Incertitude, imprécision et additivité en fusion de données : point de vue historique

Uncertainty, Imprecision and Additivity in Data Fusion: Historical Point of View

par Isabelle BLOCH
Ecole Nationale Supérieure des Télécommunications
CNRS URA 820, Département Images
46 rue Barrault, 75634 Paris Cedex 13

résumé et mots clés

Cet article présente un aperçu historique de l'évolution de la notion de probabilité. Les différentes classes de probabilités et l'antagonisme entre objectivistes (ou fréquencistes) et subjectivistes sont présentés. On montre ensuite comment certains travaux, à la suite de ceux de Cox, s'attachent à déduire ce que d'autres appellent les axiomes des probabilités d'un certain nombre de postulats de base correspondant bien à l'intuition. Ces postulats peuvent ensuite être modifiés pour faire face aux limites des probabilités et donnent lieu à d'autres approches, non probabilistes, de la fusion de données (ensembles flous, possibilités, Dempster-Shafer par exemple). Dans la théorie des croyances de Dempster-Shafer, la règle de combinaison, présentée à l'origine sans véritable justification, peut également être déduite de postulats, comme le montrent les travaux de Smets. Nous établissons les liens entre ses postulats et ceux de Cox. Enfin, un exemple de combinaison de témoignages vient illustrer les différences entre les résultats obtenus selon quelques unes des méthodes proposées dans l'histoire des probabilités pour combiner des informations.

Épistémologie, Probabilités, Ensembles flous, Théorie des croyances de Dempster-Shafer, Fusion de données, Imprécision, Incertitude, Additivité.

abstract and key words

This paper presents an historical overview of the evolution of the notion of probability. The different kinds of probability and the opposition between objectivists (or frequencists) and subjectivists are presented. Then we show how some works, following the work by Cox, aim at deducing what is called by others axioms of probabilities, from a given set of basic postulates which well fit intuition. These postulates may then be modified in order to face the limits of probability and give rise to other data fusion approaches, non probabilistic (fuzzy sets, possibility theory, Dempster-Shafer evidence theory for instance). In Dempster-Shafer evidence theory, the orthogonal combination rule, originally presented without true justification, can also be deduced from postulates, as shown in the work by Smets. We establish links between his postulates and Cox's ones. At last, we illustrate by a simple example the different computation ways for combining evidences proposed in the history of probability.

History of sciences, Probability, Fuzzy sets, Dempster-Shafer evidence theory, Data fusion, Imprecision, Uncertainty, Additivity.

1. introduction

Parmi les méthodes de fusion de données, les méthodes numériques cherchant à modéliser l'imprécision et l'incertitude des données et de la connaissance sont largement employées pour des problèmes aussi variés que l'agrégation multi-critères, la combinaison de témoignages, ou encore la fusion d'images hétérogènes. Les méthodes probabilistes sont certainement les plus populaires, mais suscitent pourtant encore de nombreuses polémiques, en particulier entre approches fréquencistes[1], objectivistes, et approches subjectivistes. Si les subjectivistes semblent prendre le pas dans de nombreux domaines, les principes fréquencistes sont souvent d'une grande utilité pratique, en particulier dès

1. On trouve aussi le terme de «fréquentiste» dans la littérature, et il semble qu'il n'y ait pas encore de consensus sur cette terminologie.

qu'il s'agit d'apprendre une loi sur de grands échantillons, par exemple pour reconnaître des cultures dans une image aérienne.

Un parcours historique des différentes acceptions des probabilités permet d'expliquer les origines de ces polémiques et montre que le choix d'une approche peut être raisonné et justifié par le problème posé et l'interprétation que l'on souhaite donner aux probabilités. La partie 2 sera consacrée à ce point de vue historique et la partie 3 à la caractérisation des différentes classes de probabilités. Nous nous sommes pour cela largement inspirés des articles de synthèse cités en référence.

Il est remarquable que l'hypothèse d'additivité des probabilités[2], aujourd'hui communément admise, n'est apparue que très tard. Cette hypothèse est posée de manière axiomatique dans la théorie de Kolmogorov. Cependant les travaux de Cox montrent que ces « axiomes » peuvent être déduits d'un certain nombre de postulats de base dictés par l'intuition (partie 4).

Dans la partie 5, nous donnerons quelques exemples montrant les limites des probabilités additives, dues aux contraintes souvent trop fortes qu'elles imposent. La modification des postulats de base pour dépasser ces limites 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 (partie 6). 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 dans la partie 7 les arguments de Smets, permettant de déduire cette règle d'axiomes plus facilement justifiables. Nous établissons ensuite les liens entre ces axiomes et ceux de Cox, du point de vue de la fusion de données, expliquant ainsi les origines des différences entre les deux théories.

Enfin, nous proposons dans la partie 8 un exemple simple de combinaison de témoignages et montrons comment il aurait été traité au cours de l'histoire par les différentes approches. On pourra constater que les théories « modernes » non probabilistes étaient, au moins dans certains cas particuliers, connues depuis le 17e siècle et ont été « oubliées » ensuite.

2. les probabilités dans l'histoire

Ce sont les travaux de Shafer, et en particulier ses remarquables articles synthétiques d'histoire des sciences [28], [29], qui ont motivé cette étude. Nous nous en sommes largement inspirés dans la présentation historique qui suit. Ils en constituent la base, et des compléments ont été apportés à partir des articles ou ouvrages [5], [6], [8], [14], [17], [19], [20], [21], [25], [36], [39].

2.1. avant 1660

Dès l'antiquité apparaît l'opposition entre connaissance et opinion, que l'on trouve en particulier chez Platon, autour des années 400 avant notre ère, et des termes comme nécessaire, possible, probable commencent à être définis. On trouve par exemple chez Aristote (autour des années – 350) des assertions du type « si un événement est nécessaire, c'est que son contraire est impossible » (il faudra attendre la théorie des possibilités pour que soit développée une théorie cohérente capable de modéliser cette phrase) ou encore « le probable est ce qui se passe habituellement » (faisant référence à la répétition de phénomènes, base de la théorie fréquenciste). Les anciens distinguaient trois catégories epistémologiques. Dans la première, une connaissance certaine est possible. Cela correspond à la notion de connaissance ou de science chez Platon. La deuxième catégorie comprend les événements pour lesquels une connaissance probable est possible. On retrouve ici la notion d'opinion de Platon, et ainsi la probabilité apparaît comme un attribut de l'opinion. La troisième catégorie, absente de la philosophie de Platon, correspond aux événements pour lesquels aucune connaissance n'est possible, donc au domaine de l'aléatoire. Ce terme s'entend dans le sens de l'exclusion de régularité statistique. Transposées en termes plus modernes, ces notions correspondent, nous semble-t-il, à celles de déduction pour la première catégorie, et d'induction pour la deuxième. La troisième correspondrait à des phénomènes ne suivant (apparemment) aucune loi prédictible ou apprise. Cette troisième catégorie semble exclure toute possibilité d'une théorie mathématique des chances.

Ces notions restent toutefois très primitives et aucune théorie n'est encore développée. Cependant leur importance est incontestable puisqu'elles sont liées à la théorie de la connaissance, considérée comme essentielle, et dont se préoccupaient beaucoup les philosophes. Le fait que les probabilités soient considérées comme « guide de la vie » par Cicéron (– 60) l'atteste. Il est difficile de résister ici au plaisir de reprendre ces quelques phrases de Sénèque, citées dans [24] :

Voici en quoi nous ne sommes pas d'accord avec les Étrusques, spécialistes de l'interprétation des foudres. Selon nous, c'est parce qu'il y a collision de nuages que la foudre fait explosion. Selon eux, il n'y a collision que pour que l'explosion se fasse. (Comme ils rapportent tout à la divinité, ils sont persuadés, non pas que les foudres annoncent l'avenir parce qu'elles ont été formées, mais qu'elles se forment parce qu'elles doivent annoncer l'avenir. (Sénèque, Questions Naturelles, II, 32).

Ces phrases illustrent bien le caractère subjectif de l'opinion : il est impossible d'imaginer une expérience qui permettrait de prouver ou de réfuter l'une ou l'autre de ces opinions. Elles illustrent aussi la différence entre causalité et liens logiques, souvent confondus.

2. La relation d'additivité exprime que pour deux événements A et B exclusifs, la probabilité de la réunion notée $A + B$ vaut $p(A + B) = p(A) + p(B)$. En particulier, on en déduit que $p(A) + p(\bar{A}) = 1$, où $\bar{A}$ désigne le contraire de A (ou son complémentaire en termes ensemblistes).

Les probabilités traduisent des liens logiques mais pas de relation de causalité [6].

Puis ces catégories épistémologiques disparaissent, sans que l'on sache vraiment l'expliquer. A la renaissance, deux notions complètement indépendantes sont manipulées : celle de chance ou d'aléatoire, et celle de probabilité, vue comme attribut de l'opinion et à laquelle n'est attachée aucune valeur numérique. La notion de chance est fortement liée à la théorie des jeux. On en trouve les prémisses dans le *Purgatoire* de Dante (1310), où sont décrites les différentes sommes que l'on peut obtenir en lançant trois dés. La théorie des jeux est ensuite largement développée à la fin du 16e siècle et dans la première moitié du 17e siècle. Cardano (1560) et Galilée (1620) recensent les différents résultats qui peuvent se produire dans un jeu et comptent les cas où chacun des résultats se produit. Apparaît ici pour la première fois la notion de « cas équiprobables ». C'est à Pascal et Fermat que l'on attribue les origines de la théorie mathématique des probabilités (bien qu'ils n'emploient pas ce terme), puisque, dans leur correspondance (autour de 1654), ils commencent à résoudre les premiers problèmes non triviaux. En 1657, est publié le premier livre sur le sujet de la théorie des jeux, écrit par Huygens. Ces trois mathématiciens et philosophes essayent de résoudre le « problème des points » : un jeu entre deux joueurs nécessitant qu'un joueur ait trois points pour gagner reste inachevé ; comment alors partager équitablement les enjeux si un joueur a un point et l'autre en a deux? Ils expliquent avec des cas équiprobables pourquoi des fréquences proportionnelles apparaissent dans une grande série d'essais. Cependant, ils sont gênés dans la résolution de leur problème à cause du déterminisme que leur impose le christianisme. Les anciens au contraire acceptaient très bien l'indéterminisme. En résumé, tous ces travaux traitent du problème des jeux et en déduisent une théorie de la chance, mais ne parlent jamais de probabilité, même s'ils introduisent un vocabulaire un peu plus épistémologique. Il demeure une grande confusion entre statistiques et connaissance a priori, conduisant au mélange de deux classes de probabilités, à partir de fréquences (liées aux statistiques) et à partir des cas équiprobables (liés à une connaissance a priori).

2.2. vers la formulation mathématique bayésienne

Le premier lien entre la théorie des jeux et les probabilités apparaît en 1662 où il est introduit par Arnault dans l'*Art de Penser*. Arnault établit une analogie entre les jeux et la vie de tous les jours et suggère qu'un point de vue épistémologique des chances permettrait d'appliquer la théorie aux probabilités (toujours considérées comme attribut de l'opinion). Il s'arrête juste avant le concept de probabilité numérique, mais ses travaux marquent clairement un tournant dans l'évolution de la notion de probabilité. L'analogie entre jeux et vie est exploitée à la fin du 17e siècle par les démographes qui calculent des tables d'espérance de vie à l'aide de la théorie des jeux, mais sans introduire la notion de probabilité.

Une contribution d'un tout autre domaine est due à Leibniz qui propose dans le *De Conditionibus* (1665) de représenter les droits légaux des personnes par des nombres. L'absence de droit est représentée par 0, un droit pur par 1 et un droit conditionnel par une fraction entre 0 et 1. Cette classification des droits repose sur la condition sur laquelle est fondée le droit : une condition impossible conduit à l'absence de droit, si elle est nécessaire le droit est pur, si elle est contingente[3], le droit est conditionnel. Ces notions de contingence ou de nécessité se retrouveront chez J. Bernoulli pour le problème de la combinaison de témoignages. Leibniz propose de relier la probabilité d'existence de la condition à la « grandeur » du droit, et semblait donc se diriger vers une conception numérique des probabilités, sans s'inspirer de la théorie des jeux. Il ne devient familier avec cette théorie que plus tard. Si ses essais sur la chance n'apportent rien de neuf du point de vue mathématique, ils reconnaissent le lien entre probabilités et théorie des jeux.

Dans le domaine de la combinaison de témoignages, les travaux de Hooper en 1699 (*A Calculation of the Credibility of Human Testimony*) conduisent à la définition de fonctions de confiance non bayésiennes, représentant la crédibilité d'un témoin, ainsi qu'à deux règles de combinaison, l'une pour des témoignages successifs, et l'autre pour des témoignages simultanés. Ces deux règles, très populaires au 18e siècle, seront complètement abandonnées au 19e siècle.

Une des plus importantes contributions à la relation entre théorie des jeux et probabilités à la fin du 17e siècle est certainement celle de J. Bernoulli, en particulier dans son ouvrage *Ars Conjectandi* (publié en 1713). Bernoulli propose dès 1680 une théorie mathématique des probabilités et de leur combinaison. Il utilise la théorie des jeux pour calculer les probabilités mais conserve aux probabilités leur aspect épistémologique et leur rôle dans le jugement des individus. Ainsi, les premiers formalismes permettant de manipuler des probabilités numériques étaient consacrés à l'étude de probabilités subjectives! La plus grande partie de la théorie de Bernoulli est consacrée à des probabilités qui sont subjectives, et qui sont des mesures de la connaissance. Elles sont calculées à partir du concept d'« argument » et leurs propriétés dépendent de la nature de ces arguments. En particulier, elles ne sont pas toujours additives. Les règles de combinaison proposées par Bernoulli, plus complètes que celles de Hooper, prennent, selon les arguments, différentes formes, dont seules certaines correspondent aux règles probabilistes usuelles. Nous reviendrons dans la partie 8 sur ces règles. Dans la dernière partie de son travail, Bernoulli établit la fameuse loi des grands nombres. Cette loi permet d'estimer a posteriori une probabilité inconnue a priori à partir de l'observation des fréquences d'occurrence. Ce théorème porte donc plus sur les probabilités aléatoires que sur les probabilités épistémologiques (selon la distinction qu'introduira Lambert) puisqu'une chance

3. contingent est employé ici dans le sens : qui peut se produire ou non.

qui ne peut être connue qu'a posteriori n'est pas initialement une caractéristique de notre connaissance.

Les successeurs de Bernoulli simplifient sa théorie et en réduisent inconsciemment la portée. Tout d'abord, ils ne sont pas convaincus par Bernoulli dont la théorie, qui leur semble souvent compliquée, n'est pas aussi bien établie que celle des jeux. De plus, ils retiennent surtout la loi des grands nombres et identifient probabilité et chance d'apparaître, conduisant à une approche essentiellement fréquenciste. Citons, parmi les successeurs de Bernoulli, Montmort (*Essai d'analyse sur les jeux de hasard*, 1708) qui essaye d'appliquer la théorie des jeux à d'autres domaines et Moivre (*De Mensura Sortis*, 1711, *Doctrine of Chances*, 1718) où l'on trouve la première règle explicite d'additivité et une représentation des probabilités entre 0 et 1. Sa définition devient la définition classique. Ces deux auteurs parlent de probabilités mais leur théorie traite surtout des chances. Les probabilités y sont définies comme le rapport du nombre de cas favorables sur le nombre de cas possibles, mais les auteurs sont confrontés au problème du comptage des cas qui n'est pas possible dans tous les domaines. Notons que Moivre avait déjà découvert la loi gaussienne.

Au 18^{e} siècle, seul Lambert poursuit les travaux de Bernoulli et distingue probabilités aléatoires et probabilités épistémologiques. Les premières sont celles qui peuvent être connues soit a priori, comme dans la théorie des jeux, soit a posteriori, données par l'expérience. Les secondes sont affectées à des événements par inférence à partir d'effets ou de circonstances, et ont un caractère subjectif. Dans *Photometrica* (1760), il traite de la théorie des erreurs et propose une méthode connue aujourd'hui sous le nom de maximum de vraisemblance. Dans le *Neues Organon* (1764), il généralise la théorie des arguments de Bernoulli, corrige et généralise ses lois de combinaison et traite le cas des jeux, des syllogismes et des témoignages de divers types. Ses lois sont un cas particulier de la règle de combinaison de Dempster (voir partie 8), et là encore les probabilités sont non additives.

Puis les travaux de Bayes (*An Essay Towards Solving a Problem in the Doctrine of Chances*, 1763) proposent une synthèse qui fait disparaître la distinction entre probabilités aléatoires et probabilités épistémologiques. Bayes inverse le théorème de Bernoulli : alors que Bernoulli estime le nombre de succès à partir de la connaissance de la probabilité, Bayes cherche à calculer la probabilité connaissant le nombre de succès dans un échantillon et l'exprime en termes de probabilités initiales, finales et de vraisemblance. Il ne manipule que des probabilités additives. Bayes n'a jamais cherché à faire connaître ses travaux. Son essai n'a été retrouvé qu'après sa mort, avec d'autres travaux (en particulier sur les corps chargés électriquement) écrits avec des abbréviations qui n'ont pu être toutes déchiffrées [16]. Cela a conduit à s'interroger sur l'origine exacte du théorème de Bayes (on pourra consulter par exemple les travaux de Stigler [37], [38], qui propose une résolution ... bayésienne de cette interrogation).

Ces travaux sont repris par Laplace (*Théorie analytique des probabilités*, 1812). Cette époque voit l'essor des probabilités inverses, vues du point de vue subjectif. Cette théorie fait apparaître la distinction entre probabilité « initiale » d'une hypothèse (ou a priori), probabilité « finale » (après l'expérience), et probabilité de « vraisemblance » (probabilité de l'expérience sachant l'hypothèse). Laplace expose également le principe de raison insuffisante : des réalisations sont considérées équiprobables si l'on n'a aucune raison de penser autrement. Ce principe sera généralement adopté jusqu'au milieu du 20^{e} siècle.

2.3. la prédominance de l'approche fréquenciste : les « objectivistes »

Au 19^{e} siècle et au début du 20^{e} siècle, l'essor des sciences physiques conduit à négliger la modélisation du raisonnement humain. A cette époque apparaît une nouvelle discipline : la statistique. La notion de probabilité est alors reliée très souvent à l'observation de phénomènes physiques, à leur répétition dans des séquences longues. On reproche alors à la théorie de Bayes et de Laplace son côté subjectif, on l'accuse de manque de rigueur et on rejette la notion de probabilité a priori qui semble trop vague.

Les travaux de Cournot (1843), Ellis (1843), Venn (1866) définissent alors des probabilités physiques, en termes de fréquences. Comme le souligne Good [14], ces travaux sont confrontés à des problèmes insolubles. Par exemple, si en jetant une pièce, on observe la séquence de Pile (P) et Face (F) suivante : PFPFPFPFPF..., on en déduit que la probabilité d'avoir Pile est 1/2, mais cela ne permet pas de conclure sur l'honnêteté du jeu.

Un des problèmes soulevés par ces approches est celui de la longueur des séquences sur lesquelles sont calculées les fréquences. Elles doivent être longues, mais de combien? Les théories sont élaborées avec des séquences infinies, mais ne disent pas comment procéder en pratique. C'est le cas de la limite de Venn, des populations hypothétiques infinies de Fisher (1912), ou encore des séquences aléatoires infinies de von Mises (1919). Von Mises pose clairement la distinction entre théorie mathématique abstraite et l'application de cette théorie : la propriété essentielle de ses séquences aléatoires infinies doit être que la probabilité de succès doit être la même quelle que soit la sous-séquence (infinie), ce qui est une notion abstraite, et il se restreint en pratique aux domaines où cette définition est raisonnable. Son argument est qu'il n'est pas nécessaire de répéter effectivement indéfiniment l'expérience pour que la probabilité existe, et il se limite donc à des probabilités physiques et des processus aléatoires, en excluant les problèmes où l'on se demande par exemple quelle est la probabilité pour que X meure à 60 ans.

Le 19^{e} siècle voit également se développer la loi gaussienne. Déjà connue de Moivre, elle est obtenue par Gauss (1823) en utilisant le principe du maximum de vraisemblance dans des problèmes d'estimation de l'erreur d'observation. Au milieu du 19^{e} siècle, elle est retrouvée d'une part par Herschel à partir de considérations géométriques pour estimer des erreurs de mesure dans la position d'une étoile, et d'autre part par Maxwell lors de ses études sur les distributions de vitesses des molécules d'un gaz [6].

Cependant, malgré le fort contexte fréquenciste de cette époque, des distinctions proches de celles faites aujourd'hui sont formulées. Par exemple, Poisson, dans ses recherches sur la probabilité des jugements (1837), distingue chance et probabilité. La chance est caractéristique de l'événement lui-même, indépendamment de notre connaissance, alors que la probabilité est relative à notre connaissance. Cette distinction est voisine de celle faite par Lambert. La distinction entre probabilité objective et probabilité subjective est également explicite dans *l'Exposition de la Théorie des Chances et des Probabilités* de Cournot (1843). Mais, même si la distinction est affichée, les théories développées au 19^e siècle permettent seulement de résoudre les problèmes liés aux probabilités physiques ou objectives.

Les travaux de Boole constituent une sorte de lien entre les approches fréquencistes et épistémologiques. Dans son ouvrage *Laws of Thought* (1854), il essaye de combiner à un niveau épistémologique des évaluations faites localement sur les divers attributs de l'information. Il maintient que les probabilités sont déduites de fréquences, mais reconnaît l'impossibilité d'estimer, dans de nombreuses situations, les fréquences jointes, qui doivent alors être générées à un niveau subjectif.

2.4. 20^e siècle : retour au subjectivisme

Au 20^e siècle, les approches classiques continuent à être développées, avec des bases mathématiques de plus en plus solides, en particulier sous l'impulsion de Kolmogorov, et l'approche fréquenciste demeure très présente et forte (en particulier en traitement du signal et des images), bénéficiant des travaux de Neyman, Pearson, Feller [11]. Parallèlement, avec la naissance de l'intelligence artificielle et son importance grandissante, le raisonnement humain et sa modélisation regagnent de l'intérêt et conduisent les chercheurs à revenir à une conception plus subjectiviste des probabilités. Deux écoles voient le jour, celle des probabilités additives, et celle des probabilités non additives.

La première école s'appuie sur des postulats de base pour arriver avec plus de rigueur et moins d'arbitraire aux probabilités, à leurs propriétés, au théorème de Bayes, etc. On y trouve les travaux de Keynes [22], Kemble [21], Cox [5], Jaynes [19], Jeffreys [20], Tribus [39]. Nous détaillerons l'approche de Cox dans la partie 4. Celle de Tribus en est directement inspirée. Jaynes et Kemble se placent dans le domaine de la mécanique statistique et montrent que l'approche subjectiviste y est indispensable. Jeffreys, par un raisonnement du même type que celui de Cox, mais où les nombres sont introduits par convention, déduit les propriétés des probabilités d'un certain nombre de principes. Ces principes en rejettent certains autres considérés comme fondamentaux dans d'autres théories (par exemple les définitions des probabilités en termes d'ensemble infini d'observations possibles, en termes de propriétés du monde, le principe de causalité, etc.). L'essence de sa théorie est qu'aucune des probabilités directes, a priori ou a posteriori, n'est une fréquence. Même si la probabilité est calculée à partir d'une fréquence, elle n'est pas identique à la fréquence et un degré de confiance raisonnable est nécessaire avant son utilisation. Le but de la théorie de Jeffreys n'est pas de justifier l'induction mais d'en assurer la cohérence mathématique. De manière très proche, un grand nombre de chercheurs s'intéressent au point de vue philosophique des probabilités subjectives par rapport aux probabilités objectives, souvent sans remettre en cause l'additivité (Keynes, Jeffreys, Ramsey, de Finetti, Koopman, Russel, Carnap, Good, Savage, etc.). En particulier, les travaux de Savage et de Finetti montrent que l'approche subjective de la théorie bayésienne est la plus légitime et la plus cohérente [12]. Dans ses travaux, de Finetti adopte une approche résolument subjective (plus que celle de Cox) et raisonne en termes de cohérence de l'opinion des individus, et même en termes de psychologie collective afin d'expliquer la coïncidence des opinions d'individus différents. Cette approche particulièrement intéressante sera décrite succinctement à la fin de la partie 4.

La deuxième école remet complètement en cause l'additivité, en s'appuyant en particulier sur les travaux des pionniers tels que Bernoulli et Lambert [14], [29]. Koopman, dans les années 40, introduit la notion de probabilités inférieure et supérieure, définissant ainsi une probabilité subjective par une inégalité et non plus comme une valeur réelle précise, en s'inspirant des travaux de Boole (*Laws of Thought*, 1854), qui annonçaient déjà cette évolution. Il est suivi dans ses recherches par plusieurs autres chercheurs (Good, Dempster, etc.). En particulier, Dempster généralise les règles de Lambert, qui ne peuvent traiter que des arguments portant sur une seule conclusion, au cas où plusieurs hypothèses sont à envisager. Des applications de ces nouvelles théories se trouvent dans le domaine de l'économie, où Shackle par exemple propose des modèles économiques faisant appel à des notions proches de la théorie des possibilités, ou encore dans le domaine de la jurisprudence, avec en particulier les travaux d'Ekelof (*Rättegang*, 1963) qui propose trois lois de combinaison de témoignages : la combinaison de témoignages successifs et celle de témoignages simultanés concordants suivent les règles de Hooper, alors que la combinaison de témoignages conflictuels est reliée à un cas particulier de la règle de Lambert.

À partir des années 60 apparaissent des théories qui ne sont plus directement reliées aux probabilités. Zadeh invente les ensembles flous en 1965 [41], Shortliffe et Buchanan construisent le système MYCIN à partir de la notion de facteurs de certitude en 1975 [31], Shafer développe la théorie des croyances (*A Mathematical Theory of Evidence*, 1976) [27], et Zadeh introduit la théorie des possibilités en 1978 [43]. Ces nouvelles théories connaissent aujourd'hui un grand développement en intelligence artificielle et en fusion de données.

3. classes de probabilités, objectivistes et subjectivistes

La partie 2 fait apparaître plusieurs classes de probabilités, que nous synthétisons ici, en nous inspirant de la classification donnée par Good [14].

1. La définition classique est issue de la théorie des jeux et repose essentiellement sur la notion de cas équiprobables. Le calcul des probabilités se fait alors par dénombrement, en comptant les cas.
2. Une version plus subjective de cette définition introduit une information supplémentaire liée à la connaissance que l'on a, par exemple une connaissance sur l'honnêteté ou non du jeu. Cette deuxième classe ne considère donc que des probabilités conditionnelles. On peut inclure également dans cette classe les probabilités subjectives de Savage ou de Finetti, estimées proportionnellement à la somme d'argent qu'une personne serait disposée à donner si ce qu'elle affirmait se révélait faux [12], [8].
3. Une troisième classe est celle des probabilités inverses, suivant Bayes et Laplace. Il s'agit de la probabilité finale d'une hypothèse (après que des expériences ont été menées) estimée à partir de la probabilité a priori (en l'absence d'expériences) et de la probabilité conditionnelle (de vraisemblance), ou probabilité des expériences étant donnée l'hypothèse.
4. Les probabilités physiques utilisées au 19^{e} siècle n'ont plus de caractère subjectif et recherchent au contraire l'objectivité en calculant des probabilités conditionnelles à des expériences réalisées.
5. L'approche purement fréquenciste calcule des fréquences d'occurrence dans des grandes séries (limite de Venn, population infinie de Fisher, etc.).
6. Enfin, la dernière classe, que Good appelle probabilité subjective « néo-classique », est la plus large. Les probabilités représentent des degrés de confiance, relatifs à un état de connaissance, prenant en compte de l'information à la fois objective et subjective. Cette définition englobe toutes les autres et peut être beaucoup plus générale. Elle s'appuie sur une théorie mathématique fondée sur quelques axiomes, permettant d'assurer la cohérence de l'ensemble de degrés de confiance. Elle peut être étendue à une théorie du comportement rationnel en introduisant des utilités. Enfin, Good suggère même de représenter ces probabilités subjectives par des inégalités, ce qui s'approche de la théorie des croyances de Dempster-Shafer par exemple.

L'opposition entre objectivistes et subjectivistes provient en fait de différences fondamentales sur le type de problèmes qu'ils cherchent à résoudre et sur leur modélisation. En effet, les fréquencistes objectivistes recherchent des fréquences dans un ensemble, ce qui suppose la possibilité de répétitions infinies dans des conditions semblables, mais donne également un moyen opérationnel de calcul. La probabilité est caractéristique de l'ensemble et n'existe pas sans lui, mais les données peuvent être hypothétiques (il n'est pas forcément nécessaire de réaliser toutes les répétitions). Cela conduit les objectivistes à refuser des problèmes car dépourvus de sens, comme des événements qui ne se produisent qu'une fois. Des énoncés sont considérés objectifs s'ils peuvent être réfutables (par des contre-exemples), même s'ils ne peuvent pas être prouvés rigoureusement [24]. Au contraire, les subjectivistes considèrent les probabilités comme des mesures de confiance, d'espérance raisonnable [5], de codage numérique d'un état de connaissance [6], d'artifice mental approprié, et peuvent donc traiter des problèmes pour lesquels il n'existe pas d'ensemble, en particulier des phénomèmes uniques. Pour de tels phénomènes, il n'existe pas de probabilité en soi, mais seulement des modèles probabilistes [24]. Les hypothèses sont évaluées en fonction de données observées et de probabilités a priori, même si la connaissance est incomplète. Les subjectivistes ne cherchent pas le meilleur comportement asymptotique comme le font les statisticiens, mais cherchent à faire la meilleure inférence possible avec les données dont ils disposent [21], [6]. Les fréquencistes traitent donc de probabilités aléatoires et les subjectivistes de probabilités épistémologiques [28], [29]. Les premières sont caractéristiques de l'événement lui-même et ne sont pas modifiées quand la connaissance change [21]. Les secondes au contraire sont toujours conditionnelles et changent avec la connaissance. Elles permettent de formuler des conclusions possibles, entre la certitude et l'impossibilité, et constituent donc une logique étendue [22]. Les subjectivistes rejettent le principe selon lequel les mêmes causes ont les mêmes effets, non parce qu'ils le jugent faux, mais parce qu'il est dénué de sens, les causes n'étant jamais identiques. Curieusement, l'objectivité a été introduite pour éliminer l'arbitraire et la subjectivité de Bayes et Laplace mais a nécessité d'utiliser des critères statistiques qui ne sont pas universels et dont le choix ajoute à nouveau de l'arbitraire [6].

Enfin, la dernière différence entre les deux approches, à la fois mathématique et de signification, est fondamentale et concerne l'additivité. Les probabilités aléatoires sont nécessairement additives, puiqu'elles sont liées à l'aspect fréquenciste. Les probabilités épistémologiques ne le sont pas nécessairement, bien que cela soit toujours sujet à controverse. Nous reviendrons sur ce point dans les parties suivantes.

En résumé, trois types de personnes se préoccupent de probabilités : les mathématiciens proposent des modèles sans s'occuper de leur adéquation à une réalité ni de l'utilisation qui va en être faite, les physiciens déduisent des lois d'observations et d'expériences, et les philosophes s'interrogent sur le sens de tout cela.

4. postulats fondamentaux pour une logique inductive

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[4], dont ils déduisent les règles des probabilités. Cette approche est due essentiellement à Cox [5], et est reprise de manière détaillée par exemple dans [39], [6], [26], où l'on pourra également trouver les démonstrations des principaux résultats. Nous présentons ici ces postulats fondamentaux et les grandes lignes du raisonnement. Nous présentons succinctement à la fin de cette partie les travaux menés par de Finetti [12]. Moins connus des traiteurs des signaux et des images, ils ont cependant un intérêt double, à la fois par leur côté fondamentalement subjectiviste, et par la simplicité de la démonstration.

4.1. postulats fondamentaux

Les postulats fondamentaux que pose Cox sont les suivants [39] (ceux proposés par Jeffreys [20] en sont très proches[5]) :

1. **Cohérence ou non-contradiction** : si on peut établir une conclusion de plusieurs manières, elles doivent toutes conduire au même résultat ; on ne doit pas avoir de conclusions contradictoires à partir des mêmes données ; de plus, à des propositions qui ont toutes la même valeur de vérité, on doit attribuer des confiances égales.
2. **Continuité de la méthode** : les opérations effectuées doivent être continues, et si un changement faible des données intervient, il ne doit pas entraîner de changement brutal dans le résultat.
3. **Universalité ou complétude** : on doit pouvoir attribuer un degré de confiance à toute proposition bien définie, et les degrés de confiance doivent pouvoir être comparés.
4. **Énoncés sans équivoque** : les propositions doivent être bien définies, c'est-à-dire qu'il doit être théoriquement possible de déterminer si une proposition est vraie ou fausse. Cela correspond à ce que Horvitz appelle la clarté [17].
5. **Pas de refus d'information** : il ne faut pas tirer de conclusion à partir d'informations partielles, c'est-à-dire que toutes les informations, expériences ou connaissances disponibles relatives à la proposition à évaluer doivent être prises en compte, et, en particulier, il est important de tenir compte de la dépendance du contexte. Ce postulat répond aux théories classiques des probabilités, où, pour atteindre l'objectivité, certains types d'informations sont écartés.

Les postulats 2 et 3 conduisent à utiliser des nombres réels pour représenter et comparer des degrés de confiance : un seul nombre réel est nécessaire et suffisant pour représenter un degré de confiance, et on passe continûment du vrai au faux.

Le postulat 1 entraîne l'existence de relations fonctionnelles entre degrés de confiance.

Le postulat 4 impose que la logique symbolique classique, déductive, se retrouve comme cas particulier.

Le postulat 5 conduit au conditionnement hypothétique : le degré de confiance dans une proposition A n'est connu que conditionnellement à un état de connaissance e qui représente des informations reliées à la confiance dans A et qui sont supposées (ou crues) vraies. 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 T reliant $[AB|e]$ (degré de confiance dans « A et B » pour l'état de connaissance e) et au moins deux des quantités $[A|e]$, $[A|Be]$, $[B|e]$, $[B|Ae]$, et qu'il existe une relation fonctionnelle S entre les degrés de confiance dans une proposition $[A|e]$ et dans sa négation $[\overline{A}|e]$.

Paris [26], dans une démonstration rigoureuse du résultat de Cox, insiste sur une hypothèse souvent omise mais indispensable pour la démonstration :

$$\forall(\alpha,\beta,\gamma)\in[0,1]^3, \forall\varepsilon>0, \exists A,B,C,D,$$

vérifiant le postulat de cohérence, tels que

$$|[D|ABC]-\alpha|<\varepsilon, |[C|AB]-\beta|<\varepsilon, |[B|A]-\gamma|<\varepsilon.$$

En particulier, cette hypothèse ne peut pas être vérifiée sur un référentiel fini.

4.2. première équation fonctionnelle

Pour la relation T, 11 fonctions sont possibles (6 de 2 arguments, 4 de 3 arguments et 1 de 4 arguments). Pour des raisons de symétrie

4. La logique inductive cherche déterminer la solution la plus vraisemblable étant donnée l'information disponible, le vrai et le faux étant les cas extrêmes, par opposition à la logique déductive pour laquelle les seuls cas possibles sont le vrai, le faux et l'ignorance totale.

5. Jeffreys, cherchant définir une méthode générale d'induction, pose les postulats suivants : toutes les hypothèses doivent être exprimées, et les conclusions déduites des hypothèses ; la théorie doit être cohérente et non contradictoire ; toute règle doit pouvoir être applicable en pratique ; la théorie doit fournir des indicateurs d'inférence éventuellement fausse, ainsi que des possibilités de révision si de nouvelles informations sont disponibles ; la théorie ne doit pas refuser a priori d'information empirique. De plus, Jeffreys suggère de s'appuyer sur les guides suivants : le nombre de postulats doit être réduit à un minimum ; la théorie doit être en accord avec le raisonnement humain ; l'induction étant plus complexe que la déduction, on ne peut pas espérer la développer plus loin que la déduction. La démarche de Jeffreys consiste alors à traduire ces postulats en axiomes plus formels, à introduire les nombres pour représenter des probablités, et enfin à démontrer les résultats classiques [20].

des rôles de A et B, ces fonctions peuvent être réduites à 7 seulement. La suite du raisonnement permet d'éliminer certaines formes de fonctions en examinant des cas particuliers conduisant à des absurdités.

Si l'état de connaissance e stipule que « A et B sont indépendants », alors $[A|e] = [A|Be]$. Parmi les 7 formes de T, celle qui est fonction de $[A|e]$ et $[A|Be]$ n'est alors plus fonction que de $[A|e]$ et ne dépend plus de B. Cette forme doit donc être éliminée.

Si $e =$ « $A = B$ », alors $[AB|e] = [B|Ae] = i$ (où i représente le degré de confiance affecté aux propositions impossibles). Cette valeur constante élimine la forme de T qui est fonction de $[A|e]$ et $[B|e]$.

Si l'on examine maintenant le cas « A est impossible », alors $[AB|e] = [A|e] = [A|Be] = i$ et $[B|Ae]$ est indéfini. Ce cas permet d'éliminer les 4 formes de T fonctions respectivement de $[A|e]$ et $[B|Ae]$, de $[A|e]$, $[A|Be]$ et $[B|Ae]$, de $[B|e]$, $[B|Ae]$ et $[A|e]$, et de $[A|e]$, $[A|Be]$, $[B|e]$ et $[B|Ae]$.

La seule forme possible est donc :

$$[AB|e] = T([A|Be], [B|e]) = T([B|Ae], [A|e]), \qquad (1)$$

dans laquelle A et B jouent des rôles interchangeables, et où le postulat de continuité impose que T soit une fonction continue.

La logique déductive classique impose que pour trois propositions A, B et C on ait $(AB)C = A(BC)$. En appliquant cette règle, on en déduit que T doit être associative. La solution générale de cette équation fonctionnelle est un produit :

$$Kf[T([A|Be], [B|e]) = f([A|Be])f([B|e]), \qquad (2)$$

où K est une constante que l'on peut prendre égale à 1 par commodité, et f est une fonction monotone. La démonstration originale de ce résultat [5] suppose que T est deux fois différentiable. Cependant les résultats d'Aczél sur les équations fonctionnelles permettent de réduire ces hypothèses [1], [2] : il suffit que T soit associative, continue, et strictement croissante par rapport à chacun des arguments ; la différentiabilité n'est pas nécessaire[6].

Si l'on suppose maintenant « $A = B$ » alors $[A|Be] = c$ (où c est le degré de confiance attribué à une proposition certaine). On en déduit que $f(c) = 1$. De manière analogue, en supposant « $A = \bar{B}$ », on trouve que $f(i)$ doit être égal à 0 ou à $+\infty$. Par convention, on choisit $f(i) = 0$. La fonction f est donc une fonction positive et croissante de 0 à 1.

4.3. deuxième équation fonctionnelle

Examinons maintenant la relation fonctionnelle S entre $[A|e]$ et $[\bar{A}|e]$. En appliquant deux fois S, on obtient $S^2 = Id$. La cohérence avec la première équation fonctionnelle entraîne que S doit vérifier l'équation :

$$yS\left[\frac{S(x)}{y}\right] = xS\left[\frac{S(y)}{x}\right] \qquad (3)$$

dont la solution générale est :

$$f([A|e])^k + f([\overline{A}|e])^k = 1 \qquad (4)$$

en supposant S deux fois différentiable.

4.4. probabilités déduites des équations fonctionnelles

On pose alors par convention $p(A|e) = f([A|e])^k$, appelée probabilité de A conditionnellement à e. Les deux équations fonctionnelles deviennent alors :

$$p(AB|e) = p(A|Be)p(B|e), \qquad (5)$$

$$p(A|e) + p(\bar{A}|e) = 1. \qquad (6)$$

On a ainsi démontré les relations imposées axiomatiquement dans l'approche classique (Kolmogorov). De plus, on manipule d'emblée des probabilités conditionnelles (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 :

$$p(A + B|e) = p(A|e) + p(B|e) - p(AB|e), \qquad (7)$$

et donc l'additivité des probabilités d'événements exclusifs (également imposée axiomatiquement dans la théorie classique). On en déduit aussi la règle de Bayes :

$$p(A|Be) = \frac{p(B|Ae)p(A|e)}{p(B|e)}. \qquad (8)$$

Il faut noter que cette approche conduit à des probabilités subjectives qui sont additives, contrairement à la pensée du 17e siècle et à celle d'une des écoles du 20e siècle.

4.5. mesure d'incertitude et théorie de l'information

Dans une démarche similaire à celle de Cox, Jaynes définit une série de critères pour en déduire une mesure d'incertitude [19]. Dans ses travaux, il cherche à rapprocher mécanique statistique et théorie de l'information. Il exprime ses recherches comme un problème de spécification de probabilités dans le cas où l'on a peu d'informations. Reprenant les deux démarches objectivistes et subjectivistes, il retient la deuxième, qui permet, en représentant un état de connaissances, de formuler des conclusions possibles si l'on n'a pas assez d'information pour avoir des conclusions

6. D. Dubois et H. Prade ont montré que si l'on accepte que T soit seulement croissante au sens large, on peut choisir $T = \min$.

certaines. Elle est donc plus générale et Jaynes l'adopte pour la mécanique statistique.

Les critères intuitifs que Jaynes demande à une mesure d'incertitude sont les suivants :

1. la mesure doit être positive et continue,
2. elle doit augmenter quand l'incertitude augmente,
3. elle doit être additive si les sources sont indépendantes.

Il en déduit une mesure unique de l'incertitude représentée par une distribution de probabilité discrète qui corresponde à ces critères intuitifs :

$$H(p_1, \ldots, p_n) = -K \sum_{i=1}^{n} p_i, \log p_i. \tag{9}$$

De même que Cox retrouve les relations probabilistes à partir de ses postulats, Jaynes retrouve ainsi à partir de ses critères l'entropie de la mécanique statistique et simultanément celle de Shannon [30].

Le principe du maximum d'entropie peut alors être considéré comme l'analogue du principe de raison insuffisante de Laplace. La différence essentielle est que le principe de Laplace a un caractère arbitraire et peut engendrer des paradoxes (la notion de cas équiprobables change si on change de variable). Au contraire, le principe du maximum d'entropie permet d'effectuer des inférences sur la base d'une information partielle de manière non biaisée. Il peut être choisi pour la bonne raison que l'entropie est déterminée de manière unique comme la valeur qui « s'engage le moins » vis-à-vis de l'information manquante, et non pour la raison négative que l'on n'a pas de raison de penser autrement. Cependant, il peut être criticable car les résultats qu'il permet d'obtenir dépendent de la manière de poser le problème. Cette critique s'applique aussi au principe de raison insuffisante et nous y reviendrons dans la partie 5.

4.6. de Finetti et la théorie du pari

L'approche proposée par de Finetti, antérieure à celle de Cox, repose également sur des axiomes simples et intuitifs dont sont déduites les propriétés des probabilités [12]. On retrouve bien sûr les axiomes de croissance et de comparaison universelle, et surtout un axiome de cohérence sur lequel s'appuie l'essentiel de la démonstration.

De Finetti développe une théorie des paris pour expliquer son raisonnement : la probabilité p attribuée par un individu à un événement E est donnée par les conditions dans lesquelles cet individu serait prêt à parier sur cet événement, c'est-à-dire dans lesquelles il miserait la somme pS pour gagner S si l'événement E se produit. À partir de cette définition, de Finetti montre d'abord que la somme des probabilités d'événements incompatibles doit être égale à 1. Soit $\{E_1, \ldots E_n\}$ une classe complète d'événements incompatibles, p_i leurs probabilités (toujours évaluées par un individu), et si les enjeux correspondant à chacun d'eux. Si l'événement E_k est réalisé, le gain G_k est défini par la différence entre l'enjeu correspondant S_k et la somme des mises, soit :

$$G_k = S_k - \sum_{i=1}^{n} p_i S_i. \tag{10}$$

On obtient n équations de ce type, correspondant aux n réalisations possibles. En considérant ces équations comme un système de n équations à n inconnues qui sont les S_i, le déterminant de ce système vaut :

$$D = \begin{bmatrix} 1-p_1 & -p_2 & \ldots & -p_n \\ -p_1 & 1-p_2 & \ldots & -p_n \\ \ldots & \ldots & \ldots & \ldots \\ -p_1 & -p_2 & \ldots & 1-p_n \end{bmatrix} = 1-(p_1+p_2+\ldots+p_n). \tag{11}$$

Si le déterminant est non nul, le système admet une solution quels que soient les G_k, y compris s'ils sont tous positifs. Cela ne serait pas cohérent avec la notion de pari. On imagine mal un jeu auquel on puisse toujours gagner ou donner la possibilité à un adversaire de gagner sûrement! La seule solution cohérente est donc celle obtenue lorsque le déterminant est nul, c'est-à-dire lorsque :

$$\sum_{i=1}^{n} p_i = 1. \tag{12}$$

Cette condition est de plus suffisante, puisque l'on a alors :

$$\sum_{i=1}^{n} p_i G_i = 0, \tag{13}$$

et donc les gains ne sont pas tous positifs.

De Finetti interprète le résultat de la manière suivante : chaque évaluation (subjective) des p_i telle que $\sum_{i=1}^{n} p_i = 1$ est une évaluation admissible, c'est-à-dire correspondant à une opinion cohérente. Le choix d'une évaluation parmi celles qui sont admissibles n'a alors plus rien d'objectif.

On déduit de l'équation 12 l'additivité des probabilités d'événements disjoints.

La deuxième étape du raisonnement est dédiée à la détermination des probabilités conditionnelles. Pour cela, on considère trois paris :

1. un pari sur E_1E_2 (E_1 et E_2), d'enjeu S_1, et de mise $p_1\ S_1$,
2. un pari sur E_2, d'enjeu S_2, et de mise p_2S_2,
3. un pari sur $E_1|E_2$, d'enjeu S, et de mise pS, le gain pour ce pari étant :

– $(1-p)S$ si $E_1|E_2$ est vrai,

– $(-p)S$ si $E_1|E_2$ est faux,

– 0 si E_2 n'a pas lieu (on considère que le jeu est nul dans ce cas, et que la mise est remboursée).

Trois réalisations sont alors possibles :

1. si E_1 et E_2 se réalisent, le gain est alors :

$$G_1 = (1 - p_1)S_1 + (1 - p_2)S_2 + (1 - p)S\,; \qquad (14)$$

2. si E_2 se réalise mais pas E_1 le gain est alors :

$$G_2 = -p_1 S_1 + (1 - p_2)S_2 - pS\,; \qquad (15)$$

3. si E_2 ne se réalise pas, le gain est alors :

$$G_3 = -p_1 S_1 - p_2 S_2. \qquad (16)$$

Considérons les équations 14, 15 et 16 comme un système de 3 équations à 3 inconnues, comme précédemment. Le déterminant vaut :

$$p_1 - pp_2 \qquad (17)$$

et, toujours pour des raisons de cohérence, doit être égal à 0. On obtient alors la relation :

$$p(E_1|E_2) = \frac{p(E_1E_2)}{p(E_2)} \qquad (18)$$

dont on déduit le théorème de Bayes.

On remarque que dans ce cas, l'espérance de gain vaut :

$$p_1G_1 + (p_2 - p_1)G_2 + (1 - p_3)G_3 = (p_1 - p_2 p)S = 0 \quad (19)$$

Donc l'espérance de gain est nulle pour tous S, S_1, S_2.

De Finetti adopte une philosophie subjectiviste dans le sens où il considère que les éléments subjectivistes, loin de devoir être éliminés comme le suggèrent les objectivistes afin de rendre la notion de probabilité plus « scientifique », sont essentiels, et inhérents à la notion de probabilité. Cela rejoint le point de vue selon lequel la probabilité exprime l'opinion d'un individu, et n'a de signification que par rapport à cet individu, qui s'oppose au point de vue objectiviste qui considère que la probabilité existe en dehors des individus et est une propriété du monde physique.

5. insuffisance des probabilités additives

La large diffusion des méthodes probabilistes en fusion de données, en particulier des méthodes bayésiennes, est due, plus qu'à la justification qu'en donne Cox, à la connaissance acquise au cours d'expériences nombreuses permettant de guider les phases de modélisation et d'apprentissage. Les utilisateurs de la théorie bayésienne l'appliquent d'ailleurs souvent de manière uniquement formelle (mathématique), sans s'occuper de l'interprétation objectiviste ou subjectiviste des probabilités qui y interviennent. Cependant, et malgré leurs bases mathématiques solides, ces méthodes souffrent de plusieurs inconvénients. Nous les regroupons dans cette partie, en soulignant toutefois que certains d'entre eux sont contestés par les inconditionnels des probabilités.

Tout d'abord, la contrainte d'additivité peut être trop forte pour certains problèmes. Prenons l'exemple que donne Smets [32], dans le domaine du diagnostic médical. Si un symptôme s est toujours présent quand un patient a une pathologie A, et que l'on observe ce symptôme s, alors la probabilité pour que le patient ait A augmente. La contrainte d'additivité impose alors que la probabilité pour que le patient n'ait pas A diminue, alors qu'il n'y a pas de raison qui justifie cela (paradoxe de Hempel), si le symptôme s peut aussi être observé dans le cas d'autres pathologies[7].

L'application des méthodes bayésiennes nécessite souvent beaucoup de connaissances sur le problème, et leur utilisation dans de bonnes conditions suppose qu'une nouvelle réflexion soit menée pour chaque problème à traiter[8]. Par exemple, le diagnostic « bayésien » peut être formalisé de la manière suivante :

$$p(A_i|O) = \frac{p(O|A_i)p(A_i)}{\Sigma_j p(O|A_j)p(A_j)} = \frac{p(O|A_i)p(A_i)}{p(O|\bar{A}_i)p(\bar{A}_i) + p(O|A_i)p(A_i)}, \qquad (20)$$

où $p(A_i|O)$ désigne la probabilité pour que le patient ait la pathologie A_i, étant donné un ensemble d'observations O (examens cliniques, images, etc.), $p(O|A_i)$ désigne la probabilité conditionnelle des observations étant donnée la pathologie, et $p(A_i)$ est la probabilité a priori de A_i. La décision est prise à partir des $p(A_i|O)$. L'utilisation de cette formule nécessite soit de connaître l'ensemble de toutes les pathologies, soit d'avoir des statistiques reliant les observations à $\bar{A}_i$ (« non pathologie A_i »). Les deux solutions semblent irréalistes. De plus, toutes les distributions de probabilités intervenant dans la formule doivent pouvoir être estimées. Le problème est alors la limite des tests statis-

7. La contrainte d'additivité et la règle de Bayes (équation 8) impliquent que si $p(sA) = 1$, alors $p(A|s) = \frac{p(A)}{p(s)}$ et donc $p(A|s) \geq p(A)$. Au contraire, $p(\bar{A}|s) = 1 - p(A|s)$, donc $p(\bar{A}|s) \leq p(\bar{A})$. Si l'argument de Smets réfutant cette inégalité peut être criticable, il peut également être interprété de la manière suivante : le modèle des probabilités additives peut être trop simpliste dans ce cas. En particulier, si parler de la probabilité d'une pathologie A peut avoir un sens, il est moins certain que parler de celle de $[\bar{A}|s]$ en ait un. En effet, $\bar{A}$ ne correspond pas à UNE pathologie, mais est un ensemble infini, mal connu, imprécis, et il est difficile d'affirmer que $[\bar{A}|s]$ est une proposition bien définie, binaire, représentant correctement la réalité. Ainsi, tout modèle permettant de déduire $p(\bar{A}|s)$ peut être facilement contestable. Nous espérons ne pas trahir la pensée de Smets par cette interprétation.

8. Cette situation s'oppose à celle des approches fréquencistes, en particulier à la théorie de Fisher, qui est essentiellement automatique : *Faced with a new situation, the working statistician can apply maximum likelihood in an automatic fashion, with little chance (in experienced hands) of going far wrong and considerable chance of providing a nearly optimal inference. In short, he does not have to think a lot about the specific situation in order to get on towards its solution* [10]. C'est une théorie d'archétypes, qui, en permettant de séparer les problèmes, permet d'obtenir des solutions raisonnables là où l'approche bayésienne, où tout doit être traité «d'un seul coup», serait trop complexe. Ce caractère automatique est certainement une des raisons qui font la popularité de la théorie de Fisher, malgré les inconvénients des approches fréquencistes.

tiques reliant les symptômes ou les observations aux pathologies et la difficulté d'avoir une estimation des probabilités a priori. Ces limites sont bien sûr plus générales et ne sont pas spécifiques à cet exemple particulier.

L'estimation des probabilités a priori est souvent délicate et est primordiale dans les cas où l'on a peu d'informations (distributions très plates des probabilités conditionnelles). Si, dans le cas du traitement d'images, les probabilités conditionnelles peuvent être souvent bien estimées par apprentissage à partir de fréquences d'occurrence, ce n'est en général pas le cas des probabilités a priori. Leur évaluation sort du cadre des probabilités fréquencistes et fait souvent appel à des concepts plus subjectifs. De plus, la combinaison bayésienne est contrainte, comme pour la modélisation, par les axiomes des probabilités, et son utilisation en pratique nécessite souvent des hypothèses simplificatrices (comme l'indépendance) rarement vérifiées. La théorie probabiliste et bayésienne combine les informations de façon conjonctive, par des produits de probabilités conditionnelles, ce qui conduit souvent en pratique à un effondrement des probabilités des événements qui sont déduits d'une longue chaîne de déduction.

La modélisation probabiliste permet de raisonner sur des singletons seulement, qui représentent les différentes hypothèses, sous la contrainte du monde clos. Nous avons vu dans l'exemple précédent du diagnostic médical que cette hypothèse ne correspond pas à la réalité. De plus, les singletons ne permettent pas de représenter des situations complexes. Prenons le cas d'images affectées par l'effet de volume partiel (situation courante en imagerie médicale). Les modèles classiques dans la littérature pour représenter ce phénomène consistent à affecter à un point des probabilités d'appartenance aux tissus dont il est composé, ces probabilités étant proportionnelles à la quantité de chaque tissu dans le volume représenté par le point. Or cela ne correspond à aucune réalité. En effet, ce type de modèle probabiliste signifie que l'on a une incertitude sur la classe à laquelle appartient le point (on sait que le point peut appartenir à plusieurs classes mais on ne sait pas laquelle), alors qu'il s'agit en fait d'une appartenance à plusieurs classes à la fois. Cela nous semble être un exemple typique où les modèles probabilistes utilisés ne modélisent pas correctement le phénomène observé. Les ensembles flous ou la théorie de Dempster-Shafer permettent des modélisations beaucoup plus proches de la réalité, et d'interprétation moins contestable. Par exemple, les ensembles flous permettent de rendre exactement compte du phénome d'appartenance partielle qui est effectivement observé. Il ne suffit pas ici de relâcher la contrainte d'additivité des probabilités pour résoudre le problème : il s'agit bien de modéliser un type de phénomène tout à fait différent.

Une autre limite provient de la difficulté d'introduire dans le système de raisonnement des connaissances qui ne se traduisent pas simplement par des probabilités. Dans le même ordre d'idées, il est difficile de modéliser l'absence de connaissances, des connaissances imprécises (au contraire des connaissances incertaines qui sont naturellement représentées par des probabilités), ou encore l'ignorance que l'on peut avoir sur un phénomène. Le principe de raison insuffisante ne suffit pas à prendre en compte l'ignorance et peut conduire à des contradictions suivant la manière dont on l'exprime. Le même type de problème se pose avec le principe du maximum d'entropie. L'exemple bien connu de Shafer sur la probabilité pour qu'il y ait de la vie sur la planète Sirius l'illustre bien [27][9]. Ces inconvénients, qui ne sont pas mieux résolus avec une version subjective des probabilités, se manifestent dès qu'il s'agit de modéliser un raisonnement humain, où les décisions sont prises à partir de données à la fois imprécises et incertaines, partielles, pas complètement fiables, conflictuelles, où les contraintes et les objectifs ne sont pas toujours très précis.

Enfin, les opérations probabilistes de fusion de données sont très simples, puisqu'elles sont fondées essentiellement sur des produits, et ne permettent pas de modéliser une large gamme de comportements.

6. modification des postulats et approches non probabilistes

À ce point de notre exposé, nous avons constaté d'une part que les relations probabilistes se démontraient à partir d'une série de postulats de base, et d'autre part que la théorie des probabilités, même vue sous sa forme subjective la plus générale, avait des limites et des faiblesses pour traiter certains types de problèmes. Comment se situent alors les théories non probabilistes de la fusion de données dans ce contexte? Qu'il s'agisse des ensembles flous ou encore de la théorie des croyances de Dempster-Shafer, les fonctions de confiance utilisées ne visent pas à modéliser simplement une incertitude probabiliste et ont donc une interprétation complètement différente. Nous n'y reviendrons pas ici (on pourra par exemple consulter [3] pour une interprétation comparée de ces théories pour la fusion d'images), mais insistons sur le fait qu'elle est fortement liée aux postulats que satisfont ces théories, qui ne sont pas exactement les mêmes que ceux des probabilités.

Partant d'une axiomatique de base, quatre processus permettent de la faire évoluer vers une nouvelle axiomatique [17] :

1. **Généralisation** : certaines propriétés sont éliminées ou affaiblies.

9. Nous ne présentons pas ici l'exemple original de Shafer, qui peut prêter à discussion, mais un exemple donné par D. Dubois qui s'en approche. L'ignorance sur l'existence de la vie sur Sirius s'exprime classiquement en probabilités par $p(\text{vie}) = p(\text{non vie}) = 0,5$. Si l'on pose le problème autrement, en supposant qu'il peut y avoir trois possibilités, vie végétale, vie animale ou pas de vie, l'expression de l'ignorance conduira à affecter une probabilité de 1/3 à chacune des trois hypothèses. On obtient alors $p(\text{vie}) = 2/3$. De manière analogue, on peut obtenir autant de valeurs différentes que de manières d'exprimer le problème. On peut trouver des exemples analogues en traitement des images, en particulier dans les problèmes de classification.

2. **Spécialisation** : on ajoute de nouvelles propriétés fondamentales ou on renforce des propriétés existantes (en les rendant plus sévères).
3. **Incohérence interne** : on peut accepter une incohérence ou incompatibilité entre certaines propriétés. L'incohérence interne peut être entraînée par la spécialisation.
4. **Substitution** : une ou plusieurs propriétés sont remplacées par une autre.

Ces quatre processus constituent, selon nous, un bon cadre de comparaison des théories.

Considérons la théorie des ensembles flous [41]. Puisqu'elle modélise l'imprécision (plus que l'incertitude), elle est capable de manipuler des propositions non claires. Le postulat 4 imposant des « énoncés sans équivoque » est réfuté par Zadeh. Il est généralisé selon la terminologie de [17]. Une telle proposition prend une forme du type « *un des prochains congrès sur le flou aura vraisemblablement lieu en Asie du Sud-Est vers la fin du printemps et durera quelques jours* [40][10], ou encore l'exemple classique de Zadeh « *une urne contient approximativement n boules de tailles diverses dont plusieurs sont grandes, quelle est la probabilité pour qu'une boule tirée au hasard soit grande?* ». La théorie des probabilités d'événements flous permet de traiter de telles propositions [42]. De manière analogue, il existe des formalismes logiques dans lesquels plusieurs interprétations des assertions peuvent exister simultanément (donc avec des équivoques) [36]. Autre modification, aux relations déduites des deux équations fonctionnelles 5 et 6 (Bayes, réunion d'événements, additivité, etc., cf. partie 4.4) sont substituées d'autres règles de combinaison, plus conformes au type d'informations traitées par les ensembles flous et en particulier à la notion d'appartenance partielle. Les ensembles flous permettent donc de pallier certaines des limites des probabilités, en particulier des connaissances ou des données imprécises trouvent cette fois une modélisation naturelle, et une grande variété d'opérateurs de fusion, couvrant toute la gamme des comportements possibles, est disponible [4].

Dans la théorie des croyances de Dempster-Shafer [27], les raisonnements sont effectués à partir de deux mesures et non une seule (la crédibilité et la plausibilité), qui permettent de représenter à la fois l'incertitude et l'imprécision, ce qui constitue donc une généralisation. Le postulat de la complétude (postulat 3) est également généralisé puisqu'il peut exister des propositions bien définies auxquelles on n'affecte pas de degré de confiance (voir partie 7.3). La dépendance du contexte, entraînant la manipulation de probabilités conditionnelles, est supprimée pour être remplacée par des relations de compatibilité. L'absence de prise en compte du contexte peut être considérée comme une faiblesse de la théorie de Dempster-Shafer. Cependant, la notion de conditionnement existe chez Shafer [27], et Smets a proposé une généralisation du théorème de Bayes dans le cas d'une combinaison disjonctive [35]. De plus, en pratique, la dépendance du contexte pourrait être introduite dès les premières étapes, lors de la modélisation du problème, du choix des éléments focaux, des hypothèses simples et des hypothèses composées. Enfin, la deuxième équation fonctionnelle sur la complémentarité (équation 6) est remplacée par une relation de dualité entre la crédibilité et la plausibilité :

$$Pls(A) = 1 - Cr(\bar{A}), \qquad (21)$$

où Pls désigne la plausibilité et Cr la crédibilité, mais il n'y a pas de relation directe générale entre $Pls(A)$ et $Pls(\bar{A})$. Cette théorie permet de dépasser les limites de la modélisation probabiliste : prise en compte d'informations imprécises, de l'ignorance (en particulier, elle ne force pas le système à estimer des valeurs qui ne peuvent pas être connues), d'informations partielles, du conflit, pas de restriction aux singletons, pas de contrainte d'additivité, etc. Elle vise à construire un ensemble le plus cohérent possible à partir de connaissances et de données imparfaites. En revanche, la règle de fusion reste imposée, avec un seul type de comportement possible. Nous reviendrons plus en détails sur cette théorie et ses postulats dans la partie 7.

La théorie des possibilités [43] peut subir une analyse similaire à celles des ensembles flous et de la théorie des croyances, puisque, fondée sur des ensembles flous, les possibilités permettent, comme les croyances, de représenter imprécision et incertitude. En effet, une distribution de possibilité π est une fonction d'appartenance à un certain ensemble flou ; la possibilité Π et la nécessité N sont des cas particuliers de la plausibilité et de la crédibilité respectivement. La substitution, au sens de [17], par rapport aux probabilités, réside dans les expressions :

$$\Pi(A \cup B) = \max[\Pi(A), \Pi(B)], \qquad (22)$$
$$N(A \cap B) = \min[N(A), N(B)], \qquad (23)$$

qui viennent remplacer $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Enfin, un exemple d'incohérence interne peut être trouvé dans le système MYCIN [31], [15], [17], [18], puisque l'interprétation probabiliste utilisée dans la construction des facteurs de certitude est partiellement incompatible avec certains modes de combinaison. Nous ne le détaillerons pas ici.

7. déduction axiomatique de la règle de combinaison de Dempster-Shafer

Il a souvent été reproché à la théorie des croyances de Dempster-Shafer d'imposer une règle de combinaison (la règle orthogonale de Dempster [27]) « ad hoc », sans justification théorique. Rappelons que cette règle permet de combiner des jeux de masses

10. Cet exemple, imaginé part J.-P. Tubach (Val Thorens, janvier 1994), lui a valu le premier prix des «jokes about fuzzy signal processing» au congrès ICASSP (Adelaïde, Australie, avril 1994).

$m_1, m_2, \dots m_n$, définis sur un même espace de discernement D selon la formule (pour $A \neq \emptyset$) :

$$(m_1 \oplus m_2 \oplus \dots \oplus m_n)(A) = \frac{\Sigma_{B_1 \cap \dots \cap B_n = A} m_1(B_1) m_2(B_2) \dots m_n(B_n)}{1 - \Sigma_{B_1 \cap \dots \cap B_n = \emptyset} m_1(B_1) m_2(B_2) \dots m_n(B_n)},$$

$$\text{et } (m_1 \oplus m_2 \oplus \dots \oplus m_n)(\emptyset) = 0 \quad (24)$$

si cette expression est définie, c'est-à-dire si :

$$\sum_{B_1 \cap \dots \cap B_n = \emptyset} m_1(B_1) m_2(B_2) \dots m_n(B_n) < 1, \quad (25)$$

les B_i étant des sous-ensembles de D. On notera $k = \Sigma_{B_1 \cap \dots \cap B_n = \emptyset} \, m_1(B_1) \dots m_n(B_n)$, qui représente le degré de conflit entre les deux jeux de masses.

Plusieurs travaux récents ont cherché à justifier cette règle, par exemple ceux de Gacôgne à partir du concept d'accentuation [13][11], ceux de Dubois et Prade [7], qui justifient mathématiquement l'emploi du produit pour combiner les masses à partir de la notion de séparabilité des sources, ou ceux de Smets à partir d'un modèle de croyance transférable (« the transferable belief model ») [33]. Les travaux de Smets proposent la justification la plus générale à notre connaissance et ce sont ses arguments que nous décrivons ici. Ils s'inscrivent de plus parfaitement dans la logique de cet article puisque sa démarche est similaire à celle de Cox. Des travaux similaires à ceux de Smets ont été effectués par Klawonn et Schwecke [23].

7.1. axiomes de Smets

La première constatation de Smets concerne le principe d'indifférence (ou principe de raison insuffisante). Affecter la même probabilité à tous les événements (simples) entraîne que des probabilités différentes sont affectées à des réunions d'événements, ce qui, à son sens, ne correspond pas à l'indifférence. Celle-ci s'exprimerait plutôt par l'existence d'une constante c positive ou nulle telle que :

$$\forall A \subset D, A \neq D, Cr(A) = c, \quad (26)$$

où D désigne l'espace de discernement. Cela est à l'évidence impossible avec des probabilités, mais l'est dans le contexte de Dempster-Shafer avec des crédibilités. En effet,

$$A \cap B = \emptyset \Rightarrow Cr(A \cup B) \geq Cr(A) + Cr(B) \quad (27)$$

d'où $c \geq 2c$ et donc $c = 0$. La fonction de masse représentant l'indifférence (ou l'ignorance totale) est donc définie par :

$$m(D) = 1 \quad \text{et} \quad \forall A \neq D, m(A) = 0 \quad (28)$$

ce qui cette fois est tout à fait satisfaisant. 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 définit 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 :

$$\begin{aligned} m'(A) &= \sum_{X \subset \bar{B}} m(A \cup X) \quad \forall A \subset B \\ &= 0 \text{ sinon} \end{aligned} \quad (29)$$

où m' désigne le nouveau jeu de masses. Ces formules peuvent éventuellement être modifiées par renormalisation[12]. Cette formule s'interprète de la manière suivante. Si l'on décompose un sous-ensemble A en réunion $A_1 \cup A_2$ avec $A_1 \subset B$ et $A_2 \subset \bar{B}$, la masse $m(A_1 \cup A_2)$ est entièrement transférée sur A_1 (d'où le nom du modèle). Dans les cas particuliers où $A_2 = \emptyset (A \subset B)$, la masse de A n'est pas modifiée, et si $A_1 = \emptyset (A \subset \bar{B})$, la masse de A devient nulle.

Dans la théorie de Shafer [27], 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[13].

11. Gacôgne montre que la règle de Dempster-Shafer (équation 24), dans le cas où l'espace de discernement est réduit à une proposition P et à son contraire $\bar{P}$, peut se déduire de la notion d'accentuation. Une fonction d'accentuation est telle qu'elle diminue les degrés de confiance inférieurs à 0,5 et augmente ceux qui sont supérieurs à 0,5, les rendant ainsi plus proches de degrés binaires (cela correspond à la notion de renforcement que l'on trouve dans la théorie algébrique des semi-groupes ordonnés ainsi que dans celle des ensembles flous). La fonction rationnelle de [0,1] dans [0,1] de plus bas degré qui soit une fonction d'accentuation est définie par : $\frac{x^2}{2x^2 - 2x + 1}$, et c'est celle qui permet de montrer l'analogie avec Dempster-Shafer. Puis cette notion est généralisée à des couples (x, y), caractérisant une proposition P, tels que $0 \leq x \leq y \leq 1$ (appelés «obligation» et «éventualité», donc proches des notions de crédibilité et plausibilité ou encore de nécessité et possibilité). L'accentuation d'un tel couple est définie par :

$$\#(x, y) = \left(\frac{2xy - x^2}{1 - 2x + 2xy}, \frac{y^2}{1 - 2x + 2xy} \right).$$

Ces deux valeurs correspondent exactement à celles que l'on obtiendrait en combinant, par Dempster- Shafer (équation 24), la crédibilité (resp. la plausibilité) d'une proposition P avec elle-même. Si maintenant on dispose de deux jeux de mesures (x, y) et (x', y') sur P, le même type de raisonnement conduit à la justification de la règle de Dempster-Shafer L13]. Il resterait à généraliser cette approche à des espaces de discernement plus complexes.

12. Notons que Smets considère que la normalisation dans Dempster-Shafer n'est pas indispensable et peut même être nuisible dans la mesure où elle masque le conflit [33]. Elle pose également des problèmes de continuité au voisinage du conflit total [7].

13. Voir note 13 page suivante.

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

A1 : $(Cr_1 \oplus Cr_2)(A)$ doit être fonction seulement des fonctions m_1 et m_2 et de A.[14]

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, c'est-à-dire :

$$\begin{aligned}(m_1 \oplus m_2)(A) &= \sum_{X \subset \bar{B}} m_1(A \cup X) \; \forall A \subset B \\ &= 0 \text{ sinon}\end{aligned} \tag{30}$$

A5 : La loi doit vérifier une propriété de symétrie interne (invariance par permutation sur les hypothèses simples).

A6 : Pour $A \neq D$, $(m_1 \oplus m_2)(A)$ ne dépend pas de $m_1(X)$ pour $X \subset \bar{A}$ (propriété d'autofonctionnalité).

A7 : Il y a au moins 3 éléments dans D.

A8 : La loi doit vérifier une propriété de continuité :

$$\begin{aligned}&m_2(A) = 1 - \varepsilon,\; m_2(D) = \varepsilon,\; m_A(A) = 1 \\ &\Rightarrow \forall X, \lim_{\varepsilon \to 0} (m_1 \oplus m_2)(X) = (m_1 \oplus m_A)(X)\end{aligned} \tag{31}$$

m_1 étant une fonction de masse quelconque ; cette propriété permet d'éliminer des cas dégénérés.

7.2. déduction de la règle de combinaison

A partir des axiomes précédents, Smets déduit la seule règle de combinaison possible satisfaisant ces axiomes. Pour cela, il raisonne sur les fonctions de communalité, définies par :

$$\forall A \subset D,\; q(A) = \sum_{A \subset X, X \subset D} m(X). \tag{32}$$

Sa démonstration, que nous ne détaillerons pas ici, repose sur les propriétés des normes triangulaires et des fonctions absolument monotones et comporte trois parties. Tout d'abord, les axiomes A1 à A4 entraînent l'existence d'une fonction f telle que le résultat de la combinaison ne dépende que de A et des communalités des sous-ensembles inclus dans A :

$$(q_1 \oplus q_2)(A) = f[A, \{q_1(X);\; X \subset A, q_2(X);\; X \subset A\}]. \tag{33}$$

Ensuite, en ajoutant les axiomes A5 et A6, il est possible de préciser la forme de f, qui ne dépend plus que de A et de $q_1(A)$ et $q_2(A)$:

$$(q_1 \oplus q_2)(A) = f[A, q_1(A), q_2(A)]. \tag{34}$$

Enfin, l'ensemble des 8 axiomes A1 à A8 permet de déterminer la forme finale de la règle de combinaison :

$$(q_1 \oplus q_2)(A) = q_1(A) q_2(A). \tag{35}$$

On retrouve bien la règle de Dempster-Shafer sur les communalités, et on en déduit la combinaison des fonctions de masse ou de crédibilité.

L'avantage de l'approche de Smets est qu'elle repose sur des axiomes dont l'interprétation correspond bien à l'intuition. Il est également plus facile de les réfuter ou de les modifier s'ils ne correspondent pas au problème que l'on se pose, de manière analogue aux modifications décrites dans la partie 6 pour les postulats de Cox.

7.3. relation avec les postulats de Cox

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.

Tout d'abord, il convient de préciser dans quel cadre nous nous plaçons pour cette comparaison. En effet, les travaux de Cox et de Smets ne traitent pas exactement du même problème, puisque ceux de Cox s'attachent à justifier les probabilités et leurs propriétés, alors que ceux de Smets s'attachent à justifier une règle de combinaison. Cependant, il est intéressant de constater certaines analogies entre les deux ensembles d'axiomes. De plus, les axiomes de Cox permettent de déduire la règle de Bayes (équation 8), qui est utilisée en traitement du signal et des images pour fusionner des informations, par l'intermédiaire de probabilités conditionnelles. Nous adoptons le point de vue de la fusion de données pour cette comparaison. Il serait également intéressant de comparer les axiomes de Cox avec ceux introduits par Smets pour justifier les fonctions de crédibilité et de plausibilité [34], mais cette comparaison ne porterait que sur les étapes de modélisation du processus de fusion, et non sur les étapes de combinaison elles-mêmes.

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 entraînait l'existence d'une

13. La règle de conditionnement sur les plausibilités

$$Pls(A|B) = \frac{Pls(A \cap B)}{Pls(B)}$$

peut elle-même être justifiée en posant

$$Pls(A \cap B) = T[Pls(A|B), Pls(B)]$$

et en appliquant une démarche analogue à celle ayant conduit à la première équation fonctionnelle (partie 4.2) [7].

14. Dans [23], les axiomes utilisés sont pour la plupart similaires à ceux de Smets. La différence essentielle réside dans l'utilisation de relations entre espaces de discernement inclus les uns dans les autres plutôt que de relations de dépendance comme le fait Smets.

relation définissant le degré de confiance dans AB n'impliquant que les propositions A et B, sous la forme des degrés de confiance affectés à $[A|B]$ et $[B]$ (ou $[B|A]$ et $[A]$) 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 4) 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 :

$$[AB|e] = T([A|Be], [B|e\,) = T([B|Ae], [A|e]). \qquad (36)$$

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 cinquième postulat de Cox. La différence essentielle est que le conditionnement y est exprimé plus comme une relation de compatibilité que comme une probabilité conditionnelle.

On ne trouve pas dans les axiomes de Smets d'équivalent au postulat 3 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[15]. 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. Du point de vue de la comparaison des degrés de confiance, elle peut être menée dans la théorie des croyances à deux niveaux, soit sur les crédibilités, soit sur les plausibilités, conduisant à des conclusions qui ne sont pas nécessairement équivalentes.

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 6). En effet, des sous-ensembles X peuvent intervenir à la fois dans $(m_1 \oplus m_2)(A)$ et dans $(m_1 \oplus m_2)(\bar{A})$. On ne peut donc pas en déduire de relation de complémentarité sur m. Celle-ci est remplacée par une relation de dualité entre Cr et Pls.

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 5 et 6 de Cox.

Ces différences entre les deux théories ont des conséquences 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 :

- d'abord au niveau de l'étape de modélisation puisque cette étape est fortement contrainte par les deux relations fonctionnelles (équations 5 et 6) 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.

8. comparaison sur un exemple

Dans cette partie, nous comparons quelques règles de combinaison ou de fusion proposées au cours de l'histoire sur un exemple simple, celui de la combinaison de témoignages simultanés indépendants. Nous nous appuyons pour cela sur les exemples donnés par Shafer [28] et par Neapolitan [25], que nous élargissons. Nous notons p la crédibilité d'un témoin, et nous sommes donc amenés à attribuer à ce qu'il dit le degré de confiance p, et q le degré de confiance que l'on attribue au contraire de ce qu'il dit. Pour l'application numérique dans le cas de deux témoins, nous prendrons $p_1 = 0,8$ et $p_2 = 0,1$. Les valeurs de q_1 et q_2 seront précisées dans chaque cas, suivant le type d'argument considéré (au sens de Bernoulli), et vaudront 0 dans le cas d'un argument pur (l'argument ne permet pas de dire quoi que ce soit sur le contraire de ce que dit le témoin) ou $1 - p_1$ (respectivement $1 - p_2$) dans le cas d'un argument mixte (ces notions seront définies dans la partie 8.2).

15. 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 $Cr(A)$ 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 que l'on n'affecte pas de degré de confiance spécifiquement à A.

8.1. méthode de Hooper

Dans son ouvrage *A Calculation of the Credibility of Human Testimony* (1699), Hooper propose la règle suivante pour combiner deux témoignages simultanés (en supposant que les témoins sont d'accord) :

$$p = 1 - (1 - p_1)(1 - p_2). \tag{37}$$

Cette règle, équivalente à la somme algébrique souvent utilisée pour calculer une réunion dans le contexte des ensembles flous, se justifie bien dans le cas où les témoins ne sont pas de connivence. En effet il est alors naturel de renforcer la confiance que l'on a dans une proposition si elle est affirmée par deux personnes[16]. Pour l'exemple, considéré, on obtient un degré de confiance résultant de 0,82.

8.2. méthode de Bernoulli

J. Bernoulli (1654-1705) élabore dans *Ars Conjectandi* (1713) une théorie fondée sur la notion d'argument. La définition d'un argument est très large et désigne un témoignage, un signe ou une circonstance qui a un lien avec la chose à prouver. Shafer interprète la notion d'argument comme un ensemble de prémisses et conclusion. Un argument est pur s'il prouve une proposition dans certains cas et ne dit rien dans les autres. Un argument est mixte s'il prouve une proposition dans certains cas et prouve le contraire dans les autres. On aura donc pour un argument pur $q = 0$, et pour un argument mixte $q = 1 - p$.

Bernoulli distingue plusieurs types d'arguments : certains existent nécessairement et prouvent de manière contingente, d'autres existent de manière contingente et prouvent nécessairement, d'autres enfin existent et prouvent de manière contingente. Cette distinction peut s'interpréter en termes de prémisses et conclusion de la manière suivante : un argument existe de manière contingente si les prémisses ne sont pas nécessairement vérifiées, et un argument prouve de manière contingente si les prémisses n'entraînent pas nécessairement la conclusion.

Les règles de combinaison sont ensuite obtenues par dénombrement des cas où les arguments prouvent la proposition affirmée par les témoins. Par exemple dans le cas de deux arguments purs, le résultat est obtenu en comptant les cas où l'argument 1 prouve et, pour les autres cas (où l'argument 1 ne dit rien), les cas où l'argument 2 prouve. Le tableau 1 résume les résultats obtenus par Bernoulli (les grandes lignes du raisonnement de Bernoulli se trouvent en annexe), et les valeurs correspondant à l'exemple numérique.

Les deux premiers cas donnent le même résultat que celui de Hooper, le dernier correspond à une combinaison probabiliste classique, avec contrainte d'additivité. On pourra noter la grande différence entre les valeurs obtenues, qui peuvent conduire à des décisions radicalement opposées.

Tableau 1. – Règles de combinaison proposées par Bernoulli dans le cas de deux témoignages simultanés non contradictoires. Les résultats numériques sont donnés pour $p_1 = 0,8$ et $p_2 = 0,1$.

Type des arguments	Combinaison	Résultat
2 arguments purs	$p = 1 - (1 - p_1)(1 - p_2)$	0,82
1 pur et 1 mixte	$p = 1 - (1 - p_1)(1 - p_2)$	0,82
2 arguments mixtes	$p = \dfrac{p_1 p_2}{p_1 p_2 + q_1 q_2}$	0,308

8.3. méthode de Lambert

J. H. Lambert (1729-1777) publie le *Neues Organon* en 1764. Il y traite des problèmes relevant de la théorie des jeux, des syllogismes et de la combinaison des témoignages. Ce sont ces derniers résultats que nous retenons ici. Lambert distingue les cas où l'on croit le témoin (cas de type a), les cas où l'on ne le croit pas mais sans rien pouvoir dire de plus (cas de type u) et les cas où l'on croit le contraire (cas de type e).

La combinaison s'effectue par multiplication. Si pour le témoin 1 on a M_1 cas de type a, N_1 cas de type u, et P_1 cas de type e, la combinaison donne, avec des notations analogues pour le témoin 2 :

$$(M_1M_2 + M_1N_2 + M_2N_1)a + N_1N_2u \\ +(P_1P_2 + P_1N_2 + P_2N_1)e, \tag{38}$$

avec la règle $aa = au = a$, $ee = eu = e$, $uu = u$, et en excluant les cas ae qui sont impossibles (on ne peut pas croire à la fois une chose et son contraire).

On obtient alors dans le cas où les témoins sont d'accord (les grandes lignes du raisonnement de Lambert sont décrites en annexe) :

$$p = \frac{p_1 + p_2 - p_1p_2 - p_1q_2 - p_2q_1}{1 - p_1q_2 - p_2q_1}. \tag{39}$$

Dans le cas de témoignages contradictoires, la crédibilité du rapport du premier témoin est obtenue par :

$$p = \frac{p_1 + q_2 - p_1q_2 - p_1p_2 - q_1q_2}{1 - p_1p_2 - q_1q_2}, \tag{40}$$

et la crédibilité du rapport du deuxième est donnée par une formule analogue.

Si l'on remplace maintenant dans ces deux équations q_1 et q_2 par leurs valeurs suivant le type d'argument ($q_i = 0$ pour un argument pur et $q_i = 1 - p_i$ pour un argument mixte), on obtient les résultats résumés dans le tableau 2.

16. Si les témoins sont au contraire de connivence, il n'est pas étonnant qu'ils disent la même chose et dans ce cas la règle se justifie moins bien.

Tableau 2. – Règles de combinaison proposées par Lambert dans le cas de deux témoignages simultanés non contradictoires et contradictoires (dans ce deuxième cas, les valeurs de confiance données sont celles du rapport du premier témoin). Les résultats numériques *sont* donnés pour $p_1 = 0,8$ et $p_2 = 0,1$.

Type des arguments	Témoins d'accord	Valeur	Témoins pas d'accord	Valeur
2 purs	$1-(1-p_1)(1-p_2)$	0,82	$\frac{p_1(1-p_2)}{1-p_1p_2}$	0,783
2 mixtes	$\frac{p_1p_2}{p_1p_2+q_1q_2}$	0,308	$\frac{p_1(1-p_2)}{p_1+p_2-2p_1p_2}$	0,973
1 pur et 1 mixte	$\frac{p_2}{1-p_1(1-p_2)}$	0,357	$\frac{1-p_2}{1-p_1p_2}$	0,978
1 mixte et 1 pur	$\frac{p_1}{1-p_2(1-p_1)}$	0,816	$\frac{p_1(1-p_2)}{1-p_1p_2}$	0,783

Si l'on retrouve ici, dans les cas où les témoins sont d'accord, les formules de Bernoulli pour le cas d'arguments du même type (voir tableau 1), les résultats sont différents quand un argument est pur et l'autre mixte. La formule de Bernoulli, dont la symétrie était étonnante, a été corrigée par Lambert. La différence provient de ce que le calcul de Bernoulli n'excluait pas les cas de type ae comme le fait Lambert. Les résultats de Lambert donnent la priorité aux arguments mixtes par rapport aux arguments purs : dans le cas où les témoins sont d'accord, on trouve une valeur faible si le deuxième argument (de faible crédibilité) est mixte et une valeur élevée si le premier (de forte crédibilité) est mixte. Cela s'explique puisqu'un argument mixte porte également sur le contraire de la proposition affirmée et lui affecte un degré de confiance non nul, contrairement à un argument pur. Ici, si le deuxième argument est mixte, la confiance accordée au contraire de la proposition est élevée puisque P_2 est faible. Si les témoins ne sont pas d'accord, les résultats correspondent bien à l'intuition puisque les valeurs de crédibilité du rapport du premier témoin sont élevées. Un calcul analogue donnerait des valeurs très faibles à la crédibilité du rapport du deuxième témoin. On accorde ainsi plus de confiance à ce que dit le témoin le plus crédible. Cet exemple montre encore que la connaissance sur le type d'argument est déterminante sur le résultat et sur la décision qui pourrait en être prise.

8.4. méthode bayésienne

L'application de la méthode proposée par Bayes (1702-1761) et développée ensuite par Laplace (1749-1827) nécessite que des informations supplémentaires soient prises en compte :

- la probabilité a priori $p(E)$ pour que la proposition soit vraie (avec nécessairement $p(\bar{E}) = 1 - p(E)$),
- la probabilité pour que ce que dit le témoin i soit vrai sachant qu'il n'est pas fiable, notée λ_i[17].

17. Dire qu'un témoin n'est pas fiable ne signifie pas qu'il dit faux mais que ce qu'il dit peut être soit vrai soit faux.

Rappelons que, ici, P_i désigne la probabilité pour que le témoin i soit fiable (correspond aux cas où l'on croit le témoin) et q_i désigne la probabilité pour que le témoin i ne soit pas fiable (ce qui correspond aux cas où l'on ne le croit pas ou l'on croit le contraire).

Cette méthode donne alors le résultat suivant lorsque les témoins sont d'accord :

$$p = p(E|1 \text{ dit } E \text{ et } 2 \text{ dit } E) = \frac{p(E)(p_1+q_1\lambda_1)(p_2+q_2\lambda_2)}{p(E)(p_1+q_1\lambda_1)(p_2+q_2\lambda_2)+p(\bar{E})q_1(1-\lambda_1)q_2(1-\lambda_2)} \quad (41)$$

Ce résultat s'obtient par la règle de Bayes (équation 8), en introduisant l'indépendance entre les deux témoignages, ce qui fait apparaître la probabilité $p(1 \text{ dit } E|E)$ (et, de manière analogue, $p(2 \text{ dit } E|E)$) :

$$p(E|1 \text{ dit } E, 2 \text{ dit } E) = \frac{p(E)p(1 \text{ dit } E|E)p(2 \text{ dit } E|E)}{p(1 \text{ dit } E \text{ et } 2 \text{ dit } E)}. \quad (42)$$

La probabilité $p(1 \text{ dit } E|E)$ s'exprime par la proposition suivante : « (1 est fiable et 1 dit vrai sachant qu'il est fiable) ou (1 n'est pas fiable et 1 dit vrai sachant qu'il n'est pas fiable) ». La première partie de cette proposition correspond à p_1 (puisque la probabilité pour que 1 dise vrai sachant qu'il est fiable est égale à 1), et la deuxième partie correspond à $q_1\lambda_1$. Par un raisonnement analogue, le dénominateur s'écrit :

$$\begin{aligned} p(1 \text{ dit } E \text{ et } 2 \text{ dit } E) &= p(E)p(1 \text{ dit } E|E)p(2 \text{ dit } E|E)p(2 \text{ dit } E|E) \\ &\quad + p(\bar{E})p(1 \text{ dit } E|\bar{E})p(2 \text{ dit } E|\bar{E}) \\ &= p(E)(p_1+q_1\lambda_1)(p_2+q_2\lambda_2)+p(\bar{E})q_1(1-\lambda_1)q_2(1-\lambda_2). \quad (43) \end{aligned}$$

Dans le cas particulier où $p(E) = p(\bar{E})$ (classiquement pris selon le principe de raison insuffisante) et $\lambda_1 = \lambda_2$, le résultat varie, suivant la valeur de λ_1, entre une borne inférieure et une borne supérieure données par :

$$P_* = \frac{p_1p_2}{p_1p_2+q_1q_2}, \quad (44)$$

$$P^* = 1, \quad (45)$$

avec $P_* = 0,308$ pour notre exemple. On retrouve le résultat obtenu par Bernoulli et Lambert pour deux arguments mixtes (c'est-à-dire dans le cas où ils manipulent des probabilités additives).

Si maintenant les témoins ne sont pas d'accord, on obtient par un calcul analogue :

$$\begin{aligned} p &= p(E|1 \text{ dit } E, 2 \text{ dit } \bar{E}) \\ &= \frac{p(E)(p_1+q_1\lambda_1)q_2(1-\lambda_2)}{p(E)(p_1+q_1\lambda_1)(q_2(1-\lambda_2)+p(\bar{E})q_1(1-\lambda_1)(p_2+q_2\lambda_2)}. \end{aligned} \tag{46}$$

Dans le cas particulier où $p(E) = p(\bar{E})$ et $\lambda_1 = \lambda_2$, le résultat varie entre les deux bornes suivantes :

$$P_* = \frac{1-p_2}{(1-p_1)+(1-p_2)}, \tag{47}$$

$$P^* = \frac{p_1(1-p_2)}{p_1+p_2-2p_1p_2}, \tag{48}$$

avec $P_* = 0,818$ et $P^* = 0,973$ pour notre exemple. P^* est obtenu ici pour la valeur $\lambda_1 = \lambda_2 = 0$, ce qui signifie que dire que le témoin n'est pas fiable est équivalent à dire que ce qu'il rapporte est faux. On retrouve donc bien le cas de deux arguments mixtes.

Si l'on ne fait aucune hypothèse sur λ_i, la probabilité peut varier entre les deux bornes $P_* = 0$ et $P^* = 1$, ce qui ne donne pas beaucoup d'information !

8.5. méthode de Demspter et Shafer

Le raisonnement effectué pour traiter cet exemple par la théorie des croyances repose sur la notion de compatibilité [17]. Par exemple, si le témoin 1 dit E et le témoin 1 dit vrai, alors la proposition E est vraie. La crédibilité de E est donc égale à p_1. De même, si le témoin 1 dit E et le témoin 1 n'est pas fiable (il peut dire vrai ou faux), alors E peut être vraie, il y a donc compatibilité et les cas correspondants interviennent dans la plausibilité, qui vaut donc $p_1 + q_1$. En résumé, on a donc :

$$Cr_1(E) = p_1, \tag{49}$$

$$Pls_1(E) = p_1 + q_1 = 1. \tag{50}$$

Notons qu'il n'y a pas de relation d'additivité entre Cr_1 et Pls_1, mais seulement une relation de dualité, c'est-à-dire que la crédibilité et la plausibilité de $\bar{E}$ se déduisent de celles de E par :

$$Cr_1(\bar{E}) = 1 - Pls_1(E) = 0, \tag{51}$$

$$Pls(\bar{E}) = 1 - Cr_1(E) = 1 - p_1. \tag{52}$$

Dans cet exemple, $D = \{E, \bar{E}\}$, et donc, par définition :

$$Cr_1(E) = m_1(E), \tag{53}$$

$$Pls_1(E) = m_1(E) + m_1(E \cup \bar{E}). \tag{54}$$

En remplaçant dans ces équations $Cr_1(E)$ et $Pls_1(E)$ par leurs expressions données par les équations 49 et 50, on obtient la fonction de masse associée au témoin 1 :

$$m_1(E) = p_1, \tag{55}$$

$$m_1(\bar{E}) = 0, \tag{56}$$

$$m_1(E \cup \bar{E}) = q_1. \tag{57}$$

On obtient bien sûr des équations similaires pour le témoin 2 et la combinaison des fonctions de masses si les témoins sont d'accord est obtenue par la règle orthogonale de Dempster (équation 24) :

$$k = m_1(E)m_2(\bar{E}) + m_1(\bar{E})m_2(E) = 0. \tag{58}$$

$$\begin{aligned} (m_1 \oplus m_2)(E) &= \frac{1}{1-k}\Big[m_1(E)m_2(E) \\ &\quad + m_1(E)m_2(E \cup \bar{E}) + m_2(E)m_1(E \cup \bar{E})\Big] \\ &= p_1p_2 + p_1q_2 + p_2q_1 = 1 - (1-p_1)(1-p_2). \end{aligned} \tag{59}$$

$$(m_1 \oplus m_2)(\bar{E}) = 0. \tag{60}$$

$$\begin{aligned} (m_1 \oplus m_2)(E \cup \bar{E}) &= m_1(E \cup \bar{E})m_2(E \cup \bar{E}) \\ &= q_1q_2 = (1-p_1)(1-p_2). \end{aligned} \tag{61}$$

On en déduit la crédibilité et la plausibilité :

$$Cr(E) = (m_1 \oplus m_2)(E) = 1 - (1-p_1)(1-p_2), \tag{62}$$

$$Pls(E) = (m_1 \oplus m_2)(E) + (m_1 \oplus m_2)(E \cup \bar{E}) = 1, \tag{63}$$

avec $Cr(E) = 0,82$. On retrouve le résultat obtenu pour des arguments purs par Hooper, Bernoulli et Lambert.

Si les témoins ne sont pas d'accord, supposons que le témoin 2 dise $\bar{E}$. S'il dit vrai, alors $\bar{E}$ est vraie. Ces cas interviennent donc dans la crédibilité de $\bar{E}$. On obtient alors pour le témoin 2 :

$$Cr_2(\bar{E}) = p_2, \tag{64}$$

$$Pls_2(\bar{E}) = 1. \tag{65}$$

La fonction de masse correspondante est :

$$m_2(E) = O, \tag{66}$$

$$m_2(\bar{E}) = p_2, \tag{67}$$

$$m_2(E \cup \bar{E}) = q_2. \tag{68}$$

La combinaison par la règle de Dempster (équation 24) donne alors :

$$k = m_1(E)m_2(\bar{E}) + m_1(\bar{E})m_2(E) = p_1p_2. \tag{69}$$

$$\begin{aligned} (m_1 \oplus m_2)(E) &= \frac{1}{1-k}\Big[m_1(E)m_2(E) + m_1(E)m_2(E \cup \bar{E}) \\ &\quad + m_2(E)m_1(E \cup \bar{E})\Big] = \frac{p_1q_2}{1-p_1p_2} = \frac{p_1(1-p_2)}{1-p_1p_2}, \end{aligned} \tag{70}$$

d'où les valeurs de la crédibilité et de la plausibilité de E :

$$Cr(E) = \frac{p_1(1-p_2)}{1-p_1p_2}, \tag{71}$$

$$Pls(E) = \frac{1-p_2}{1-p_1p_2}. \tag{72}$$

Pour notre exemple, on obtient $Cr(E) = 0,783$ et $Pls(E) = 0,978$, qui correspondent aux valeurs trouvées par la méthode de Lambert pour des arguments de types différents .

9. conclusion

Nous avons illustré sur un exemple comment différents modèles proposés dans l'histoire permettent de combiner des témoignages. L'application montre d'une part une divergence entre les résultats selon les types de témoignages (et donc entre les décisions qui seraient prises à l'issue de la combinaison), ce qui montre l'importance du « modèle de témoin » et du « modèle d'argument » (au sens de Bernoulli) utilisés. En effet, chaque méthode utilise implicitement ou explicitement des modèles, et le choix d'une méthode ne pourra se faire sans vérifier que les modèles correspondent à la situation à traiter. D'autre part, les résultats montrent une analogie remarquable entre des modèles anciens et des modèles « modernes ». Notons que la variété d'opérateurs proposés par la théorie des ensembles flous (non illustrés ici) permet également une plus grande variété de modèles de combinaison (si l'on interprète par exemple une fonction d'appartenance comme une distribution de possibilité), dont certains correspondent à ceux présentés ici. Ce sont alors les connaissances sur le problème et les propriétés requises qui fournissent un guide de choix [4]. Dans le cas de la théorie de Dempster- Shafer, il existe également des variantes ensemblistes de la règle de combinaison. Shafer se place dans le cas d'information disjontive. Si ce n'est pas le cas, l'intersection qui apparaît dans la règle de Dempster peut être remplacée par une réunion ou encore une moyenne, comme cela a été développé dans [9].

L'exemple de la combinaison de témoignages a ici un caractère historique puisqu'il correspond aux types de problèmes que traitaient Hooper, Bernoulli, Lambert, Laplace. Des exemples similaires peuvent aisément être trouvés en traitement du signal et des images, en particulier en fusion de données. L'information fournie par un capteur, celui-ci pouvant être fiable ou non, peut être interprétée comme un témoignage, nous apportant une information sur une proposition à évaluer (par exemple la présence de tel objet dans une image, ou l'appartenance d'un point ou d'une région à telle classe), et un raisonnement analogue à celui que nous décrivons ici peut être effectué.

Ces problèmes de combinaison de témoignages, ainsi que les méthodes permettant de les résoudre, connaissent d'ailleurs aujourd'hui un regain d'intérêt, non seulement en intelligence artificielle, mais aussi en fusion de données issues de signaux ou d'images.

Le bref parcours historique des différentes classes de probabilités que nous avons présenté montre que les théories « nouvelles » de la fusion avaient déjà des germes dans les théories du 18^{e} siècle : par exemple, les règles de combinaison de Lambert sont des cas particuliers de la règle orthogonale de Dempster. Il montre aussi que l'additivité des probabilités aujourd'hui communément admise n'est apparue que très tard.

À notre avis, l'apport essentiel du 20^{e} siècle est double. D'une part, des relations posées de manière axiomatique peuvent être démontrées à partir de postulats correspondant à l'intuition et que l'on est libre d'accepter ou de refuser selon les problèmes à résoudre (dans ce cas, on obtient d'autres théories, comme l'a montré la comparaison entre les postulats de Cox pour déduire les probabilités et ceux de Smets pour déduire la théorie des croyances). D'autre part, une analyse plus approfondie du raisonnement humain, en particulier en fusion de données, a conduit à des théories non probabilistes, permettant la modélisation de facteurs non probabilistes et allant contre les limites reconnues des probabilités.

Remerciements

Nous remercions Guy Demoment, qui nous a confié une version préliminaire de son cours de DEA [6] ainsi que les références sur Bayes [16], [37], [38], [10], Didier Dubois, qui nous a également suggéré plusieurs références et qui nous a en particulier fait connaître les travaux de De Finetti et de Paris, et les rapporteurs. Par leurs commentaires riches et constructifs, ils nous ont permis d'améliorer et de préciser plusieurs points de cet article.

10. annexe

Règles de combinaison de Bernoulli et Lambert

Raisonnement de Bernoulli

Bernoulli déduit ses règles de combinaison d'un dénombrement de cas. Considérons le cas de deux arguments A_1 et A_2, et notons b (resp. e) le nombre de cas dans lesquels A_1 (resp. A_2) prouve la proposition considérée, c (resp. f) le nombre de cas où il ne prouve rien ou prouve le contraire, avec $b + c = a$ (resp. $e + f = d$).

Prenons d'abord le cas d'arguments purs. Si A_1 est le seul argument dont nous disposons, la confiance que nous accordons à la proposition considérée est :

$$p_1 = \frac{b}{a} = \frac{a-c}{a}. \tag{73}$$

Si l'on dispose des deux arguments, ce degré de confiance peut alors être calculé de la manière suivante : dans e cas, l'argument A_2 suffit à prouver la proposition ; dans les f autres cas, c'est l'argument A_1 qui décide, et il prouve dans une proportion $\frac{a-c}{a}$ des cas. Cela donne au total :

$$p = \frac{e + f\frac{a-c}{a}}{d} = 1 - \frac{cf}{ad} = 1 - (1-p_1)(1-p_2). \tag{74}$$

Notons que cette expression est symétrique et ne dépend donc pas de l'ordre dans lequel sont considérés les arguments lors du raisonnement.

Cette expression se généralise par un raisonnement analogue au cas de n arguments purs :

$$p = 1 - (1 - p_1)(1 - p_2)\ldots(1 - p_n). \tag{75}$$

Considérons maintenant le cas d'arguments mixtes. Pour A_1 seul, le rapport des degrés de confiance de la proposition et de son contraire est égal au rapport entre le nombre de cas qui prouvent et le nombre de cas qui prouvent le contraire :

$$\frac{p_1}{q_1} = \frac{b}{c}. \tag{76}$$

Dans le cas de plusieurs arguments mixtes, le rapport $\frac{p}{q}$ global est obtenu comme le produit des $\frac{p_i}{q_i}$:

$$\frac{p}{q} = \frac{p_1}{q_1}\frac{p_2}{q_2} = \frac{b}{c}\frac{e}{f}. \tag{77}$$

En introduisant l'hypothèse d'additivité $p+q = 1$ (et $p_i+q_i = 1$), Bernoulli obtient donc :

$$p = \frac{be}{be + cf} = \frac{p_1p_2}{p_1p_2 + q_1q_2}, \tag{78}$$

et :

$$q = \frac{cf}{bd + cf} = \frac{q_1q_2}{p_1p_2 + q_1q_2}. \tag{79}$$

Enfin, dans le cas ou certains arguments sont purs et d'autres mixtes, Bernoulli commence par combiner les arguments de même type entre eux, selon les formules 75 et 78, pour obtenir p_{purs} et p_{mixtes} puis regroupe les résultats selon le raisonnement suivant : dans les cas où les arguments purs prouvent, les arguments mixtes n'interviennent pas, et dans les cas où les arguments purs ne disent rien, ce sont les arguments mixtes qui décident :

$$p = p_{purs} + (1 - p_{purs})p_{mixtes}. \tag{80}$$

Dans le cas où A_1 est pur et A_2 est mixte, on obtient :

$$p = 1 - (1 - p_1)(1 - p_2). \tag{81}$$

On retouve dans ce cas la règle de combinaison obtenue pour deux arguments purs (équation 74), et c'est ce que Lambert va remettre en cause.

Raisonnement de Lambert

Le raisonnement de Lambert sépare plus clairement les cas où l'on ne croit pas le témoin (sans pour autant croire le contraire de ce qu'il dit) des cas où l'on croit le contraire de ce qu'il dit. Il traite donc 3 types de cas :

– les cas de type a (on croit le témoin i), en nombre M_i,
– les cas de type u (on ne le croit pas), en nombre N_i,
– les cas de type e (on croit le contraire), en nombre P_i.

Le nombre total de cas pour le témoin i est noté sous la forme suivante :

$$M_ia + N_iu + P_ie. \tag{82}$$

La combinaison de deux témoignages est évaluée par dénombrement de tous les cas possibles, avec les règles suivantes :

– cas de type a pour les deux témoins : résultat de type a,
– cas de type u pour les deux témoins : résultat de type u,
– cas de type e pour les deux témoins : résultat de type e,
– cas de type a pour un témoin et u pour l'autre : résultat de type a,
– cas de type e pour un témoin et u pour l'autre : résultat de type e,
– cas de type a pour un témoin et e pour l'autre : ces cas sont impossibles et sont donc éliminés dans la combinaison.

Avec ces règles de combinaison, le nombre de cas total résultant de la combinaison de deux témoignage indépendants est alors (avec les notations définies par l'équation 82) :

$$(M_1M_2+M_1N_2+M_2N_1)a+N_1N_2u+(P_1P_2+P_1N_2+P_2N_1)e. \tag{83}$$

Revenons aux degrés de confiance p et q. Pour chaque témoin, on a :

$$p_i = \frac{M_i}{M_i + N_i + P_i}, \tag{84}$$

et

$$q_i = \frac{P_i}{M_i + N_i + P_i}. \tag{85}$$

Notons qu'il n'y a ici aucune hypothèse a priori d'additivité. On retrouve les cas de Bernoulli : lorsqu'un argument est pur, $P_i = 0$ et donc $q_i = 0$, et lorsqu'il est mixte, $N_i = 0$ et donc $q_i = 1 - p_i$. La formulation de Lambert laisse la place à une classe intermédiaire d'argument, où il peut y avoir à la fois des cas de type u et des cas de type e.

Dans les cas où les témoins sont d'accord, le degré de confiance total dans ce qu'ils disent s'obtient comme somme des cas où les deux témoins sont de type a et des cas où l'un est de type a et l'autre de type u, divisé par tous les cas non impossibles, soit :

$$\begin{aligned} p &= \frac{p_1p_2 + p_1(1 - p_2 - q_2) + (1 - p_1 - q_1)p_2}{1 - p_1q_2 - p_2q_1} \\ &= \frac{p_1 + p_2 - p_1p_2 - p_1p_2 - q_1p_2}{1 - p_1q_2 - p_2q_1}. \end{aligned} \tag{86}$$

Par un raisonnement analogue, q vaut :

$$q = \frac{q_1 + q_2 - q_1q_2 - p_1q_2 - q_1p_2}{1 - p_1q_2 - p_2q_1}. \tag{87}$$

Dans les cas où les témoins ne sont pas d'accord, cette fois les cas où le premier témoin est de type a et le second de type e sont des cas possibles. Le degré de confiance total attribué à ce que dit le premier témoin s'obtient alors comme somme des cas où le premier témoin est de type a (ou u) et le second e (ou u), divisé par tous les cas non impossibles (ici : tous les cas sauf ceux où les témoins sont d'accord) :

$$\frac{p_1q_2 = p_1(1-p_2-q_2)+(1-p_1-q_1)q_2}{1-p_1p_2-q_1q_2} = \frac{p_1+q_2-p_1q_2-p_1p_2-q_1q_2}{1-p_1p_2-q_1q_2}. \qquad (88)$$

On obtient une formule analogue pour le degré de confiance dans ce que dit le deuxième temoln :

$$\frac{q_1+p_2-q_1p_2-p_1p_2-q_1q_2}{1-p_1p_2-q_1q_2}. \qquad (89)$$

Les résultats du tableau 2 sont simplement obtenus en remplaçant q_1 et q_2 dans les formules 86 et 88 par leurs valeurs suivant le type d'argument (on ne considère pas ici le cas d'arguments intermédiaires) :

– 2 arguments purs : $q_1 = q_2 = 0$,

– 2 arguments mixtes : $q_1 = 1 - p_1$ et $q_2 = 1 - p_2$,

– argument 1 pur et argument 2 mixte : $q_1 = 0$ et $q_2 = 1 - p_2$,

– argument 1 mixte et argument 2 pur : $q_1 = 1 - p_1$ et $q_2 = 0$.

BIBLIOGRAPHIE

[1] J. Aczél, *Sur les opérations définies pour nombres réels*, Bull. Soc. Math. Franç., 76, 59-64, 1948.

[2] J. Aczél, *Lectures on Functional Equations and Their Applications*, Academic Press, New York, 1966.

[3] I. Bloch, H. Maître, *Fusion de données en traitement d'images : modèles d'information et décisions*, Traitement du Signal, Vol. 11, N° 6, 1994, 435-446.

[4] I. Bloch, *Information Combination Operators for Data Fusion : A Comparative Review with Classification*, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 26, N° 1, January 1996, 52-67.

[5] R. T. Cox, *Probability, Frequency and Reasonable Expectation, American Journal of Physics*, Vol. 14, N° 1, 1946, 1-14.

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

[7] D. Dubois, H. Prade, *On the Unicity of Dempster Rule of Combination*, International Journal of Intelligent Systems, Vol. 1, 1986, 133-142.

[8] D. Dubois, H. Prade, *Théorie des possibilités, applications à la représentation des connaissances en informatique*, Masson, Paris, 1988.

[9] D. Dubois, H. Prade, *Representation and Combination of Uncertainty utith Belief Functions and Possibility Measures*, Compu. Intell., Vol. 4, 1988, 244-264.

[10] B. Efron, *Why isn't Everyone a Bayesian?*, The American Statistician, Vol. 40, No-1, 1986, 1-11.

[11] W. Feller, *An Introduction to Probability Theory and its Applications*, New York, Wiley, 1966.

[12] B. de Finetti, *La prévision : ses lois logiques, ses sources subjectives*, Annales de l'Institut Henri Poincaré, Vol. 7, N° 1, 1937, 1-68.

[13] L. Gacôgne, *About a Foundation of Dempster's Rule*, Rapport Laforia 93/27, 1993.

[14] I. J. Good, *Kinds of Probability*, Science, Vol. 129, N° 3347, February 1959, 443-447.

[15] D. Heckerman, *Probabilistic Interpretation for Mycin's Certainty Factors*, Uncertainty in Artificial Intelligence, L. N. Kanal and J. F. Lemmer (Ed.), Elsevier Science Publishers, North-Holland, 1986, 167-196.

[16] J. D. Holland, *The Reverend Thomas Bayes, F.R.S. (1702-61)*, J. Roy. Stat. Soc. (A), Vol. 125, 1962, 451-461.

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

[18] E. Horvitz, D. Heckerman, *The Inconsistent Use of Measures of Certainty in Artificial Intelligence Research*, Uncertainty in Artificial Intelligence, L. N. Kanal and J. F. Lemmer (Ed.), Elsevier Science Publishers, North-Holland, 1986, 137-151.

[19] E. T. Jaynes, *Information Theory and Statistical Mechanics*, Physical Review, Vol. 106, N° 4, 1957, 620-630.

[20] R. Jeffreys, *Theory of Probability*, Oxford University Press, 1961.

[21] E. C. Kemble, *Is the Frequency Theory of Probability Adequate for All Scientific Purposes?* Am. J. Physics, vol. 10, 1942 6-16.

[22] J. M. Keynes, *A Treatise on Probability*, Macmillan, London, 1929.

[23] F. Klawonn, E. Schwecke, *On the Axiomatic Justification of Dempster's Rule of Combination*, International Journal of Intelligent Systems, Vol. 7, 1992, 49-478.

[24] G. Matheron, *Estimer et choisir – Essai sur la pratique des probabilités*, École Nationale Supérieure des Mines de Paris, Centre de Géostatistique, 1978.

[25] R. E. Neapolitan, *A Survey of Uncertain and Approximate Inference*, In "Fuzzy Logic for the Management of Uncertainty", L. Zadeh, J. Kacprzyk Eds., J. Wiley & Sons, New York, 1992, 55-82.

[26] J. B. Paris, *The Uncertain Reasoner's Companion, a Mathematical Perspective*, Cambridge University Press, 1995.

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

[28] G. Shafer, *Non-Additive Probabilities in the Work of Bernoulli and Lambert*, Archive for History of Exact Sciences, Vol. 19, 1978, 309-370.

[29] G. Shafer, *The Combination of Evidence*, International Journal of Intelligent Systems, Vol. 1, 1986, 155-179.

[30] C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, University of Illinois Press, 1959.

[31] E. H. Shortliffe, B. G. Buchanan, *A model of Inexact Reasoning in Medicine*, Mathematical Biosciences 23, 1975, 351-379.

[32] P. Smets, *Medical Diagnosis : Fuzzy Sets and Degree of Belief*, Colloque International sur la Théorie et les Applications des Sous-Ensembles Flous, Marseille, September 1978.

[33] P. Smets, *The Combination of Evidence in the Transferable Belief Model*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, N° 5, 1990, 447-45.

[34] P. Smets, *Quantifying Beliefs by Belief Functions : an Axiomatic Justification*, 13th International Joint Conference on Artificial Intelligence, Chambéry, France, 1993, 598-603.

[35] P. Smets, *Belief Function : The Disjunctive Rule of Combination and the Generalized Bayesian Theorem*, International Journal of Approximate Reasoning, Vol. 9, 1993, 1-35.

[36] Léa Sombé, *Raisonnements sur des informations incomplètes en intelligence artificielle*, Teknea, Marseille, 1989.

[37] S. M. Stigler, *Thomas Bayes's Bayesian Inference*, J. Roy. Stat. (A), Vol. 145, 1982, 250-258.

[38] S. M. Stigler, *Who Discovered Bayes's Theorem?*, The American Statistician, Vol. 37, No-4, 1983, 290-296.

[39] M. Tribus, *Rational, Descriptions, Decisions and Designs*, Pergamon Press Inc., 1969.

[40] J.-P. Tubach, Communication personnelle, Val-Thorens, 1994.

[41] L. A. Zadeh, *Fuzzy Sets*, Inform. and Control 8, 1965, 338-353.

[42] L. A. Zadeh, *Probability Measures of Fuzzy Events*, Journal of Mathematical Analysis and Applications, Vol. 23, 1968, 421-427.

[43] L. A. Zadeh, *Fuzzy Sets as a Basis for a Theory of Possibility*, Fuzzy Sets and Systems 1, 1978, 3-28.

Manuscrit reçu le 11 mai 1995.

L'AUTEUR

Isabelle BLOCH

Isabelle Bloch est née en 1964. Ingénieur Civil des Mines de Paris en 1986, elle soutient sa thèse de doctorat en 1990 à l'ENST et son Habilitation à Diriger des Recherches en 1995 à l'université Paris 5. Elle est maître de conférences à l'ENST. Son activité de recherche est consacrée au traitement d'images et d'objets 3D, aux ensembles flous, à la morphologie mathématique 3D et floue, à la fusion de données en traitement d'images, à la théorie de la décision, à la théorie des croyances de Dempster-Shafer, à l'imagerie médicale et satellitaire.