notion d'ensemble négligeable au sens de LEBESGUE. développe-
ment
 b -adique

Soit b un nombre entier strictement positif. Tout nombre réel x s'écrit de façon unique

$$x = x_0 + \sum_{i=1}^{+\infty} \frac{x_i}{b^i}$$

où $x_0 \in \mathbb{N}$, $x_i \in \{0, 1, 2, \dots, b-1\}$ pour $i \geq 1$ et la suite (x_i) n'est pas constante égale à $b-1$ à partir d'un certain rang^{10.2}.

Si I est un intervalle réel, nous noterons $|I|$ sa longueur. Un sous-ensemble E de la droite réelle \mathbb{R} est *négligeable* (ou *négligeable au sens de LEBESGUE*) si, pour tout $\epsilon > 0$, il existe une famille dénombrable d'intervalles $(I_k)_{k \geq 1}$ telle que ensemble
négligeable
au sens de
Lebesgue

$$E \subset \bigcup_{k \geq 1} I_k \quad \text{et} \quad \sum_{k \geq 1} |I_k| < \epsilon .$$

Si E est négligeable, on dit que presque tout nombre réel appartient au complémentaire de E .

Notons Ω l'ensemble $\{0, 1, \dots, b-1\}^{\mathbb{N}^*}$ des suites d'entiers entre 0 et $b-1$ et $\omega = (\omega_n)_{n \geq 1}$ les éléments de Ω . Notons Ω' le sous-ensemble

^{10.2} La notion de *développement b -adique* est exposée dans de nombreux ouvrages, par exemple au début du troisième tome du *Cours de mathématiques spéciales* de MM. Ramis, Deschamps et Odoux, éd. Masson, 1988.

formé des suites qui ne sont pas constantes égales à $b - 1$ à partir d'un certain rang. On munit l'ensemble fini $\{0, 1, \dots, b - 1\}$ de la probabilité uniforme $p_0 = p_1 = \dots = p_{b-1} = \frac{1}{b}$ et, suivant la construction résumée dans le paragraphe 10.3., cela définit une notion de sous-ensemble négligeable de l'ensemble Ω . Le complémentaire de Ω' dans Ω est dénombrable, donc il est négligeable. Considérons l'application Φ de Ω' dans l'intervalle $[0, 1[$ définie par

$$\Phi(\omega) = \sum_{i=1}^{+\infty} \frac{\omega_i}{b^i}.$$

PROPOSITION 10.7. — *L'application Φ est bijective et un sous-ensemble A de Ω' est négligeable dans Ω si et seulement si son image $\Phi(A)$ est négligeable au sens de LEBESGUE.*

mesure

Pour définir la mesure $|C|$ d'un ensemble C qui est réunion finie d'intervalles, nous utiliserons le lemme élémentaire suivant.

LEMME 10.8. — *Si C est une réunion finie d'intervalles réels, alors C peut s'écrire comme une réunion finie $C = \cup_k I_k$ où les I_k sont des intervalles deux à deux disjoints. Cette écriture n'est pas unique mais la somme $\sum_k |I_k|$ des longueurs des intervalles ne dépend que de C .*

Avec les notations du lemme, on pose $|C| = \sum_k |I_k|$.

Démonstration (Proposition 10.7) — Le fait que l'application Φ soit bijective est connu. Soient $n > 0$ et $(a_1, a_2, \dots, a_n) \in \Omega_1^n$. Si l'on note $\alpha := \sum_{i=1}^n \frac{a_i}{b^i}$, alors

$$\Phi\left(\{\omega \in \Omega' : \omega_i = a_i, 1 \leq i \leq n\}\right) = \left[\alpha, \alpha + \frac{1}{b^n}\right].$$

L'image par Φ d'un événement de type fini dans Ω est donc une réunion finie d'intervalles b -adiques, c'est-à-dire d'intervalles dont les bornes sont rationnelles du type $\frac{a}{b^j}$, avec a et j entiers. De plus, on remarque que

$$P\left(\{\omega \in \Omega' : \omega_i = a_i, 1 \leq i \leq n\}\right) = \frac{1}{b^n} = \left| \left[\alpha, \alpha + \frac{1}{b^n}\right] \right|.$$

On en déduit que, pour tout événement de type fini B , on a $P(B) = |\Phi(B)|$. Il est à présent clair que l'image par Φ d'un événement négligeable est négligeable au sens de LEBESGUE.

Pour démontrer la réciproque, commençons par remarquer que tout intervalle peut être approché par des intervalles b -adiques : pour tout intervalle réel I , pour tout $\epsilon > 0$ il existe un intervalle b -adique fermé à gauche et ouvert à droite I' tel que $I \subset I'$ et $|I'| < |I| + \epsilon$. Soit E un sous-ensemble de $[0, 1[$ négligeable au sens de LEBESGUE. En utilisant l'affirmation précédente on peut vérifier que pour tout $\epsilon > 0$, il existe une famille dénombrable d'intervalles b -adiques fermés à gauche et ouverts à droite (I'_k) telle que $E \subset \cup I'_k$ et $\sum |I'_k| < \epsilon$. Posons $A_k = \Phi^{-1}(I'_k)$. Pour chaque k , l'ensemble A_k est un événement de type fini et $P(A_k) = |I'_k|$. (La vérification de ce point est aisée ; il suffit d'écrire l'intervalle I'_k comme réunion finie d'intervalles deux à deux disjoints de la forme $[\alpha, \alpha + b^{-j}[$, avec $b^j \alpha \in \mathbb{N}$, et d'utiliser les considérations qui précèdent.) On constate que $\Phi^{-1}(E) \subset \cup A_k$ et $\sum P(A_k) < \epsilon$. On en déduit que $\Phi^{-1}(E)$ est négligeable dans Ω . — QED (Proposition 10.7)

Suivant BOREL, nous dirons qu'un nombre réel est *normal en base b* si, dans son développement en base b , tout bloc de chiffres apparaît avec une fréquence asymptotique égale à $b^{-\ell}$ où ℓ est la longueur de ce bloc ; autrement dit, $x = x_0 + \sum_{i=1}^{+\infty} \frac{x_i}{b^i}$ est normal en base b si et seulement si, pour tout $\ell > 0$ et tout $a = (a_1, a_2, \dots, a_\ell) \in \{0, 1, \dots, b-1\}^\ell$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{card}\{i : 1 \leq i \leq n, (x_i, x_{i+1}, \dots, x_{i+\ell-1}) = a\} = \frac{1}{b^\ell}.$$

Un nombre réel est dit *normal* s'il est normal en toute base.

Du corollaire 10.5 et de la proposition 10.7, on déduit que presque tout nombre réel est normal en base b . De plus, de la même façon que nous l'avons vu dans Ω , on peut montrer qu'une réunion dénombrable de sous-ensembles négligeables de \mathbb{R} est encore négligeable (au sens de LEBESGUE). Cela permet d'affirmer que *presque tout nombre réel est normal en toute base*. Un paradoxe troublant est que l'on n'en connaît aucun exemple ! Les nombres rationnels ne conviennent certainement pas puisque, dans toute base, leur développement est périodique à partir d'un certain rang. En ce qui concerne les nombres irrationnels les plus populaires comme par exemple $\sqrt{2}$, $\ln 2$ ou π , on ignore s'ils sont normaux ! La solution de ces problèmes semble encore un défi insurmontable pour les mathématiques contemporaines.

10.5. *Les lemmes de BOREL-CANTELLI.* — Les travaux de BOREL pour établir la loi forte des grands nombres ont été prolongés par CANTELLI^{10.3} pour aboutir en particulier aux lemmes qui portent leurs noms et qui sont indispensables pour préciser la loi des grands nombres.

L'argument utilisé pour démontrer le théorème 10.3 peut être résumé ainsi : l'estimation des grands écarts implique la convergence d'une série de probabilités d'événements, et on en déduit qu'un certain événement est négligeable. Ce principe est d'un emploi constant en Calcul des Probabilités et nous allons en donner un énoncé général (Proposition 10.9) et des applications. Nous revenons au modèle mathématique décrit dans le paragraphe 10.1.

Si $(A_n)_{n \geq 1}$ est une suite de sous-ensembles de Ω , on note « A_n infiniment souvent » l'ensemble $\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$. Ainsi on a $\omega \in$ « A_n infiniment souvent » si et seulement si ω appartient à A_n pour une infinité de valeurs de l'indice n .

PROPOSITION 10.9 (Premier lemme de BOREL-CANTELLI). — *Soit $(A_n)_{n \geq 1}$ une suite d'événements. Si $\sum_{n \geq 1} P(A_n) < \infty$ alors « A_n infiniment souvent » est un événement négligeable.*

Démonstration. — L'ensemble « A_n infiniment souvent » est contenu, pour tout $n \geq 1$, dans la réunion d'événements $\bigcup_{k \geq n} A_k$ et l'hypothèse de convergence de série nous assure que la somme des probabilités des A_k , pour $k \geq n$, peut être rendue arbitrairement petite à condition de choisir n assez grand. Le résultat se déduit donc directement de la définition d'événement négligeable. — QED (Proposition 10.9)

Nous allons illustrer ce résultat par trois exemples, puis par l'énoncé d'une loi des grands nombres assez générale (Théorème 10.12).

EXEMPLE. — Soient $(k_n)_{n \geq 1}$ et $(\ell_n)_{n \geq 1}$ deux suites de nombres entiers positifs. Pour chaque $n \geq 1$, notons A_n l'événement « on observe ℓ_n succès consécutifs à partir de la k_n -ième expérience élémentaire », autrement dit

$$A_n = \{ \omega \in \Omega : \omega_{k_n} = \omega_{k_n+1} = \dots = \omega_{k_n+\ell_n-1} = 1 \} .$$

Si $\sum_{n \geq 1} p^{\ell_n} < +\infty$, alors, presque sûrement, il n'y a qu'un nombre fini d'événements A_n qui se réalisent. —◇

^{10.3} F.P. Cantelli, *Sulla probabilità come limite della frequenza*, Rendiconti d. r. Acad. d. Lincei, vol. 26, p. 39-45, 1917.

EXEMPLE. — Si trois parties de pile ou face se déroulent en parallèle et indépendamment les unes des autres, alors, presque sûrement, les parties ne seront simultanément équilibrées qu'un nombre fini de fois. (On dit qu'une partie de pile ou face est équilibrée à l'instant n si au bout de n lancers, il est apparu autant de « piles » que de « faces ».) Cette affirmation est une conséquence directe du premier lemme de BOREL-CANTELLI et sa démonstration est un bon exercice ; cet exemple est à relier à l'étude de la récurrence des marches aléatoires, qui est développée dans le chapitre 12. —◇

EXEMPLE. — On utilise ici les notions introduites dans la partie 10.4. Voici une façon naturelle d'associer à un nombre réel une suite de nombres entiers que l'on lit dans son développement décimal. Si x est un nombre réel, notons $x = x_0 + \sum_{i=1}^{+\infty} x_i 10^{-i}$ son développement décimal et considérons, pour chaque entier positif n , le nombre entier $L_n(x) = \sum_{i=1}^n x_i 10^i$.

Soit (ϵ_n) une suite réelle telle que $\sum_n \epsilon_n < +\infty$. Pour presque tout nombre réel x (c'est-à-dire pour tout nombre réel en dehors d'un ensemble négligeable au sens de LEBESGUE), il existe $n(x) \in \mathbb{N}$ tel que, pour tout $n \geq n(x)$, on ait $L_n(x) \geq \epsilon_n 10^{n+1}$. Voici une brève indication pour justifier cette affirmation : dans l'espace Ω des développements décimaux, l'événement $(L_n \leq \epsilon_n 10^{n+1})$ a une probabilité de l'ordre de ϵ_n . —◇

La démonstration de la loi forte des grands nombres (Théorème 10.3) a été basée sur le fait que, pour chaque $m > 0$, l'ensemble « $|\frac{1}{n} S_n - p| > \frac{1}{m}$ infiniment souvent » est un événement négligeable. Ce principe est généralisé dans la prochaine proposition. Mais avant de l'énoncer nous devons définir les notions de variable aléatoire, d'indépendance et d'espérance. Ces notions sont des prolongements naturels de celles introduites dans les chapitres 2 et 3, ainsi que dans le paragraphe 10.1. Nous considérons toujours l'espace Ω des suites de 0 et de 1, et le paramètre p , représentant la probabilité du 1, est fixé.

Une *variable aléatoire de type fini* est une fonction réelle définie sur Ω et ne dépendant que d'un nombre fini de coordonnées. Autrement dit une application X de Ω dans \mathbb{R} est une variable aléatoire de type fini si elle ne prend qu'un nombre fini de valeurs et si, pour chaque $x \in \mathbb{R}$, l'ensemble

$$(X = x) := \{\omega \in \Omega : X(\omega) = x\}$$

est un événement de type fini. Toute variable aléatoire de type fini s'écrit comme une combinaison linéaire finie de fonctions indicatrices d'événements

ments de type fini.

Nota bene : nous ne considérons dans cet ouvrage que des variables aléatoires de type fini ; nous parlerons simplement de « variables aléatoires ».

Si X et Y sont deux variables aléatoires, alors les applications $X + Y$, XY , $\max(X, Y)$, $\min(X, Y)$ sont encore des variables aléatoires. Si f est une application quelconque de \mathbb{R} dans \mathbb{R} et si X est une variable aléatoire, alors $f(X) := f \circ X$ est encore une variable aléatoire. Pour chaque $n \geq 1$, S_n est une variable aléatoire.

Remarquons qu'une application X de Ω dans \mathbb{R} est une variable aléatoire (de type fini) si et seulement si il existe un nombre entier $n = n(X)$ et une variable aléatoire X' définie sur l'espace fini Ω_n tels que, pour tout $\omega \in \Omega$, on ait

$$X(\omega) = X'(\omega^{(n)}) \quad \text{où } \omega^{(n)} := (\omega_1, \omega_2, \dots, \omega_n). \quad (10.3)$$

Bien sûr, cet entier n n'est pas unique et chaque entier plus grand convient encore.

Une famille de variables aléatoires $(X_i)_{i \in I}$ est appelée *famille de variables aléatoires indépendantes* si, pour tout choix de nombres réels x_i , $i \in I$, les événements $(X_i = x_i)$, $i \in I$, sont indépendants. Si $(X_i)_{i \in I}$ est une famille finie de variables aléatoires, on peut, suivant (10.3), choisir un entier $n = n(X_i)$ commun à toutes ces variables et on remarque que les $(X_i)_{i \in I}$ sont indépendantes si et seulement si les variables aléatoires $(X'_i)_{i \in I}$, définies sur l'espace probabilisé fini (Ω_n, P_n) , sont indépendantes. De la proposition 3.1, on déduit alors le fait suivant : si $(X_i)_{i \in I}$ est une famille de variables aléatoires indépendantes, si J et K sont deux sous-ensembles finis disjoints de l'ensemble d'indices I et si Y , respectivement Z , est une fonction réelle de $(X_i)_{i \in J}$, respectivement de $(X_i)_{i \in K}$, alors Y et Z sont deux variables aléatoires indépendantes.

Si X est une variable aléatoire, notons $X(\Omega)$ l'ensemble des valeurs prises par X . Par définition l'*espérance* de X est le nombre

$$E[X] := \sum_{x \in X(\Omega)} xP(X = x).$$

En utilisant la présentation (10.3), on peut écrire $E[X] = E_n[X']$.

Les propriétés de l'espérance énoncées dans les chapitres 2 et 3 dans le cadre des espaces probabilisés finis s'étendent mot pour mot aux variables aléatoires (de type fini) définies sur Ω .

La proposition suivante généralise la méthode que nous avons utilisée pour établir la loi forte des grands nombres.

PROPOSITION 10.10. — Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires.

$$\text{Si pour tout } \epsilon > 0 \text{ on a } \sum_{n=1}^{+\infty} P(X_n > \epsilon) < \infty$$

$$\text{alors, presque sûrement, } \lim_{n \rightarrow +\infty} X_n = 0.$$

Démonstration. — La proposition 10.9, avec $A_n := (X_n > \epsilon)$, nous donne : pour tout $\epsilon > 0$, presque sûrement, il existe $n_0(\omega, \epsilon) \geq 0$ tel que, pour tout $n \geq n_0(\omega, \epsilon)$, $|X_n(\omega)| \leq \epsilon$. En considérant une réunion dénombrable d'événements négligeables, on en déduit que : presque sûrement, pour tout nombre entier $m > 0$, il existe $n_0(\omega, 1/m) \geq 0$ tel que, pour tout $n \geq n_0(\omega, 1/m)$, $|X_n(\omega)| \leq 1/m$. Cela signifie exactement que la suite (X_n) tend vers zéro presque sûrement. — QED (Proposition 10.10)

De cette proposition et de l'inégalité de Markov, on déduit directement le critère de convergence suivant, qui est très souvent utilisé.

COROLLAIRE 10.11. — Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires.

Si $\sum_{n=0}^{+\infty} E[|X_n|] < \infty$, alors la suite (X_n) tend vers zéro presque sûrement.

Pour illustrer ce critère de convergence presque sûre, nous allons donner une nouvelle preuve de la loi forte des grands nombres de BOREL, due à CANTELLI. Puis nous présenterons une généralisation des lois fortes des grands nombres précédemment énoncées.

Pour chaque nombre entier positif n , notons X_n la variable aléatoire définie par $X_n(\omega) = \omega_n - p$ si $\omega = (\omega_1, \omega_2, \omega_3, \dots)$. Chaque X_n est d'espérance nulle. L'ensemble de ces variables aléatoires forme une famille de variables aléatoires indépendantes (on dit dans ce cas que $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes).

suite de
variables
aléatoires
indépendantes

On a

$$\begin{aligned} E \left[\left(\frac{S_n}{n} - p \right)^4 \right] &= \frac{1}{n^4} E \left[\left(\sum_{i=1}^n X_i \right)^4 \right] \\ &= \frac{1}{n^4} \sum_{1 \leq i, j, k, l \leq n} E [X_i X_j X_k X_l] . \end{aligned}$$

Si i est distinct de j, k et ℓ , alors les variables aléatoires X_i et $X_j X_k X_\ell$ sont indépendantes et $\mathbf{E}(X_i X_j X_k X_\ell) = \mathbf{E}(X_i) \mathbf{E}(X_j X_k X_\ell) = 0$. Cela permet d'éliminer un grand nombre de termes du développement précédent et on obtient :

$$\mathbf{E} \left[\left(\frac{S_n}{n} - p \right)^4 \right] = \frac{1}{n^4} \left(\sum_{1 \leq i \leq n} \mathbf{E} [X_i^4] + 6 \sum_{1 \leq i < j \leq n} \mathbf{E} [X_i^2 X_j^2] \right).$$

On a $|X_i| \leq 1$, donc $\mathbf{E} [X_i^4] \leq 1$ et $\mathbf{E} [X_i^2 X_j^2] \leq 1$. D'où

$$\mathbf{E} \left[\left(\frac{S_n}{n} - p \right)^4 \right] \leq \frac{1}{n^4} (n + 3n(n-1)) = O \left(\frac{1}{n^2} \right),$$

qui est le terme général d'une série convergente. Du corollaire 10.11, on déduit alors que la suite $\left(\left(\frac{S_n}{n} - p \right)^4 \right)$ tend presque sûrement vers zéro, ce qui est la conclusion de la loi forte des grands nombres de BOREL.

La loi des grands nombres pour le jeu de pile ou face, que nous avons exprimée sous diverses formes, est l'archétype d'une large classe de théorèmes du Calcul des Probabilités. L'énoncé suivant est un exemple simple de loi des grands nombres applicable dans de nombreuses situations. Il contient les théorèmes 5.1 et 10.3.

THÉORÈME 10.12. — Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires deux à deux indépendantes et vérifiant : pour chaque n , $\mathbf{E}(X_n) = 0$ et $\sup_{n \geq 1} \mathbf{E} (X_n^2) < \infty$. En notant $R_n = \sum_{i=1}^n X_i$, on a

$$\text{pour tout } \epsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbf{P} \left(\left| \frac{R_n}{n} \right| \geq \epsilon \right) = 0, \quad (10.4)$$

et

$$\text{presque sûrement,} \quad \lim_{n \rightarrow +\infty} \frac{R_n}{n} = 0. \quad (10.5)$$

Remarquons que si les X_n forment une suite de variables aléatoires indépendantes, alors elles sont *a fortiori* deux à deux indépendantes. (Et nous avons déjà remarqué dans le chapitre 3 que la réciproque est fausse.)

Démonstration. — Posons $M := \sup_{n \geq 1} \mathbf{E} [X_n^2]$. Si $i \neq j$, alors $\mathbf{E}[X_i X_j] = \mathbf{E}[X_i] \mathbf{E}[X_j] = 0$. Par un développement direct, on en déduit

que

$$\mathbb{E} \left[\left(\frac{R_n}{n} \right)^2 \right] = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} [X_i^2] \leq \frac{M}{n}. \quad (10.6)$$

On en déduit (10.4) en utilisant l'inégalité de BIENAYMÉ-TCHEBYCHEV. Le résultat de convergence presque sûre est plus subtil. De (10.6), on déduit que

$$\sum_{n \geq 1} \mathbb{E} \left[\left(\frac{R_{n^2}}{n^2} \right)^2 \right] < \infty$$

et donc, d'après le corollaire 10.11,

$$\text{presque sûrement, } \lim_{n \rightarrow +\infty} \frac{R_{n^2}}{n^2} = 0.$$

Si n est un nombre entier positif, notons m la partie entière de \sqrt{n} . On a $m^2 \leq n < (m+1)^2$. D'une part

$$\text{presque sûrement, } \lim_{n \rightarrow +\infty} \frac{R_{m^2}}{n} = 0$$

d'autre part,

$$\begin{aligned} \mathbb{E} \left[\left(\frac{R_n}{n} - \frac{R_{m^2}}{n} \right)^2 \right] &= \frac{1}{n^2} \mathbb{E} \left[\left(\sum_{i=m^2+1}^n X_i \right)^2 \right] = \frac{1}{n^2} \sum_{i=m^2+1}^n \mathbb{E} [X_i^2] \\ &\leq \frac{n - m^2}{n^2} M = O(n^{-3/2}) \end{aligned}$$

est le terme général d'une série (d'indice n) convergente. Une nouvelle utilisation du corollaire 10.11 donne

$$\text{presque sûrement, } \lim_{n \rightarrow +\infty} \frac{R_n}{n} - \frac{R_{m^2}}{n} = 0.$$

— QED (Théorème 10.12)

second
lemme de
Borel-
Cantelli

PROPOSITION 10.13 (Second lemme de BOREL-CANTELLI). — Soit $(A_n)_{n \geq 1}$ une suite d'événements indépendants. Si $\sum_{n \geq 1} \mathbf{P}(A_n) = +\infty$, alors « A_n infiniment souvent » est un événement presque sûr.

Ce second lemme de BOREL-CANTELLI nous donne, sous l'hypothèse très forte de l'indépendance des événements, deux renseignements : d'une

part on a la réciproque du premier lemme de BOREL-CANTELLI, d'autre part l'ensemble « A_n infiniment souvent » est soit négligeable, soit presque sûr. C'est un cas particulier de la loi 0-1 de KOLMOGOROV dont l'exposé complet est réservé à des ouvrages plus avancés.

Démonstration (Second lemme de BOREL-CANTELLI) — Le complémentaire de l'ensemble « A_n infiniment souvent » est $\bigcup_{n \geq 1} \bigcap_{k \geq n} A_k^c$. Pour montrer que cet ensemble est un événement négligeable, il suffit de montrer que, pour tout n fixé, l'ensemble $\bigcap_{k \geq n} A_k^c$ est un événement négligeable. Pour chaque $m \geq n$, on note B_m l'événement $B_m := \bigcap_{n \leq k \leq m} A_k^c$. On a $\bigcap_{k \geq n} A_k^c \subset B_m$ et, grâce à la propriété d'indépendance

$$\begin{aligned} P(B_m) &= \prod_{k=n}^m P(A_k^c) = \prod_{k=n}^m (1 - P(A_k)) \\ &\leq \prod_{k=n}^m \exp(-P(A_k)) = \exp\left(-\sum_{k=n}^m P(A_k)\right), \end{aligned}$$

qui, par hypothèse, peut être rendu arbitrairement petit à condition de choisir m assez grand. Cela prouve que $\bigcap_{k \geq n} A_k^c$ est un événement négligeable.

— QED (Second lemme de BOREL-CANTELLI)

EXEMPLE. — Reprenons le cadre de l'exemple qui suit immédiatement l'énoncé du premier lemme de BOREL-CANTELLI (page 68). Si les suites (k_n) et (ℓ_n) vérifient, pour tout n , $k_n + \ell_n \leq k_{n+1}$, alors les événements A_n sont indépendants et, si de plus $\sum_{n \geq 1} p^{\ell_n} = +\infty$, alors l'événement « A_n infiniment souvent » est presque sûr. —◇

EXEMPLE. — Illustrons aussi le second lemme de BOREL-CANTELLI par une propriété des tirages aléatoires de nombres réels. Comment tirer un nombre réel au hasard, suivant une loi uniforme, entre 0 et 1 ? C'est une question délicate, dont l'étude complète requiert l'utilisation de la « mesure de LEBESGUE », mais nous pouvons en donner une première approche avec le jeu de pile ou face. Fixons le paramètre p égal à $1/2$. Pour associer un nombre réel U de l'intervalle $[0, 1]$ à une suite $\omega \in \Omega$ on procède comme dans le paragraphe précédent et on pose

$$U = \sum_{k=1}^{+\infty} \frac{\omega_k}{2^k}.$$

Comment tirer au hasard une suite de nombres réels indépendants ? Suivant une idée de STEINHAUS, on peut procéder de la façon suivante : on considère une famille infinie $(E_n)_{n \geq 1}$ de sous-ensembles infinis et deux à deux disjoints de

\mathbb{N}^* (par exemple E_1 est l'ensemble des nombres pairs, E_2 est l'ensemble des multiples impairs de 3, E_3 est l'ensemble des multiples de 5 non multiples de 2 ou 3, etc...). On note $E_n = \{a(n, 1), a(n, 2), a(n, 3), \dots\}$ et on pose

$$U_n = \sum_{k=1}^{+\infty} \frac{\omega_{a(n,k)}}{2^k}.$$

On obtient ainsi une suite (U_n) de *nombres réels aléatoires, indépendants et uniformément distribués entre 0 et 1*. Soit (ϵ_n) une suite de nombres réels positifs. Le premier lemme de BOREL-CANTELLI permet d'affirmer que si $\sum_n \epsilon_n < \infty$, alors, presque sûrement, pour tout n assez grand, $U_n > \epsilon_n$. Le second lemme de BOREL-CANTELLI permet d'affirmer que si $\sum_n \epsilon_n = +\infty$, alors, presque sûrement, pour une infinité d'entiers n , $U_n < \epsilon_n$.

(Dans cet exemple nous avons considéré des variables aléatoires U qui ne sont pas de type fini, nous évadant ainsi quelques instants hors du cadre théorique fixé dans cet opuscule.) —◇

11. La loi du logarithme itéré

11.1. Introduction. — Nous étudions dans ce chapitre la vitesse de convergence dans la loi forte des grands nombres. Le cadre reste celui du chapitre précédent, le modèle du jeu de pile ou face, avec une pièce éventuellement faussée. Pour chaque expérience élémentaire (un lancer de la pièce) la probabilité de succès est p , et le nombre de succès en n expériences élémentaires indépendantes est noté S_n . La loi forte des grands nombres de BOREL affirme que, presque sûrement, la suite (S_n/n) tend vers p quand n tend vers l'infini. Rappelons également que le théorème limite central affirme que, avec une forte probabilité, $S_n - np$ est de l'ordre de grandeur de \sqrt{n} quand n est grand.

La *loi du logarithme itéré* de KHINTCHINE est un renforcement considérable de la loi forte des grands nombres. Elle donne une estimation très précise de la hauteur des fluctuations presque sûres de la suite $(S_n - np)$. Elle stipule que, presque sûrement, pour tout $\epsilon > 0$,

il existe une infinité de n tels que $S_n - np > (1 - \epsilon)\sqrt{2p(1-p)n \ln \ln n}$,

et que,

pour tout n assez grand, $S_n - np < (1 + \epsilon)\sqrt{2p(1-p)n \ln \ln n}$.

On voit apparaître ici la fonction *logarithme itéré* $\ln \ln$, composée du logarithme népérien avec lui-même, qui est le prototype d'une fonction qui tend vers l'infini, mais extrêmement lentement, quand la variable tend vers l'infini (essayez sur votre calculette !).

Comme le théorème limite central, la loi du logarithme itéré illustre bien le fait, réellement fascinant, que le hasard le plus absolu obéit à des lois extrêmement précises.

Pour présenter ce résultat et les méthodes en jeu nous procédons par étapes, suivant un ordre historique et, nous l'espérons, pédagogique. Nous démontrerons successivement :

- L'estimation de HAUSDORFF^{11.1} : presque sûrement, pour tout $\epsilon > 0$,

$$S_n - np = O\left(n^{\epsilon+1/2}\right).$$

- L'estimation de HARDY et LITTLEWOOD^{11.2} : presque sûrement,

$$S_n - np = O\left(\sqrt{n \ln n}\right)$$

- La loi du logarithme itéré de KHINTCHINE^{11.3}.

Notre modèle mathématique, basé sur la donnée du couple (Ω, \mathbb{P}) reste celui décrit dans le paragraphe 10.1. Si $\omega = (\omega_n)_{n \geq 1} \in \Omega$, on pose

$$X_n(\omega) = \omega_n - p \quad \text{et} \quad R_n(\omega) = \sum_{k=1}^n X_k(\omega) = S_n(\omega) - np.$$

Ainsi, (X_n) est une suite de variables aléatoires de type fini, indépendantes, de même loi et d'espérance nulle. Avec les estimations des grands écarts et des écarts modérés, c'est tout ce que nous utiliserons dans les démonstrations à venir.

^{11.1} F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1913.

^{11.2} G.H. Hardy et J.E. Littlewood, *Some problems of Diophantine approximations*, *Acta Mathematica*, vol. 37, p. 155-339, 1914.

^{11.3} A. Khintchine, *Über einen Satz der Wahrscheinlichkeitsrechnung*, *Fundamenta Mathematicae*, vol. 6, p. 9-20, 1924.

11.2. L'estimation de HAUSDORFF. —

PROPOSITION 11.1. — *Presque sûrement, pour tout $\epsilon > 0$,*

$$S_n - np = O\left(n^{\epsilon+1/2}\right) \quad \text{quand } n \rightarrow +\infty.$$

Démonstration. — Nous allons prolonger la méthode de preuve de la loi forte des grands nombres due à CANTELLI, présentée page 71.

Si i_1, i_2, \dots, i_k sont des entiers positifs, on a $E[X_{i_1} X_{i_2} \dots X_{i_k}] \leq 1$, et si l'un des i_j est différent de tous les autres, alors $E[X_{i_1} X_{i_2} \dots X_{i_k}] = 0$.

Soit k un entier positif. On a

$$E[R_n^{2k}] = \sum_{1 \leq i_1, i_2, \dots, i_{2k} \leq n} E[X_{i_1} X_{i_2} \dots X_{i_{2k}}] \leq N(k, n)$$

où $N(k, n)$ désigne le nombre d'applications de l'ensemble $\{1, 2, \dots, 2k\}$ dans l'ensemble $\{1, 2, \dots, n\}$ telles que chaque valeur prise une fois est prise au moins deux fois. Notons $M(k)$ le nombre de partitions de l'ensemble $\{1, 2, \dots, 2k\}$ en sous-ensembles contenant chacun au moins deux éléments. Soit P une telle partition ; elle possède au plus k éléments. Le nombre d'applications de $\{1, 2, \dots, 2k\}$ dans $\{1, 2, \dots, n\}$ constante sur chaque élément de P est inférieur à n^k . On en déduit que $N(k, n) \leq n^k M(k)$.

Soit $\epsilon > 0$. On a

$$E\left[\left(n^{-\epsilon-1/2} R_n\right)^{2k}\right] \leq n^{-2k\epsilon-k} N(k, n) \leq n^{-2k\epsilon} M(k).$$

À condition d'avoir choisi $k > \frac{1}{2\epsilon}$, on en déduit que

$$\sum_{n \geq 1} E\left[\left(n^{-\epsilon-1/2} R_n\right)^{2k}\right] < \infty,$$

ce qui entraîne, d'après le corollaire 10.11, que la suite $(n^{-\epsilon-1/2} R_n)$ tend presque sûrement vers zéro quand n tend vers l'infini.

Pour chaque $\epsilon > 0$, il y a un événement négligeable en dehors duquel $n^{-\epsilon-1/2} R_n \rightarrow 0$. Pour démontrer la proposition 11.1, il suffit de considérer une famille dénombrable de valeurs de ϵ . En excluant une réunion dénombrable d'événements négligeables, on obtient : en dehors d'un événement négligeable, pour chaque $\epsilon > 0$, on a $n^{-\epsilon-1/2} R_n \rightarrow 0$.

— QED (Proposition 11.1)

11.3. *L'estimation de HARDY et LITTLEWOOD.* — C'est une conséquence directe de l'estimation des grands écarts.

PROPOSITION 11.2. — *Presque sûrement, quand $n \rightarrow +\infty$*

$$S_n - np = O\left(\sqrt{n \ln n}\right).$$

Démonstration. — Le théorème 6.1 nous dit que

$$\mathbb{P}\left(R_n > \sqrt{n \ln n}\right) \leq \exp\left(-nh_+\left(\sqrt{\frac{\ln n}{n}}\right)\right),$$

où, quand $\epsilon \rightarrow 0$,

$$h_+(\epsilon) = \frac{\epsilon^2}{2p(1-p)} + O(\epsilon^3).$$

Quand n tend vers l'infini, on a

$$h_+\left(\sqrt{\frac{\ln n}{n}}\right) = \frac{\ln n}{2p(1-p)n} + o\left(\frac{1}{n}\right)$$

et

$$\exp\left(-nh_+\left(\sqrt{\frac{\ln n}{n}}\right)\right) \sim \exp\left(-\frac{\ln n}{2p(1-p)}\right) = n^{-1/(2p(1-p))},$$

qui est le terme général d'une série convergente puisque $\frac{1}{2p(1-p)} \geq 2$.

Ainsi, on a

$$\sum_{n \geq 1} \mathbb{P}\left(R_n > \sqrt{n \ln n}\right) < \infty.$$

Grâce au premier lemme de BOREL-CANTELLI (Proposition 10.9), on conclut que, presque sûrement, pour tout n assez grand, $R_n \leq \sqrt{n \ln n}$.

— QED (Proposition 11.2)

11.4. La loi du logarithme itéré de KHINTCHINE. —

THÉORÈME 11.3. — *On a presque sûrement*

$$\limsup_{n \rightarrow \infty} \frac{S_n - np}{\sqrt{2p(1-p)n \ln \ln n}} = +1,$$

et

$$\liminf_{n \rightarrow \infty} \frac{S_n - np}{\sqrt{2p(1-p)n \ln \ln n}} = -1.$$

Pour démontrer ce résultat, nous utilisons les deux lemmes suivants. Le premier se déduit de l'estimation des grands écarts et de l'estimation des écarts modérés. Le second est un exemple simple d'*inégalité maximale* ; ces inégalités jouent un rôle fondamental pour établir des résultats asymptotiques presque sûrs ; elles sont attachées au nom de KOLMOGOROV.

Pour tout nombre entier $n > 1$, notons

$$\alpha(n) := \sqrt{2p(1-p)n \ln \ln n}.$$

LEMME 11.4. — *Pour tous nombres positifs a et δ , pour tout n assez grand,*

$$(\ln n)^{-a^2(1+\delta)} < \mathbf{P}(R_n \geq a\alpha(n)) < (\ln n)^{-a^2(1-\delta)}.$$

LEMME 11.5. — *Soit $(Y_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, d'espérance nulle et de variance σ^2 .* inégalité maximale

Posons $T_n := Y_1 + Y_2 + \dots + Y_n$. Pour tout $b \in \mathbb{R}$, on a

$$\mathbf{P}\left(\max_{1 \leq k \leq n} T_k \geq b\right) \leq \frac{4}{3} \cdot \mathbf{P}(T_n \geq b - 2\sigma\sqrt{n}).$$

Démonstration (Lemme 11.4) — D'après le théorème 6.1, on a

$$\mathbf{P}(R_n \geq a\alpha(n)) \leq \exp\left(-nh_+\left(\frac{a\alpha(n)}{n}\right)\right) \quad (11.1)$$

et, puisque la suite $\left(\frac{\alpha(n)}{n}\right)$ tend vers zéro, la proposition 6.2 permet d'affirmer que

$$h_+\left(\frac{a\alpha(n)}{n}\right) = \frac{a^2}{2p(1-p)} \left(\frac{\alpha(n)}{n}\right)^2 + O\left(\left(\frac{\alpha(n)}{n}\right)^3\right).$$

Ainsi, on a

$$h_+ \left(\frac{a\alpha(n)}{n} \right) = a^2 \frac{\ln \ln n}{n} + O \left(\left(\frac{\ln \ln n}{n} \right)^{3/2} \right)$$

et, pour tout n assez grand,

$$nh_+ \left(\frac{a\alpha(n)}{n} \right) \geq a^2(1 - \delta) \ln \ln n .$$

Revenant à l'inégalité (11.1), on conclut que, pour tout n assez grand,

$$\mathbf{P}(R_n \geq a\alpha(n)) \leq (\ln n)^{-a^2(1-\delta)} .$$

D'après le théorème 8.1, qui s'applique puisque $\sqrt{\ln \ln n} = o(n^{1/6})$, on a

$$\begin{aligned} \mathbf{P}(R_n \geq a\alpha(n)) &= \mathbf{P} \left(\frac{S_n}{n} - p \geq \sqrt{\frac{p(1-p)}{n}} a \sqrt{2 \ln \ln n} \right) \\ &\sim \frac{1}{\sqrt{2\pi} a \sqrt{2 \ln \ln n}} \exp(-a^2 \ln \ln n) = \frac{1}{2a\sqrt{\pi \ln \ln n}} (\ln n)^{-a^2} . \end{aligned}$$

Puisque $\sqrt{\ln \ln n} = o((\ln n)^{a^2\delta})$, on peut conclure que, pour tout n assez grand,

$$\mathbf{P}(R_n \geq a\alpha(n)) \geq (\ln n)^{-a^2(1+\delta)} .$$

— QED (Lemme 11.4)

Démonstration (Lemme 11.5) — Les variables aléatoires Y_n étant indépendantes, on a, pour $1 \leq k \leq n$, $\text{var}(T_n - T_k) = (n - k)\sigma^2$. D'après l'inégalité de BIENAYMÉ-TCHEBYCHEV,

$$\mathbf{P}(|T_n - T_k| \leq 2\sigma\sqrt{n}) \geq 1 - \frac{\text{var}(T_n - T_k)}{4\sigma^2 n} = 1 - \frac{n - k}{4n} \geq \frac{3}{4} .$$

Les propriétés élémentaires de la probabilité permettent d'écrire

$$\mathbf{P} \left(\max_{1 \leq k \leq n} T_k \geq b \right) = \sum_{k=1}^n \mathbf{P}(T_1 < b, T_2 < b, \dots, T_{k-1} < b \text{ et } T_k \geq b) ,$$

qui est majoré par

$$\frac{4}{3} \sum_{k=1}^n \mathbf{P}(T_1 < b, \dots, T_{k-1} < b \text{ et } T_k \geq b) \cdot \mathbf{P}(|T_n - T_k| \leq 2\sigma\sqrt{n}) .$$

Or, pour chaque k , la variable aléatoire $T_n - T_k$ est indépendante des variables aléatoires T_1, T_2, \dots, T_k . La quantité précédente s'écrit donc :

$$\frac{4}{3} \sum_{k=1}^n \mathbf{P} (T_1 < b, \dots, T_{k-1} < b, T_k \geq b \text{ et } |T_n - T_k| \leq 2\sigma\sqrt{n}) .$$

Mais l'événement $(T_1 < b, \dots, T_{k-1} < b, T_k \geq b \text{ et } |T_n - T_k| \leq 2\sigma\sqrt{n})$ est contenu dans $(T_1 < b, \dots, T_{k-1} < b, T_k \geq b \text{ et } T_n \geq b - 2\sigma\sqrt{n})$, d'où

$$\begin{aligned} & \mathbf{P} \left(\max_{1 \leq k \leq n} T_k \geq b \right) \\ & \leq \frac{4}{3} \sum_{k=1}^n \mathbf{P} (T_1 < b, \dots, T_{k-1} < b, T_k \geq b \text{ et } T_n \geq b - 2\sigma\sqrt{n}) . \end{aligned}$$

Une nouvelle utilisation des propriétés élémentaires de la probabilité permet de conclure :

$$\mathbf{P} \left(\max_{1 \leq k \leq n} T_k \geq b \right) \leq \frac{4}{3} \mathbf{P} (T_n \geq b - 2\sigma\sqrt{n}) .$$

— QED (Lemme 11.5)

Nous allons à présent aborder la démonstration de la loi du logarithme itéré. Il nous suffit bien sûr de démontrer que

$$\limsup_{n \rightarrow \infty} \frac{S_n - np}{\alpha(n)} = +1 ,$$

le calcul de la limite inférieure s'en déduisant par simple symétrie : en remplaçant S_n par $n - S_n$ on retrouve la même situation, après interversion de p et $1 - p$.

Nous procéderons en deux étapes, en montrant que, pour tout $\eta > 0$:

- (i) la limite supérieure est presque sûrement inférieure à $1 + \eta$,
- (ii) la limite supérieure est presque sûrement supérieure à $1 - \eta$.

Nous pourrions conclure en faisant tendre η vers zéro : à chaque η , on associe un ensemble presque sûr et, en considérant une suite de η qui tend vers zéro, on utilise le fait qu'une réunion dénombrable d'ensembles négligeables est négligeable.

En guise d'introduction à la démonstration, donnons un argument simplifié qui donne une réponse partielle. Fixons un nombre $\gamma > 1$ et notons $n_k = [\gamma^k]$ la partie entière de γ^k . D'après le lemme 11.4, pour tout $\delta > 0$ et pour tout n assez grand, on a

$$\mathbf{P} (R_{n_k} \geq (1 + \eta)\alpha(n_k)) < (\ln n_k)^{-(1+\eta)^2(1-\delta)} ,$$

donc

$$\mathbb{P} \left(R_{n_k} \geq (1 + \eta)\alpha(n_k) \right) = O \left(k^{-(1+\eta)^2(1-\delta)} \right).$$

Si δ a été choisi assez petit pour que $(1 + \eta)^2(1 - \delta) > 1$, on voit que

$$\sum_{k \geq 0} \mathbb{P} \left(R_{n_k} \geq (1 + \eta)\alpha(n_k) \right) < \infty.$$

et le premier lemme de BOREL-CANTELLI (Proposition 10.9) nous permet de conclure que, presque sûrement,

$$\limsup_{k \rightarrow +\infty} \frac{R_{n_k}}{\alpha(n_k)} \leq 1 + \eta.$$

Nous avons obtenu l'estimation attendue le long de sous-suites à croissance exponentielle. Il faudra travailler un peu pour obtenir l'estimation complète. Pour l'estimation par en dessous (ii), la tâche sera facilitée par le fait qu'il nous suffira de considérer une sous-suite, mais sera rendue plus difficile par le fait que nous devons utiliser le second lemme de BOREL-CANTELLI (Proposition 10.13), qui fait appel à une propriété d'indépendance.

Démonstration (Théorème 11.3) — Fixons $\eta > 0$. Pour un nombre réel $\gamma > 1$ dont nous préciserons le choix plus tard et pour tout $k \in \mathbb{N}$, notons $n_k := \lceil \gamma^k \rceil$. Nous allons démontrer que

$$\sum_{k \geq 0} \mathbb{P} \left(\max_{n \leq n_{k+1}} R_n \geq (1 + \eta)\alpha(n_k) \right) < \infty. \quad (11.2)$$

D'après le lemme 11.5, on a

$$\begin{aligned} \mathbb{P} \left(\max_{n \leq n_{k+1}} R_n \geq (1 + \eta)\alpha(n_k) \right) \\ \leq \frac{4}{3} \mathbb{P} \left(R_{n_{k+1}} \geq (1 + \eta)\alpha(n_k) - 2\sqrt{n_{k+1}p(1-p)} \right). \end{aligned}$$

Quand $k \rightarrow +\infty$, on a $\sqrt{n_{k+1}} = o(\alpha(n_k))$ et donc, pour tout k assez grand,

$$2\sqrt{n_{k+1}p(1-p)} < \frac{1}{2}\eta\alpha(n_k).$$

Pour un tel k , on a

$$\mathbb{P} \left(\max_{n \leq n_{k+1}} R_n \geq (1 + \eta)\alpha(n_k) \right) \leq \frac{4}{3} \mathbb{P} \left(R_{n_{k+1}} \geq (1 + \eta/2)\alpha(n_k) \right).$$

On a $\alpha(n_{k+1}) \sim \sqrt{\gamma}\alpha(n_k)$ et on choisit γ tel que $(1 + \eta/2) > (1 + \eta/4)\sqrt{\gamma}$. On a alors, pour tout k assez grand, $(1 + \eta/2)\alpha(n_k) > (1 + \eta/4)\alpha(n_{k+1})$, et donc

$$\mathbb{P} \left(\max_{n \leq n_{k+1}} R_n \geq (1 + \eta)\alpha(n_k) \right) \leq \frac{4}{3} \mathbb{P} \left(R_{n_{k+1}} \geq (1 + \eta/4)\alpha(n_{k+1}) \right) .$$

Nous utilisons à présent le lemme 11.4 pour majorer cette expression (avec le choix $a = (1 - \delta)^{-1} = (1 + \eta/4)$) et nous obtenons, pour tout k assez grand,

$$\mathbb{P} \left(\max_{n \leq n_{k+1}} R_n \geq (1 + \eta)\alpha(n_k) \right) \leq \frac{4}{3} (\ln n_{k+1})^{-(1+\eta/4)} .$$

Enfin, on a $(\ln n_{k+1})^{-(1+\eta/4)} \sim (\ln \gamma)^{-(1+\eta/4)} k^{-(1+\eta/4)}$, qui est le terme général d'une série convergente.

Nous avons démontré (11.2). Le premier lemme de BOREL-CANTELLI permet alors d'affirmer que, presque sûrement, pour tout k assez grand,

$$\max_{n \leq n_{k+1}} R_n < (1 + \eta)\alpha(n_k) ,$$

et, en particulier,

$$\max_{n_k \leq n < n_{k+1}} R_n < (1 + \eta)\alpha(n_k) .$$

Cela entraîne bien que, presque sûrement, pour tout n assez grand

$$R_n < (1 + \eta)\alpha(n) .$$

L'estimation par au dessus (i) est établie.

Pour démontrer l'estimation par en dessous (ii), il nous suffit de démontrer qu'il existe une sous-suite (n_k) de la suite de tous les entiers telle que « $(R_{n_k} \geq (1 - \eta)\alpha(n_k))$ *infinitement souvent* » soit un événement presque sûr. Nous choisissons une sous-suite de la forme $n_k = \gamma^k$ où γ est à présent un nombre entier suffisamment grand. Les conditions imposées à γ seront précisées plus tard. Nous allons montrer que

$$\sum_{n \geq 1} \mathbb{P} \left(R_{\gamma^n} - R_{\gamma^{n-1}} \geq \left(1 - \frac{\eta}{2}\right) \alpha(\gamma^n) \right) = \infty , \quad (11.3)$$

et que

$$\text{presque sûrement, pour tout } n \text{ assez grand, } R_{\gamma^{n-1}} \geq -\frac{\eta}{2}\alpha(\gamma^n) . \quad (11.4)$$

La variable aléatoire $R_{\gamma^n} - R_{\gamma^{n-1}}$ suit la même loi que la variable aléatoire $R_{\gamma^n - \gamma^{n-1}}$. On a donc

$$\mathbb{P} \left(R_{\gamma^n} - R_{\gamma^{n-1}} \geq \left(1 - \frac{\eta}{2}\right) \alpha(\gamma^n) \right) = \mathbb{P} \left(R_{\gamma^n - \gamma^{n-1}} \geq \left(1 - \frac{\eta}{2}\right) \alpha(\gamma^n) \right) .$$

Quand n tend vers l'infini, on a

$$\frac{\alpha(\gamma^n - \gamma^{n-1})}{\alpha(\gamma^n)} \sim \sqrt{\frac{\gamma - 1}{\gamma}} .$$

On impose à γ la condition $(1 - \eta/2) \sqrt{\gamma} < (1 - \eta/4) \sqrt{\gamma - 1}$, et on a alors, pour tout n assez grand,

$$(1 - \eta/2) \alpha(\gamma^n) < (1 - \eta/4) \alpha(\gamma^n - \gamma^{n-1}) .$$

Pour un tel n , on a

$$\begin{aligned} \mathbb{P} \left(R_{\gamma^n} - R_{\gamma^{n-1}} \geq \left(1 - \frac{\eta}{2}\right) \alpha(\gamma^n) \right) \\ \geq \mathbb{P} \left(R_{\gamma^n - \gamma^{n-1}} \geq \left(1 - \frac{\eta}{4}\right) \alpha(\gamma^n - \gamma^{n-1}) \right) . \end{aligned}$$

Nous utilisons à présent le lemme 11.4 pour minorer cette expression (avec le choix $a = (1 + \delta)^{-1} = (1 - \eta/4)$) et nous obtenons, pour tout n assez grand,

$$\mathbb{P} \left(R_{\gamma^n} - R_{\gamma^{n-1}} \geq \left(1 - \frac{\eta}{2}\right) \alpha(\gamma^n) \right) \geq (\ln(\gamma^n - \gamma^{n-1}))^{-1+\eta/4} .$$

Mais $(\ln(\gamma^n - \gamma^{n-1}))^{-1+\eta/4} \sim (n \ln(\gamma))^{-1+\eta/4}$ est le terme général d'une série divergente. Cela démontre (11.3).

Quand n tend vers l'infini, $\alpha(\gamma^n) \sim \sqrt{\gamma} \alpha(\gamma^{n-1})$. On impose à γ la condition $\eta \sqrt{\gamma} > 4$, et on a alors, pour tout n assez grand, $\eta \alpha(\gamma^n) > 4 \alpha(\gamma^{n-1})$. Pour un tel n , on a

$$\left(R_{\gamma^{n-1}} \leq -\frac{\eta}{2} \alpha(\gamma^n) \right) \subset \left(-R_{\gamma^{n-1}} \geq 2 \alpha(\gamma^{n-1}) \right) .$$

Or l'estimation par au dessus (i) établie dans la première partie de cette démonstration, appliquée à la suite $(-R_n)$, permet d'affirmer que, presque sûrement, pour tout n assez grand, $-R_{\gamma^{n-1}} < 2 \alpha(\gamma^{n-1})$. Cela démontre (11.4).

La suite de variables aléatoires $(R_{\gamma^n} - R_{\gamma^{n-1}})$ est une suite de variables aléatoires indépendantes. Le second lemme de BOREL-CANTELLI permet d'affirmer, à partir de (11.3) : presque sûrement, il existe une infinité d'entiers n tels que $R_{\gamma^n} - R_{\gamma^{n-1}} > \left(1 - \frac{\eta}{2}\right) \alpha(\gamma^n)$. Associé à (11.4), cela

entraîne que, presque sûrement, il existe une infinité d'entiers n tels que $R_{\gamma^n} > (1 - \eta) \alpha(\gamma^n)$. L'estimation par en dessous (ii) est établie.

— QED (Théorème 11.3)

12. Récurrence des marches aléatoires

12.1. Introduction et formalisme. — Nous abordons dans ce chapitre la question de la récurrence des marches aléatoires dans l'espace discret \mathbb{Z}^N , de dimension $N \geq 1$. Une *marche aléatoire* décrit les emplacements successifs d'un promeneur qui se déplace, à chaque instant entier, dans une direction et d'une distance aléatoires. On suppose que chacun de ces pas aléatoires est indépendant des déplacements qui l'ont précédé et que tous ces pas aléatoires suivent la même loi de probabilité. On peut donner un sens à la notion de marche aléatoire dans un groupe quelconque. Nous nous limitons ici au cas des groupes additifs \mathbb{Z}^N , et nous supposons que la longueur maximale des pas est bornée, ce qui est naturel en ce qui concerne le promeneur.

En imaginant que la promenade n'est pas limitée dans le temps, nous nous posons les questions suivantes : le promeneur est-il sûr de revenir à son point de départ ? le promeneur est-il sûr de visiter tous les points de l'espace ? le promeneur est-il sûr de revenir infiniment souvent à son point de départ ?

Nous verrons que les réponses à ces questions dépendent essentiellement de deux paramètres : le déplacement moyen du promeneur au cours d'un pas et la dimension N de l'espace dans lequel il se déplace. Pour dire les choses rapidement nous observerons les phénomènes suivant : si l'espérance du pas est non nulle, alors, presque sûrement, le promeneur part à l'infini ; si l'espérance du pas est nulle, alors les réponses aux questions précédentes sont positives à condition que le promeneur se déplace dans un espace de dimension inférieure ou égale à 2.

Le modèle mathématique associé à cette expérience sera celui abondamment décrit dans les chapitres précédents, puisqu'il s'agit bien de décrire une suite d'expériences élémentaires identiques et indépendantes (chaque pas du promeneur).

Commençons par la description du modèle pour *la marche aléatoire aux plus proches voisins sur \mathbb{Z}* : c'est le trajet suivi le long d'un chemin par un promeneur qui, à chaque instant entier, effectue un pas de longueur fixée, soit vers l'avant, soit vers l'arrière. On retrouve bien sûr ici le jeu de pile ou

face : à chaque instant entier le promeneur lance une pièce pour décider s'il avance ou recule. Formellement, on reprend l'espace (Ω, P) décrit dans le paragraphe 10.1. Si $\omega = (\omega_n)_{n \geq 1} \in \Omega = \{0, 1\}^{\mathbb{N}^*}$, on pose $Y_n(\omega) = 2\omega_n - 1$. Ainsi (Y_n) est une suite de variables aléatoires indépendantes telles que

$$P(Y_n = 1) = p \quad \text{et} \quad P(Y_n = -1) = 1 - p.$$

La *marche aléatoire* est alors la suite de variables aléatoires $(M_n)_{n \geq 1}$ définie par

$$M_n := Y_1 + Y_2 + \dots + Y_n.$$

Le paramètre p , nombre réel entre 0 et 1, est la probabilité que le pas soit effectué vers l'avant. Bien sûr, nous retrouvons ici l'objet que nous avons étudié dans les chapitres précédents. Nos diverses notations sont reliées par $M_n = 2S_n - n = 2R_n + (2p - 1)n$.

Plus généralement, une marche aléatoire sur \mathbb{Z}^N , avec nombre fini de pas possibles, peut être décrite de la façon suivante : soient (e_1, e_2, \dots, e_d) une famille finie d'éléments de \mathbb{Z}^N et (p_1, p_2, \dots, p_d) une famille de nombres réels positifs dont la somme est égale à 1. (Les e_i désignent les pas admissibles et p_i est la probabilité associée au pas e_i .) On considère l'espace probabilisé (Ω, P) décrit dans le paragraphe 10.3., et, pour chaque $\omega = (\omega_n) \in \Omega = \{1, 2, \dots, d\}^{\mathbb{N}^*}$, on pose $Y_n(\omega) = e_{\omega_n}$. La *marche aléatoire* est alors la suite de variables aléatoires $(M_n)_{n \geq 1}$ définie par

$$M_n := Y_1 + Y_2 + \dots + Y_n.$$

Il est temps de définir précisément la notion de récurrence. Une marche aléatoire $(M_n)_{n \geq 1}$ est dite *récurrenente* si elle revient presque sûrement une infinité de fois à son point de départ, c'est-à-dire si l'ensemble « $(M_n = 0)$ *infiniment souvent*» est un événement presque sûr. Cette marche aléatoire est dite *transitoire* si elle ne revient presque sûrement qu'un nombre fini de fois à son point de départ, c'est-à-dire si l'ensemble « $(M_n = 0)$ *infiniment souvent*» est un événement négligeable.

Une autre définition nous sera utile : la marche aléatoire est dite *centrée* si l'espérance du pas est nulle, c'est-à-dire si $E[Y_n] := \sum_{i=1}^d p_i e_i = 0$.

La suite de ce chapitre sera constituée de trois paragraphes. Dans le premier on décrit les propriétés de récurrence de la marche aléatoire aux plus proches voisins sur \mathbb{Z} ; cette marche est une copie fidèle du jeu de pile ou face, et nous pourrons facilement utiliser les résultats des précédents chapitres. Dans un deuxième paragraphe, on étudie les propriétés

des marches aléatoires dans un cadre général. Les propriétés de récurrence des marches aléatoires sur \mathbb{Z}^N sont décrites dans le dernier paragraphe.

12.2. La marche aléatoire aux plus proches voisins sur \mathbb{Z} . —

Les variables aléatoires Y_n prennent ici les valeurs $+1$ et -1 avec probabilités respectives p et $1 - p$, et $M_n = Y_1 + Y_2 + \dots + Y_n$. Si $p > 1/2$ (respectivement $p < 1/2$), la loi forte des grands nombres assure que, presque sûrement, $M_n \rightarrow +\infty$ (respectivement $-\infty$) quand $n \rightarrow \infty$. On en déduit immédiatement que si $p \neq 1/2$ alors la marche aléatoire est transitoire.

Si $p = 1/2$, cette marche aléatoire est appelée la *marche aléatoire simple* sur \mathbb{Z} . Cette marche aléatoire est centrée. On peut démontrer de diverses façons qu'elle est récurrente, par exemple comme conséquence de la loi des retours en zéro (Théorème 9.2) ou de la loi du logarithme itéré (Théorème 11.3). En effet, grâce à cette loi du logarithme itéré, on sait que, presque sûrement, $\limsup M_n = +\infty$ et $\liminf M_n = -\infty$. La longueur des pas étant toujours égale à 1, cela entraîne que, presque sûrement, la marche aléatoire visite infiniment souvent tous les points de \mathbb{Z} .

marche aléatoire simple sur \mathbb{Z} .

Ces arguments se généralisent-ils à d'autres marches aléatoires ? Pour partie. Si la marche aléatoire n'est pas centrée, la loi forte des grands nombres (Théorème 10.12) assure immédiatement que cette marche est transitoire. Par contre l'étude de la récurrence est plus délicate. Considérons par exemple la marche aléatoire (M_n) sur \mathbb{Z} de loi $\mathbb{P}(Y_n = -3) = 2/5$ et $\mathbb{P}(Y_n = 2) = 3/5$; cette marche est centrée, et la loi du logarithme itéré nous assure que, presque sûrement, $\limsup M_n = +\infty$ et $\liminf M_n = -\infty$, mais cela ne suffit pas pour conclure que la marche est récurrente. De même pour une marche aléatoire centrée quelconque sur \mathbb{Z}^N , avec $N \geq 2$.

12.3. Généralités. —

La propriété générale la plus frappante est que toute marche aléatoire est soit récurrente, soit transitoire. Cette *loi du tout ou rien* fait l'objet du théorème 12.1. Le lien entre le nombre de retours de la marche aléatoire à son point de départ et le nombre de passages en un point quelconque est précisé dans la proposition qui suit.

THÉORÈME 12.1. — *Toute marche aléatoire $(M_n)_{n \geq 1}$ est soit récurrente, soit transitoire. Il y a équivalence entre :*

- (i) *la marche aléatoire est récurrente ;*
- (ii) $\lim_{n \rightarrow +\infty} \mathbb{P}(\text{il existe } k \leq n \text{ tel que } M_k = 0) = 1 ;$

$$(iii) \sum_{n=1}^{+\infty} P(M_n = 0) = +\infty.$$

On peut vérifier que la propriété (ii) entraîne que le retour de la marche aléatoire au moins une fois à son point de départ est un événement presque sûr. En fait, la réciproque est vraie : la propriété (ii) est équivalente au retour presque sûr de la marche en zéro. Mais ce fait ne sera pas discuté ici.

On notera l'analogie entre l'implication (iii) \Rightarrow (ii) et le second lemme de BOREL-CANTELLI. Mais ce lemme ne s'applique pas directement puisque les variables aléatoires M_n ne sont pas indépendantes.

Si m, s et t sont des nombres entiers positifs, avec $s \leq t$, nous notons $A_{s,t}^m$ l'événement «la marche aléatoire repasse au moins m fois par son point de départ entre les instants s et t ». Autrement dit

$$\omega \in A_{s,t}^m \iff \text{card}\{n : s \leq n \leq t \text{ et } M_n(\omega) = 0\} \geq m.$$

Nous utiliserons les deux lemmes suivants.

LEMME 12.2. — Pour tous $m, t > 0$, on a $P(A_{1,t}^m) \leq (P(A_{1,t}^1))^m$.

LEMME 12.3. — Pour tous $m, t > 0$, on a $(P(A_{1,t}^1))^m \leq P(A_{1,mt}^m)$.

Démonstration (Lemme 12.2) — On a bien sûr $P(A_{1,t}^m) = 0$ si $m > t$. Le lemme, trivial pour $m = 1$, peut être démontré par récurrence sur m . Afin de ne pas manipuler des expressions trop lourdes, nous ne détaillons que le premier pas de cette récurrence. Quand il est compris, les étapes suivantes ne présentent pas de difficulté. Démontrons donc la formule annoncée pour $m = 2$. En considérant les deux premiers temps de retour de la marche aléatoire à son point de départ, on écrit l'événement $A_{1,t}^2$ comme une réunion finie d'événements deux à deux disjoints. On a

$$P(A_{1,t}^2) = \sum_{1 \leq j < k \leq t} P((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0) \\ \text{et } (M_i \neq 0, j < i < k) \text{ et } (M_k = 0)),$$

donc

$$P(A_{1,t}^2) = \sum_{1 \leq j < k \leq t} P((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0) \\ \text{et } (M_i - M_j \neq 0, j < i < k) \text{ et } (M_k - M_j = 0)).$$

Mais l'événement $((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0))$ est indépendant de

l'événement $((M_i - M_j \neq 0, j < i < k) \text{ et } (M_k - M_j = 0))$, d'où

$$\begin{aligned} P(A_{1,t}^2) &= \sum_{1 \leq j < k \leq t} P((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0)) \\ &\quad \times P((M_i - M_j \neq 0, j < i < k) \text{ et } (M_k - M_j = 0)) . \end{aligned}$$

Grâce à la propriété d'invariance par décalage (cf. §10.1.), on a

$$\begin{aligned} P((M_i - M_j \neq 0, j < i < k) \text{ et } (M_k - M_j = 0)) \\ = P((M_{i-j} \neq 0, j < i < k) \text{ et } (M_{k-j} = 0)) , \end{aligned}$$

d'où

$$\begin{aligned} P(A_{1,t}^2) &= \sum_{1 \leq j < k \leq t} P((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0)) \\ &\quad \times P((M_i \neq 0, 1 \leq i < k - j) \text{ et } (M_{k-j} = 0)) . \end{aligned} \tag{12.1}$$

On en déduit que

$$\begin{aligned} P(A_{1,t}^2) &\leq \sum_{j=1}^t \sum_{\ell=1}^t P((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0)) \\ &\quad \times P((M_i \neq 0, 1 \leq i < \ell) \text{ et } (M_\ell = 0)) . \end{aligned}$$

Mais suivant une décomposition similaire à celle utilisée au départ de ce calcul, on a

$$P(A_{1,t}^1) = \sum_{j=1}^t P((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0)) ,$$

On peut conclure que

$$P(A_{1,t}^2) \leq (P(A_{1,t}^1))^2 .$$

— QED (Lemme 12.2)

Démonstration (Lemme 12.3) — Ce lemme peut être démontré par récurrence sur m . De nouveau, nous ne décrivons que le premier pas de cette récurrence en nous limitant au cas $m = 2$.

La relation (12.1) s'écrit

$$\begin{aligned} P(A_{1,2t}^2) &= \sum_{1 \leq j < k \leq 2t} P((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0)) \\ &\quad \times P((M_i \neq 0, 1 \leq i < k - j) \text{ et } (M_{k-j} = 0)) . \end{aligned}$$

On en déduit que

$$P(A_{1,2t}^2) \geq \sum_{j=1}^t \sum_{k=j+1}^{j+t} P((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0)) \\ \times P((M_i \neq 0, 1 \leq i < k-j) \text{ et } (M_{k-j} = 0)) ,$$

c'est-à-dire, en posant $\ell = k - j$,

$$P(A_{1,2t}^2) \geq \sum_{j=1}^t P((M_i \neq 0, 1 \leq i < j) \text{ et } (M_j = 0)) \\ \times \sum_{\ell=1}^t P((M_i \neq 0, 1 \leq i < \ell) \text{ et } (M_\ell = 0)) ,$$

ce qui s'écrit simplement

$$P(A_{1,2t}^2) \geq (P(A_{1,t}^1))^2 .$$

— QED (Lemme 12.3)

Démonstration (Théorème 12.1) — Si $\sum_{n \geq 1} P(M_n = 0) < +\infty$, alors, d'après le premier lemme de BOREL-CANTELLI, « $M_n = 0$ infiniment souvent » est un événement négligeable ; la marche aléatoire est transitoire, donc non récurrente. Cela démontre l'implication (i) \Rightarrow (iii).

Démontrons l'implication (iii) \Rightarrow (ii). La suite $(P(A_{1,n}^1))_{n \geq 1}$ est croissante et majorée par 1. Posons

$$\rho = \lim_{n \rightarrow \infty} P(A_{1,n}^1) .$$

La condition (ii) s'écrit simplement $\rho = 1$. Par linéarité de l'espérance, on a

$$\sum_{k=1}^n P(M_k = 0) = E \left[\sum_{k=1}^n \mathbf{1}_{(M_k=0)} \right] ,$$

ce qui s'écrit encore

$$\sum_{k=1}^n P(M_k = 0) = E [\text{card}\{j : 1 \leq j \leq n \text{ et } M_j = 0\}] \\ = \sum_{j=1}^n j P(\text{la suite } (M_k)_{1 \leq k \leq n} \text{ s'annule exactement } j \text{ fois}) . \\ = \sum_{j=1}^n j (P(A_{1,n}^j) - P(A_{1,n}^{j+1})) = \sum_{j=1}^n P(A_{1,n}^j)$$

En utilisant le lemme 12.2, on en déduit que

$$\sum_{k=1}^n \mathbf{P}(M_k = 0) \leq \sum_{j=1}^n (\mathbf{P}(A_{1,n}^1))^j \leq \sum_{j=1}^n \rho^j,$$

et donc

$$\sum_{k=1}^{+\infty} \mathbf{P}(M_k = 0) = +\infty \Rightarrow \rho = 1.$$

Cela prouve que (iii) \Rightarrow (ii).

Il nous reste à montrer que (ii) \Rightarrow (i). Soit m un nombre entier positif. Notons $A_{1,\infty}^m$ l'ensemble des ω tels que la suite $(M_n(\omega))_{n \geq 1}$ s'annule au moins m fois. Pour chaque entier $t > 0$, on a

$$A_{1,\infty}^m \supset A_{1,mt}^m.$$

Le lemme 12.3 nous dit que

$$\mathbf{P}(A_{1,mt}^m) \geq (\mathbf{P}(A_{1,t}^1))^m.$$

Supposons la condition (ii) satisfaite, c'est-à-dire $\lim_{t \rightarrow \infty} \mathbf{P}(A_{1,t}^1) = 1$. L'ensemble $A_{1,\infty}^m$ contient des événements de probabilité arbitrairement proche de 1. Donc cet ensemble est un événement presque sûr.

L'ensemble des ω tels que la suite $(M_n(\omega))_{n \geq 1}$ s'annule une infinité de fois est l'intersection des $A_{1,\infty}^m$ quand m varie dans \mathbb{N} . On sait qu'une intersection dénombrable d'événements presque sûrs est un événement presque sûr (Proposition 10.1), et on conclut que, presque sûrement, la suite $(M_n)_{n \geq 1}$ s'annule une infinité de fois. La marche aléatoire est récurrente.

Les équivalences annoncées sont établies, et nous avons vu que la marche aléatoire est soit récurrente, soit transitoire, suivant que la série $\sum_n \mathbf{P}(M_n = 0)$ diverge ou converge. — QED (Théorème 12.1)

Étudions à présent le nombre de visites aux autres points de l'espace. Notons S le semi-groupe engendré par l'ensemble $E := \{e_1, e_2, \dots, e_d\}$ des pas possibles de la marche aléatoire (on suppose ici que les probabilités p_i sont toutes > 0). L'ensemble S est l'ensemble des éléments de \mathbb{Z}^N qui s'écrivent comme somme finie d'éléments de E , toutes les répétitions étant bien sûr autorisées. Nous remarquons que S est exactement l'ensemble des points qui peuvent être visités par la marche aléatoire.

PROPOSITION 12.4. — *Si la marche aléatoire est récurrente, alors S est un groupe et, presque sûrement, tous les points de S sont visités infiniment souvent par la marche aléatoire.*

Si la marche est transitoire, alors, presque sûrement, chaque point de S n'est visité qu'un nombre fini de fois, autrement dit, presque sûrement, $\lim_{n \rightarrow \infty} |M_n| = +\infty$.

Dans cet énoncé nous avons noté $|\cdot|$ une norme quelconque sur l'espace \mathbb{R}^N . Pour l'énoncé suivant nous choisissons la *norme infinie* $|(x_1, x_2, \dots, x_N)| = \max \{|x_1|, |x_2|, \dots, |x_N|\}$.

Dans la démonstration de cette proposition, nous utilisons le lemme suivant.

LEMME 12.5. — *La marche aléatoire (M_n) dans l'espace \mathbb{Z}^N vérifie : pour tout $x \in \mathbb{Z}^N$.*

$$\sum_{n=1}^{+\infty} \mathbf{P}(M_n = x) \leq 1 + \sum_{n=1}^{+\infty} \mathbf{P}(M_n = 0) ,$$

et, pour tout $m > 0$,

$$\sum_{n=1}^{+\infty} \mathbf{P}(|M_n| \leq m) \leq (2m + 1)^N \left(1 + \sum_{n=1}^{+\infty} \mathbf{P}(M_n = 0) \right) .$$

Démonstration (Lemme 12.5) — Par décomposition de l'événement $(M_n = x)$ comme une réunion d'événements deux à deux disjoints suivant l'instant de premier passage de la marche au point x , on obtient

$$\begin{aligned} \mathbf{P}(M_n = x) \\ = \sum_{k=1}^n \mathbf{P}((M_i \neq x, 1 \leq i < k) \text{ et } (M_k = x) \text{ et } (M_n - M_k = 0)) . \end{aligned}$$

En utilisant, comme dans la démonstration du lemme 12.2, l'indépendance des accroissements de la marche aléatoire et l'invariance de la probabilité par décalage, on arrive à

$$\mathbf{P}(M_n = x) = \sum_{k=1}^n \mathbf{P}((M_i \neq x, 1 \leq i < k) \text{ et } (M_k = x)) \cdot \mathbf{P}(M_{n-k} = 0) .$$

D'où l'on déduit, en notant $M_0 = 0$, que

$$\begin{aligned} \sum_{n=1}^{+\infty} \mathbf{P}(M_n = x) &= \sum_{k=1}^{+\infty} \sum_{n=k}^{+\infty} \mathbf{P}((M_i \neq x, 1 \leq i < k) \text{ et } (M_k = x)) \cdot \mathbf{P}(M_{n-k} = 0) \\ &= \sum_{k=1}^{+\infty} \mathbf{P}((M_i \neq x, 1 \leq i < k) \text{ et } (M_k = x)) \cdot \sum_{n=0}^{+\infty} \mathbf{P}(M_n = 0) . \end{aligned}$$

Les événements $((M_i \neq x, 1 \leq i < k) \text{ et } (M_k = x))$ étant deux à deux disjoints (quand k varie), la somme de leurs probabilités est inférieure ou égale à 1. La première inégalité de l'énoncé est établie. La seconde s'en déduit immédiatement puisque

$$\mathbf{P}(|M_n| \leq m) = \sum_{x \in \mathbb{Z}^N, |x| \leq m} \mathbf{P}(M_n = x)$$

et que, pour tout m entier > 0 , le cardinal de $\{x \in \mathbb{Z}^N, |x| \leq m\}$ est égal à $(2m+1)^N$. — QED (Lemme 12.5)

Démonstration (Proposition 12.4) — Supposons la marche aléatoire (M_n) transitoire. D'après le théorème 12.1, la série $\sum_{n \geq 1} \mathbf{P}(M_n = 0)$ est convergente. Le lemme précédent nous permet alors d'affirmer que, pour tout $m > 0$, la série $\sum_{n \geq 1} \mathbf{P}(|M_n| \leq m)$ est convergente. Le premier lemme de BOREL-CANTELLI nous donne donc la seconde affirmation de la proposition.

Supposons la marche aléatoire (M_n) récurrente.

Soit $x \in S$. Il existe $k \geq 0$ tel que $\mathbf{P}(M_k = x) > 0$. Fixons un tel k . Presque sûrement, il existe $n > k$ tel que $M_n = 0$. Un événement presque sûr et un événement (de type fini) de probabilité strictement positive ne peuvent pas être disjoints. Il existe donc $\omega \in \Omega$ tel que $M_k(\omega) = x$ et $M_n(\omega) = 0$, avec $k < n$. En conséquence, $-x \in S$. Cela suffit à prouver que S est un groupe.

Nous voulons montrer à présent que si un site x de l'espace est visité par la marche aléatoire avec probabilité positive, alors il sera presque sûrement visité une infinité de fois. L'idée est la suivante : nous sommes sûrs que la marche repassera une infinité de fois par zéro. À chacun de ces passages on peut imaginer qu'une nouvelle marche aléatoire démarre, indépendante de ce qui l'a précédée. On réalise donc une infinité d'expériences identiques

et indépendantes, avec pour chacune d'entre elles une probabilité positive d'atteindre x . Nous sommes alors sûrs d'atteindre x une infinité de fois. Formalisons cette idée.

D'après le théorème précédent, on a $\lim_{t \rightarrow +\infty} \mathbb{P}(A_{1,t}^1) = 1$. Grâce au lemme 12.3, on a, pour tout $m > 0$, $\lim_{t \rightarrow +\infty} \mathbb{P}(A_{1,mt}^m) = 1$, et donc

$$\lim_{t \rightarrow +\infty} \mathbb{P}(A_{1,t}^m) = 1.$$

Si m, s et t sont des nombres entiers positifs, on a bien sûr $A_{s,t}^m \supset A_{1,t}^{m+s}$, et donc

$$\lim_{t \rightarrow +\infty} \mathbb{P}(A_{s,t}^m) = 1.$$

Soit $x \in S$. On fixe $k \geq 1$ tel que $\mathbb{P}(M_k = x) > 0$ et on note $\delta := 1 - \mathbb{P}(M_k = x) < 1$. Soit $\epsilon > 0$. On construit par récurrence une suite $(n_j)_{j \geq 0}$ telle que

$$n_0 = 1, \quad n_j - n_{j-1} > k \quad \text{et} \quad \mathbb{P}(A_{n_{j-1}, n_j - k}^1) > 1 - 2^{-j} \epsilon.$$

Pour $j \geq 1$, notons B_j l'événement «la marche aléatoire passe au moins une fois par zéro entre les instants n_{j-1} et $n_j - k$, mais ne passe pas par x entre les instants n_{j-1} et n_j », autrement dit

$$B_j := A_{n_{j-1}, n_j - k}^1 \cap (M_n \neq x, n_{j-1} \leq n \leq n_j).$$

Soient J, K deux nombres entiers tels que $0 < J < K$. Majorons la probabilité de l'événement «le site x n'est pas visité entre les instants n_{J-1} et n_K ».

$$\mathbb{P}(M_n \neq x, n_{J-1} \leq n \leq n_K) \leq \mathbb{P}\left(\bigcup_{j=J}^K (A_{n_{j-1}, n_j - k}^1)^c\right) + \mathbb{P}\left(\bigcap_{j=J}^K B_j\right).$$

On a

$$\mathbb{P}\left(\bigcup_{j=J}^K (A_{n_{j-1}, n_j - k}^1)^c\right) \leq \sum_{j=J}^K 2^{-j} \epsilon \leq 2^{1-J} \epsilon.$$

Notons ℓ_j un nombre entier autorisé à varier entre n_{j-1} et $n_j - k$. Si l'événement B_j est réalisé, alors il existe un unique entier ℓ_j tel que

$$(M_\ell \neq 0, n_{j-1} \leq \ell < \ell_j), \quad M_{\ell_j} = 0 \quad \text{et} \quad M_{\ell_j + k} \neq x.$$

On en déduit que

$$\mathbb{P} \left(\bigcap_{j=J}^K B_j \right) \leq \sum_{\ell_J, \dots, \ell_K} \mathbb{P} \left(\bigcap_{j=J}^K \left(\left(A_{n_{j-1}, \ell_{j-1}}^1 \right)^c \cap (M_{\ell_j} = 0) \cap (M_{\ell_j+k} \neq x) \right) \right).$$

Mais l'accroissement $M_{\ell_K+k} - M_{\ell_K}$ étant indépendant de ce qui l'a précédé dans la marche aléatoire, on a

$$\begin{aligned} & \mathbb{P} \left(\bigcap_{j=J}^K \left(\left(A_{n_{j-1}, \ell_{j-1}}^1 \right)^c \cap (M_{\ell_j} = 0) \cap (M_{\ell_j+k} \neq x) \right) \right) \\ &= \mathbb{P} \left(\bigcap_{j=J}^{K-1} \left(\left(A_{n_{j-1}, \ell_{j-1}}^1 \right)^c \cap (M_{\ell_j} = 0) \cap (M_{\ell_j+k} \neq x) \right) \right. \\ & \quad \left. \cap \left(A_{n_{K-1}, \ell_{K-1}}^1 \right)^c \cap (M_{\ell_K} = 0) \right) \times \mathbb{P} (M_{\ell_K+k} - M_{\ell_K} \neq x). \end{aligned}$$

Et, grâce à la propriété d'invariance par décalage,

$$\mathbb{P} (M_{\ell_K+k} - M_{\ell_K} \neq x) = \mathbb{P} (M_k \neq x) = \delta.$$

On obtient ainsi l'inégalité

$$\begin{aligned} & \sum_{\ell_J, \dots, \ell_K} \mathbb{P} \left(\bigcap_{j=J}^K \left(\left(A_{n_{j-1}, \ell_{j-1}}^1 \right)^c \cap (M_{\ell_j} = 0) \cap (M_{\ell_j+k} \neq x) \right) \right) \\ & \leq \delta \sum_{\ell_J, \dots, \ell_{K-1}} \mathbb{P} \left(\bigcap_{j=J}^{K-1} \left(\left(A_{n_{j-1}, \ell_{j-1}}^1 \right)^c \cap (M_{\ell_j} = 0) \cap (M_{\ell_j+k} \neq x) \right) \right), \end{aligned}$$

qui entraîne, par récurrence,

$$\sum_{\ell_J, \dots, \ell_K} \mathbb{P} \left(\bigcap_{j=J}^K \left(\left(A_{n_{j-1}, \ell_{j-1}}^1 \right)^c \cap (M_{\ell_j} = 0) \cap (M_{\ell_j+k} \neq x) \right) \right) \leq \delta^{K-J+1}.$$

En rassemblant les inégalités précédentes, on arrive à

$$\mathbb{P} (M_n \neq x, n_{J-1} \leq n \leq n_K) \leq 2^{1-J} \epsilon + \delta^{K-J+1}.$$

Pour chaque $J > 0$ fixons $K = K(J) > 0$ tel que $\delta^{K-J+1} \leq 2^{1-J}\epsilon$. L'ensemble E des ω tels que « $M_n(\omega)$ n'est égal à x que pour au plus un nombre fini de valeurs de n » est contenu dans la réunion d'événements

$$\bigcup_{J>0} (M_n \neq x, n_{J-1} \leq n \leq n_{K(J)})$$

et on a

$$\sum_{J>0} P(M_n \neq x, n_{J-1} \leq n \leq n_{K(J)}) < 4\epsilon.$$

Cela prouve que E est un événement négligeable.

Ainsi, pour tout $x \in S$, la marche aléatoire passe presque sûrement infiniment souvent au point x . L'ensemble S étant dénombrable, on en déduit que, presque sûrement, la marche aléatoire passe infiniment souvent en chacun des points de S . La première partie de la proposition est démontrée.
— QED (Proposition 12.4)

12.4. Récurrence des marches aléatoires sur \mathbb{Z}^N . —

PROPOSITION 12.6. — *Une marche aléatoire sur \mathbb{Z} est récurrente si et seulement si elle est centrée.*

Démonstration. — La loi forte des grands nombres (Théorème 10.12, (10.5)) nous dit que, presque sûrement,

$$\lim_{n \rightarrow +\infty} \frac{M_n}{n} = E[Y_1].$$

On en déduit immédiatement que si la marche aléatoire n'est pas centrée, alors $\lim_{n \rightarrow +\infty} |M_n| = +\infty$, et donc la marche est transitoire.

Supposons à présent la marche centrée, et excluons le cas trivial des variables aléatoires identiquement nulles. Si les variables aléatoires Y_n ne peuvent prendre que deux valeurs, nous pouvons utiliser des arguments variés pour établir la récurrence. Par exemple, le théorème de MOIVRE-LAPLACE (Théorème 7.3) nous dit que, à une constante multiplicative près,

$$P(M_{2n} = 0) \sim \frac{1}{\sqrt{n}},$$

ce qui entraîne que la série de terme général $P(M_n = 0)$ est divergente. Le théorème 12.1 permet alors de conclure. Un autre argument est basé sur la loi du logarithme itéré qui nous assure que, presque sûrement,

$$\limsup_{n \rightarrow +\infty} M_n = +\infty \quad \text{et} \quad \liminf_{n \rightarrow +\infty} M_n = -\infty.$$

Les accroissements de la marche étant bornés, il est alors impossible que $\lim_{n \rightarrow +\infty} |M_n| = +\infty$ et, en utilisant le théorème 12.1 et la proposition 12.4, on conclut que la marche est récurrente.

Détaillons à présent un joli argument^{12.1} qui s'applique à toutes les marches aléatoires centrées sur \mathbb{Z} , et qui n'utilise que la loi faible des grands nombres.

D'après le lemme 12.5, on a

$$1 + \sum_{n=1}^{+\infty} \mathbf{P}(M_n = 0) \geq \frac{1}{2m+1} \sum_{n=1}^{+\infty} \mathbf{P}(|M_n| \leq m) .$$

On en déduit que, pour tout nombre entier $a > 0$,

$$1 + \sum_{n=1}^{+\infty} \mathbf{P}(M_n = 0) \geq \frac{1}{2m+1} \sum_{n=1}^{ma} \mathbf{P}(|M_n| \leq m) ,$$

puis que

$$1 + \sum_{n=1}^{+\infty} \mathbf{P}(M_n = 0) \geq \frac{1}{2m+1} \sum_{n=1}^{ma} \mathbf{P}\left(|M_n| \leq \frac{n}{a}\right) .$$

Or la loi faible des grands nombres (Théorème 10.12, (10.4)) nous assure que

$$\lim_{n \rightarrow +\infty} \mathbf{P}\left(|M_n| \leq \frac{n}{a}\right) = 1 ,$$

et cela entraîne que

$$\lim_{m \rightarrow +\infty} \frac{1}{2m+1} \sum_{n=1}^{ma} \mathbf{P}\left(|M_n| \leq \frac{n}{a}\right) = \frac{a}{2} .$$

Finalement on a

$$1 + \sum_{n=1}^{+\infty} \mathbf{P}(M_n = 0) \geq \frac{a}{2} ,$$

et, a étant arbitraire, cela établit que la série de terme général $\mathbf{P}(M_n = 0)$ est divergente. La marche aléatoire est récurrente.

—— QED (Proposition 12.6)

^{12.1} K.I. Chung et D. Ornstein, *On the recurrence of sums of random variables*, Bulletin of the American Mathematical Society, vol. 103, p.30-32, 1962.

Étudions à présent les marches aléatoires en dimension $N > 1$. Nous noterons $Y_n = (Y_n^1, Y_n^2, \dots, Y_n^N) \in \mathbb{Z}^N$ les pas de cette marche et $\sum_{k=1}^n Y_k = M_n = (M_n^1, M_n^2, \dots, M_n^N)$ la marche elle-même.

Si la marche aléatoire n'est pas centrée, alors il existe i entre 1 et N tel que $E[Y_n^i] \neq 0$ et la loi forte des grands nombres nous assure que, presque sûrement, $\lim_{n \rightarrow +\infty} |M_n^i| = +\infty$. Une telle marche est transitoire.

La question de savoir si les marches aléatoires centrées sont récurrentes fut résolue au milieu du siècle dernier^{12.2}. Appelons *dimension de la marche aléatoire* la dimension du sous-espace vectoriel de \mathbb{R}^N engendré par l'ensemble des pas admissibles $\{e_1, e_2, \dots, e_d\}$. Les résultats sont les suivants : toute marche aléatoire centrée de dimension inférieure ou égale à 2 est récurrente ; par contre, toute marche aléatoire de dimension supérieure ou égale à 3 est transitoire.

Le cas de la dimension 1 est résolu dans la proposition précédente. Faut de disposer des outils d'analyse nécessaires (en particulier en analyse de FOURIER), nous ne démontrerons pas ici les résultats précédents dans toute leur généralité. Nous nous limiterons à l'étude des marches aléatoires aux plus proches voisins.

Dans l'espace \mathbb{Z}^N chaque point a $2N$ plus proches voisins. Notons (b_1, b_2, \dots, b_N) la base canonique de \mathbb{R}^N . La *marche aléatoire aux plus proches voisins sur \mathbb{Z}^N* est définie par $d = 2N$ et

$$(e_1, e_2, \dots, e_d) = (b_1, -b_1, b_2, -b_2, \dots, b_N, -b_N).$$

Cette marche est centrée si et seulement si, pour tout i entre 1 et n , on a $p_{2i-1} = p_{2i}$. On supposera que, pour chaque i , on a p_{2i-1} ou p_{2i} non nul, ce qui assure que la dimension de cette marche est bien égale à N . Dans le cas particulier où tous les p_i sont égaux (et valent donc $1/2N$), on parle de la « marche aléatoire simple ». Ces marches simples sont celles qui ont été étudiées dans le travail précurseur de György POLYA^{12.3}.

PROPOSITION 12.7. — Soit $(M_n)_{n \geq 1}$ une marche aléatoire aux plus proches voisins dans l'espace \mathbb{Z}^N . Avec les notations précédentes, on a,

^{12.2} K.L. Chung et W. H. J. Fuchs, *On the distribution of values of sums of random variables*, Memoirs of the American Mathematical Society, No 6, 1951.

^{12.3} G. Polya, *Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz*, Mathematische Annalen, vol. 84, p. 149-160, 1921.

pour tout $n \geq 1$,

$$P(M_{2n-1} = 0) = 0,$$

et

$$P(M_{2n} = 0) = \sum_{k_1+k_2+\dots+k_N=n} \frac{(2n)!}{(k_1!k_2!\dots k_N!)^2} \prod_{i=1}^N (p_{2i-1} p_{2i})^{k_i},$$

où les k_i sont des nombres entiers naturels.

PROPOSITION 12.8. — *Toute marche aléatoire aux plus proches voisins, centrée, sur \mathbb{Z}^2 est récurrente.*

PROPOSITION 12.9. — *En dimension $N \geq 3$, toute marche aléatoire aux plus proches voisins est transitoire.*

Démonstration (Proposition 12.7) — Si ω est un élément de Ω tel que $M_m(\omega) = 0$, alors il existe k_1, k_2, \dots, k_N tels que la suite finie $(Y_1(\omega), Y_2(\omega), \dots, Y_m(\omega))$ prenne exactement k_i fois la valeur b_i et k_i fois la valeur $-b_i$, $1 \leq i \leq N$. Pour qu'il existe un tel ω , il est nécessaire que m soit pair et, en posant $m = 2n$, on aura $k_1 + k_2 + \dots + k_N = n$.

Soit $(E_1, F_1, E_2, F_2, \dots, E_N, F_N)$ une partition de l'ensemble des entiers entre 1 et $2n$ telle que, pour chaque i entre 1 et N , on ait $\text{card } E_i = \text{card } F_i = k_i$. Grâce à la propriété d'indépendance, l'événement « les indices j dans E_i sont ceux pour lesquels $Y_j = b_j$ et les indices j dans F_i sont ceux

pour lesquels $Y_j = -b_j$ » a une probabilité égale à $\prod_{i=1}^N (p_{2i-1} p_{2i})^{k_i}$.

La suite (k_1, k_2, \dots, k_N) étant fixée, combien existe-t-il de telles partitions ? Il en existe exactement

$$\binom{2n}{k_1} \cdot \binom{2n-k_1}{k_2} \cdot \binom{2(n-k_1)}{k_2} \cdot \binom{2(n-k_1)-k_2}{k_2} \dots \binom{2(n-k_1-k_2-\dots-k_{N-1})}{k_N}.$$

En effet $\binom{2n}{k_1}$ est le nombre de façon de choisir E_1 , puis, ce choix étant effectué, $\binom{2n-k_1}{k_2}$ est le nombre de façon de choisir F_1 , puis, ces choix étant effectués, $\binom{2(n-k_1)}{k_2}$ est le nombre de façon de choisir E_2 , etc... Une simplification évidente nous indique que le produit de coefficients binomiaux ci-dessus est égal à

$$\frac{(2n)!}{(k_1!k_2!\dots k_N!)^2}.$$

L'événement « la suite finie $(Y_1(\omega), Y_2(\omega), \dots, Y_n(\omega))$ prend exactement k_i fois la valeur b_i et k_i fois la valeur $-b_i$, $1 \leq i \leq N$ » a donc une probabilité égale à

$$\frac{(2n)!}{(k_1!k_2!\dots k_N!)^2} \prod_{i=1}^N (p_{2i-1} p_{2i})^{k_i},$$

et la conclusion en découle.

— QED (Proposition 12.7)

Dans l'étude de la dimension 2, nous donnerons une autre formule explicite pour les probabilités de retour en zéro. Cette formule sera déduite de résultats classiques sur les polynômes de LEGENDRE. Ces polynômes et leurs propriétés sont abondamment décrits dans des traités de mathématiques classiques, et nous nous limiterons ici à l'exposé de la formule que nous utiliserons et de sa justification rapide. Nous noterons, pour tout $n \in \mathbb{N}$,

$$g_n := \frac{(2n)!}{(2^n n!)^2}.$$

formule
de Rodrigues
polynômes
de Legendre

Nous adopterons la définition suivante, connue sous le nom de *formule de RODRIGUES* : pour chaque $n \in \mathbb{N}$, le *polynôme de LEGENDRE* de degré n , noté L_n , est le polynôme dérivé d'ordre n du polynôme $(x^2 - 1)^n$. Autrement dit

$$L_n(x) = \frac{d^n}{dx^n} (x^2 - 1)^n.$$

PROPOSITION 12.10. — Soit $n \in \mathbb{N}$. Pour tout nombre y non nul, on a

$$L_n\left(\frac{1}{2}\left(y + \frac{1}{y}\right)\right) = 2^n n! \sum_{k=0}^n g_k g_{n-k} y^{2k-n}.$$

Le point remarquable du développement donné dans le précédent énoncé est la positivité des coefficients.

Démonstration (Proposition 12.10) — Nous ne donnons ici que les étapes d'une vérification, laissant les détails calculatoires et classiques au lecteur.

La formule du binôme nous donne

$$(x^2 - 1)^n = \sum_{k=0}^n (-1)^k \binom{n}{k} x^{2(n-k)},$$

dont on déduit, par dérivation terme à terme, et en notant m la partie entière de $n/2$,

$$L_n(x) = n! \sum_{k=0}^m (-1)^k \frac{(2(n-k))!}{k!(n-k)!(n-2k)!} x^{n-2k}. \quad (12.2)$$

Le développement en série entière de la fonction $u \mapsto (1-u)^{-1/2}$ au voisinage de zéro s'écrit

$$\frac{1}{\sqrt{1-u}} = \sum_{n=0}^{+\infty} g_n u^n.$$

En posant $u = 2tx - t^2$ dans cette formule, en effectuant un petit calcul et en utilisant (12.2), on constate que développement en série entière de la fonction $t \mapsto (1-2xt+t^2)^{-1/2}$ au voisinage de zéro s'écrit

$$\frac{1}{\sqrt{1-2xt+t^2}} = \sum_{n=0}^{+\infty} \frac{1}{2^n n!} L_n(x) t^n. \quad (12.3)$$

Soient $y \neq 0$ et $x = \frac{1}{2} \left(y + \frac{1}{y} \right)$. On a

$$1 - 2xt + t^2 = (1 - yt) \left(1 - \frac{1}{y}t \right),$$

et donc

$$\frac{1}{\sqrt{1-2xt+t^2}} = \left(\sum_{n=0}^{+\infty} g_n y^n t^n \right) \left(\sum_{n=0}^{+\infty} g_n y^{-n} t^n \right).$$

En effectuant ce produit et en identifiant avec (12.3), on obtient

$$\frac{1}{2^n n!} L_n(x) = \sum_{k=0}^n g_k g_{n-k} y^{2k-n}.$$

— QED (Proposition 12.10)

Étudions la marche aléatoire centrée aux plus proches voisins en dimension $N = 2$. Sa loi est déterminée par l'unique paramètre p_1 puisque $p_2 = p_1$ et $p_3 = p_4 = \frac{1}{2} - p_1$. Notons simplement $p := p_1$ et $q := p_3$.

PROPOSITION 12.11. — Pour cette marche aléatoire $(M_n)_{n \geq 1}$, on a, pour tout n ,

$$\mathbf{P}(M_{2n} = 0) = g_n \sum_{k=0}^n g_k g_{n-k} (4p - 1)^{2k}.$$

Démonstration. — Partons de la proposition 12.7, qui s'écrit ici

$$\mathbf{P}(M_{2n} = 0) = \sum_{k=0}^n \frac{(2n)!}{(k!(n-k)!)^2} p^{2k} q^{2(n-k)} = 2^{2n} g_n \sum_{k=0}^n \binom{n}{k}^2 p^{2k} q^{2(n-k)}.$$

En observant les formules

$$(p^2 + x)^n = \sum_{k=0}^n \binom{n}{k} p^{2k} x^{n-k} \quad \text{et} \quad (q^2 + x)^n = \sum_{k=0}^n \binom{n}{k} q^{2(n-k)} x^k,$$

on constate que $\mathbf{P}(M_{2n} = 0)$ est égal au coefficient de x^n dans le polynôme

$$Q(x) := 2^{2n} g_n (p^2 + x)^n (q^2 + x)^n.$$

Dans le cas où $p = q (= \frac{1}{4})$, on en déduit immédiatement que

$$\mathbf{P}(M_{2n} = 0) = 2^{2n} g_n \binom{2n}{n} 16^{-n} = g_n^2.$$

Supposons $p \neq q$. On a $Q(x) = 2^{2n} g_n (p^2 q^2 + (p^2 + q^2)x + x^2)^n$ et

$$\mathbf{P}(M_{2n} = 0) = \frac{1}{n!} Q^{(n)}(0).$$

En mettant le trinôme $p^2 q^2 + (p^2 + q^2)x + x^2$ sous forme canonique, on arrive à

$$Q(x) = 2^{2n} g_n \left(\frac{p^2 - q^2}{2} \right)^{2n} \left(\left(\frac{2x + p^2 + q^2}{p^2 - q^2} \right)^2 - 1 \right)^n.$$

Le changement de variable affine $x' = \frac{2x + p^2 + q^2}{p^2 - q^2}$ nous ramène à la définition des polynômes de LEGENDRE et on obtient

$$Q^{(n)}(0) = 2^{2n} g_n \left(\frac{p^2 - q^2}{2} \right)^{2n} L_n \left(\frac{p^2 + q^2}{p^2 - q^2} \right).$$

Puisque $q = \frac{1}{2} - p$, cela s'écrit aussi

$$Q^{(n)}(0) = 2^{2n} g_n \left(\frac{4p-1}{8} \right)^n L_n \left(\frac{1}{2} \left(4p-1 + \frac{1}{4p-1} \right) \right).$$

La proposition 12.10 nous permet de conclure que

$$Q^{(n)}(0) = n! g_n \sum_{k=0}^n g_k g_{n-k} (4p-1)^{2k}.$$

— QED (Proposition 12.11)

Nous sommes maintenant en mesure de démontrer la récurrence de la marche centrée en dimension 2. Le principe de notre raisonnement sera le suivant : pour la marche aléatoire simple ($p = 1/4$) l'expression des probabilités de retour en zéro donne directement la récurrence ; quand $p \neq 1/4$, les probabilités de retour en zéro sont supérieures.

Démonstration (Proposition 12.8) — La proposition 12.11 nous dit en particulier que

$$\mathbf{P}(M_{2n} = 0) \geq g_n^2 = \frac{((2n)!)^2}{(2^n n!)^4}.$$

On remarque que g_n est la probabilité de retour en zéro en $2n$ pas pour la marche aléatoire simple en dimension 1. Et cette quantité a déjà été estimée par la formule de STIRLING. On a $g_n \sim 1/\sqrt{n\pi}$, et donc g_n^2 est le terme général d'une série divergente. Ainsi

$$\sum_{n=1}^{+\infty} \mathbf{P}(M_{2n} = 0) = +\infty$$

et le théorème 12.1 permet de conclure que la marche est récurrente.

— QED (Proposition 12.8)

Démonstration (Proposition 12.9) — Soit (M_n) une marche aléatoire centrée aux plus proches voisins de dimension $N \geq 3$. Nous allons montrer que

$$\mathbf{P}(M_{2n} = 0) = O\left(n^{-N/2}\right).$$

Cela assure que $\sum_n \mathbf{P}(M_n = 0) < +\infty$ et prouve que la marche est transitoire.

Détaillons l'argument pour $N = 3$, son extension au cas général étant immédiate. Nous utiliserons les deux estimations suivantes :

$$\text{pour tout } n \in \mathbb{N}, \quad \frac{1}{(n!)^2} \leq \frac{2^{2n}}{(2n)!}, \quad (12.4)$$

et

$$\text{il existe } c > 0 \text{ tel que, pour tout } n \in \mathbb{N}^*, \quad \frac{1}{(n!)^2} \leq c \frac{2^{2n}}{\sqrt{n}(2n)!}. \quad (12.5)$$

L'inégalité (12.4) est évidente puisque $(1+1)^{2n} \geq \binom{2n}{n}$. L'inégalité (12.5) est une conséquence de la formule de STIRLING qui assure que, quand n tend vers l'infini, $\frac{(2n)!}{2^{2n}(n!)^2} \sim \frac{1}{\sqrt{n\pi}}$.

Les paramètres de la marche aléatoire centrée en dimension $N = 3$ seront notés $p = p_1 = p_2$, $q = p_3 = p_4$ et $r = p_5 = p_6$. On a $p, q, r > 0$ et $p + q + r = \frac{1}{2}$. Nous savons que

$$P(M_{2n} = 0) = \sum_{i+j+k=n} \frac{(2n)!}{(i!j!k!)^2} p^{2i} q^{2j} r^{2k}.$$

Notons

$$I(p) := \left\{ i \in \mathbb{N} : 0 \leq i \leq n \text{ et } \left| p - \frac{i}{2n} \right| < \frac{p}{2} \right\},$$

$$A := \sum_{i+j+k=n, i \notin I(p)} \frac{(2n)!}{(i!j!k!)^2} p^{2i} q^{2j} r^{2k},$$

et

$$B := \sum_{i+j+k=n, i \in I(p), j \in I(q), k \in I(r)} \frac{(2n)!}{(i!j!k!)^2} p^{2i} q^{2j} r^{2k}.$$

En utilisant (12.4), on obtient

$$A \leq \sum_{i+j+k=n, i \notin I(p)} \frac{(2n)! 2^{2n}}{(2i)!(2j)!(2k)!} p^{2i} q^{2j} r^{2k},$$

ce qui peut s'écrire

$$A \leq \sum_{0 \leq i \leq n, i \notin I(p)} \frac{(2n)! 2^{2n}}{(2i)!} p^{2i} \sum_{j=0}^{n-i} \frac{1}{(2j)!(2(n-i-j))!} q^{2j} r^{2(n-i-j)}.$$

La formule du binôme nous assure que

$$\sum_{j=0}^{n-i} \frac{1}{(2j)!(2(n-i-j))!} q^{2j} r^{2(n-i-j)} \leq \frac{1}{(2(n-i))!} (q+r)^{2(n-i)},$$

d'où l'on déduit que

$$A \leq \sum_{0 \leq i \leq n, i \notin I(p)} \binom{2n}{2i} (2p)^{2i} (2(q+r))^{2(n-i)}. \quad (12.6)$$

En rappelant que si $i \notin I(p)$ alors $|\frac{2i}{2n} - 2p| \geq p$, on reconnaît dans le majorant de A une probabilité contrôlée par l'estimation des grands écarts : si on note S'_n une variable aléatoire qui suit la loi binomiale de paramètres n et $2p$, la majoration (12.6) entraîne que

$$A \leq P \left(\left| 2p - \frac{S'_n}{2n} \right| \geq p \right).$$

D'après le théorème 6.1 cette quantité tend vers zéro à vitesse exponentielle quand $n \rightarrow \infty$: il existe une constante $d > 0$ telle que $A \leq e^{-nd}$. Les sommes de $\frac{(2n)!}{(i!j!k!)^2} p^{2i} q^{2j} r^{2k}$ pour $j \notin I(q)$ ou pour $k \notin I(r)$ se traitent de la même façon. Il nous reste donc à étudier la quantité B .

Dans la somme définissant la quantité B , les indices de sommation i, j, k vérifient $i > np, j > nq$ et $k > nr$. En utilisant (12.5), on obtient

$$B \leq \sum_{i+j+k=n} (2n)! \frac{c^{2i}}{\sqrt{i}(2i)!} \frac{c^{2j}}{\sqrt{j}(2j)!} \frac{c^{2k}}{\sqrt{k}(2k)!} p^{2i} q^{2j} r^{2k},$$

et, en minorant les racines carrées,

$$B \leq \frac{c^3}{\sqrt{pqr}n^{3/2}} \sum_{i+j+k=n} \frac{(2n)!}{(2i)!(2j)!(2k)!} (2p)^{2i} (2q)^{2j} (2r)^{2k}.$$

Or, par une extension facile de la formule du binôme de NEWTON, on a

$$\sum_{i'+j'+k'=2n} \frac{(2n)!}{i'!j'!k'!} (2p)^{i'} (2q)^{j'} (2r)^{k'} = (2p+2q+2r)^{2n} = 1,$$

et on en déduit finalement que

$$B \leq \frac{c^3}{\sqrt{pqr}n^{3/2}}.$$

— QED (Proposition 12.9)

13. Épilogue

Certaines des lois mathématiques qui sont décrites dans *Pile ou Face* peuvent faire l'objet de vérifications expérimentales. C'est en particulier le cas de la loi faible des grands nombres, du théorème limite central et de la loi d'Arcsinus. Comment peut-on réaliser des expériences ? Il est vite fastidieux de lancer soit-même un grand nombre de fois une pièce de monnaie. Mais on peut utiliser des données existantes ou « faire jouer un ordinateur à pile ou face » : il existe des programmes qui imitent les tirages au hasard en produisant des suites de nombres *pseudo-aléatoires*. Ces *générateurs de nombres pseudo-aléatoires* et leurs utilisations sont présentés par exemple dans les livres de Donald Knuth et de Nicolas Bouleau cités dans la bibliographie.

Ici s'achève cet opuscule. Il ne constitue bien sûr qu'une étape dans la découverte du Calcul des Probabilités. Pour aller plus loin dans cette découverte, il est nécessaire d'adopter le formalisme de la théorie ensembliste de la mesure. Depuis les travaux fondamentaux de Andreï KOLMOGOROV^{13.1} cette théorie est en effet le socle sur lequel est bâtie la théorie des probabilités. Une fois ce formalisme adopté, le contenu de cet opuscule a trois prolongements naturels :

- il existe de nombreux autres théorèmes limites ;
- ces théorèmes s'étendent à des situations beaucoup plus générales que celle des suites de variables aléatoires indépendantes de BERNOULLI ;
- l'étude du jeu de pile ou face, et des marches aléatoires qui lui sont associées est une première approche du mouvement brownien, qui est leur analogue « à temps continu ».

^{13.1} A. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, 1933.

BIOGRAPHIES

Les pages se rapportant aux auteurs sont indiquées après leurs biographies, qui sont classées suivant les dates de naissance. Nous présentons les auteurs suivants, pour lesquels nous disposons de sources : Bernoulli (5), Bernstein (26), Bienaymé (14), Boole (16), Borel (22), Cantelli (23), Cesàro (20), Cramer (31), Fourier (10), Gauss (11), Hardy (25), Hausdorff (21), Huygens (3), Khintchine (32), Kolmogorov (33), Laplace (8), Lebesgue (24), Legendre (9), Lévy (28), Littlewood (27), Markov (19), de Moivre (6), Newton (4), Pascal (2), Poisson (12), Polya (30), Riemann (18), Rodrigues (13), Steinhaus (29), Stirling (7), Tchebychev (17), Wallis (1), Weierstrass (15).

1. WALLIS, JOHN : [Ashford (Angleterre), 1616–Oxford, 1703] — Au collège Emmanuel de Cambridge, il étudie la divinité et les mathématiques. Il est ordonné prêtre en 1640 et devient professeur à Oxford en 1649. Ses travaux mathématiques ouvrent la voie à la création du calcul infinitésimal par NEWTON. Il est le premier à définir correctement les exposants négatifs et fractionnaires. Les intégrales qui portent son nom apparaissent dans ses recherches de l'aire du disque. Il développe une approche analytique de l'étude des coniques. *cf.* page(s) 29.

2. PASCAL, BLAISE : [Clermont-Ferrand, 1623–Paris, 1662] — Profitant dès sa jeunesse de rencontres avec les plus grands savants de son temps, il montre un génie mathématique précoce : son *Essay pour les coniques* date de 1640. En 1642, il donne le premier modèle d'une machine à calculer, ancêtre de nos calculateurs modernes. Il se livre à des expériences sur la nature du vide. La découverte du *triangle arithmétique* qui porte aujourd'hui son nom date de 1653. Mais son génie est aussi littéraire et philosophique : *Les Pensées* et *les Provinciales* sont deux chefs-d'œuvre de la littérature française. Une question posée par son ami le chevalier de MÉRÉ le conduit à l'étude des probabilités, et, dans des échanges épistolaires avec FERMAT, il fonde les prémices du Calcul des Probabilités. *cf.* page(s) 11.

3. HUYGENS, CHRISTIAAN : [La Haye, 1629–La Haye, 1695] — Astronome, physicien et technicien, il découvre les anneaux de Saturne, travaille sur le pendule et la chute des corps, et invente le ressort spirale pour la fabrication des montres. En optique, il énonce la première théorie ondulatoire de la lumière. Son œuvre mathématique est aussi importante, en particulier ses travaux sur les courbes où il crée la théorie des enveloppes de familles de droites et la théorie des développantes, étudiant en particulier

la *chaînette* et la *tractrice*. Il contribue à la création du Calcul des Probabilités par ses échanges épistolaires avec PASCAL et FERMAT et la rédaction du premier traité sur le sujet (*De ratiociniis in ludo aleae*, titre que l'on peut traduire par *Raisonnement sur les jeux de dés*). A Paris en 1666, il est l'un des fondateurs de l'Académie des sciences. cf. page(s) 5, 6,

4. NEWTON, ISAAC : [Woolsthorpe (Angleterre), 1642–Kensington, 1727] — Savant précoce, il fait ses principales découvertes mathématiques avant 25 ans. Il est le fondateur, parallèlement à LEIBNITZ, du calcul différentiel et intégral. Ses recherches sur les fonctions et sur les courbes, en particulier les coniques et les cubiques sont aussi de première importance. La publication de ces travaux est plus tardive que leur découverte : *Philosophiæ naturalibus principia mathematica* est publié en 1687. Bien sûr, son œuvre en physique est tout aussi fondamentale : en optique, il décrit la nature de la lumière blanche ; en mécanique, il établit les lois de la gravitation universelle. cf. page(s) 11, 12, 17, 105.

5. BERNOULLI, JAKOB (OU JACQUES) : [Bâle, 1654–Bâle, 1705] — Avec son frère cadet, ses neveux et petits-neveux, il représente une grande famille de mathématiciens suisses d'origine hollandaise. Après avoir voyagé en Europe et rencontré d'éminents scientifiques de son époque, il devient professeur à l'université de Bâle en 1687. Il s'intéresse au calcul infinitésimal et à son application à l'étude des courbes. Il introduit la rigueur mathématique dans l'étude des convergences et donne la première démonstration de la loi faible des grands nombres pour le jeu de pile ou face. Son nom est resté attaché à des équations différentielles, une courbe (*la lemniscate*), des polynômes et une loi de probabilité. cf. page(s) 12, 14.

6. DE MOIVRE, ABRAHAM : [Vitry-le-François, 1667–Londres, 1754] — D'origine française, il s'installe à Londres à dix-huit ans. Il y rencontre NEWTON et devient vite un mathématicien reconnu. Ces principaux travaux portent sur le Calcul des Probabilités. Le premier, il caractérise l'indépendance de deux événements par le fait que la probabilité de leur conjonction est le produit de leurs probabilités. Il découvre la formule de STIRLING et en déduit l'expression limite de la loi binomiale décrite dans cet opuscule. Il développe l'utilisation de la représentation polaire des nombres complexes (d'où la formule qui porte son nom). Enfin, il s'intéresse aux applications des mathématiques en finance et en démographie. cf. page(s) 20, 24, 27, 30, 32, 36, 38, 96.

7. STIRLING, JAMES : [Garden (Écosse), 1692–Edimbourg, 1770] — Après des études à Oxford, il enseigne à Venise entre 1715 et 1725, puis à Londres à partir de 1726. On lui doit des résultats sur les courbes algébriques (une courbe de degré n est caractérisée par la donnée de $n(n+3)/2$ points) et des développements asymptotiques, en particulier celui de $\ln(n!)$. cf. page(s) 18, 19, 27, 30, 31, 43, 51, 53, 55, 103, 104.

8. LAPLACE, PIERRE SIMON : [Beaumont-en-Auge (Normandie), 1749–Paris, 1827] — Mathématicien et physicien, il participe à la création de l'École Polytechnique. Ses contributions à l'astronomie et à la théorie de la gravitation sont considérables : il

étudie en particulier l'origine et la stabilité du système solaire. Ses principaux travaux mathématiques portent sur la théorie des équations différentielles et des équations aux dérivées partielles, et sur la théorie des probabilités. Il y développe l'utilisation des densités continues et participe à la fondation des outils statistiques. Son monumental traité *Théorie analytique des probabilités* (1812), dans lequel il reprend et amplifie les travaux de ses prédécesseurs, aura une influence considérable tout au long du siècle. LAPLACE joua un rôle public important sous le premier Empire et sous la Restauration. cf. page(s) 20, 24, 27, 30, 32, 36, 38, 96.

9. LEGENDRE, ADRIEN-MARIE : [Toulouse, 1752–Paris, 1833] — Il est élève de l'abbé MARIE au collège Mazarin à Paris, puis professeur à l'École militaire de 1775 à 1780. Il gagne en 1782 le grand prix de l'Académie de Berlin pour un mémoire sur la trajectoire des projectiles dans les milieux résistants. À partir de 1795, il enseigne à l'École Normale Supérieure. Après ses travaux sur la mécanique, il apporte des contributions fondamentales à la théorie des nombres (cf. Pierre Damphousse, *Découvrir l'arithmétique*, Opuscule 1). Il fait aussi œuvre de pédagogue en remaniant les *Éléments* d'EUCLIDE en un ouvrage traduit en anglais qui fait abandonner l'usage du texte d'EUCLIDE dans les écoles de l'Amérique d'alors. Il est célèbre pour son travail sur les fonctions elliptiques, (et d'autres traités de géométrie, de mécanique, d'astronomie). Il est homme de grande honnêteté, et quand les travaux d'ABEL et de JACOBI sur les fonctions elliptiques sont connus, il admet leur supériorité. Pour son opposition aux pressions des politiques sur le monde scientifique, on lui retire toute retraite, et il finit sa vie dans la misère et l'abandon. cf. page(s) 99, 100, 102.

10. FOURIER, JEAN-BAPTISTE JOSEPH : [Auxerre, 1768–Paris, 1830] — «Fourier est un pur produit de la révolution française, et sa vie une sorte de film de l'histoire de France entre 1770 et 1830» (Jean-Pierre Kahane, *Séries de Fourier*, Cassini, 1998). Issu d'un milieu modeste, ce qui l'empêcha de réaliser la carrière militaire qu'il convoitait, il est engagé dans la Révolution et participe à la première promotion de l'École Normale Supérieure en 1794. Dès l'année suivante il enseigne dans la toute nouvelle École Polytechnique, en particulier la théorie des probabilités. Il participe à la campagne d'Égypte avec Napoléon et occupe des responsabilités politiques sous le Consulat et l'Empire. Ses principaux travaux mathématiques portent sur la résolution des équations différentielles qui régissent la propagation de la chaleur dans un solide. Pour ce faire il fonde la théorie du développement d'une fonction périodique en somme de série trigonométrique. L'analyse de FOURIER est devenue une branche importante des mathématiques. cf. page(s) 27, 98.

11. GAUSS, CARL FRIEDRICH : [Brunswick, 1777–Göttingen, 1855] — Un des plus grands mathématiciens de tous les temps, un savant dont on peut comparer la stature à celle de NEWTON, il est le père des mathématiques modernes en Allemagne. En plus de traités fondamentaux d'astronomie et de physique, il apporte des contributions majeures à presque tous les domaines des mathématiques. Il donne la première démonstration complète du *théorème fondamental de l'algèbre*. Ses principaux ouvrages *Disquisi-*

tiones arithmeticae et *Disquisitiones generales circa superficies curvas* portent sur la théorie des nombres et la géométrie différentielle. cf. page(s) 20, 21, 22.

12. POISSON, SIMÉON DENIS : [Pithiviers, 1781–Sceaux, 1840] — Ancien élève de l'École Polytechnique, où il fut remarqué par ses pairs LAGRANGE et LAPLACE, il est considéré comme un des fondateurs de la physique mathématique. Maîtrisant les techniques de l'analyse mathématique de son époque il en donne des applications à divers domaines : mécanique des fluides, mouvement des planètes, élasticité, théorie de la chaleur, électrostatique, probabilités. Dans ce dernier domaine, il montre en particulier comment, dans certaines conditions, la loi binomiale peut être approchée par la loi, dite de POISSON, définie par $P(S_n = k) = e^{-np}(np)^k/k!$ (cf. chap. 4). Il connut les honneurs académiques et politiques : professeur à l'École Polytechnique dès 1806, puis à la faculté des sciences de Paris, il entre à l'Académie des sciences en 1812, et il sera fait baron par Louis XVIII. cf. page(s) 14.

13. RODRIGUES, BENJAMIN OLINDE : [Bordeaux, 1794–Paris, 1851] — Ancien élève de l'École Normale Supérieure, ses principales contributions portent sur la géométrie : décomposition des déplacements de l'espace euclidien, théorie des surfaces. Sa formule décrivant les polynômes de LEGENDRE se généralise aux autres familles classiques de polynômes orthogonaux. cf. page(s) 100.

14. BIENAYMÉ, IRÉNÉE JULES : [Paris, 1796–Paris, 1878] — Statisticien, inspecteur général des Finances, il applique la théorie des probabilités aux calculs financiers et entre à l'Académie des sciences en 1852. cf. page(s) 7, 14, 16, 73, 80.

15. WEIERSTRASS, KARL THEODOR WILHELM : [Osterfelde (Westphalie), 1815–Berlin, 1897] — Ce mathématicien allemand, grand pédagogue, a une influence considérable. Il introduit dans l'analyse mathématique le langage et la rigueur modernes. Ses premiers travaux portent sur les intégrales elliptiques et les fonctions abéliennes. Il s'intéresse à la construction des nombres réels, dégage la notion de convergence uniforme et démontre son célèbre théorème d'approximation polynomiale. Il construit de nouvelles fonctions de la variable réelle ou complexe comme sommes de séries ou produits infinis. En algèbre linéaire il fonde la théorie moderne des déterminants. cf. page(s) 14.

16. BOOLE, GEORGE : [Lincoln (Angleterre), 1815–Cork (Irlande), 1864] — Fils d'un boutiquier anglais, autodidacte, il se lance à vingt ans dans la lecture de LAPLACE et LAGRANGE et travaille sur les équations différentielles. Il obtient un poste de professeur au Queens College de Cork en 1849. Il est le fondateur de la Logique en tant que branche des Mathématiques indépendante de la Philosophie. Il aperçoit très tôt un lien entre sa théorie de logique formelle et le Calcul des Probabilités. cf. page(s) 58.

17. TCHEBYCHEV, PAFNOUTI LVOVITCH : [Okatovo (Russie), 1821–Saint-Pétersbourg, 1894] — Sa thèse, soutenue en 1846, porte sur la théorie des probabilités. Professeur à l'université de Saint-Pétersbourg à partir de 1847, ses principaux travaux sont consacrés à la théorie des nombres. Il fait des avancées décisives sur l'étude de la répartition des nombres premiers parmi tous les entiers ; il démontre en particulier la conjecture de

BERTRAND : pour tout nombre entier $n \geq 1$, il existe un nombre premier entre n et $2n$. Il contribue aussi à la théorie de l'approximation des fonctions, en étudiant les polynômes qui portent aujourd'hui son nom. cf. page(s) 7, 14, 16, 73, 80.

18. RIEMANN, GEORG FRIEDRICH BERNHARD : [Breselenz (Allemagne), 1826–Selasca (Italie), 1866] — Un des génies mathématiques, dont l'œuvre est encore retentissante. Il étudie à Berlin et à Göttingen, où il passe sa thèse en 1851, sur les fondements de la théorie des fonctions d'une variable complexe. Ses travaux en analyse et en théorie des nombres sont fondamentaux et sa conjecture, baptisée *Hypothèse de RIEMANN*, est le problème le plus célèbre des mathématiques au début du troisième millénaire. Il définit la première notion d'intégrale pour les fonctions non nécessairement continues. Enfin son œuvre géométrique fonde une grande part des mathématiques et de la physique théorique modernes, en particulier la relativité et la topologie. cf. page(s) 1, 27, 32, 38, 39, 43, 56.

19. MARKOV, ANDREÏ ANDREÏEVITCH : [Riazan (Russie), 1856–Petrograd (Saint-Petersbourg), 1922] — Élève de TCHEBYCHEV à Saint-Petersbourg, il travaille d'abord en théorie des nombres et en analyse, avant d'aborder la théorie des probabilités. Il invente la notion de *chaînes de MARKOV*, qui sont des suites de variables aléatoires soumises à une condition généralisant l'indépendance. Il établit des théorèmes limites pour ces chaînes. Cette notion s'est révélée absolument capitale dans les applications des probabilités. Il est considéré comme le fondateur de la théorie des processus stochastiques. cf. page(s) 6, 17.

20. CESÀRO, ERNESTO : [Naples, 1859–Naples, 1906] — Mathématicien italien aux centres d'intérêt variés, professeur à l'université de Naples. Il étudie les liens entre l'arithmétique et les probabilités : il démontre que la probabilité pour que deux nombres entiers tirés au hasard soient premiers entre eux est égale à $6/\pi^2$ (énoncé à préciser !). Dans l'étude de la convergence des séries entières, il introduit la notion de convergence qui porte aujourd'hui son nom. cf. page(s) 53.

21. HAUSDORFF, FELIX : [Breslau, 1868–Bonn, 1942] — Mathématicien allemand, enseignant aux universités de Leipzig, puis Greisswald et Bonn, ses contributions à la théorie des ensembles, l'analyse fonctionnelle, la topologie et la théorie de la mesure sont fondamentales. Il laisse son nom à l'axiome de séparation en topologie, à une notion de distance entre parties compactes d'un espace métrique et à une notion très fine de mesure des ensembles. Sa fin sera tragique : menacé de camp d'extermination par le régime nazi, il se suicide avec sa femme et sa belle-sœur. cf. page(s) 76.

22. BOREL, ÉMILE : [Saint Affrique, 1871–Paris, 1956] — Ancien élève de l'École Normale Supérieure, il est un des créateurs de la théorie de la mesure, qui permet à LEBESGUE de fonder la théorie moderne de l'intégration des fonctions. Il est à l'origine d'une approche des probabilités basée sur la théorie de la mesure, qui est aujourd'hui unanimement adoptée. Entouré de nombreux collaborateurs, il dirige la rédaction d'un important *Traité du Calcul des Probabilités et de ses applications* (Gauthier-Villars,

Paris, années 20). Il donne des contributions importantes aux théories des fonctions d'une variable réelle et de la sommation des séries numériques, et il est un des fondateurs de la théorie des jeux. On lui doit aussi des écrits sur l'histoire des sciences, la philosophie, la psychologie, la pédagogie et l'économie politique. Député, ministre sous la troisième République et opposant au régime de Vichy, son activité politique a également été importante. cf. page(s) 6, 57, 61, 62, 67, 68, 69, 71, 72, 73, 74, 75, 78, 82, 83, 84, 88, 90, 93.

23. CANTELLI, FRANCESCO : [Palerme, 1875–1966] — Mathématicien italien, il enseigne à l'École Économique et Commerciale de Rome. Ses travaux portent sur le calcul des probabilités et la statistique mathématique. Il participe activement au débat sur les fondements mathématiques des probabilités, entre les points de vue fréquentiste et axiomatique. cf. page(s) 57, 68, 69, 71, 73, 74, 75, 77, 78, 82, 83, 84, 88, 90, 93.

24. LEBESGUE, HENRI : [Beauvais, 1875–Paris, 1941] — Élève de l'École Normale Supérieure, il suit les cours de BOREL. Dans sa thèse intitulée *Intégrale, longueur, aire*, soutenue en 1901, il fonde la théorie moderne de l'intégration. Cette théorie permettra à KOLMOGOROV de bâtir l'axiomatique moderne des probabilités. Ses contributions à la théorie des séries trigonométriques (sous quelles conditions une fonction périodique est-elle somme de sa série de FOURIER ?) sont importantes. Enseignant successivement à Nancy, Rennes et Poitiers, puis au Collège de France à partir de 1912, son influence de pédagogue est largement reconnue. cf. page(s) 65, 66, 67, 69, 74.

25. HARDY, GODFREY HAROLD : [Cranleigh (Angleterre), 1877–Cambridge, 1947] — Mathématicien anglais, professeur à Cambridge, son domaine de prédilection est la théorie analytique des nombres, c'est-à-dire l'application de la théorie des fonctions d'une variable réelle ou complexe à la théorie des nombres. Sa collaboration avec LITTLEWOOD est particulièrement fructueuse. Militant pour les *mathématiques pures*, il écrit, après plusieurs traités mathématiques, une intéressante autobiographie (*A Mathematician's Apology*, Cambridge, 1940. Ed. Belin, 1985, pour une traduction en français). cf. page(s) 76.

26. BERNSTEIN, SERGEÏ NATANOVITCH : [Odessa (Ukraine), 1880–Moscou, 1968] — Mathématicien ukrainien, il enseigne à l'université de Kharkov et à l'université de Paris, puis à celle de Leningrad (Saint-Pétersbourg) jusqu'en 1941. Ses travaux portent sur la théorie des probabilités, mais aussi sur les équations différentielles et l'analyse fonctionnelle. cf. page(s) 8, 14, 15, 16.

27. LITTLEWOOD, JOHN EDENSOR : [Rochester (Angleterre), 1885–Cambridge, 1977] — Mathématicien anglais, professeur à Cambridge, ses principales contributions portent sur la théorie analytique des nombres et l'étude des séries trigonométriques. Sa collaboration avec HARDY est particulièrement fructueuse. cf. page(s) 76.

28. LÉVY, PAUL : [Paris, 1886–Paris, 1971] — Issu d'une famille de mathématiciens, élève, puis professeur, à l'École Polytechnique, il est une figure majeure des mathématiques françaises du 20ème siècle. Ses travaux portent sur l'analyse fonctionnelle

et la théorie des équations différentielles, mais il est surtout connu pour sa formidable contribution au Calcul des Probabilités. Il systématise l'utilisation de la fonction caractéristique (c'est-à-dire la transformée de FOURIER des probabilités) pour l'étude des convergences du type théorème limite central ou loi d'Arcsinus, dont il donne des généralisations. Il étudie les processus stochastiques et introduit la notion de *martingale*. Avec WIENER, il est le père de la théorie moderne du mouvement brownien. Ses principaux livres, d'une très grande richesse, sont *Leçons d'analyse fonctionnelle* (1922), *Calcul des probabilités* (1925-1951), *Théorie de l'addition des variables aléatoires* (1937-1954), *Processus stochastiques et mouvement brownien* (1948). Ses réflexions autobiographiques (*Quelques aspects de la pensée d'un mathématicien*) sont également d'un grand intérêt. cf. page(s) 41. 42.

29. STEINHAUS, HUGO DYONIZY : [Jaslo (Galicie), 1887–Wroclaw (Pologne), 1972] — Mathématicien polonais, élève de HILBERT et très influencé par les idées de LEBESGUE. À la fin de la première guerre mondiale il crée avec BANACH la fructueuse école mathématique polonaise moderne, fondatrice de l'analyse fonctionnelle. Ses travaux portent sur les séries trigonométriques, la théorie de la mesure, la topologie et la théorie des jeux. En 1923, il propose la première modélisation rigoureuse du jeu de pile ou face infini basée sur la théorie de la mesure. cf. page(s) 74.

30. POLYA, GYÖRGY : [Budapest, 1887–Palo Alto (Californie), 1985] — D'origine hongroise, il fait ses études à Budapest, puis enseigne à Zurich de 1914 à 1940, date à laquelle il émigre aux États-Unis. Ses principaux travaux portent sur la théorie des nombres et la combinatoire, grandes spécialités hongroises, l'analyse fonctionnelle et les probabilités, avec l'étude des marches aléatoires et des fonctions caractéristiques. Plusieurs de ses ouvrages ont connu un très grand succès : *Inequalities* (cosigné avec HARDY et LITTLEWOOD), des livres d'exercices, et *How to solve it* une étude sur la démarche de résolution des problèmes. cf. page(s) 98.

31. CRAMER, KARL HARALD : [Stockholm, 1893–Stockholm, 1985] — Ce mathématicien suédois fut parallèlement enseignant-chercheur en mathématiques, et actuaire auprès d'une société d'assurance sur la vie. Ses travaux portent sur les probabilités, les statistiques et la théorie des jeux. Il est considéré comme un fondateur de la théorie mathématique des statistiques, rigoureusement bâtie sur la théorie des probabilités. cf. page(s) 36.

32. KHINTCHINE, ALEXANDRE IAKOBLEVITCH : [Kondrovo (Russie), 1894–Moscou 1959] — Un des grands maîtres de l'école probabiliste soviétique, il énonce la loi du logarithme itéré et développe, souvent en concurrence avec LÉVY, la théorie des processus stochastiques. Introduisant la notion de processus stationnaire, il est un des créateurs de la théorie ergodique. Ses travaux portent aussi sur la théorie de la mesure et la théorie des nombres, en particulier l'approximation diophantienne et les fractions continues. cf. page(s) 75, 76.

33. KOLMOGOROV, ANDREÏ NIKOLAÏEVICH : [Tambov (Russie), 1903–Moscou, 1987] — Un des grands maîtres des mathématiques soviétiques, il est célèbre dès 1922 pour ses

résultats sur les séries trigonométriques. La liste des domaines auxquels il s'intéresse est immense, et son apport est fondamental, en particulier en théorie des ensembles, en théorie des systèmes dynamiques (stabilité du système solaire) et en théorie de la complexité (informatique théorique). En Calcul des Probabilités, il est sans doute la figure majeure du siècle : il fonde la théorie sur des bases mathématiques solides et développe la théorie des processus stochastiques, en particulier des processus de MARKOV. Il s'intéresse aussi beaucoup à l'enseignement des mathématiques et joue un rôle important dans la politique éducative et scientifique de l'URSS. Il obtient les honneurs les plus élevés dans son pays et devient membre des plus grandes académies scientifiques internationales. *cf.* page(s) 74, 79, 106.

ÉLÉMENTS DE BIBLIOGRAPHIE

Le point de vue élémentaire et rigoureux adopté dans cet opuscule se retrouve dans le premier tome du livre de William Feller et dans le petit livre de Yakov Sinai. On trouvera de plus dans le livre de Feller des références bibliographiques précises.

Un survol plaisant de la théorie des probabilités est présenté dans le « Que-sais-je ? » de Paul Deheuvels.

Il existe de nombreux traités modernes et plus avancés sur les probabilités. Parmi ces ouvrages citons l'excellent petit livre de John Lamperti, les livres de William Feller, Leo Breiman, Patrick Billingsley et Richard Durrett. Des textes en français sont aussi disponibles en particulier le livre de Alfred Rényi et celui de Marie Duflo et Didier Dacunha-Castelle. Tous ces traités présentent les bases mathématiques du Calcul des Probabilités, l'exposé d'une large part de la théorie et donnent des références bibliographiques pour aller plus loin. Le livre de Frank Spitzer est consacré à l'étude des marches aléatoires.

- [1] BILLINGSLEY, P., *Probability and Measure*, Wiley, New York, 1979.
- [2] BOULEAU, N., *Probabilités de l'ingénieur, Variables aléatoires et simulation*, Hermann, Paris, 1986.
- [3] BREIMAN, L., *Probability*, Addison-Welsey, Reading (Massachusetts), 1968.
- [4] DACUNHA-CASTELLE, D. ET DUFLO, M., *Probabilités et Statistiques*, Masson, Paris, 1982.
- [5] DEHEUVELS, P., *La probabilité, le hasard et la certitude, Que sais-je ?*, Presses Universitaires de France, Paris, 1982.
- [6] DURRETT, R., *Probability : Theory and Examples*, Wadsworth & Brooks, Pacific Grove (California), 1991.
- [7] FELLER, W., *An Introduction to Probability Theory and Its Applications*, Wiley, New York, 1957.
- [8] KNUTH, D., *The Art of Computer Programming*, Vol. 2, *Random Numbers*, Addison-Welsey, Reading (Massachusetts), 1969.

- [9] LAMPERTI, J., *Probability, A Survey of the Mathematical Theory*, Benjamin, New York, 1966.
- [10] RÉNYI, A., *Calcul des Probabilités*, Dunod, Paris, 1966.
- [11] SINAI, Y., *Probability Theory, An Introduction Course*, Springer Textbook, Springer-Verlag, Berlin, 1992.
- [12] SPITZER, F., *Principles of Random Walks*, Van Nostrand, Princeton, 1964.

INDEX

- algèbre de Boole, [p. 58]
- Arcsinus (loi de —), [p. 41]
- Bernoulli (loi de —), [p. 12]
- Bienaymé-Tchebychev (inégalité de —), [p. 7]
- binomiale (loi —), [p. 11]
- binomiaux (coefficients —), [p. 11]
- centrée (marche aléatoire —), [p. 86]
- clan, [p. 58]
- développement b -adique, [p. 65]
- écarts modérés, [p. 35]
- écart type d'une variable aléatoire, [p. 7]
- espace probabilisé, [p. 3]
- espérance, [p. 5, p. 70]
- événement, [p. 2]
- événement de type fini, [p. 58]
- événement négligeable, [p. 58]
- événement presque sûr, [p. 59]
- événements indépendants, [p. 60]
- éventualités (ensemble des —), [p. 2]
- fonction indicatrice, [p. 2]
- Gauss (courbe de), [p. 21]
- Gauss (convergence vers la loi de —), [p. 20]
- grands écarts (ou grandes déviations), [p. 16]
- indépendantes (variables aléatoires —), [p. 9, p. 70]
- indépendantes (suite de variables aléatoires —), [p. 71]
- indépendants (événements —), [p. 7]
- indépendants (famille d'événements —), [p. 8]
- inégalité maximale, [p. 79]
- issue (probabilité d'une —), [p. 2]
- issues (ensemble des —), [p. 2]
- lemme de Borel-Cantelli (premier —), [p. 68]
- lemme de Borel-Cantelli (second —), [p. 73]
- loi des grands nombres, [p. 13]
- loi du logarithme itéré, [p. 75]

loi d'une variable aléatoire , [p. 5]
loi forte des grands nombres , [p. 57]
marche aléatoire, [p. 86]
marche aléatoire aux plus proches voisins sur \mathbb{Z} , [p. 86]
marche aléatoire aux plus proches voisins sur \mathbb{Z}^N , [p. 98]
marche aléatoire simple, [p. 42]
marche aléatoire simple sur \mathbb{Z} , [p. 87]
marche aléatoire simple sur \mathbb{Z}^N , [p. 98]
Markov (inégalité de —), [p. 6]
mesure, [p. 66]
Moivre-Laplace (théorème de —), [p. 30]
négligeable au sens de Lebesgue (ensemble —), [p. 65]
nombre normal en base b , [p. 67]
nombre normal , [p. 67]
polynômes d'approximation de Bernstein, [p. 14]
polynômes de Legendre , [p. 100]
principe de réflexion, [p. 44]
probabilité, [p. 3]
probabilité d'un événement, [p. 2]
probabilité d'un événement de type fini, [p. 58]
probabilité produit, [p. 5]
probabilité uniforme, [p. 3]
récurrente (marche aléatoire —), [p. 86]
retours en zéro (loi du nombre de —), [p. 42]
Rodrigues (formule de —), [p. 100]
Stirling (formule de —), [p. 27]
théorème limite central , [p. 21]
théorème limite central avec bornes mobiles , [p. 38]
transitoire (marche aléatoire —), [p. 86]
variable aléatoire, [p. 5]
variable aléatoire de type fini , [p. 69]
variance d'une variable aléatoire, [p. 7]
Wallis (intégrales de —), [p. 29]

Achevé d'imprimer en juin 2001
sur les presses de Normandie Roto Impression s.a.
à Lonrai (Orne)
N° d'imprimeur : 011433
Dépôt légal : juin 2001

Imprimé en France

OPUSCULES se compose de petits ouvrages portant chacun sur des points des programmes des universités et des classes préparatoires. La collection vise la solidité du développement mathématique, en se souciant du développement historique, niveau par niveau, dans des opuscules de références autonomes. Elle se veut sobre, pertinente, destinée aux étudiants, aux candidats des concours et aux esprits curieux des mathématiques.

Pile ou Face décrit les lois mathématiques que l'on doit observer quand on répète des expériences aléatoires, comme dans les jeux de hasard. Cet opuscule est une introduction au calcul des probabilités, qui est aujourd'hui une branche importante des mathématiques.

Le niveau requis pour aborder *Pile ou Face* est celui du cours d'analyse des deux premières années universitaires (ou des classes préparatoires scientifiques). Ce petit livre s'adresse aux étudiants désirant aborder le calcul des probabilités, aux futurs et actuels enseignants de mathématiques, mais aussi aux esprits curieux car le sujet traité allie utilité et esthétique.

Après une introduction rapide au formalisme du calcul des probabilités, les principaux "théorèmes limites" sont énoncés et démontrés, dans le cadre restreint mais déjà riche du jeu de pile ou face. On trouvera les attendues "lois des grands nombres", le "théorème limite central", la précise "loi du logarithme itéré", ainsi que les estimations des grands écarts, les lois d'arcsinus et l'étude de la récurrence des marches aléatoires, suivant leur dimension.