

Représentation et synthèse de l'information experte

- Journées des thèses IRSN, 25-27 septembre 2006 -

Sébastien Destercke

1ère année de thèse

Méthodes de synthèse de l'information probabiliste imprécise

2 novembre 2005

thèse financée par l'IRSN

Directeur de thèse : Didier Dubois
Université Toulouse 3

Tuteur IRSN : Eric Chojnacki
DPAM/SEMIC/LEMI

1 Contexte et objectifs

Dans le domaine de l'analyse du risque, les paramètres d'un modèle (constantes, variables d'entrée) sont souvent connus de manière incertaine. Cette incertitude est de deux ordres : la variabilité et l'imprécision. La variabilité appartient à la nature aléatoire du phénomène et est en principe irréductible : c'est par exemple le temps de défaillance d'un composant électronique ou le pH d'un sol. L'imprécision est due à un manque de connaissance, d'expérience ou de données et peut être réduite par l'arrivée de nouvelles informations : c'est par exemple l'imprécision d'un capteur. Dans la pratique, ces deux types d'incertitude coexistent souvent (il suffit de penser au temps de défaillance d'un composant jamais testé).

Lorsque l'imprécision concernant un paramètre est très grande du fait de la pauvreté (voire de l'absence) de données le concernant, il est souvent fait appel à des experts qui donnent leurs opinions sur la valeur de ce paramètre. De par leur expérience et leurs connaissances, les experts apportent des informations précieuses qu'il aurait souvent été très coûteux ou impossible d'obtenir par d'autres moyens. De plus, les experts sont capables de prendre en compte des informations très disparates (e.g. des données géo-politiques pour un expert financier) et de donner des opinions dans des situations parfois mal définies. Cependant, les experts sont soumis à des biais non-systématiques inconnus des sources matérielles (biais psychologiques, pression sociale, ...).

Afin de pouvoir les utiliser dans un modèle, il est donc nécessaire de pouvoir représenter mathématiquement les opinions d'experts, d'évaluer leur qualité et de les synthétiser en une représentation utilisable pratiquement. Usuellement, c'est la théorie des probabilités qui est choisie pour réaliser ces trois opérations.

Cependant, la représentation, l'évaluation et la synthèse d'opinions d'experts au moyen de distributions uniques de probabilités pose des problèmes et limite fortement les possibilités d'action sur les opinions, en particulier lorsque ces opinions sont divergentes.

Dans la section 2, nous étudions la capacité de représentation de la théorie des probabilités et des autres théories présentées dans cet article. La section 3 présente ensuite la notion de pondération des experts. Enfin, la section 4 s'intéresse à la synthèse (i.e. fusion) des opinions de différents experts et à la manière dont cette synthèse est réalisable dans les différentes théories.

2 Représentation des connaissances

Nous considérons que l'expert exprime son opinion et transmet ses informations sous forme de quantiles ou d'intervalles de confiance. Les théories des probabilités, des possibilités et des fonctions de croyance peuvent être utilisées pour représenter ce type connaissance. Nous rappelons brièvement leur cadre avant de voir en quoi l'utilisation d'autres théories que celle des probabilités est utile.

2.1 Représentation des incertitudes

Les différentes théories permettant de représenter des incertitudes sont présentées de manière informelle. Pour leurs aspects théoriques, nous renvoyons le lecteur aux références bibliographiques.

2.1.1 Théorie des probabilités

L'incertitude est représentée par une distribution de probabilité. En cas d'imprécision ou d'ignorance partielle, c'est la distribution la moins informative parmi les distributions possibles qui est choisie (principe de maximum d'entropie).

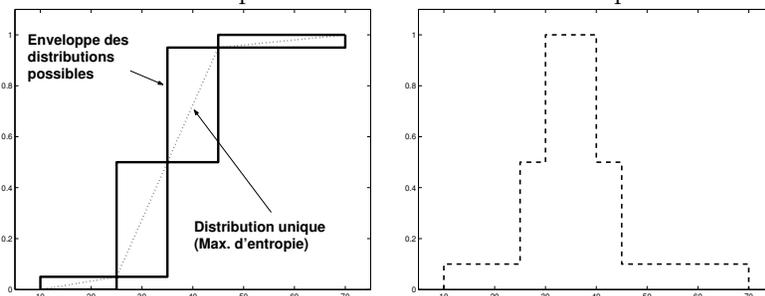
2.1.2 Théorie des possibilités [3]

L'incertitude est représentée par une distribution de possibilité, qui correspond à une opinion exprimée sous forme d'intervalles de confiances emboîtés.

2.1.3 Théorie des fonctions de croyance [5]

L'incertitude est représentée par des masses attribuées à des ensembles de valeurs (et non plus forcément à des valeurs uniques, ce qui est le cas des probabilités). Le formalisme des fonctions de croyance est très général et englobe ceux des possibilités et des probabilités.

FIG. 1 – Représentation des informations expertes



2.2 Représentation de l'information experte

Une fois que l'expert a exprimé son opinion sous forme de quantile ou d'intervalles de confiance, il faut représenter mathématiquement ces informations.

La représentation de ces informations par la distribution de probabilité la moins informative qui leur correspond revient à introduire des hypothèses supplémentaires et subjectives par rapport aux informations transmises par l'expert.

Prenons l'exemple d'un expert fournissant son avis sur la période d'épuration d'un sol par lavage. Il donne respectivement les valeurs 25, 35 et 45 ans pour les quantiles à 5%, 50% et 95%. Il existe évidemment une infinité de distributions de probabilités qui correspondent à ces informations. La figure 1.A montre les limites des distributions possibles ainsi que la distribution unique obtenue après application du principe de maximum d'entropie. On peut constater que choisir une distribution revient à ajouter beaucoup d'informations, alors qu'utiliser la théorie des fonctions de croyances permet de modéliser fidèlement la famille des distributions délimitée par l'information experte.

Imaginons que le même expert nous donne son avis sous forme d'intervalles de confiance : il donne l'intervalle $[30, 40]$ dont il est certain à 50%, et conserve l'intervalle $[25, 45]$ dont il est certain à 90%. Interrogé sur les valeurs maximales et minimales, il déclare ne pas pouvoir être certain de ces dernières, et préfère ne pas en fournir. A nouveau, son avis correspond à une infinité de distributions de probabilité, mais à une seule distribution de possibilité. La figure 1.B montre cette distribution de possibilité.

3 Pondération des experts

Pondérer les experts consiste à leur attribuer un poids qui évaluera la qualité de leurs opinions. Pour ce faire, une technique couramment utilisée consiste à demander aux experts, en plus de leurs opinions sur les paramètres inconnus, leur opinion sur des paramètres témoins dont la valeur est inconnue d'eux mais est (ou sera) connue du décideur.

Le poids à attribuer à un expert est ensuite calculée en fonction de son informativité (i.e. un expert qui sera très précis sera préféré à un expert très imprécis) et de sa calibration (i.e. l'adéquation entre ses opinions et les connaissances sur les paramètres témoins). Selon le formalisme adopté, ces deux notions se traduisent par des indices différents.

Ces poids seront ensuite utilisés pour départager les experts lorsque ceux-ci seront en conflit. Par manque de place, nous ne rentrerons pas dans les détails des opérations de pondération.

4 Synthèse (fusion) des opinions expertes

Une fois les opinions de différents experts obtenues et les experts pondérés, il est souhaitable de les synthétiser en un résumé facilement utilisable. Si cette synthèse est facile à réaliser lorsque les opinions expriment un certain consensus, ce n'est pas le cas lorsqu'elles sont conflictuelles. Comparativement aux probabilités, les théories non-probabilistes de l'incertain permettent de gérer le conflit de façon plus souple et diversifiée (ce qui permet au décideur d'adopter des comportements plus adaptés à ses besoins et à la situation). Nous allons illustrer ce fait en détaillant les approches probabilistes et possibilistes.

4.1 Approche probabiliste

Dans le cas des probabilités, les distributions données par les experts doivent être résumées en une seule distribution de probabilité. S'il existe de nombreuses méthodes pour fusionner des distributions [1], la plus utilisée en pratique est la somme pondérée : elle est simple, possède de bonnes propriétés et a prouvé son efficacité dans de nombreuses applications.

Soit k opinions expertes p_i ayant des poids w_i t.q. $\sum_{i=0}^k w_i = 1$, la distribution de probabilité résultant de la somme pondérée s'écrit alors $p(x) = \sum_{i=0}^k w_i p_i(x)$. Lorsque les experts n'ont pas pu être pondérés, les poids deviennent $w_i = 1/k$.

Prenons comme exemple un deuxième expert qui fournit comme opinion les valeurs 40, 50, 60 pour les trois quantiles à 5, 50, 95 %. Par manque de paramètres témoins, des poids égaux sont attribués aux deux experts.

FIG. 2 – Somme pondérée de distributions de probabilités

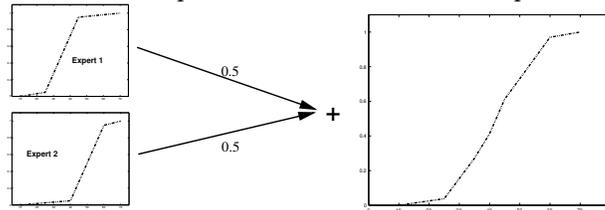
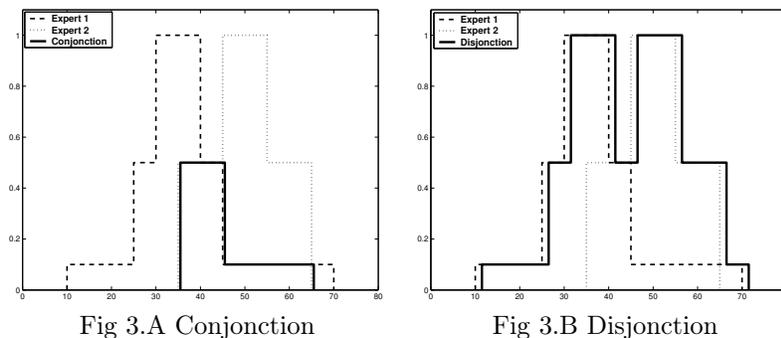


FIG. 3 – Modes de fusion non-probabilistes



Comme le montre la figure 2, le résultat de la combinaison est une distribution qui constitue un "compromis" entre les deux opinions expertes, résultat qui semble pertinent.

Il est intéressant de noter que si le deuxième expert avait donné comme opinion les valeurs 50, 55, 60, le résultat serait resté assez proche de celui de la figure 2 : le conflit entre les deux experts, pourtant manifeste, n'apparaît pas clairement dans le résultat, et peut passer inaperçu. Il serait donc souhaitable d'avoir d'autres méthodes mettant en évidence ce conflit, ce qui est très difficile si l'on veut rester au sein du domaine des probabilités.

4.2 Approches non-probabilistes

Supposons à présent que le deuxième expert nous ait donné les intervalles $[45, 55]$ et $[35, 65]$, dont il est respectivement certain à 50 et 100%. Les opinions des deux experts pouvant être considérées comme des ensembles de distributions de probabilité, elles peuvent être fusionnées selon trois modes principaux :

- La conjonction : elle consiste à prendre l'intersection de toutes les opinions. Le résultat est généralement très précis, mais peut aboutir à l'ensemble vide en cas de conflit important. Cette méthode devrait donc être utilisée lorsque les opinions expriment un certain consensus. La figure 3.A correspond à la conjonction des deux opinions.
- La disjonction : elle consiste à prendre l'union de toutes les opinions. Le résultat existe toujours, même en cas de conflit, mais est souvent très imprécis. La disjonction revient à faire l'hypothèse qu'au moins une opinion est juste, sans savoir laquelle exactement. La figure 3.B correspond à la disjonction des deux opinions.
- La moyenne pondérée : c'est l'équivalent de la moyenne réalisée pour les probabilités.

Comme nous le voyons, les notions de disjonction et de conjonction (absente en théorie des probabilités) permettent de mieux cerner et traiter les conflits.

Il est de plus possible de combiner disjonctions et conjonctions pour obtenir de nouveaux modes de fusion des opinions.

5 Conclusion et perspectives

Nous avons montré que, contrairement aux théories des probabilités, les théories non-probabilistes de l'incertain permettent de représenter plus fidèlement l'imprécision contenue dans les opinions expertes.

De même, en considérant des ensembles de probabilités et non plus des distributions uniques, il est possible d'utiliser bien plus de méthodes de fusion des opinions, ce qui permet de gérer le conflit entre opinions avec plus de souplesse.

Dans le cadre des opinions expertes dans l'analyse du risque, les théories non-probabilistes de l'incertain sont encore peu utilisées (contrairement aux probabilités [2]), et ce malgré des résultats encourageants [4]. Le travail qui consiste à confronter ces méthodes à des problèmes réels reste donc en grande partie à réaliser.

Notons également que les méthodes présentées ici peuvent être adaptées à d'autres types "d'experts" (e.g. codes de calcul) que les experts humains qui étaient le sujet principal de cet article.

Références

- [1] R. Clemen and R. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2) :187–203, 1999.
- [2] R. Cooke. *Experts in uncertainty*. Oxford University Press, Oxford, UK, 1991.
- [3] D. Dubois and H. Prade. *Possibility Theory : An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.
- [4] S. Sandri, D. Dubois, and H. Kalfsbeek. Elicitation, assessment and pooling of expert judgments using possibility theory. *IEEE transactions on fuzzy systems*, 3(3) :313–335, August 1995.
- [5] G. Shafer. *A mathematical Theory of Evidence*. Princeton University Press, 1976.