Fondements des probabilités et des croyances : une discussion des travaux de Cox et Smets

Isabelle Bloch
Ecole Nationale Supérieure des Télécommunications, Département Images
46 rue Barrault, 75634 Paris Cedex 13
Tél : (1) 45 81 75 85 - Fax : (1) 45 81 37 94 - Email : bloch@ima.enst.fr

RÉSUMÉ

Cet article est dédié à la comparaison des théories des probabilités et des croyances du point de vue des postulats qui les gouvernent. Ces postulats, qui permettent de justifier, selon les théories de Cox et de Smets respectivement, des propriétés souvent imposées axiomatiquement, ont des conséquences importantes sur les différentes étapes de la fusion numérique et permettent d'expliquer les différences entre les deux théories.

ABSTRACT

This paper aims at comparing probability and evidence theories from the point of view of the postulates which govern these theories. Following the work by Cox and by Smets respectively, these postulates lead to a justification of the main steps involved in numerical fusion and to an explanation of the differences between both theories.

1 Introduction

Les méthodes de fusion numérique, qu'elles soient probabilistes ou non probabilistes, diffèrent aux trois niveaux qui constituent classiquement le processus de fusion : celui de la modélisation des fonctions de confiance, celui de la combinaison de ces fonctions déterminées à partir d'informations fournies par plusieurs sources, et celui de la décision finale. Ces différences peuvent pour la plupart être expliquées par des différences fondamentales entre les théories, portant sur les postulats qui les gouvernent et qui imposent des contraintes sur les trois niveaux précédents. La comparaison entre les théories numériques de la fusion est rarement effectuée au niveau des postulats. Nous nous restreignons ici aux théories des probabilités et des croyances.

Selon la conception moderne des probabilités, les « contraintes » (d'additivité en particulier) peuvent être déduites d'un certain nombre de postulats de base dictés par l'intuition [1]. La modification des postulats de base pour dépasser les limites des probabilités conduit à des théories numériques différentes, ne satisfaisant plus les mêmes propriétés, et on retrouve ainsi des approches telles que les ensembles flous ou la théorie des croyances de Dempster-Shafer [3]. Il a souvent été reproché à cette dernière que la règle de combinaison orthogonale de Dempster n'avait pas de justification théorique. Plusieurs auteurs ont répondu à ces critiques, et nous présentons ici les arguments de Smets [5], permettant de déduire cette règle d'axiomes plus facilement justifiables. Nous établissons ensuite les liens entre ces axiomes et ceux de Cox [1], expliquant ainsi les origines des différences entre les deux théories.

2 Les axiomes de Cox

Plutôt que d'accepter les « axiomes » des probabilités tels qu'ils sont présentés par exemple dans l'approche classique de Kolmogorov, les approches plus subjectivistes partent de postulats intuitifs, directement liés à ce que l'on attend d'une logique inductive, dont les règles des probabilités sont ensuite déduites. Cette approche est due essentiellement à Cox [1], et est reprise de manière détaillée par exemple dans [6], [2]. Les postulats fondamentaux que pose Cox sont les suivants :

P1. cohérence ou non-contradiction,
P2. continuité de la méthode,
P3. universalité ou complétude,
P4. énoncés sans équivoque,
P5. pas de refus d'information et prise en compte

de la dépendance du contexte.

Les postulats P2 et P3 conduisent à utiliser des nombres réels pour représenter et comparer des degrés de confiance.

Le postulat P1 entraîne l'existence de relations fonctionnelles entre degrés de confiance.

Le postulat P4 impose que la logique symbolique classique, déductive, se retrouve comme cas particulier.

Le postulat P5 conduit au conditionnement hypothétique : le degré de confiance dans une proposition A n'est connu que conditionnellement à un état de connaissance e. Un tel degré de confiance est noté $[A|e]$.

Le postulat de cohérence et le conditionnement hypothétique imposent alors qu'il existe une équation fonctionnelle reliant $[AB|e]$ (degré de confiance dans A et B sachant e) et au moins deux des quantités $[A|e]$, $[A|Be]$, $[B|e]$, $[B|Ae]$, et qu'il existe une relation fonctionnelle entre les degrés de confiance dans une proposition $[A|e]$ et dans sa négation $[\bar{A}|e]$. Ces deux équations fonctionnelles sont :

$$p(AB|e) = p(A|Be)p(B|e), \tag{1}$$

$$p(A|e) + p(\bar{A}|e) = 1. \tag{2}$$

Ces relations, démontrées dans cette approche à partir des axiomes, étaient imposées axiomatiquement dans l'approche classique (Kolmogorov). De plus, on manipule d'emblée des probabilités conditionnelles (qui sont relatives à un état de connaissance), alors que celles-ci ne sont introduites qu'après dans la théorie classique. Enfin, on en déduit la relation donnant la probabilité de la réunion de deux événements et donc l'additivité des probabilités d'événements exclusifs (également imposée axiomatiquement dans la théorie classique). On en déduit aussi la règle de Bayes.

3 Les axiomes de Smets

Plusieurs travaux récents ont cherché à justifier la règle de combinaison de Dempster, en particulier ceux de Smets à partir d'un modèle de croyance transférable, ou communicable (« the transferable belief model ») [5].

La première constatation de Smets concerne le principe d'indifférence (ou principe de raison insuffisante). Les problèmes que ce principe pose en probabilités sont aisément levés dans le contexte de Dempster-Shafer avec des crédibilités. En effet, la fonction de masse définie par :

$$m(D) = 1 \text{ et } \forall A \subset D, A \neq D, \ m(A) = 0$$

(où D est l'espace de discernement), représente parfaitement l'indifférence (ou l'ignorance totale). Cette fonction de masse particulière jouera un rôle important dans la règle de combinaison puisqu'elle en est l'élément neutre, ce qui confirme son interprétation en termes d'ignorance totale, qui ne peut modifier aucune autre fonction de masse.

La deuxième idée de Smets est celle du modèle de croyance transférable, qui permet de définir le conditionnement. Le problème se pose de la manière suivante : étant donnée une information nouvelle permettant d'affirmer que la vérité se trouve dans un sous-ensemble B de l'espace de discernement D, comment modifier un jeu de masses m pour prendre en compte cette nouvelle information ? La formulation que propose Smets est la suivante :

$$\begin{aligned} m'(A) &= \sum_{X \subset \bar{B}} m(A \cup X) \ \ \forall A \subset B \\ &= 0 \ \text{ sinon} \end{aligned} \tag{3}$$

où m' désigne le nouveau jeu de masses.

Dans la théorie de Shafer [4], la formule de conditionnement est déduite de la règle de combinaison, alors qu'ici elle la précède et est simplement construite par des considérations logiques.

Dans une troisième étape, Smets définit des axiomes qu'il veut voir vérifiés par la règle de combinaison, notée $\oplus$:

A1 : $(Bel_1 \oplus Bel_2)(A)$ doit être une fonction de Bel_1, Bel_2 et de A seulement (où Bel_i désigne la crédibilité associée à la fonction de masse i) ;
A2 : $\oplus$ doit être commutative ;
A3 : $\oplus$ doit être associative ;
A4 : Si $m_2(B) = 1$, alors $m_1 \oplus m_2$ doit vérifier la loi du conditionnement (équation 3) ;
A5 : La loi doit vérifier une propriété de symétrie interne (invariance par permutation sur les événements) ;
A6 : Pour $A \neq D$, $(m_1 \oplus m_2)(A)$ ne dépend pas de $m_1(X)$ pour $X \subset \bar{A}$;
A7 : Il y a au moins 3 éléments dans D ;
A8 : La loi doit vérifier une propriété de continuité.

À partir de ces axiomes, Smets déduit la seule règle de combinaison possible satisfaisant ces axiomes en raisonnant sur les fonctions de communalité :

$$(q_1 \oplus q_2)(A) = q_1(A)q_2(A),$$

avec :

$$q_i(A) = \sum_{A \subset X} m_i(X).$$

Sa démonstration repose sur les propriétés des normes triangulaires et des fonctions absolument monotones. On retrouve bien la règle de Dempster-Shafer sur les communalités, et on en déduit la combinaison des fonctions de masse :

$$\begin{aligned}(m_1 \oplus m_2)(A) &= \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B) m_2(C)} \\ \text{pour } A \neq \emptyset, & \\ \text{et } (m_1 \oplus m_2)(\emptyset) &= 0\end{aligned}$$

et celle des fonctions de crédibilité.

4 Liens entre les axiomes de Cox et ceux de Smets

Dans cette partie, nous tâchons d'établir des liens entre les postulats de Cox et les axiomes de Smets, afin de montrer pourquoi ils conduisent à des théories différentes.

L'axiome A1, exprimant la dépendance entre des degrés de confiance et leur combinaison, est moins strict que les postulats de Cox. En effet, le postulat de cohérence P1 entraînait l'existence d'une relation définissant le degré de confiance dans AB n'impliquant que les degrés de confiance affectés à A et B mais pas à d'autres propositions. L'axiome de Smets, plus général, correspond à la possibilité offerte par la théorie de Dempster-Shafer de travailler sur des sous-ensembles et non plus simplement sur des singletons.

Les axiomes A2, A3 et A5 correspondent à des propriétés de la logique classique des propositions. Les postulats de Cox (en particulier le postulat P4) impliquent également que la logique déductive soit retrouvée comme cas particulier. Les deux approches coïncident donc sur ce point. Ces axiomes sont utilisés dans la démarche de Cox pour éliminer certaines formes de relations fonctionnelles entre $[AB|e]$ et les autres degrés de confiance, pour ne retenir que la seule forme cohérente avec la logique déductive :

$$[AB|e] = T([A|Be],[B|e]) = T([B|Ae],[A|e]).$$

De même ces axiomes sont utilisés dans la démonstration de Smets pour éliminer des dépendances et montrer que $(q_1 \oplus q_2)(A)$ ne dépend dans un premier temps que de A, et de $q_1(X)$ et $q_2(X)$ pour $X \subset A$, puis dans un deuxième temps que de $q_1(A)$ et $q_2(A)$.

L'axiome A4 (conditionnement) traduit une idée très proche de celle du conditionnement hypothétique déduite du postulat P5 de Cox. La différence essentielle est que le conditionnement y est exprimé plus comme une relation de compatibilité que comme une probabilité conditionnelle. Cette propriété de conditionnement est fortement utilisée dans les démonstrations de Smets, souvent conjointement avec l'associativité, pour éliminer des dépendances, grâce à des fonctions de masse particulières n'ayant qu'un élément focal conditionnant le résultat de la combinaison. Cette condition est plus souple que dans le cas des probabilités puisque le conditionnement par un sous-ensemble B ne limite pas les événements impliqués dans l'évaluation de $[A|B]$ à B mais fait intervenir tous les sous-ensembles inclus dans B.

On ne trouve pas dans les axiomes de Smets d'équivalent au postulat P3 de Cox (universalité). Cela se justifie par la base même de la théorie des croyances, où l'on ne caractérise plus les propositions par un seul nombre mais plutôt par deux (crédibilité et plausibilité), et où l'on accepte de ne pas affecter de degré de confiance à une proposition bien définie[1]. Cette souplesse permet de résoudre facilement des problèmes liés au manque d'information : si une source n'est pas capable de donner d'informations sur A mais qu'elle en donne par exemple sur $A \cup B$, cette situation est naturellement prise en compte par la théorie des croyances en affectant une masse à $A \cup B$ et pas à A, alors qu'elle nécessite souvent l'introduction d'hypothèses ou de modèles dans la théorie des probabilités pour pouvoir affecter un degré de confiance à A.

La déduction de Cox pour les probabilités implique un postulat de comparaison universelle des degrés de confiance, conduisant à l'utilisation d'un nombre réel pour quantifier le degré de confiance dans chaque proposition [2]. Dans la théorie des croyances, la comparaison peut être menée à deux niveaux, soit sur les crédibilités, soit sur les plausibilités, conduisant à des conclusions qui ne sont

1. Cela peut se faire par exemple en affectant une masse nulle à cette proposition A. Cela ne signifie pas pour autant qu'une confiance nulle soit attribuée à A, puisque la crédibilité $Bel(A)$ et la plausibilité $Pls(A)$ ne sont pas nécessairement nulles, des masses non nulles pouvant être affectées à des propositions B telles que $A \cap B \neq \emptyset$. Cela signifie simplement qu'on n'affecte pas de degré de confiance spécifiquement à A.

pas nécessairement équivalentes, en particulier dans l'étape de décision. Là encore, la théorie des croyances est moins contraignante.

L'axiome A6 de Smets n'entraîne pas que A et $\bar{A}$ soient interchangeables, alors que cette propriété est utilisée explicitement par Cox pour déduire la deuxième équation fonctionnelle (équation 2). En effet, des sous-ensembles X peuvent intervenir à la fois dans $(m_1 \oplus m_2)(A)$ et dans $(m_1 \oplus m_2)(\bar{A})$, et les contraintes sont donc beaucoup plus faibles que dans la théorie des probabilités. On ne peut donc pas en déduire de relation de complémentarité sur m. Celle-ci est remplacée par la relation de dualité entre la crédibilité et la plausibilité :

$$Bel(A) = 1 - Pls(\bar{A}).$$

Enfin, les axiomes A7 et A8 sont considérés par Smets lui-même comme des axiomes techniques permettant de faire les démonstrations. La régularité imposée aux fonctions est à mettre en parallèle avec les hypothèses de régularité faites pour les deux équations fonctionnelles de Cox (équations 1 et 2).

5 Conclusion

Cette comparaison explique pourquoi les deux séries d'axiomes conduisent à des théories différentes, et montre que le choix d'une théorie peut être raisonné et justifié en fonction des propriétés imposées par le problème posé et par l'interprétation que l'on souhaite donner aux degrés de confiance.

Partant d'une axiomatique de base come celle de Cox, quatre processus permettent de la faire évoluer vers une nouvelle axiomatique [3] : généralisation, spécialisation, incohérence interne, ou substitution. Ces quatre processus constituent, selon nous, un bon cadre de comparaison des théories. Comme nous l'avons vu, dans la théorie des croyances de Dempster-Shafer, les raisonnements sont effectués à partir de deux mesures et non d'une seule, qui permettent de représenter à la fois l'incertitude et l'imprécision, ce qui constitue donc une généralisation. Le postulat de la complétude (P3) est également généralisé puisqu'il peut exister des propositions bien définies auxquelles on n'affecte pas de degré de confiance. La dépendance du contexte est supprimée pour être remplacée par des relations de compatibilité. Enfin, la deuxième équation fonctionnelle sur la complémentarité (2) est remplacée par une relation de dualité entre la crédibilité et la plausibilité, mais il n'y a pas de relation directe générale entre $Bel(A)$ et $Bel(\bar{A})$.

Ces différences ont des conséquences d'abord au niveau de l'étape de modélisation puisque cette étape est fortement contrainte par les deux relations fonctionnelles (1 et 2) dans la fusion probabiliste alors que la théorie des croyances permet de s'adapter avec souplesse à beaucoup de situations (nous avons cité l'exemple de capteurs ne donnant d'information que sur la réunion de deux classes, sans les différencier). Au niveau de la combinaison des fonctions de confiance, les postulats imposent la règle de Bayes d'une part, la règle de Dempster d'autre part, et leurs différences proviennent en particulier des contraintes plus souples imposées par le conditionnement de Smets que par le conditionnement hypothétique de Cox. Enfin, au niveau de la prise de décision, étape ultime du processus de fusion, les différences proviennent surtout de la comparaison des degrés de confiance, laissant la place à plusieurs types de décision dans la théorie de Dempster-Shafer.

Références

[1] **R. T. Cox :** *Probability, Frequency and Reasonable Expectation, AmericanJournal of Physics*, Vol. 14, No-1, 1-14, 1946.

[2] **G. Demoment :** *Probabilités, modélisation des incertitudes, inférence logique, et traitement des données expérimentales*, Cours de l'Université de Paris-Sud, Centre d'Orsay, septembre 1993.

[3] **E. J. Horvitz, D. E. Heckerman, C. P. Langlotz :** *A Framework for Comparing Alternative Formalisms for Plausible Reasoning*, Proc. of the National Conference on Artificial Intelligence, 210-214, 1986

[4] **G. Shafer :** *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[5] **P. Smets :** *The Combination of Evidence in the Transferable Belief Model*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No-5, 447-458, 1990.

[6] **M. Tribus :** *Rational, Decriptions, Decisions and Designs*, Pergamon Press Inc., 1969.