Chapitre 3

Représentations de l'incertitude en intelligence artificielle

De par ses préoccupations de représentation des connaissances, l'intelligence artificielle a été amenée à s'intéresser à différents cadres de traitement de l'incertitude : la théorie des probabilités, mais aussi les théories plus récentes des possibilités, des fonctions de croyance, et des probabilités imprécises. Ce chapitre en offre un panorama introductif qui fait ressortir les spécificités de chaque cadre représentationnel, tout en identifiant les principales questions posées par la représentation de l'incertain. C'est aussi l'occasion de situer brièvement les théories des ensembles flous, et des ensembles approximatifs, motivées respectivement par la prise en compte de la gradualité de certaines propriétés, et la granularité des représentations induites par les langages de description. De plus, cette vue d'ensemble inclut des présentations succinctes d'autres cadres théoriques comme ceux de l'analyse formelle de concepts, des objets conditionnels, des fonctions de rang, ou de la logique possibiliste, en relation avec les autres approches présentées.

3.1 Eléments historiques et intérêt de nouveaux cadres

Le problème de la prise en compte et donc de la modélisation de l'incertain n'est pas propre à l'intelligence artificielle. Historiquement, cette préoccupation apparaît au XVII^{e} siècle avec notamment les travaux pionniers de Huyghens, de Pascal, du chevalier de Méré, et de Jacques Bernoulli. On distingue alors la notion objective de *chance*, liée à l'analyse des jeux de hasard, de la notion subjective de *probabilité* liée à la question de la fiabilité des témoignages devant les tribunaux. Ainsi chez J. Bernoulli, les chances, qui sont reliées à la notion de fréquence sont naturellement additives, mais les probabilités subjectives sont non-additives, comme c'est encore le cas au milieu du XVIII^{e} siècle chez Lambert (qui propose une règle de combinaison qui s'avère rétrospectivement être un cas particulier de la règle de Dempster ; voir (Shafer, 1978; Martin, 2006)). Cependant avec le développement de la physique et de l'actuariat, on cessera de s'intéresser pour longtemps au versant subjectif des probabilités et la vision additive

Auteurs : S. BENFERHAT (CRIL), TH. DENOEUX (Heudiasyc), D. DUBOIS (IRIT), H. PRADE (IRIT)

prend définitivement le pas sur l'autre, avec les travaux de Laplace, que les probabilités soient fréquentistes ou non. Ainsi au milieu du XX^{e} siècle, en économie, les théories de la décision en présence de risques de von Neumann et Morgenstern, ou de la décision dans l'incertain subjectif de Savage seront basées sur des probabilités additives.

C'est l'émergence de l'informatique qui a reposé les questions de la représentation des connaissances et de la formalisation des raisonnements, en présence d'imprécision, d'incertitude, de contradictions, indépendamment de la théorie des probabilités, et longtemps sans lien avec une problématique de décision. L'accent a plutôt été mis au départ en intelligence artificielle sur des formalismes de type logique, éventuellement qualitatifs, ainsi que sur la modélisation des informations linguistiques (avec la théorie des ensembles flous, aussi appelée logique floue).

En effet l'information disponible est souvent incertaine, et cela vaut aussi pour la connaissance des experts qui sont susceptibles de fournir des règles qui conduisent à des conclusions incertaines même à partir de faits certains. Le besoin de traitement de l'incertitude est donc apparu dès les premiers systèmes experts, au début des années 1970. Un des plus connus et des tout premiers, MYCIN (Shortliffe, 1976; Buchanan et Shortliffe, 1984), proposait d'ailleurs un calcul empirique de propagation de l'incertitude entièrement original, à base de degrés de croyance et de défiance. Nous ne le rappellerons pas ici, faute de place, car il est aujourd'hui dépassé à cause de ses insuffisances en matière de gestion des exceptions notamment. Mais ses règles de calcul préfigurent les nouveaux cadres de représentation de l'incertitude qui allaient apparaître. Voir (Dubois et Prade, 1989) sur ce point ; voir aussi (Lucas et van der Gaag, 1991).

3.2 Imprécision, incertitude, gradualité, granularité

Avant de dresser une vue d'ensemble des différents cadres de représentation (voir (Halpern, 2003; Dubois et Prade, 2006; Liu, 2001; Parsons, 2001) pour plus de détails sur certains aspects), il convient de préciser un peu le vocabulaire employé. On appelle *information* toute collection de symboles ou de signes produits soit par l'observation de phénomènes naturels ou artificiels, soit par l'activité cognitive humaine. Une première distinction apparaît entre *informations* dites *objectives*, issues de mesures ou de capteurs ou plus généralement de l'observation directe de phénomènes, et celles dites *subjectives* exprimées par des individus et conçues sans le recours à l'observation directe du réel. L'information peut prendre une forme numérique, notamment l'information objective (mesures issues de capteurs, processus de comptage) ou qualitative (notamment l'information subjective, par exemple exprimée en langage naturel). Néanmoins, ce partage n'est pas aussi strict : l'information subjective peut être numérique, et l'information objective qualitative (une couleur perçue par un capteur symbolique, par exemple). L'information numérique peut prendre de nombreuses formes : nombres, intervalles, fonctions. L'information symbolique structurée est souvent codée dans une représentation logique ou graphique. Il y a des représentations hybrides telles les logiques pondérées, ou les réseaux bayésiens (Pearl, 1988). Enfin, une autre distinction importante doit être faite entre *information singulière* et *information générique*. L'information singulière concerne un fait particulier, résultant d'une observation ou d'un témoignage. L'information générique se réfère à une classe de situations (ce peut être une loi physique, un modèle statistique issu d'un ensemble représentatif d'observations, ou encore une connaissance de sens commun comme « les oiseaux volent »).

Information imprécise

Pour représenter l'état épistémique d'un agent, il faut disposer d'une représentation des états du monde, selon le point de vue de cet agent, c'est-à-dire en formaliser les aspects pertinents à l'aide de certains attributs. Soit v le vecteur d'attributs pertinents pour l'agent et S le domaine de v (pas toujours connu explicitement). S est l'ensemble des (descriptions des) états possible du monde. Un sous-ensemble A de S est un événement, que l'on peut voir comme une proposition qui affirme $v \in A$.

Une information est dite *imprécise* si elle est insuffisante pour permettre à l'agent de répondre à une question qu'il se pose. L'imprécision correspond à l'idée d'information *incomplète* (manquante). La question à laquelle l'agent cherche à répondre est de la forme *quelle est la valeur de* v ? ou plus généralement est-ce que v satisfait une certaine propriété ? La notion d'imprécision n'est pas absolue. Si on s'intéresse à l'âge d'une personne, le terme *mineur* est précis si le référentiel est $S = \{mineur, majeur\}$ et que la question est : est-ce que la personne a le droit de vote ? En revanche, s'il s'agit de connaître son âge et si $S = \{0, 1, \ldots, 150\}$ (en nombre d'années), le terme *mineur* est (très) imprécis. La forme type d'une information imprécise est $v \in A$ où A est un *sous-ensemble* de S *contenant plus d'un élément*. Une remarque importante est le fait que les éléments de A, vus comme valeurs possibles de v sont mutuellement exclusifs (car la variable n'a qu'une seule valeur). Donc une information imprécise prend la forme d'une *disjonction* de valeurs mutuellement exclusives. Ainsi, dire que *Jean a entre 20 et 22 ans*, soit $v =$ *âge(Jean)* $\in \{20, 21, 22\}$, c'est supposer $v = 20$ ou $v = 21$ ou $v = 22$. Une forme extrême d'information imprécise est l'*ignorance totale* : la valeur de v est complètement inconnue. En logique classique, l'imprécision apparaît explicitement comme une disjonction. Un ensemble utilisé pour représenter une information imprécise est dit *ensemble disjonctif*. Il s'oppose à la vision conjonctive usuelle d'un ensemble, vu comme une collection d'éléments. Un *ensemble conjonctif* représente une information précise. Ainsi, si on parle des soeurs de Jean, $v = \{\text{Emilie, Stéphanie}\}$ est une information précise, qui signifie que les soeurs de Jean sont Emilie *et* Stéphanie. En effet, le référentiel S est ici l'ensemble de tous les prénoms féminins. Dans ce cadre, une information imprécise correspondrait à une *disjonction de sous-ensembles conjonctifs* de prénoms. On peut comparer deux informations imprécises en termes de contenu informationnel : une information $v \in A_1$ est dite *plus spécifique* qu'une information $v \in A_2$ si et seulement si A_1 est un sous-ensemble propre de A_2.

Information incertaine

Une *information* est dite *incertaine* pour un agent lorsque l'agent ne sait pas si cette information est vraie ou fausse. Une information élémentaire exprimée par une proposition est modélisée par un sous-ensemble de valeurs possibles, de la forme $v \in A$, et on affecte à cette information un marqueur d'incertitude. Ce marqueur se situe au méta-niveau par rapport aux informations elles-mêmes. Il peut être numérique ou linguistique, comme dans les énoncés « la probabilité que l'opération prenne plus d'une heure est 0.7 », ou « il n'est pas absolument certain que Jean vienne à la réunion ». La représentation la plus courante de l'incertitude consiste à attribuer à chaque proposition ou événement A, sous-ensemble de S, un nombre $g(A)$ dans l'intervalle unité. $g(A)$ mesure la confiance de l'agent dans la vérité de la proposition $v \in A$. Cette proposition n'est par convention que vraie ou fausse, même si l'agent peut ignorer cette valeur de vérité. Les conditions suivantes sont naturellement requises :

$$g(\emptyset) = 0;\ \ g(S) = 1;\ \ \text{si } A \subseteq B \text{ alors } g(A) \leq g(B). \tag{3.1}$$

En effet, la proposition contradictoire $\emptyset$ est impossible, et la tautologie S est certaine. De plus si A est plus spécifique que B (et donc l'implique), l'agent ne peut pas avoir plus confiance en A qu'en B. Quand S est infini, on ajoute des propriétés de continuité par rapport à des suites monotones d'ensembles. Une telle fonction g est appelée tantôt *capacité* (Choquet, 1953), tantôt *mesure floue* (Sugeno, 1977), ou encore « fonction de plausibilité » (Halpern, 2001) (à ne pas confondre avec le dual des fonctions de croyance, voir la section 5 de ce chapitre). Une conséquence importante de (3.1) est :

$$g(A \cap B) \leq \min(g(A), g(B));\quad g(A \cup B) \geq \max(g(A), g(B)). \tag{3.2}$$

Une idée naturelle est alors de s'intéresser aux mesures de confiance g telles que $g(A \cup B)$ ne dépendent que de $g(A)$ et de $g(B)$ (quand A et B sont mutuellement exclusifs), c.-à-d.

$$\text{si } A \cap B = \emptyset \text{ alors } g(A \cup B) = g(A) * g(B). \tag{3.3}$$

Une mesure *conjuguée* $\overline{g}(A) = 1 - g(\overline{A})$ satisfait la condition duale $\overline{g}(A \cap B) = \overline{g}(A) \perp \overline{g}(B)$ si $A \cup B = S$ avec $a \perp b = 1 - (1 - a) * (1 - b)$ (Dubois et Prade, 1982). Les mesures g et $\overline{g}$ sont dites *décomposables*. La compatibilité avec l'algèbre de Boole des évènements incite à prendre des opérations $*$ et $\perp$ associatives, ce qui conduit à ce que $\perp$ et $*$ soient respectivement des *normes* et des *co-normes* dites *triangulaires* (Schweizer et Sklar, 1963) (appelées ainsi du fait de leur rôle dans l'expression de l'inégalité triangulaire en géométrie stochastique). Les principaux choix possibles pour $a \perp b$ (resp. $a * b$) sont les opérateurs $\min(a, b)$, produit ($a \times b$), et $\max(0, a + b - 1)$ (resp. $\max(a, b)$, $a + b - a \times b$, et $\min(1, a + b)$). On obtient les mesures de *probabilité* pour $a * b = \min(1, a + b)$ (et pour $a \perp b = \max(0, a + b - 1)$), et les mesures de *possibilité* et de *nécessité* respectivement pour $a * b = \max(a, b)$ et $a \perp b = \min(a, b)$. Des mesures décomposables plus sophistiquées peuvent faire sens (Dubois *et al.*, 2000b).

Gradualité et ensemble flou

La représentation d'une proposition comme un énoncé susceptible d'être vrai ou faux (ou d'un événement comme pouvant se produire ou non) n'est qu'une convention commode. Elle n'est pas toujours idéale. Certains types d'information ne s'y prêtent pas naturellement. C'est notamment le cas des énoncés comportant des propriétés *graduelles*, comme, par exemple, la proposition *Jean est jeune* qui peut n'être ni complètement vraie, ni complètement fausse : elle est plus vraie si Jean a vingt ans que s'il en a trente (même si dans ce dernier cas Jean est encore jeune). De plus on peut moduler le terme *jeune* avec des adverbes d'intensité : on peut dire *très jeune, plus trop jeune*, etc. Ce qui ne serait pas le cas d'une propriété booléenne comme *célibataire*. Autrement dit la proposition *Paul est jeune* n'est pas booléenne, car la propriété *jeune* est graduelle et suggère un ordre implicite entre les valeurs d'attributs auxquels elle se réfère. Ce type d'information est pris en compte par la notion d'*ensemble flou* (Zadeh, 1965). Un ensemble flou F est une application de S dans un ensemble totalement ordonné L qui est souvent l'intervalle $[0, 1]$. $F(s)$ est le degré d'appartenance de l'élément s à F. C'est une mesure de l'adéquation entre la situation s et le prédicat F.

Il est naturel d'utiliser des ensembles flous quand on a affaire à une information en langage naturel qui se réfère à un attribut numérique. Zadeh (1975) a introduit la notion de *variable linguistique* à valeurs dans un ensemble fini ordonné de termes. Chacun de ces termes représente un sous-ensemble de l'échelle numérique associée à l'attribut et ces sous-ensembles correspondent à une partition de cette échelle. Par exemple, l'ensemble de termes $\{$*jeune, adulte, vieux*$\}$ forme le domaine de la variable linguistique *âge(Jean)* et partitionne le domaine de cet attribut. Néanmoins on conçoit que les transitions entre les zones d'âge correspondant aux termes soient graduelles plutôt que nettes. Dans un tel cas, il paraît quelque peu arbitraire de fixer sur une échelle continue un seuil précis $s_\star$ tel que $F(s) = 0$ si $s > s_\star$ et $F(s) = 1$ sinon, au-delà duquel on cesse brutalement d'être jeune. La fonction d'appartenance de l'ensemble flou à valeur sur l'échelle [0, 1], représentant ici la propriété graduelle *jeune*, ne fait que refléter l'échelle continue de l'attribut (ici : l'âge). Les opérations ensemblistes d'union, d'intersection, et de complémentation s'étendent de manière naturelle aux ensembles flous en posant $(F \cup G)(s) = F(s) * G(s)$; $(F \cap G)(s) = F(s) \perp G(s)$; $\overline{F}(s) = 1 - F(s)$ où $*$ et $\perp$ sont les co-normes et normes rencontrées plus haut ; $* = \max$ et $\perp = \min$ correspond au choix le plus usuel. Avec ce choix, les formules de De Morgan sont préservées, ainsi que l'idempotence de $\cup$ et $\cap$ et les distributivités de l'un par rapport à l'autre, mais le tiers exclu ($A \cup \overline{A} = S$) et la contradiction ($A \cap \overline{A} = \emptyset$) ne le sont pas. En choisissant $* = \min(1, \cdot + \cdot)$ et $\perp = \max(0, \cdot + \cdot - 1)$ on rétablirait le tiers exclu et la contradiction, au prix de la perte de l'idempotence et des distributivités. Quant à l'inclusion entre deux ensembles flous elle est usuellement définie par la condition $F \subseteq G \Leftrightarrow \forall s, F(s) \leq G(s)$. Une inclusion plus drastique requiert l'inclusion du support de F (les éléments s tels que $F(s) > 0$) dans le noyau de G (les éléments s tels que $G(s) = 1$). Il existe aussi différents degrés d'inclusion de la forme $d(F \subseteq G) = \min_s F(s) \rightarrow G(s)$, où $\rightarrow$ est une implication dans une logique multivalente. Pour ce qui est du raisonnement approché en termes d'ensembles flous, le lecteur pourra consulter la section du chapitre **??** sur l'interpolation.

Ne pas confondre degré de certitude et degré de vérité

Il est important de distinguer entre degré d'adéquation (souvent appelé *degré de vérité*) et degré de confiance (ou de / (d'in)certitude). Déjà, dans le langage naturel les phrases *Jean est très jeune* et *Jean est probablement jeune* n'ont clairement pas le même sens. Dans le premier cas, le degré d'appartenance de *âge(Jean)* à $F =$ *Jeune* est certainement élevé ; dans l'autre cas il n'est pas totalement exclu que Jean soit vieux. Un degré d'appartenance est vu comme degré d'adéquation si on connaît la valeur *âge(Jean)* $= s$ que l'on veut qualifier linguistiquement : Le qualificatif *jeune* convient alors au degré $F(s)$.

Les degrés de vérité et les degrés de certitude correspondent à des notions distinctes apparaissant dans des situations différentes avec des contenus sémantiques n'ayant rien à voir entre eux. De plus, ils sont gouvernés par des cadres mathématiques qu'il importe de ne pas confondre, en dépit de leurs ressemblances superficielles quant aux opérateurs rencontrés. En effet, comme on l'a vu dans les expressions mentionnées plus haut, les degrés de vérité (ou d'adéquation) peuvent être *compositionnels* à la fois par rapport à la disjonction F ou G représentée par $F \cup G$, à la conjonction F et G représentée par $F \cap G$, et par rapport à la négation représentée par $\overline{F}$, où F et G sont des ensembles flous. Par contre, ceci est impossible pour les degrés de certitude. Cela provient du fait que l'algèbre de Boole des événements n'est pas compatible avec une structure de treillis sur un ensemble à plus de deux valeurs tel que [0, 1]

(Dubois et Prade, 2001). De fait, sur des événements booléens, les probabilités ne sont compositionnelles que par rapport à la négation ($Prob(\overline{A}) = 1 - Prob(A)$), et comme on le verra dans la suite les possibilités (resp. les nécessités) ne sont compositionnelles que par rapport à la disjonction (resp. conjonction) ; en effet, on peut avoir pleinement confiance en la vérité de $A \cup B$ (en particulier si $B = \overline{A}$!), sans n'être ni certain de la vérité de A, ni de la vérité de B.

Mentionnons par ailleurs l'approche proposée par Ginsberg (1990) qui organise les valeurs de vérité selon une structure de bi-treillis, qui est équipée de deux ordres partiels (l'un reflétant la vérité, l'autre la quantité d'information), et d'une opération de négation qui en renverse un laissant l'autre inchangé. Ceci permet d'obtenir un cadre unificateur au plan sémantique pour différentes formes d'inférence non monotone (cf. le chapitre **??** et la sous-section 3.4.2).

Ensembles approximatifs

Dans les paragraphes précédents, on n'a pas discuté les hypothèses qui sous-tendent la définition de l'ensemble S des états du monde. L'ensemble S est rarement donné a priori. L'approche logique en intelligence artificielle part d'un ensemble de propositions exprimées dans un certain langage, et auxquelles un agent attribue des degrés de confiance. Alors S est l'ensemble des états engendré par ces propositions (mathématiquement, ces sous-ensembles forment la plus petite algèbre de Boole contenant ces propositions). Cette façon de voir a des conséquences importantes pour la représentation et la mise à jour des informations. Par exemple, si une nouvelle proposition est ajoutée, il se peut qu'elle fasse changer l'ensemble S en le raffinant. Cela s'appelle *un changement de granularité* de la représentation.

Le cas le plus simple de changement de granularité est le suivant : soit un ensemble Ω d'entités décrites à l'aide d'attributs $V_1, V_2, \ldots, V_k$ de domaines respectifs $D_1, D_2, \ldots, D_k$. Alors S est le produit cartésien $D_1 \times D_2 \times \cdots \times D_k$. Chaque élément de S sera raffiné en plusieurs éléments si on ajoute un $(k + 1)$ième attribut. Supposons un ensemble d'objets Ω décrits par de tels attributs. Rien n'empêche des objets différents d'avoir la même description en termes de ces attributs : ils seront de ce fait indiscernables pour ce langage de description. Soit Θ un sous-ensemble de Ω. On ne peut pas en général le décrire avec S. En effet, soit R la relation d'équivalence sur Ω définie par l'identité des descriptions des éléments ω de Ω : $\omega_1 R \omega_2$ si et seulement si $V_i(\omega_1) = V_i(\omega_2), \forall i = 1, \ldots, k$. Soit $[\omega]_R$ la classe d'équivalence de ω. Chaque élément de S correspond à une classe d'équivalence dans Ω. Alors l'ensemble Θ ne peut souvent être qu'approché par le langage de S et non décrit exactement. Soient Θ^* et Θ_* les approximations supérieures et inférieures de Θ définies comme suit

$$\Theta^* = \{\omega \in \Omega : [\omega]_R \cap \Theta \neq \emptyset\}; \;\; \Theta_* = \{\omega \in \Omega : [\omega]_R \subseteq \Theta\} \tag{3.4}$$

La paire (Θ^*, Θ_*) est appelée *ensemble approximatif* (en anglais : *rough set*) (Pawlak, 1991; Pawlak et Skowron, 2007a,b,c). Seuls les ensembles Θ^* et Θ_* d'individus peuvent être parfaitement décrits par des combinaisons de valeurs d'attributs $V_1, V_2, \ldots, V_k$ correspondant à des sous-ensembles de S. Notons que les histogrammes et les images numérisées correspondent à cette même idée d'indiscernabilité et de granularité, les classes d'équivalences correspondant respectivement aux supports des bandes verticales de l'histogramme et aux pixels. L'idée d'ensemble approximatif est ainsi liée à l'indiscernabilité alors que celle d'ensemble flou est liée à la gradualité. Il est cependant possible de définir des hybridations mutuelles (Dubois et Prade, 1992) si Θ devient un ensemble flou, ou si la relation R (ou la partition induite) devient floue. Les ensembles approximatifs sont utilisés en particulier en apprentissage (voir chapitre **??**).

3.3 Le cadre probabiliste

La théorie des probabilités est la plus ancienne des théories de l'incertain, et à ce titre la mieux développée mathématiquement et la plus établie. On peut envisager la présentation de la théorie des probabilités d'un point de vue purement mathématique. Dans ce cas, on part d'un ensemble d'épreuves Ω et d'une variable aléatoire, entendue comme une application V de Ω dans S (souvent les réels). Dans le cas le plus simple, on suppose que S est un ensemble fini, ce qui définit une partition finie de Ω. Soit $\mathcal{B}$ l'algèbre de Boole engendrée par cette partition. On définit un espace probabilisé comme le triplet $(\Omega, \mathcal{B}, \mathcal{P})$, où P est une mesure de probabilité, c'est-à-dire une application de $\mathcal{B}$ dans $[0, 1]$ telle que

$$P(\emptyset) = 0; \quad P(\Omega) = 1; \tag{3.5}$$

$$\text{si } A \cap B = \emptyset \text{ alors } P(A \cup B) = P(A) + P(B). \tag{3.6}$$

Les éléments de $\mathcal{B}$ sont dits ensembles mesurables. La distribution de probabilité associée à V est alors caractérisée par une pondération $p_1, p_2, \ldots, p_{card(S)}$, définie par $p_i = P(V^{-1}(s_i))$, et telle que

$$\sum_{i=1}^{card(S)} p_i = 1.$$

On peut étendre les probabilités à des événements flous $-$ la formule de base (dans le cas fini) est $P(F) = \sum_i p_i \cdot F(s_i)$ qui généralise à un ensemble flou F la formule $P(A) = \sum_{s_i \in A} p_i$ de la probabilité d'un événement classique A à partir de sa distribution (Zadeh, 1968).

Derrière le modèle mathématique des probabilités, se cachent des visions très différentes de ce que peut signifier une mesure de probabilité (Fine, 1983). Dans ce qui suit on aborde brièvement quelques-uns de ces points de vue en mettant l'accent sur les limitations de la représentation de l'incertain par une distribution unique. On complète la section par un bref aperçu sur les objets conditionnels, contrepartie logique de l'idée d'événement conditionnel, et sur un type très particulier de probabilités, dites à grandes marches, qui jouent un rôle remarquable dans la représentation des règles par défaut.

Fréquentisme et subjectivisme

On considère la théorie des probabilités comme un outil de représentation de l'information. On doit alors donner aux probabilités une interprétation. Il y a au moins trois interprétations des mesures de probabilité.

La plus simple est en termes de dénombrements. On considère que Ω est fini et p_i est proportionnel au nombre d'éléments dans $V^{-1}(s_i)$. On compte le nombre de cas favorables sur le nombre de cas possibles pour évaluer la probabilité de chaque événement. La validité de cette approche repose sur des considérations de symétrie (de type principe d'indifférence) ou des hypothèses de phénomènes *réellement* aléatoires (par exemple, des dés non pipés, etc...) motivant des distributions uniformes.

La plus courante des interprétations est fréquentiste. On suppose qu'on accumule des observations (qui forment un échantillonnage de l'ensemble Ω, soit un sous-ensemble fini $\Omega(n)$ à n éléments). On peut calculer les fréquences d'observation de $V = s_i$

$$f_i = \frac{card(V^{-1}(s_i) \cap \Omega(n))}{n}$$

(ou si S est infini, construire un histogramme associé à la variable aléatoire V en considérant les fréquences des éléments d'une partition finie de S). On suppose alors que lorsque le nombre d'observations augmente, $\Omega(n)$ devient représentatif de Ω, que les fréquences f_i convergent, vers des valeurs de probabilité définies comme $p_i = \lim_{n\to\infty} f_i$. Pour utiliser cette définition des probabilités, il faut disposer d'un nombre suffisant d'observations (idéalement un nombre infini) du phénomène observé. Ceci interdit d'attribuer des probabilités à des événements non répétables.

Ce qui joue le rôle des fréquences pour les événements non-répétables, ce sont des sommes d'argent misées sur l'occurrence ou la non-occurrence d'événements. On définit le degré de confiance d'un agent en l'événement A comme le prix $P(A)$ que cet agent accepterait de payer pour acheter un billet de loterie qui lui fait gagner 1 euro si l'événement A se produit. Plus l'agent croit en l'occurrence de A, moins il estime risqué d'acheter un billet de loterie à un prix proche de 1 euro. Il faut de plus supposer que celui qui vend les billets de loterie (le banquier) refuse de vendre si on lui propose des prix trop bas, afin de forcer l'agent à donner un juste prix : s'il estime le prix d'achat trop bas il peut imposer à l'agent un échange de rôles, c'est-à-dire obliger l'agent à lui vendre un billet de loterie au prix $P(A)$ et à lui payer 1 euro si l'événement A se produit. L'approche repose sur un principe de cohérence qui stipule que l'agent est rationnel, c'est-à-dire qu'il cherche à éviter les pertes d'argent sûres. Supposons que l'agent achète deux billets de loterie relatifs à deux propositions contraires A et $\overline{A}$. Le principe de cohérence impose que $P(A) + P(\overline{A}) = 1$. En effet, un seul des deux événements A ou $\overline{A}$ peut se produire. Donc $P(A) + P(\overline{A}) \leq 1$, sinon l'agent perd sûrement $P(A) + P(\overline{A}) - 1$ euros. Mais si l'agent propose des prix tels que $P(A) + P(\overline{A}) < 1$ alors le banquier prend sa place. De même, avec trois propositions mutuellement exclusives A, B et $\overline{A \cup B}$, on montre que seul $P(A) + P(B) + P(\overline{A \cup B}) = 1$ est rationnel, et comme $P(\overline{A \cup B}) = 1 - P(A \cup B)$, on en conclut que $P(A \cup B) = P(A) + P(B)$.

L'approche subjectiviste semble une simple réinterprétation du calcul des probabilités fréquentistes. En fait, comme le montrent (De Finetti, 1974) et ses élèves (Coletti et Scozzafava, 2002), ce n'est pas si simple. Dans l'approche subjectiviste, il n'y a pas d'espace d'épreuves. Le point de départ est un ensemble de propositions booléennes $\{A_j : j = 1, n\}$ auxquelles sont attachés des degrés de confiance c_i, et un ensemble de contraintes logiques entre ces propositions. On construit l'ensemble des états S induit par ces propositions et ces contraintes. On suppose, en vertu du principe de cohérence, que l'agent a affecté ces degrés de confiance selon une mesure de probabilité P telle que $c_j = P(A_j), \forall j = 1, \ldots, n$. Alors que l'approche fréquentiste amène à poser l'existence d'une mesure de probabilité unique (obtenue par estimation à partir de données statistiques) qui modélise le phénomène étudié, on voit qu'il n'en est rien ici. Il y a éventuellement plusieurs mesures de probabilité telles que $c_j = P(A_j), \forall j = 1, \ldots, n$. Chacune est rationnelle, mais les informations dont on dispose ne permettent pas forcément de l'isoler. Il peut aussi n'y en avoir aucune si l'agent n'est pas cohérent. Pour calculer la probabilité $P(A)$ d'un événement quelconque A sur la base des paires $\{(A_j, c_j) : i = 1, n\}$, on est amené à résoudre un problème de programmation linéaire (dont les variables de décision sont les probabilités élémentaires p_i) de la forme : maximiser (ou minimiser) $\sum_{s_i \in A} p_i$ sous les contraintes $c_j = \sum_{s_k \in A_j} p_k, \forall j = 1, \ldots, n$. En ce sens, l'approche subjectiviste des probabilités est une extension de l'approche logique de la représentation des connaissances et de la déduction. Il y a d'autres différences entre probabilités subjectives et fréquentistes quand on aborde la notion de conditionnement.

Probabilités conditionnelles

En prenant S comme référentiel, on fait implicitement l'hypothèse que S représente l'ensemble des états du monde possibles. Cela suggère que l'on écrive la probabilité $P(A)$ sous la forme $P(A \mid S)$ pour mettre ce fait en évidence. Si par la suite l'agent obtient de nouvelles informations qui l'amènent à restreindre plus avant l'ensemble des états du monde, les probabilités vont changer de contexte. Soit $C \subset S$ le nouveau contexte, et soit $P(A \mid C)$ la probabilité de A dans ce contexte. Le passage de $P(A)$ à $P(A \mid C)$ consiste essentiellement à renormaliser les probabilités affectées aux états de C, soit

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)} \tag{3.7}$$

On retrouve bien la valeur de $P(A)$ sous la forme $P(A \mid S)$. Cette définition est facile à justifier dans le cas fréquentiste, car $P(A \mid C)$ est la limite d'une fréquence relative.

Dans un cadre subjectiviste c'est beaucoup moins trivial. La probabilité $P(A \mid C)$ est attachée à l'occurrence de l'événement conditionnel $A \mid C$. Elle est considérée comme primitive (et non plus dérivée de la mesure de probabilité). $A \mid C$ représente l'occurrence de l'événement A dans le contexte hypothétique où C est vrai. Dire que la distribution de probabilité P est connue, c'est disposer de toutes les valeurs $P(A \mid C)$ dans tous les contextes. L'agent ne fait que choisir la probabilité conditionnelle adaptée à ses connaissances sur la situation courante, une vision très différente de celle du changement de mesure de probabilité suite à une révision de connaissances. La quantité $P(A \mid C)$ est alors encore interprétée comme une somme d'argent misée sur A, mais on suppose de plus que cette somme est remboursée au joueur si l'événement C n'a pas lieu (De Finetti, 1974). Dans ce cadre opérationnel, on montre que l'identité $P(A \cap C) = P(A \mid C) \cdot P(C)$ prend tout son sens. La définition de la probabilité conditionnelle sous la forme d'un quotient présuppose $P(C) \neq 0$, ce qui peut parfois s'avérer trop restrictif. En effet dans le cadre de De Finetti où les probabilités recueillies peuvent concerner n'importe quel événement conditionnel, on peut imaginer que l'ensemble de connaissances dont on dispose pour raisonner prenne la forme d'un ensemble de probabilités conditionnelles $\{P(A_i \mid C_j), i = 1, \ldots, m; j = 1, \ldots, n\}$ correspondant à divers contextes potentiels dont la probabilité peut parfois être nulle. La définition de la probabilité conditionnelle comme toute solution de l'équation $P(A \cap C) = P(A \mid C) \cdot P(C)$ fait encore sens si $P(C) = 0$ (voir (Coletti et Scozzafava, 2002)).

Certains justifient la probabilité conditionnelle en termes de révision (Gärdenfors, 2008). La quantité $P(A \mid C)$ est alors vue comme la nouvelle probabilité de A lorsque l'agent apprend que C s'est produit. Un principe fondamental de la révision des croyances est le changement minimal : l'agent révise ses croyances au minimum de façon à absorber l'information nouvelle interprétée par la contrainte $P(C) = 1$. Il est à noter que dans ce cadre on ne peut faire la différence entre la simple *observation factuelle* qu'on se trouve dans une situation où C est vrai pour laquelle on cherche à faire une prédiction, et la situation où on apprendrait que C est *toujours* vrai (Dubois et Prade, 1997b). Ainsi dans un problème de la prédiction, face à une situation particulière, représentée par une proposition C (en médecine, les résultats d'analyse d'un patient), on tente de formuler des affirmations A sur le monde courant avec leurs degrés de croyance associés (prédire la maladie du patient). On utilise alors la probabilité conditionnelle $P(A \mid C)$ (qui est par exemple la fréquence d'observation de A dans le contexte C). Par contre le scénario de la révision (voir chapitre **??**) est différent : étant donné une distribution de probabilité P (qui peut être de l'information générique ou non), on apprend que la probabilité d'un événement C est 1 (et non $P(C) < 1$ comme on le supposait auparavant). Il s'agit alors de déterminer la nouvelle mesure de probabilité P', telle que $P'(C) = 1$, la plus proche de P, pour satisfaire au principe de changement minimal (Domotor, 1985). On montre alors qu'en utilisant une mesure d'information relative appropriée, $P'(A) = P(A \mid C), \forall A$.

La propriété d'additivité des probabilités permet de déduire deux résultats remarquables sur les probabilités conditionnelles :

– Le théorème des probabilités totales :
  Si $\{C_1, \ldots, C_k\}$ forment une partition de S, alors $P(A) = \sum_{i=1}^{k} P(A \mid C_i)P(C_i)$.
– Le théorème de Bayes :

$$P(C_j \mid A) = \frac{P(A|C_j)P(C_j)}{\sum_{i=1}^{k} P(A|C_i)P(C_i)}.$$

Le premier permet d'induire la probabilité d'un événement dans un contexte général S connaissant les probabilités de cet événement dans des contextes particuliers, pourvu que ces contextes forment une partition des états possibles, et pourvu que l'on dispose des probabilités de chacun de ces contextes. Le théorème de Bayes traite le problème de classification suivant : on dispose de k classes d'objets qui forment une partition de S. Si on suppose connue la probabilité d'apparition $P(A \mid C_j)$ de la propriété A pour les objets de chaque classe C_j, ainsi que la probabilité a priori qu'un objet soit de classe C_j, alors si on dispose d'un nouvel objet pour lequel on observe la propriété A, on sait calculer la probabilité $P(C_j \mid A)$ que cet objet appartienne à la classe C_j. Dans les problèmes de diagnostic, on peut remplacer classe par dysfonctionnement, et propriété observée par symptôme. L'exploitation des probabilités conditionnelles dans le cadre des réseaux bayésiens (Pearl, 1988) est discutée au chapitre **??** de cet ouvrage.

En 1946, R. T. Cox (1946) proposa de justifier la notion de probabilité comme mesure de croyance, en s'appuyant sur la structure d'algèbre de Boole des événements, à partir des postulats suivants, où $g(A|B) \in [0, 1]$ est un degré de croyance conditionnel, A, B étant des évènements d'une algèbre de Boole, avec $B \neq \emptyset$:

i) $g(A \cap C|B) = F(g(A|C \cap B), g(C|B))$ (si $C \cap B \neq \emptyset$) ;
ii) $g(\overline{A}|B) = n(g(A|B))$, $B \neq \emptyset$, où $\overline{A}$ est le complémentaire de A ;
iii) la fonction F est deux fois différentiable, avec une dérivée seconde continue, et la fonction n est deux fois différentiable,

Sur cette base, Cox affirma que $g(A|B)$ doit être isomorphe à une mesure de probabilité. Ce résultat a été répété à l'envie pour justifier les mesures de probabilité comme la seule façon raisonnable de représenter numériquement des degrés de croyance (Horvitz *et al.*, 1986; Cheeseman, 1988; Jaynes, 2003). Outre que la démonstration originale de Cox s'est avérée fautive – voir (Paris, 1994) pour une autre version du théorème avec des conditions iii) modifiées : il suffit que F soit strictement monotone croissante pour chaque argument), et voir (Halpern, 1999a,b) qui montre que le résultat ne tient pas avec des ensembles finis, et requièrt un postulat technique additionnel pour le valider dans le cas infini) – il est à souligner que cette approche exclut d'emblée par son postulat ii) toutes les autres approches de la représentation de l'incertain considérées dans ce chapitre, empéchant la représentation de l'information incomplète, ce qui réduit singulièrement la portée du résultat.

Les probabilités et la logique classique sont des cadres qu'il n'est pas simple de combiner. Rappelons que la barre de conditionnement n'est pas un connecteur logique, et que $Prob(q|p)$

et $Prob(p \to q) = Prob(\neg p \vee q)$ ont en général des valeurs complètement différentes, et ne coïncident que si elles sont égales à 1. De plus, un ensemble de propositions auxquelles sont associées une même borne inférieure de probabilité n'est en général pas clos déductivement si cette borne est plus petite que 1 (Kyburg, Jr. et Teng, 2012). Par ailleurs, au premier ordre, il convient de ne pas confondre une conjecture universelle incertaine (Gaifman et Snir, 1982) (par exemple, $Prob(\forall x, P(x) \to Q(x)) = \alpha$) avec un énoncé universel relatif à une probabilité (par exemple, $\forall x, Prob(P(x) \to Q(x)) = \alpha$, ou $\forall x, Prob(Q(x)|P(x)) = \alpha$). Des extensions de réseaux bayésiens à des langages du premier ordre ont été développés (Milch et Russell, 2007). Bornons-nous par ailleurs à mentionner des travaux qui se sont efforcés de réconcilier probabilités et logiques (propositionnelle, du premier ordre, mais aussi modale) de différentes manières : (Halpern, 1990; Bacchus, 1991; Nilsson, 1993; Abadi et Halpern, 1994; Marchioni et Godo, 2004; Jaeger, 2001; Halpern et Pucella, 2002, 2006; Jaeger, 2006), ainsi que des travaux visant à développer une version probabiliste qualitative des réseaux bayésiens (Renooij et van der Gaag, 1999; Parsons, 2001; Bolt *et al.*, 2005; Renooij et van der Gaag, 2008).

Probabilité unique et représentation de l'ignorance

L'approche subjectiviste dite *bayésienne* des probabilités subjectives pose l'unicité de la mesure de probabilité comme préalable à toute modélisation (voir par exemple (Lindley, 1982)). De fait si l'agent décide d'attribuer des probabilités subjectives aux éléments de S directement, le principe de cohérence l'oblige à définir une distribution de probabilité unique. Si les connaissances dont on dispose sont insuffisantes pour la caractériser, l'approche bayésienne fait souvent appel à des principes de sélection, tels que principe d'indifférence pour exploiter les symétries, ou le principe de maximum d'entropie (Jaynes, 1979; Paris, 1994). Le recours au principe de maximum d'entropie dans un cadre subjectiviste est contestable car il sélectionne dès que possible la distribution uniforme.

Le credo bayésien est que tout état épistémique d'un agent est représentable par une distribution de probabilité a priori. L'emploi systématique d'une probabilité unique comme outil universel de représentation de l'incertitude pose néanmoins de sérieux problèmes. En particulier, on ne fait plus la différence entre l'information incomplète sur une situation et le cas où cette situation est réellement aléatoire. Dans le cas du jet d'un dé, il est difficile d'interpréter de façon non ambiguë l'affectation d'une distribution uniforme de probabilité aux faces du dé. En effet il se peut que l'agent sache que le dé n'est pas pipé (aléatoire pur) et que la distribution de fréquence limite doit être uniforme. Mais si l'agent ignore tout du dé, qu'il n'a pas pu tester, alors la distribution uniforme obtenue n'est que le résultat du principe de symétrie (l'agent n'a aucune raison de penser qu'il gagnera plus d'argent en pariant sur une face plutôt que sur une autre). Par ailleurs, le choix d'un référentiel dépend souvent de la source d'information, et plusieurs points de vue ou plusieurs langages peuvent coexister pour un même problème. Une distribution uniforme sur un référentiel ne correspondra pas à une distribution uniforme sur un autre. Prenons l'exemple de l'existence de la vie extra-terrestre (Shafer, 1976) : l'agent ne sait pas s'il y en a ou pas. Si v représente l'affirmation de la vie, et $\neg v$ son contraire, $P_1(v) = P_1(\neg v) = \frac{1}{2}$ sur $S_1 = \{v, \neg v\}$. Mais on peut aussi distinguer entre vie animale (va), et vie végétale seulement (vv), et considérer le référentiel $S_2 = \{va, vv, \neg v\}$, et l'agent ignorant va alors proposer $P_2(va) = P_2(vv) = P_2(\neg v) = \frac{1}{3}$. Comme v est la disjonction de va et vv, les distributions P_1 et P_2 sont incompatibles alors qu'elles sont censées représenter le même état de connaissance. Enfin il y a un problème de mesurage dans le cas des probabilités subjectives. Il est difficile d'affirmer que l'agent puisse fournir, même au travers d'estimations de prix, des valeurs de probabilité très précises.

Ces remarques ont motivé le développement d'autres approches de l'incertitude. Pour certaines, on abandonne le cadre numérique au profit de structures ordinales, qu'on peut alors considérer comme sous-jacentes aux représentations numériques subjectivistes. Pour d'autres, on injecte de l'incomplétude dans le modèle probabiliste, obtenant diverses approches de degrés de généralité mathématique divers. Dans les deux cas on retrouve la théorie des possibilités (qualitative ou quantitative, respectivement (Dubois et Prade, 1998)) comme fournissant le modèle non trivial de l'incertain non-probabiliste le plus simple de tous.

Objets conditionnels et probabilités à grandes marches

On peut considérer la probabilité conditionnelle $P(A \mid C)$ comme la probabilité d'un *tri-événement* $A \mid C$ que l'on lira « si la connaissance courante est synthétisée par C alors conclure A », où A et C représentent des propositions classiques (des sous-ensembles de S). Ce qui a été proposé par (De Finetti, 1936) pour la première fois. Un tri-événement $A \mid C$ (on parle aussi d'« objet conditionnel »), partitionne alors en trois morceaux l'ensemble des états $s \in S$:

- soit $s \in A \cap C$; on dit que s est un exemple de la règle « si C alors A ». Le tri-événement est vrai (valeur 1) dans l'état s ;
- soit $s \in \overline{A} \cap C$; on dit que s est un contre-exemple de la règle « si C alors A ». Le tri-événement est faux (valeur 0) dans s ;
- soit $s \in \overline{C}$; on dit que la règle « si C alors A » ne s'applique pas à s. Le tri-événement prend une troisième valeur de vérité (I) dans s.

Un tri-événement $A \mid C$ est ainsi interprété comme une paire $(A \cap C, \overline{A} \cap C)$ d'ensembles disjoints. Il existe une relation de conséquence naturelle entre deux objets conditionnels, définie par

$$B \mid A \vDash D \mid C \Leftrightarrow A \cap B \vDash C \cap D \text{ and } C \cap \overline{D} \vDash A \cap \overline{B}$$

qui exprime que les exemples de $B \mid A$ sont des exemples de $D \mid C$ et que les contre-exemples de $D \mid C$ sont des contre-exemples de $B \mid A$. Une telle représentation des règles « si ... alors » a le mérite d'éviter les paradoxes de la confirmation (si on convient que l'observation de A et B vrais confirme la règle « si A alors B », alors en logique classique où l'implication matérielle est contraposable, on sera amené à dire que la vue d'un cygne blanc confirme que « tous les corbeaux sont noirs », ce qui est impossible avec les objets conditionnels ($B \mid A$ n'équivaut pas à $\overline{A} \mid \overline{B}$ puisqu'ils ne correspondent pas aux mêmes exemples quoiqu'ils aient les mêmes contre-exemples (Benferhat *et al.*, 2008)).

Un calcul trivalué a été développé pour les objets conditionnels (Dubois et Prade, 1994), qui permet par exemple d'établir une contrepartie qualitative de la règle de Bayes : $(A \cap B) \mid S = (B \mid A) \wedge (A \mid S)$ (où $\wedge$ est ici une conjonction trivaluée), et de proposer une sémantique simple pour le système d'inférence préférentielle (Kraus *et al.*, 1990) modélisant le raisonnement tolérant les exceptions, l'objet conditionnel $B \mid A$ représentant la règle « (généralement) si A alors B » (voir la section sur l'inférence non monotone du chapitre **??** et la sous-section 3.4.2). En effet, les deux objets conditionnels $B \mid A$ et $\overline{B} \mid (A \cap C)$ peuvent coexister dans la même base sans créer d'incohérence (à la différence d'une modélisation en logique propositionnelle). Dans ce cadre, à partir d'une situation où tout ce qu'on sait est que E est vrai, on peut tirer toute conclusion F telle que $F \mid E$ soit en relation de conséquence avec la conjonction (définie de manière appropriée) de tous les objets conditionnels de la base représentant la connaissance générique disponible (Benferhat *et al.*, 1997).

Une condition minimale pour que A soit assimilable à une croyance acceptée en termes probabilistes est que $P(A) > P(\overline{A})$ (ce qui équivaut à $P(A) > 1/2$). Mais un ensemble de contraintes $P(A_i) > \alpha$ $(i = 1, n)$ n'est pas en général clos déductivement si $\alpha < 1$, comme on l'a rappelé. Cependant, la propriété de clôture déductive est préservée pour un type très particulier de probabilités, dites *à grandes marches* qui sont définies sur un référentiel fini par la condition

$$\forall i < n-1, p_i > \sum_{j=i+1,\ldots,n} p_j \text{ où } p_i = P(s_i) \text{ avec } p_1 > \ldots > p_{n-1} \geq p_n > 0.$$

Un exemple d'une telle distribution, pour $n = 5$, est $p_1 = 0.6, p_2 = 0.3, p_3 = 0.06, p_4 = 0.03, p_5 = 0.01$. Ce type de probabilités (qui sont complètement à l'opposé des probabilités uniformes) offre une sémantique probabiliste au système d'inférence préférentielle pour le raisonnement en présence d'exceptions (Benferhat *et al.*, 1999b; Snow, 1999).

Mentionnons enfin que puisque $P(A|B)$ peut s'écrire comme une fonction de $P(A \cap B)$ et de $P(A \cup \overline{B})$, une approche à la Cox peut être envisagée en remplaçant le premier postulat par $g(A|B) = h(g(A \cap B), g(A \cup \overline{B}))$, et en ajoutant le postulat $g((A|B)|C) = g(A|B \cap C)$, si $B \cap C \neq \emptyset$ pour étudier théoriquement le conditionnement dans le cadre des probabilités imprécises (Dubois *et al.*, 2010).

3.4 Théorie des possibilités

Les éléments de base de la théorie des possibilités (Zadeh, 1978; Dubois et Prade, 1987b) ont été proposés par Zadeh, complètement indépendamment des travaux d'un économiste anglais (Shackle, 1961) qui avait ébauché une théorie similaire (en termes de degré de surprise – ce qui correspond à un degré d'impossibilité) pour offrir un modèle plus fidèle que les probabilités de la manière dont les agents humains appréhendent l'incertain. Les mesures de *possibilité* s'avèrent être des mesures max-décomposables pour la disjunction. D'abord conçues par Zadeh en relation avec la représentation d'éléments d'information partielle exprimée linguistiquement (typiquement, « quelle est la possibilité que Paul ait plus de 30 ans sachant qu'il est jeune »), il est ensuite apparu que la brique de base utilisée par Zadeh pour bâtir la théorie des possibilités, à savoir la notion de distribution de possibilité, n'avait pas nécessairement pour origine en pratique la représentation de propriétés graduelles (telles que « jeune » dans l'exemple ci-dessus), mais permettait de formaliser tout état épistémique nuancé où l'on assigne un degré de possibilité aux interprétations induites par un langage propositionnel quelconque. Les mesures de possibilité se sont vues ensuite associées par dualité à des mesures de *nécessité* de manière naturelle, pour être finalement complétées par deux autres fonctions d'ensemble. C'est ce qui est d'abord rappelé dans la présentation du cadre général qui suit, avant de distinguer les possibilités quantitatives des possibilités qualitatives, et de présenter succinctement la logique possibiliste. Cette section est complétée par la relation entre les possibilités qualitatives, la représentation des règles par défaut et l'inférence non monotone, et se termine par une brève discussion de l'analyse formelle de concepts qui quoique développée dans une perspective totalement différente, s'avère offrir un parallèle formel remarquable avec la théorie des possibilités.

3.4.1 Le cadre général

Soit π_x une fonction du référentiel S sur une échelle L, qui peut être l'intervalle $[0,1]$, un sous-ensemble fini tel que par exemple $\{0, 0.1, 0.2, \cdots, 0.9, 1\}$, ou un ensemble quelconque totalement ordonné (de niveaux symboliques). Plus généralement, L peut être remplacé par un treillis avec un plus grand et un plus petit élément. Dans la suite, L est un ensemble totalement ordonné ayant un plus grand et un plus petit élément, qu'on notera 1 et 0 respectivement. Le degré de possibilité $\pi_x(s)$ est d'autant plus grand que la valeur s est davantage plausible pour la variable x supposée se rapporter à un attribut (l'âge de Paul dans l'exemple plus haut). Notre information sur l'état du monde est représentée par π_x, qui est appelée *distribution de possibilité* sur S. Notons que si la distribution de possibilité est induite au travers de termes linguistiques graduels, on évalue alors la plausibilité de s en termes de distance à des situations idéalement plausibles, et non en termes de fréquence d'occurrence par exemple. Les valeurs s telles que $\pi_x(s) = 0$ sont considérées comme impossibles pour x. Les valeurs s telles que $\pi_x(s) = 1$ sont donc les plus plausibles pour x (il peut y en avoir plusieurs) ; de plus, si l'information est cohérente, il doit y en avoir au moins une, on dit alors que la distribution est *normalisée*. Il y a deux cas extrêmes d'information imprécise : 1) l'ignorance totale : en l'absence d'information, on ne peut affirmer que la tautologie, qui prend ici la forme $x \in S$, représentée par la distribution de possibilité $\pi_x^?(s) = 1, \forall s \in S$; 2) l'information précise : elle prend la forme $x = s_0$ pour un état $s_0 \in S$, représentée par la distribution de possibilité $\pi_x^{s_0}(s) = 1$ si $s = s_0$ et 0 sinon. Notons que c'est la valeur 0 qui apporte l'information.

Une distribution de possibilité exprime une restriction sur les valeurs plus ou moins possibles d'une variable x. Une distribution π_x est dite plus *spécifique* pour x que π'_x (c'est-à-dire plus restrictive) si $\forall s, \pi_x(s) \leq \pi'_x(s)$. Quand on représente l'information disponible sur les valeurs possibles d'une variable, il importe donc d'être aussi spécifique que possible (pour ne pas être imprécis), mais surtout pas plus spécifique que permis par l'information pour ne pas tomber dans une précision illusoire ou arbitraire. Des mesures de spécificité ont été définies de manière analogue à l'entropie probabiliste (Higashi et Klir, 1982).

Les deux mesures de base

Des évaluations de vraisemblance et de certitude induite par l'information représentée par la distribution π_x sur la proposition $x \in A$ peuvent alors être calculées en termes de degré de possibilité et de nécessité de l'événement A :

$$\Pi(A) = \max_{s \in A} \pi_x(s); \quad N(A) = 1 - \Pi(\overline{A}) = \min_{s \notin A} 1 - \pi_x(s) \tag{3.8}$$

La notation $1 - \cdot$ ne doit pas suggérer qu'on est nécessairement sur une échelle numérique, c'est juste la fonction de renversement de l'ordre sur L. Quand la distribution π_x ne prend que des valeurs sur $\{0, 1\}$, c'est-à-dire qu'il existe un ensemble $E \subseteq S$ tel que $\pi_x(s) = 1 \Leftrightarrow s \in E$, il est facile de voir que $\Pi(A) = 1$ si et seulement si la proposition $x \in A$ n'est pas incohérente avec l'information $x \in E$, c'est-à-dire si $A \cap E \neq \emptyset$, et que $N(A) = 1$ si et seulement si la proposition $x \in A$ est impliquée par l'information $v \in E$ (on a $E \subseteq A$). $\Pi(A) = 0$ signifie que A est impossible si $x \in E$ est vrai. $N(A) = 1$ exprime que A est certain si $x \in E$ est vrai.

De manière générale, les fonctions N et Π sont totalement liées entre elles par la propriété de dualité $N(A) = 1 - \Pi(\overline{A})$, laquelle différencie nettement les mesures de nécessité et de possibilité, des probabilités qui sont auto-duales au sens où $P(A) = 1 - P(\overline{A})$. L'évaluation de l'incertitude de type possibiliste est implicitement à l'oeuvre en logique classique. Si K est une base propositionnelle dont E forme l'ensemble des modèles, et p est la forme syntaxique de la proposition $x \in A$, alors $N(A) = 1$ si et seulement si K implique p, et $\Pi(A) = 0$ si et seulement si $K \cup \{p\}$ est logiquement incohérent. Notons que la présence de p dans K signifie que $N(A) = 1$, alors que la présence de la négation $\neg p$ dans K signifie que $\Pi(A) = 0$. En revanche on ne sait pas exprimer dans K que $N(A) = 0$ ni que $\Pi(A) = 1$. Pour le faire, il faut utiliser le formalisme de la logique modale (voir chapitre **??**), qui préfixe les propositions par les modalités du possible ($\Diamond$) et du nécessaire ($\Box$) : Dans une base modale K^{mod}, $\Diamond p \in K^{mod}$ exprime $\Pi(A) = 1$, et $\Box p \in K^{mod}$ exprime $N(A) = 1$ (déjà exprimé par $p \in K$ en logique classique). La relation de dualité entre Π et N est bien connue en logique modale, où elle s'écrit $\Diamond p = \neg\Box\neg p$. Une telle logique modale (un fragment élémentaire de la logique KD), appelée MEL, avec une sémantique en termes de distributions de possibilité booléennes, a été définie par Banerjee et Dubois (2009) (une idée semblable avait déjà été suggérée par Mongin (1994)). L'approche possibiliste distingue donc trois états épistémiques extrêmes :

- la certitude que $x \in A$ est vrai : $N(A) = 1$, donc $\Pi(A) = 1$;
- la certitude que $x \in A$ est faux : $\Pi(A) = 0$, donc $N(A) = 0$:
- l'ignorance quant à $x \in A$: $\Pi(A) = 1$, et $N(A) = 0$.

qui peuvent être affinés dès que L comporte au moins un élément intermédiaire entre 0 et 1 par des états épistémiques tels que $0 < N(A) < 1$ ou $0 < \Pi(A) < 1$.

Il est facile de vérifier dans le cas général que les mesures de possibilité et de nécessité saturent chacune l'une des inégalités (3.2) :

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)). \tag{3.9}$$

$$N(A \cap B) = \min(N(A), N(B)). \tag{3.10}$$

Les mesures de possibilité sont dites *maxitives* et sont caractérisées (dans le cas fini) par la propriété de maxitivité (3.9), et les mesures de nécessité sont dites *minitives* et sont caractérisées (dans le cas fini) par la propriété de minitivité (3.9), même quand ces mesures sont à valeurs dans $[0, 1]$. En général, les mesures de possibilité et de nécessité sont distinctes. On ne peut avoir les propriétés de maxitivité et de minitivité pour tous les événements, sauf si $N = \Pi$ correspond à une information précise ($E = \{s_0\}$), et coïncide aussi avec une mesure de probabilité dite de Dirac. En général, $N(A \cup B) > \max(N(A), N(B))$ et $\Pi(A \cap B) < \min(\Pi(A), \Pi(B))$. Dans ces inégalités, l'écart peut être maximal. Il est facile de vérifier que si on ignore si A est vrai ou faux (car $A \cap E \neq \emptyset$ et $\overline{A} \cap E \neq \emptyset$), alors $\Pi(A) = \Pi(\overline{A}) = 1$ et $N(A) = N(\overline{A}) = 0$; mais par construction $\Pi(A \cap \overline{A}) = \Pi(\emptyset) = 0$ et $N(A \cup \overline{A}) = N(S) = 1$.

Deux fonctions d'ensemble décroissantes. Bipolarité

Une autre fonction d'ensemble Δ et sa duale ∇ (introduites en 1991, voir par exemple (Dubois et Prade, 1998)) peuvent être associées à la distribution π_x de manière naturelle dans ce cadre :

$$\Delta(A) = \min_{s \in A} \pi_x(s); \quad \nabla(A) = 1 - \Delta(\overline{A}) = \max_{s \notin A} 1 - \pi_x(s) \tag{3.11}$$

Observons tout d'abord qu'à la différence de Π et de N, Δ et ∇ sont des fonctions *décroissantes* par rapport à l'inclusion ensembliste (et donc par rapport à la relation de conséquence logique). Pour cette raison, on parle quelquefois d'anti-mesure. Δ est appelé mesure de *possibilité forte* ou *garantie* puisque sur A, la possibilité n'est jamais inférieure à $\Delta(A)$ (tandis que Π qui n'évalue qu'un degré de cohérence est une mesure de possibilité faible) ; ∇ est une mesure de nécessité faible, tout comme N est une mesure de nécessité forte. De fait, on a toujours l'inégalité :

$$\forall A, \max(\Delta(A), N(A)) \leq \min(\Pi(A), \nabla(A)) \tag{3.12}$$

pourvu que π_x et $1 - \pi_x$ soient normalisés. Les mesures Δ et ∇ sont caractérisées par les propriétés suivantes :

$$\Delta(A \cup B) = \min(\Delta(A), \Delta(B)); \;\; \Delta(\emptyset) = 1. \tag{3.13}$$

$$\nabla(A \cap B) = \max(\nabla(A), \nabla(B)); \;\; \nabla(S) = 0. \tag{3.14}$$

D'un point de vue représentation des connaissances, il est intéressant de considérer le cas où la distribution de possibilité π_x ne prend qu'un nombre fini de valeurs distinctes $\alpha_1 = 1 > \cdots > \alpha_n > \alpha_{n+1} = 0$. Elle peut alors être décrite par n sous-ensembles emboîtés $E_1 \subseteq \cdots \subseteq E_i \subseteq \cdots \subseteq E_n$ où $\pi_x(s) \geq \alpha_i \Leftrightarrow s \in E_i$. On peut alors vérifier que $\Delta(E_i) \geq \alpha_i$, que $N(E_i) \geq 1-\alpha_{i+1}$ pour $i = 1, n$, et que $\pi_x(s) = \max_{E_i \ni s} \Delta(E_i) = \min_{E_i \not\ni s}(1-N(E_i))$ (avec les conventions $\max_\emptyset = 0$ et $\min_\emptyset = 1$). Une distribution π_x peut donc être vue à la fois comme une disjonction d'ensembles E_i pondérée au sens de Δ, et comme une conjonction d'ensembles E_i pondérée au sens de N. La première « lecture » en termes de Δ propose une vue positive de la distribution de possibilité qui dit à quel point chaque valeur est possible, tandis que la seconde en terme de N dit à quel point chaque valeur est non impossible (en effet s est d'autant plus impossible que s n'appartient qu'à un nombre plus restreint de E_i).

Ces caractères positif et négatif attachés respectivement à Δ et à N, sont à la base d'une représentation *bipolaire* de l'information en théorie des possibilités (Benferhat *et al.*, 2008). L'idée de bipolarité réfère à un traitement explicite des aspects positifs ou négatifs que peut présenter l'information (Dubois et Prade, 2008). Il existe plusieurs formes de bipolarité ; on ne décrit ici, brièvement, que le cas où la bipolarité est induite par deux formes d'information distinctes. Dans le cadre possibiliste, on utilise *deux* distributions de possibilité δ_v et π_x pour représenter respectivement les valeurs garanties possibles pour x et les valeurs non impossibles (car non rejetées). L'idée de représentation bipolaire s'applique alors aussi bien aux connaissances qu'aux préférences (cf. chapitre **??**). Ces deux distributions s'interprètent alors différemment : $\delta_x(s) = 1$ signifie que s est assurément possible car cette valeur ou cet état a été observé dans le cas des connaissances (ou qu'elle/il est réellement satisfaisant(e) si on parle de préférences). Par contre $\delta_x(s) = 0$ signifie juste qu'on ne sait encore rien sur cette valeur s'il s'agit de connaissances (ou qu'elle n'a rien de particulièrement attractif dans le cas des préférences). A l'inverse, $\pi_x(s) = 1$ signifie que s n'a rien d'impossible, mais $\pi_x(s) = 0$ signifie que s est définitivement exclue (ou non acceptable pour des préférences). Comme ce qui est garanti possible doit être parmi ce qui est non exclu, on doit imposer la condition de cohérence $\delta_x \leq \pi_x$ (qui correspond à une inclusion standard d'ensembles flous). En logique possibiliste (voir ci-après) la distribution π_x s'obtient à partir de contraintes de la forme $N(A_i) \geq \eta_i$ et δ_x à partir de contraintes de la forme $\Delta(B_j) \geq \delta_j$ où $A_i \subseteq S, B_j \subseteq S$, et $\eta_i \in L, \delta_j \in L$. L'idée de représentation bipolaire n'est pas propre à la théorie des possibilités, même si jusqu'à présent elle a été assez peu considérée dans d'autres cadres, voir cependant (Dubois *et al.*, 2000a).

Evaluation possibiliste d'événements flous

Les fonctions d'ensemble Π, N, Δ et ∇ s'étendent à des ensembles flous. La possibilité d'un événement flou F est définie (Zadeh, 1978) par

$$\Pi(F) = \sup_s \min(F(s), \pi_x(s)).$$

La nécessité est toujours liée par la dualité $N(F) = 1 - \Pi(\overline{F}) = \inf_s \max(F(s), 1 - \pi_x(s))$. Π et N continuent de satisfaire les propriétés de décomposition 3.9 et 3.10. $\Pi(F)$ et $N(F)$ s'avèrent être des intégrales de Sugeno (Dubois et Prade, 1980). La possibilité et la nécessité d'événements flous sont utiles pour évaluer à quel point une condition flexible est satisfaite par une donnée mal connue (Cayrol *et al.*, 1982) ; en particulier si $\pi_x = F$, on a seulement $N(F) \geq 1/2$, car pour avoir $N(F) = 1$, il faut $\forall s\ \pi_x(s) > 0 \Rightarrow F(s) = 1$, c'est-à-dire l'inclusion du support de π dans le noyau de F de façon à ce que toute valeur un tant soit peu possible satisfasse pleinement F. Ces mesures sont à la base d'une approche des problèmes de diagnostic de pannes permettant un traitement qualitatif de l'incertain où on peut distinguer entre des manifestations qui sont (plus ou moins) certainement présentes (ou absentes) et des manifestations qui sont (plus ou moins) possiblement présentes (ou absentes) quand une panne se produit (Cayrac *et al.*, 1996; Dubois *et al.*, 2001a). Les fonctions Δ et ∇ s'étendent de la même façon, $\Delta(F) = \inf_s \max(1 - F(s), \pi_x(s))$, et on préserve $\nabla(F) = 1 - \Delta(\overline{F})$, ainsi que 3.13 et 3.14.

Les fonctions d'ensemble N et Δ permettent de représenter des règles floues (voir aussi chapitre **??**) du type « plus x est F plus il est *certain* que y soit G », et du type « plus x est F plus il est *possible* que y soit G » respectivement, où F mais aussi éventuellement G sont des propriétés graduelles représentées par des ensembles flous (Dubois et Prade, 1996). En effet, les deux types de règles correspondent respectivement à des contraintes de la forme $N(G) \geq F(s)$ et de la forme $\Delta(G) \geq F(s)$ dont les solutions sont des distributions de possibilité telles que

$$\pi_{x,y}(s,t) \leq \max(1 - F(s), G(t)) \text{ et } \pi_{x,y}(s,t) \geq \min(F(s), G(t))$$

Pour cela, comme $c \leq \max(a, 1-b) \Leftrightarrow (1-a) \rightarrow (1-c) \geq b$ (où $\rightarrow$ est l'implication de Gödel $u \rightarrow v = 1$ si $u \leq v$, $u \rightarrow v = v$ sinon), il faut prendre des extensions floues particulières de N et Δ (quand G est flou) : $N(G) = \inf_s (1 - F(s)) \rightarrow (1 - \pi_x(s))$ et $\Delta(G) = \inf_s F(s) \rightarrow \pi_x(s)$. Ces nécessité et possibilité garantie d'événements flous coïncident avec les définitions du paragraphe précédent quand G est non flou, mais sont telles que si $N(G) = 1$ alors $\pi_x = G$ (en effet si on dit qu'on est complètement certain que « Paul est jeune », on affirme que « Paul est jeune » ($\pi_{age(Paul)} = jeune$). Ceci étend la vision bipolaire d'une règle R de la forme « si $x \in A$ alors $y \in B$ » qui sur un produit cartésien de domaines $S \times T$ peut être représentée par la contrainte $R(s,t) \geq (A \times B)(s,t)$ du point de vue de ses exemples, et par la contrainte $\overline{R}(s,t) \geq (A \times \overline{B})(s,t) \Leftrightarrow R(u,v) \leq (\overline{A} + B)(s,t)$ où la barre exprime la complémentation et $A + B = \overline{\overline{A} \times \overline{B}}$, du point de vue de ses contre-exemples. On retrouve l'idée de la règle « si A alors B » représentée par l'objet conditionnel $B|A$.

Possibilités qualitatives et possibilités quantitatives. Conditionnement

Puisque les opérateurs à l'œuvre en théorie des possibilités sont le maximum, le minimum, et une opération de renversement d'échelle (la complémentation à 1, $1-(\cdot)$, peut être remplacée sur une échelle finie ordonnée $\{\alpha_0, \cdots, \alpha_m\}$ par la fonction n définie par $n(\alpha_k) = \alpha_{m-k}$), on peut soit utiliser une échelle numérique telle que $[0, 1]$, et on parlera de possibilité quantitative, soit une échelle finie et on parlera de possibilité qualitative (Dubois et Prade, 1998). Dans les deux cas, la théorie des possibilités (qualitative ou quantitative) fournit un modèle très simple, mais non trivial, de l'incertain non-probabiliste permettant la représentation d'états d'ignorance partielle ou totale. La différence principale entre les cadres qualitatif et quantitatif réside dans la définition du conditionnement. En accord avec la contrepartie de la règle de Bayes déjà mentionnée pour les objets conditionnels : $(A \cap B) \mid S = (A \mid B) \wedge (B \mid S)$, le conditionnement en théorie des possibilités non-numériques, à valeurs sur une échelle finie L, est défini par

$$\Pi(A \cap B) = \min(\Pi(A \mid B), \Pi(B)). \tag{3.15}$$

Cette équation n'a pas de solution unique. Néanmoins, dans l'esprit de la théorie des possibilités on est amené à choisir la solution la moins informative, à savoir, pour $B \neq \emptyset$, et $A \neq \emptyset$:

$$\Pi(A \mid B) = 1 \text{ si } \Pi(A \cap B) = \Pi(B), \text{ et } \Pi(A \cap B) \text{ sinon} \tag{3.16}$$

Elle est semblable à la probabilité conditionnelle, mais on ne divise plus $\Pi(A \cap B)$. Si $\Pi(B) = 0$, alors $\Pi(A \mid B) = 1$ dès que $A \neq \emptyset$. Conditionner sur un événement impossible peut détruire l'information. La mesure de nécessité conditionnelle est alors définie par $N(A \mid B) = 1 - \Pi(\overline{A} \mid B)$. Elle coïncide avec la nécessité de l'implication matérielle sauf si $\Pi(A \cap B) = \Pi(B)$. Notons que l'équation duale $N(A \cap B) = \min(N(A \mid B), N(B))$ n'a pas grand intérêt, car sa solution minimale est $N(A \mid B) = N(A \cap B) = \min(N(A), N(B))$, ce qui revient à poser $\Pi(A \mid B) = \Pi(\overline{B} \cup A)$. En revanche la solution issue de l'équation (3.15) capture bien le conditionnement ordinal car on vérifie que $N(A \mid B) > 0 \iff \Pi(A \cap B) > \Pi(\overline{A} \cap B)$ quand $\Pi(B) > 0$. Ceci veut dire qu'une proposition A est acceptée comme vraie dans le contexte B si elle est plus plausible que son contraire dans ce contexte.

Dans le cas des possibilités numériques, le manque de continuité de la fonction définissant $\Pi(A \mid B)$ dans 3.16 conduit De Cooman (1997) à préférer (1997) la définition suivante basée sur le produit, qui on le verra coïncide avec la règle de conditionnement de Dempster pour les fonctions de croyance :

$$\Pi(A \mid B) = \frac{\Pi(A \cap B)}{\Pi(B)} \text{ pourvu que } \Pi(B) \neq 0.$$

Comme dans le cas probabiliste, où le conditionnement est à la base d'une représentation sous forme de réseau bayésien, chacune de ces deux formes, qualitative ou quantitative, de conditionnement donnent naissance à des représentations graphiques possibilistes (voir le chapitre **??**). Tandis que l'indépendance stochastique entre événements de probabilité non nulle s'avère symétrique puisque $Prob(B|A) = Prob(B)$ équivaut à $Prob(A \cap B) = Prob(A) \cdot Prob(B)$, ce n'est plus le cas pour l'indépendance possibiliste entre événements dont il existe plusieurs formes (telles que l'indépendance absolue de B par rapport à A si $N(B|A) = N(B) > 0$, ou le fait d'être non informatif si $N(B|A) = N(B) = N(\overline{B}|A) = N(\overline{B}) = 0$) (Dubois *et al.*, 1999). Il existe aussi plusieurs options pour la définition de l'indépendance possibiliste entre variables, dont certaines sont utilisées pour les réseaux possibilistes ; voir (Ben Amor *et al.*, 2002) pour le cadre qualitatif. L'indépendance possibiliste conditionnelle entre variables dans le cadre quantitatif est défini comme $\forall x, y, z, \Pi(x, y|z) = \Pi(x|z) \cdot \Pi(y|z)$ qui équivaut à $\forall x, y, z, \Pi(x|y, z) = \Pi(x|z)$.

Possibilités quantitatives et probabilités

Dans sa version quantitative, les possibilités peuvent être rapprochées des probabilités, et ce de plusieurs manières. Nous évoquerons brièvement les trois principales : les possibilités (resp. nécessités) comme bornes supérieures (resp. inférieures) de probabilités, le comportement possibiliste des probabilités extrêmes, l'interprétation d'une distribution de possibilité comme une fonction de vraisemblance (« likelihood function ») en statistique non-bayésienne. Commençons par cette dernière.

Une distribution de possibilité numérique peut en effet être interprétée comme une *fonction de vraisemblance* (Smets, 1982), (Dubois *et al.*, 1997). Dans le cadre d'un problème d'estimation, on s'intéresse à la détermination de la valeur d'un paramètre $\theta \in \Theta$ qui définit une distribution de probabilité $P(\cdot \mid \theta)$ sur S. Supposons qu'on ait fait une observation A. La fonction $P(A \mid \theta), \theta \in \Theta$ n'est pas une distribution de probabilité, mais une fonction de vraisemblance $Vr(\theta)$: Une valeur a de θ est considérée d'autant plus plausible que $P(A \mid a)$ est élevé, et l'hypothèse $\theta = a$ sera rejetée si $P(A \mid a) = 0$ (ou inférieure à un seuil de pertinence). Souvent, on renormalise cette fonction pour que son maximum vaille 1. On peut poser $\pi(a) = P(A \mid a)$ (moyennant cette renormalisation) et interpréter cette fonction de vraisemblance comme un degré de possibilité. En particulier on vérifie que $\forall B \subseteq \Theta$, des bornes sur la valeur de $P(A \mid B)$ peuvent être calculées comme :

$$\min_{\theta \in B} P(A \mid \theta) \leq P(A \mid B) \leq \max_{\theta \in B} P(A \mid \theta)$$

ce qui montre que l'axiome de maxitivité fournit un calcul optimiste de $P(A \mid B)$ interprété comme $\Pi(B)$, tandis que la possibilité garantie correspond à une évaluation pessimiste. Il est facile de vérifier que poser $P(A \mid B) = \max_{\theta \in B} P(A \mid \theta)$ est la seule façon de construire une mesure de confiance sur Θ à partir de $P(A \mid \theta), \theta \in \Theta$. En effet, la monotonie dans l'inclusion de Vr impose $P(A \mid B) \geq \max_{\theta \in B} P(A \mid \theta)$ (Coletti et Scozzafava, 2003).

Les *fonctions conditionnelles ordinales* (aussi appelées « *fonctions de rang* »), proposées par Spohn (1988; 2012) offrent un cadre de représentation de l'incertain extrêmement proche de celui de la théorie des possibilités, si ce n'est que chaque interprétation s est associée avec un degré $\kappa(s)$ non plus dans $[0, 1]$, mais dans les entiers $\mathbb{N}$, voire les ordinaux. Plus $\kappa(s)$ est petit, plus s est possible. $\kappa(s) = +\infty$ signifie que s est impossible, tandis que $\kappa(s) = 0$ signifie qu'absolument rien ne s'oppose à ce que s soit l'état réel du monde. A partir d'une « kappa » distribution ou fonction de rang, on définit la fonction d'ensemble :

$$\kappa(A) = \min_{s \in A} \kappa(s) \text{ and } \kappa(\emptyset) = +\infty.$$

Le conditionnement est défini par Spohn de la façon suivante :

$$\kappa(s \mid B) = \begin{cases} \kappa(s) - \kappa(B) & \text{si } s \in B \\ +\infty & \text{sinon.} \end{cases}$$

Spohn propose d'interpréter $\kappa(s)$ comme l'exposant d'une probabilité infinitésimale, ce qui est en accord avec le fait que $\kappa(A \cup B) = \min(\kappa(A), \kappa(B))$; $\kappa(s \mid B)$ est alors l'exposant de la probabilité infinitésimale conditionnelle $\mathrm{P}(s \mid B)$.

Il est facile de passer des fonctions de rang à la théorie des possibilités, par les transformations suivantes (Dubois et Prade, 1991) :

$$\pi_\kappa(s) = 2^{-\kappa(s)}, \Pi_\kappa(A) = 2^{-\kappa(A)}.$$

π_κ et Π_κ prennent donc leurs valeurs sur un sous-ensemble de rationnels de $[0, 1]$. Π_κ est bien une mesure de possibilité : $\Pi_\kappa(A \cup B) = 2^{-\min(\kappa(A),\kappa(B))} = \max(\Pi_\kappa(A), \Pi_\kappa(B))$. De plus pour la conditionnelle, on a $\forall s$, $\pi_{\kappa(s|B)} = 2^{-\kappa(s)-\kappa(B)} = \frac{2^{-\kappa(s)}}{2^{-\kappa(B)}} = \frac{\pi_\kappa(s)}{\Pi_\kappa(B)}$, qui est le conditionnement possibiliste basé sur le produit. La transformation inverse est seulement possible quand $\kappa(s) = -log_2(\pi(s))$ prend des valeurs entières. L'avantage de l'échelle $[0, 1]$ est sa capacité de pouvoir introduire autant de niveaux intermédiaires que nécessaire.

Les possibilités numériques peuvent aussi être considérées comme des bornes supérieures de probabilités, comme Zadeh en avait eu l'intuition dès le départ (Zadeh, 1978). En effet, considérons une suite croissante d'ensembles emboîtés $E_1 \subset E_2 \subset, \ldots, \subset E_k$. Soit $\nu_1 \leq \nu_2 \leq, \ldots, \leq \nu_k$, des bornes inférieures de probabilité, et soit $\mathcal{P} = \{P, P(E_i) \geq \nu_i, \forall i = 1, \ldots, k\}$. Ce type d'information est typiquement fourni par un expert s'exprimant de façon imprécise sur la valeur d'un paramètre : Il suggère que $x \in E_i$ avec un degré de confiance au moins égal à ν_i. Alors $P_*(A) = \inf_{P \in \mathcal{P}} P(A)$ est une mesure de nécessité ; et $P^*(A) = \sup_{P \in \mathcal{P}} P(A)$ est une mesure de possibilité, engendrée par la distribution de possibilité (Dubois et Prade, 1992) :

$$\forall s \in S, \pi(s) = \min_{i=1,\ldots,k} \max(E_i(s), 1 - \nu_i). \tag{3.17}$$

avec $E_i(s) = 1$ si $s \in E_i$ et 0 sinon. Voir (De Cooman et Aeyels, 1999) pour l'extension de ce résultat au cas infini. Dans ce cadre, chaque E_i est une sorte d'ensemble de confiance (un intervalle dans le cas où S est les nombres réels) et la probabilité d'appartenance à cet ensemble est au moins ν_i. La probabilité de non-appartenance à E_i est donc au plus $1 - \nu_i$. Cet ensemble de confiance pondéré par un degré de certitude correspond à la distribution de possibilité $\max(E_i(s), 1 - \nu_i)$. L'équation ci-dessus effectue la conjonction de ces distributions locales. Il est clair que la distribution π code de façon très économique la famille de probabilités $\mathcal{P}$. Inversement, étant donné une distribution de possibilité π, l'ensemble des mesures de probabilité qu'elle représente est $\mathcal{P}(\pi) = \{P, P(A) \leq \Pi(A), \forall A \text{ mesurable}\} = \{P, P(A) \geq N(A), \forall A \text{ mesurable}\}$. Dans le cas où les E_i ne sont pas emboîtés, on n'obtient plus qu'une approximation de la famille de probabilités ; de meilleures approximations peuvent être obtenues au moyen de *paires* de distributions de possibilité qui enserrent la famille des fonctions de probabilités cumulées (Destercke *et al.*, 2008).

Signalons enfin l'existence de transformations possibilité / probabilité (Dubois *et al.*, 2004) qui permettent de passer d'une représentation d'un type à une autre de l'autre type compatible avec la précédente, mais non équivalente. Dans cette perspective, il est intéressant de noter que l'idée de spécificité possibiliste a une contrepartie, appelée en anglais « peakedness », en probabilités (Dubois et Hüllermeier, 2007). Elle permet de comparer les distributions de probabilité en termes de dispersion.

3.4.2 Logique possibiliste, incohérence, et non monotonie

La logique possibiliste (Dubois *et al.*, 1994; Dubois et Prade, 2004) manipule des paires constituées d'une formule (bien formée) de la logique classique (propositionnelle, ou du premier ordre), et d'une pondération qui peut être qualitative ou numérique (appartenant en général à une échelle complètement ordonnée, mais éventuellement seulement à un treillis muni d'un plus petit et d'un plus grand élément).

Syntaxe et sémantique

Dans sa version de base, on ne considère que des conjonctions de paires de la forme (p, α) de formules de la logique propositionnelle p associées à une pondération α appartenant à l'intervalle $(0, 1]$, interprétée comme une borne inférieure d'une mesure de nécessité, c'est-à-dire que (p, α) code une contrainte de la forme $N(p) \geq \alpha$. Cela correspond soit à un élément de connaissance (on est certain au niveau α que p est vrai), soit à une préférence (p représente alors un but avec une priorité α). La propriété de décomposabilité des mesures de nécessité 3.10 permet de ne pas faire de différence entre $(p \wedge q, \alpha)$ et $(p, \alpha) \wedge (q, \alpha)$, et ainsi de ramener les bases possibilistes à des conjonctions de clauses pondérées.

Soit $\mathcal{B}^N = \{(p_j, \alpha_j) \mid j = 1, \ldots, m\}$ une base de logique possibiliste. Elle est associée à la distribution de possibilité

$$\pi_{\mathcal{B}}^N(s) = \min_{j=1,\ldots,m} \pi_{(p_j,\alpha_j)}(s)$$

sur les interprétations, où $\pi_{(p_j,\alpha_j)}(s) = 1$ si $s \in M(p_j)$, et $\pi_{(p_j,\alpha_j)}(s) = 1 - \alpha_j$ si $s \notin M(p_j)$, et $M(p)$ est l'ensemble des interprétations induites par le langage propositionnel sous-jacent pour lesquelles p est vraie. Ainsi une interprétation s est d'autant plus possible qu'elle ne viole aucune formule p_j ayant un niveau de priorité α_j élevé.

La logique possibiliste de base est associée à la règle d'inférence

$$(\neg p \vee q, \alpha); (p \vee r, \beta) \vdash (q \vee r, \min(\alpha, \beta)).$$

Cette règle est saine et complète pour la réfutation, par rapport à la sémantique possibiliste. Il est à noter que la règle probabiliste analogue

$$Prob(\neg p \vee q) \geq \alpha; Prob(p \vee r) \geq \beta \vdash Prob(q \vee r) \geq \max(0, \alpha + \beta - 1)$$

est saine, mais pas complète par rapport à la sémantique probabiliste. Cela est à mettre en relation avec le fait qu'un ensemble de formules possibilistes $\{(p_j, \beta_j) \text{ with } \beta_j \geq \alpha\}_{j=1,n}$ possède une clôture déductive qui ne contient que des formules de niveau au moins α, alors que c'est faux pour l'ensemble $\{p_j | Prob(p_j) \geq \alpha\}_{j=1,n}$ (sauf si $\alpha = 1$).

La distribution de possibilité $\pi_{\mathcal{B}}^N(s) = \min_{j=1,\ldots,m} \max(M(p_j)(s), 1 - \alpha_j)$ (où $M(p_j)(s) = 1$ si $s \in M(p_j)$ et $M(p_j)(s) = 0$ sinon) s'exprime sous la forme d'une combinaison min-max, il s'agit donc d'une description « par en dessus ». On peut donc toujours voir B^N comme une conjonction de clauses pondérées, c.-à-d., une extension de la forme normale conjonctive. Une représentation duale de la logique possibiliste est basée sur les mesures de possibilité garantie. Une formule est alors une paire $[q, \beta]$, interprétée comme la contrainte $\Delta(q) \geq \beta$, où Δ est une (anti-)mesure de possibilité garantie. Elle exprime donc que *tous* les modèles de q sont au moins possibles, au moins satisfaisants au niveau β. Une base $\mathcal{B}^\Delta = \{[q_i, \beta_i] \mid i = 1, \ldots, n\}$ est alors associée à la distribution

$$\pi_{\mathcal{B}}^\Delta(s) = \max\nolimits_{i=1,\ldots,n} \pi_{[q_i,\beta_i]}(s)$$

avec $\pi_{[q_i,\beta_i]}(s) = \beta_i$ si $s \in M(q_i)$ et $\pi_{[q_i,\beta_i]}(s) = 0$ sinon. Il s'agit d'une description « par en dessous » de $\pi_{\mathcal{B}}^\Delta$. On peut toujours transformer une base possibiliste duale $\mathcal{B}^\Delta$ en une base où les formules q_i sont des conjonctions de littéraux (des cubes) sans altérer $\pi_{\mathcal{B}}^\Delta$. On peut donc aussi voir une base en logique possibiliste comme une conjonction de « cubes pondérés », c.-à-d., une extension de la forme normale disjonctive (mais attention la conjonction $[p, \alpha]$ et $[q, \alpha]$ est équivalente à $[p \vee q, \alpha]$ à cause de la propriété 3.13, tandis qu'en logique possibiliste standard la conjonction (p, α) et (q, α) est équivalente à $(p \wedge q, \alpha)$). Une base $\mathcal{B}^{\Delta}$ en logique possibiliste exprimée en termes de mesure de possibilité garantie peut toujours être réécrite de manière équivalente en termes de logique possibiliste standard $\mathcal{B}^{N}$ basée sur les mesures de nécessité (Benferhat et Kaci, 2003; Benferhat *et al.*, 2008) et vice-versa de façon à ce que $\pi^{N}_{\mathcal{B}} = \pi^{\Delta}_{\mathcal{B}}$. A noter cependant que la logique possibiliste en termes de mesure Δ obéit à une toute autre règle d'inférence : $[\neg p \wedge q, \alpha]; [p \wedge r, \beta] \vdash [q \wedge r, \min(\alpha, \beta)]$ qui propage la possibilité garantie en accord avec la monotonie décroissante de Δ (en effet, si $r = \top$, et $q \vdash p$, alors $\alpha = 1$ car $\Delta(\bot) = 1$, et cette règle conclut $[q, \beta]$ à partir de $[p, \beta]$).

Une information possibiliste (avec un nombre fini de niveaux de possibilité) peut donc être représentée par une distribution de possibilité, mais aussi de manière plus compacte sous la forme d'un ensemble fini de formules associées soit à un niveau de priorité, soit à un niveau de satisfaction garantie. De plus il existe des représentations graphiques équivalentes en termes de réseaux possibilistes (basés soit sur le conditionnement qualitatif, soit sur le quantitatif), avec des passages possibles d'un type de représentation à l'autre (Benferhat *et al.*, 2002).

Il existe différentes types d'extension de la logique possibiliste qui permettent d'associer à des formules logiques notamment des bornes inférieures de mesures de possibilité (ce qui permet d'exprimer des formes d'ignorance), ou des périodes temporelles où on est plus ou moins certain qu'une formule est vraie, ou des ensembles de sources plus ou moins certaines d'un énoncé ; voir (Dubois et Prade, 2004) pour des références. Une autre forme d'extension, appelée « *logique possibiliste généralisée* » permet d'appliquer des négations et des disjonctions à des formules possibilistes de base. Il en résulte alors une logique à deux niveaux puisque les connecteurs peuvent intervenir à l'intérieur ou à l'extérieur des formules possibilistes de base, dont la sémantique est en termes d'*ensembles* de distributions de possibilité. On a pu montrer que la logique possibiliste généralisée (Dubois *et al.*, 2012) permettait de clarifier la sémantique de la programmation logique par ensembles-réponses (cf. chapitre **??**) au travers de celle de la « logique d'équilibre » (Pearce, 2006). En fait, c'est une généralisation conjointe de la logique modale des possibilités MEL et de la logique possibiliste standard, en complet accord avec la théorie des possibilités.

Les possibilités qualitatives peuvent être généralisées à des possibilités imprécises, de la même manière qu'on s'intéresse aux probabilités imprécises 3.6. Signalons à ce sujet le récent résultat montrant que toute capacité, c.-à-d. toute fonction d'ensemble monotone croissante, peut être caractérisée sur un référentiel fini par un ensemble de mesures de possibilité ; ceci permet à la fois de trouver une contrepartie aux capacités en termes de logiques modales dites non régulières, qui permettent un traitement de la paraconsistance, et de définir une contrepartie qualitative des fonctions de croyance et de leur règles de combinaison (Dubois *et al.*, 2013)).

Gestion de l'incohérence et raisonnement non monotone

Un aspect important de la logique possibiliste est sa capacité à faire face à l'incohérence. Le niveau d'incohérence $inc(\mathcal{B})$ d'une base possibiliste $\mathcal{B}$ est défini comme $inc(\mathcal{B}) = \max\{\alpha | \mathcal{B} \vdash (\bot, \alpha)\}$. Toutes les formules dont le niveau est strictement plus grand que $inc(\mathcal{B})$ ne peuvent contribuer à l'incohérence. On peut montrer que $1 - inc(\mathcal{B})$ est la hauteur $h(\pi_{\mathcal{B}})$ de $\pi_{\mathcal{B}}$, définie par $h(\pi_{\mathcal{B}}) = \max_s \pi_{\mathcal{B}}(s)$ ($\pi_{\mathcal{B}}$ étant la distribution de possibilité associée à $\mathcal{B}$). On montre que $inc(\mathcal{B}) = 0$ si et seulement si l'ensemble des formules de $\mathcal{B}$, en oubliant leur pondération, sont cohérentes au sens classique. Toutes les formules de $\mathcal{B}$ dont le niveau est plus petit ou égal à $inc(\mathcal{B})$ sont ignorées dans le mécanisme d'inférence possibiliste standard ; on dit qu'elles sont « noyées ». Cependant, il existe d'autres formes d'inférence qui prolongent l'inférence possibiliste tout en prenant en compte les formules qui sont en dessous, ou au niveau d'incohérence, mais qui ne participent aucunement à des sous-ensembles incohérents de formules, voir (Benferhat *et al.*, 1999a) pour une vue d'ensemble de ces inférences.

L'application de règles par défaut ayant des exceptions potentielles (par exemple, « les oiseaux volent »), à des situations particulières (par exemple, « Cornélius est un oiseau ») sur lesquelles on est incomplètement informé, peut conduire à des conclusions (par exemple, « Cornélius vole ») qui deviennent incohérentes avec de nouvelles conclusions qui pourront être obtenues quand plus d'information deviendra disponible sur ces situations (par exemple, « Cornélius est une autruche »). Le caractère non monotone de la possibilité conditionnelle qualitative, qui autorise d'avoir à la fois, $N(B \mid A) > 0$ et $N(\overline{B} \mid A \cap A') > 0$, c'est-à-dire que l'arrivée de l'information A' conduit à rejeter la proposition B préalablement acceptée dans le contexte où on ne sait que A, permet de traiter ce problème.

En effet, une règle par défaut « si A alors généralement B », peut se représenter par la contrainte possibiliste $\Pi(B \cap A) > \Pi(\overline{B} \cap A)$ qui exprime qu'il est davantage possible d'avoir B vrai que B faux dans le contexte où A est vrai. Une base de règles par défaut est alors représentée par un ensemble de telles contraintes qui déterminent un ensemble de mesures de possibilité qui les satisfait. A partir d'une telle base de règles, deux types d'inférence sont concevables afin de déduire de nouvelles règles applicables à une situation où on sait *exactement* K (c'est-à-dire les règles de la forme « si K alors généralement C », permettant de conclure (provisoirement) C pour cette situation K). Un premier type d'inférence, prudent, requiert que la contrainte associée à $C|K$ soit satisfaite par *toutes* les mesures de possibilité qui satisfont les contraintes (supposées cohérentes) associées à la base de règles par défaut. Une seconde, plus hardie, restreint cette exigence à la plus grande (la moins spécique) des distributions de possibilité, solution de ces dernières contraintes (on peut montrer que cette distribution est unique). On peut établir que la première inférence correspond exactement à l'inférence préférentielle obéissant au système P (Kraus *et al.*, 1990) de postulats pour le raisonnement non monotone (voir le chapitre **??**), tandis que la seconde n'est autre que l'inférence dite par « fermeture rationnelle » (Lehmann et Magidor, 1992). Ces deux types d'inférence peuvent être également justifiées dans d'autres sémantiques telles que celles des objets conditionnels, des probabilités infinitésimales, des systèmes Z et Z^+ (Pearl, 1990; Goldszmidt et Pearl, 1991), des logiques modales conditionnelles ; voir (Benferhat *et al.*, 1997) pour une vue d'ensemble et des références. Des sémantiques en termes de probabilités à grandes marches (Benferhat *et al.*, 1999b), et de probabilités conditionnelles au sens de De Finetti (1974) (Coletti et Scozzafava, 2002) (dans ce dernier cas une règle « si A alors généralement B » correspond tout simplement à une contrainte $Prob(B|A) = 1$ où $Prob(B|A)$ continue de faire sens quand $Prob(A) = 0$ (0 ne veut plus dire ici impossible, mais plutôt quelque chose comme « négligeable dans un premier temps »), grâce à un traitement hiérarchisé des contraintes induit par un partitionnement de l'ensemble des interprétations (Biazzo *et al.*, 2002). Le cadre de la logique possibiliste permet donc un traitement pratique du raisonnement à partir de règles par défaut présentant des exceptions potentielles (Benferhat *et al.*, 1998), tout comme le raisonnement à partir d'informations incertaines qualitatives ; il est même possible de combiner les deux (Dupin de Saint-Cyr et Prade, 2008).

La théorie de la révision des croyances (Gärdenfors, 2008) (voir chapitre **??**), qui est étroitement liée au raisonnement non monotone, repose sur la notion d'enracinement épistémique, qui permet de baser le processus de révision sur un ordre de remise en cause des éléments d'information. Il est intéressant de noter qu'une relation d'enracinement épistémique n'est autre qu'une relation de nécessité qualitative (Dubois et Prade, 1991) (les relations de possibilité et de nécessité qualitatives ayant pour seule contrepartie sur une échelle ordonnée les mesures de possibilité et de nécessité (Dubois, 1986)). On peut d'ailleurs donner sens à l'intuition que les propositions de la base de croyances qui sont indépendantes de l'information entrante doivent demeurer après révision, dans le cadre possibiliste (Dubois *et al.*, 1999). Signalons aussi que l'idée de filtrage au sens de Kalman, qui conjugue mise à jour et révision, peut avoir une expression qualitative dans le cadre de la logique possibiliste (Benferhat *et al.*, 2000).

Mentionnons par ailleurs, qu'un modèle de la perception causale, où un agent, en présence d'une suite d'événements qui ont lieu, pose des liens causaux entre certains de ces événements sur la base de ses croyances sur le cours normal des choses, a été développé par (Bonnefon *et al.*, 2008). Dans ce problème, *la causalité* tient un rôle différent de celui qu'elle a dans les logiques de l'action (chapitre **??**) ou le diagnostic (chapitre **??**), où les relations de causalité sont supposées connues, alors que ces relations sont posées ici par l'agent sur la base de ses croyances. Les croyances sont représentées par des règles par défaut (obéissant aux postulats du système P), et on privilégie les événements « anormaux » comme causes que peut éventuellement retenir l'agent, dans le modèle proposé, qui par ailleurs a fait l'objet d'une comparaison détaillée avec le modèle probabiliste (Bonnefon *et al.*, 2012).

3.4.3 Théorie des possibilités et analyse formelle de concepts

L'analyse formelle de concepts (AFC) est un formalisme de représentation des connaissances qui est à la base de la méthodologie pour la fouille de données (cf. chapitre chapitre **??**) en fournissant un cadre théorique pour l'apprentissage de hiérarchies de concepts. En tant que cadre de représentation, il a été montré récemment qu'il pouvait être rapproché de la théorie des possibilités, et aussi dans une certaine mesure de celle des ensembles approximatifs. C'est ce qui explique la petite place – sans doute inattendue – qui lui est faite dans ce chapitre.

En analyse formelle de concepts (Barbut et Montjardet, 1970; Ganter et Wille, 1999), on se donne au départ une relation $\mathcal{R}$ binaire, appelée *contexte formel*, entre un ensemble d'objets $\mathcal{O}$ et un ensemble de propriétés $\mathcal{P}$. Etant donné un objet x et une propriété y, soit $R(x) = \{y \in \mathcal{P} \mid x\mathcal{R}y\}$ l'ensemble des propriétés possédées par l'objet x ($x\mathcal{R}y$ signifie que x possède la propriété y) et soit $R(y) = \{x \in \mathcal{O} \mid x\mathcal{R}y\}$ l'ensemble des objets possédant la propriété y. On définit en AFC des correspondances entre les ensembles $2^{\mathcal{O}}$ et $2^{\mathcal{P}}$. Ces correspondances sont appelées opérateurs de dérivation de Galois. L'opérateur de Galois à la base de l'AFC, noté ici $(.)^{\Delta}$ (pour des raisons qui seront claires dans la suite), permet d'exprimer l'ensemble des propriétés satisfaites par *tous* les objets de $X \subseteq \mathcal{O}$ comme :

$$\begin{aligned} X^{\Delta} = \quad & \{y \in \mathcal{P} \mid \forall x \in \mathcal{O}\ (x \in X \Rightarrow x\mathcal{R}y)\} \\ = \quad & \{y \in \mathcal{P} \mid X \subseteq R(y)\} = \textstyle\bigcap_{x \in X} R(x) \end{aligned}$$

Cela permet aussi d'exprimer de manière duale l'ensemble des objets satisfaisant toutes les propriétés de Y comme :

$$\begin{aligned} Y^{\Delta} = & \quad \{x \in \mathcal{O} \mid \forall y \in \mathcal{P}\ (y \in Y \Rightarrow x\mathcal{R}y)\} \\ = & \quad \{x \in \mathcal{O} \mid Y \subseteq R(x)\} = \bigcap_{y \in Y} R(y) \end{aligned}$$

La paire duale d'opérateurs $((.)^{\Delta}, (.)^{\Delta})$ appliqués respectivement sur $2^{\mathcal{O}}$ et $2^{\mathcal{P}}$ constitue une connexion de Galois qui permet d'induire des concepts formels. Un *concept formel* est une paire (X, Y) telle que

$$X^{\Delta} = Y \text{ et } Y^{\Delta} = X.$$

Autrement dit, X est l'ensemble maximal d'objets satisfaisant toutes les propriétés déjà satisfaites par tous les objets de X. L'ensemble X (resp. Y) est appelé *extension* (resp. *intension*) du concept. De manière équivalente, (X, Y) est un concept formel si et seulement si c'est une paire maximale au sens de l'inclusion telle que

$$X \times Y \subseteq \mathcal{R}.$$

L'ensemble de tous les concepts formels est naturellement équipé d'une relation d'ordre (notée $\preccurlyeq$) et définie comme : $(X_1, Y_1) \preceq (X_2, Y_2)$ ssi $X_1 \subseteq X_2$ (ou $Y_2 \subseteq Y_1$). Cet ensemble muni de la relation d'ordre $\preccurlyeq$ forme un treillis complet. C'est l'exploitation de ce treillis qui permet de trouver des règles d'association entre ensembles de propriétés (Pasquier *et al.*, 1999).

Sur la base d'un parallèle avec la *théorie des possibilités* (en effet $X^{\Delta} = \bigcap_{x \in X} R(x)$ est le pendant de la définition d'une mesure de possibilité garantie $\Delta(F) = \min_{x \in F} \pi(x)$ où π est une distribution de possibilités), d'autres opérateurs ont été introduits (Dubois et Prade, 2012) : A savoir l'opérateur de possibilité (noté $(.)^{\Pi}$) et son dual l'opérateur de nécessité (noté $(.)^{N}$), ainsi que l'opérateur $(.)^{\nabla}$ dual de l'opérateur $(.)^{\triangle}$ à la base de l'AFC, définis comme suit :

- X^{Π} est l'ensemble des propriétés satisfaites par au moins un objet dans X :
  $X^{\Pi} = \{y \in \mathcal{P} \mid \exists x \in X,\ x\mathcal{R}y\} = \{y \in \mathcal{P} \mid X \cap R(y) \neq \emptyset\} = \bigcup_{x \in X} R(x)$;
- X^{N} est l'ensemble des propriétés que seuls les objets de X ont :
  $X^{N} = \{y \in \mathcal{P} \mid \forall x \in \mathcal{O}\ (x\mathcal{R}y \Rightarrow x \in X)\} = \{y \in \mathcal{P} \mid R(y) \subseteq X\} = \bigcap_{x \notin X} \overline{R}(x)$,
  (où $\overline{R}(x)$ est l'ensemble des propriétés que n'a pas x) ;
- X^{∇} est l'ensemble des propriétés qui ne sont pas satisfaites par au moins un objet en dehors de X :
  $X^{\nabla} = \{y \in \mathcal{P} \mid \exists x \in \overline{X},\ x\overline{\mathcal{R}}y\} = \{y \in \mathcal{P} | R(y) \cup X \neq \mathcal{O}\} = \bigcup_{x \notin X} \overline{R}(x)$.

Les opérateurs Y^{Π}, Y^{N}, Y^{∇} sont obtenus de manière duale. On montre que les paires (X, Y) telles que $X^{N} = Y$ et $Y^{N} = X$ (ou de manière équivalente $X^{\Pi} = Y$ et $Y^{\Pi} = X$) caractérisent des sous-contextes indépendants (i.e. qui n'ont en commun ni objets ni propriétés) à l'intérieur du contexte initial (Dubois et Prade, 2012). Les paires (X, Y) telles que $X^{N} = Y$ et $Y^{N} = X$ sont telles que :

$$(X \times Y) \cup (\overline{X} \times \overline{Y}) \supseteq \mathcal{R}.$$

Quant à $X^{\nabla} = Y$ et $Y^{\nabla} = X$, cela constitue une autre caractérisation des concepts formels. On peut montrer aussi que les quatre ensembles X^{Π}, X^{N}, X^{Δ}, X^{∇} constituent des informations complémentaires, toutes nécessaires à une analyse complète de la situation de X dans le contexte formel $\mathcal{K} = (\mathcal{O}, \mathcal{P}, \mathcal{R})$. En pratique, on suppose qu'à la fois $R(x) \neq \emptyset$ et $R(x) \neq \mathcal{P}$, c'est-à-dire que tout objet doit avoir au moins une propriété dans $\mathcal{P}$, mais aucun objet n'a toutes les propriétés de $\mathcal{P}$. Sous cette hypothèse de bi-normalisation, la relation d'inclusion suivante est satisfaite : $R^N(Y) \cup R^\Delta(Y) \subseteq R^\Pi(Y) \cap R^\nabla(Y)$, contrepartie d'une relation qui tient en théorie des possibilités pourvu que la distribution prenne les valeurs 1 et 0 (cf. équation 3.12).

Mentionnons par ailleurs qu'il existe une extension de l'AFC aux propriétés graduelles (Belohlavek, 2002), ainsi que son extension à des contextes formels comportant des informations incomplètes ou incertaines (Burmeister et Holzer, 2005; Djouadi *et al.*, 2011). Une autre extension concerne la possibilité d'associer les objets non plus à de simples propriétés, mais à des descriptions structurées, éventuellement imprécises, ou logiques, grâce aux « structures de patrons » (Ganter et Kuznetsov, 2001; Ferré et Ridoux, 2004), ce qui est toujours en accord avec le parallèle possibiliste (Assaghir *et al.*, 2010).

On peut observer que les idées duales d'intersection non vide et d'inclusion, sont à la base des théories des ensembles approximatifs, des possibilités, et des concepts formels. On va à nouveau les trouver à l'œuvre dans la théorie des fonctions de croyances.

3.5 Théorie des fonctions de croyances

Le modèle des fonctions de croyance (Shafer, 1976; Yager et Liu, 2008; Guan et Bell, 1991) probabilise l'approche ensembliste de l'imprécis. On passe d'une représentation de la forme $x \in E$ où E est un ensemble de valeurs possibles de x, à une distribution de probabilité discrète sur les divers énoncés possibles de la forme $x \in E$ (en supposant le référentiel S fini). On note m une distribution de probabilité sur l'ensemble 2^S des parties de S. On appelle m *fonction de masse*, $m(E)$ la *masse de croyance* affectée à l'ensemble E, et *ensemble focal* tout sous-ensemble E de S tel que $m(E) > 0$. On note $\mathcal{F}$ la famille des ensembles focaux. En général, on n'affecte pas de masse positive à l'ensemble vide (on suppose $m(\emptyset) = 0$) mais le modèle des croyances transférables (Smets et Kennes, 1994) relâche cette contrainte. La masse $m(\emptyset)$ représente alors le degré de contradiction interne de la fonction de masse. La condition $m(\emptyset) = 0$ est une forme de normalisation. Comme m est une distribution de probabilité, la condition $\sum_{E \subseteq S} m(E) = 1$ est vérifiée.

Dans cette représentation hybride de l'incertain il est important de comprendre le sens de la fonction de masse, et de ne pas confondre $m(E)$ avec la probabilité d'occurrence de l'événement E. Shafer dit que $m(E)$ est la masse de croyance affectée à E seul et à aucun de ses sous-ensembles (Shafer, 1976). Une explication plus claire est de dire que $m(E)$ est la probabilité pour que l'agent ne sache rien de plus que $x \in E$. Il y a donc une modalité épistémique implicite dans $m(E)$, mais absente dans $P(E)$. Cela explique que la fonction m puisse être non-monotone par rapport à l'inclusion : on peut avoir $m(E) > m(E') > 0$ quand $E \subset E'$, si l'agent est suffisamment sûr que ce qu'il sait est de la forme de la forme $x \in E$. En particulier, $m(S)$ est la probabilité que l'agent ne sache rien. Une fonction de masse m modélise un état de connaissance.

3.5.1 Fonction de masse induite par une fonction multivoque

On peut interpréter une fonction de masse en considérant que l'information fournie par une source (un élément d'évidence) est assimilable à un message codé dont la signification est aléatoire (Shafer, 1981). Plus précisément, supposons la source envoie un message codé en choisissant au hasard un code parmi un ensemble $C = \{c_1, \ldots, c_n\}$ avec des probabilités $p_1, \ldots, p_n$. En décodant le message avec le code c_i, on obtient une information de la forme $x \in \Gamma(c_i) = A_i$. La probabilité que le message signifie $x \in A$ est donc

$$m(E) = \sum_{\{1 \leq i \leq n : A_i = E\}} p_i. \tag{3.18}$$

En particulier, la probabilité que le message soit vide, c'est-à-dire ne contienne aucune information sur x, est $m(S)$. Le triplet (C, P, Γ), où P est une mesure de probabilité sur C, définit un ensemble aléatoire (Nguyen, 2006). L'équivalence formelle entre les notions d'ensemble aléatoire et de fonction de croyance a été démontrée pour la première fois par (Nguyen, 1978).

> **Exemple** : Considérons une montre qui peut être en panne. On connaît la probabilité ϵ qu'elle tombe en panne. L'ensemble C décrit l'état de la montre $C = \{panne, marche\}$, mais l'agent s'intéresse à l'heure qu'il est. Donc S est l'ensemble des heures possibles. Supposons que la montre indique l'heure h. Dans ce cas, la correspondance Γ est telle que $\Gamma(marche) = \{h\}$ (si la montre marche elle donne la bonne heure), et $\Gamma(panne) = S$ (si elle est en panne on ne sait pas l'heure qu'il est). La fonction de masse induite sur S est donc $m(h) = 1 - \epsilon$ et $m(S) = \epsilon$ qui est bien la probabilité de ne pas savoir l'heure qu'il est.

La fonction de masse obtenue dans l'exemple ci-dessus est dite à support simple parce que la masse est répartie entre un seul sous-ensemble E de S et S lui-même. C'est le cas d'une source non-fiable affirmant $x \in E$, que l'agent croit non-pertinente avec la probabilité ϵ. Cette valeur est attribuée à S alors que $m(E) = 1 - \epsilon$.

Cette façon d'engendrer une fonction de masse à partir d'une fonction multivoque a été initialement proposée par Dempster (1967) dans le contexte de l'inférence statistique, puis généralisée par Shafer (1976; 1981). Parmi les théories de l'incertain, la théorie des fonctions de croyance présente la particularité de mettre l'accent sur les éléments d'évidence qui engendrent un état de connaissance, comme le montre le titre de l'ouvrage fondateur de Shafer (1976) : *A Mathematical Theory of Evidence*.

3.5.2 Fonctions de base

Une fonction de masse m induit deux fonctions d'ensemble, respectivement une fonction de croyance Bel (pour '*belief*') et une fonction de plausibilité Pl, définies par

$$Bel(A) = \sum_{E \subseteq A, E \neq \emptyset} m(E); \quad Pl(A) = \sum_{E \cap A \neq \emptyset} m(E). \tag{3.19}$$

Quand $m(\emptyset) = 0$, il est clair que $Bel(S) = Pl(S) = 1, Pl(\emptyset) = Bel(\emptyset) = 0$, et $Bel(A) = 1 - Pl(\overline{A})$ de sorte que ces fonctions sont duales l'une de l'autre, comme le sont la nécessité et la possibilité. Le degré de croyance $Bel(A)$ peut être interprété comme le degré de *prouvabilité* de A à partir des connaissances disponibles représentées par m. Dans le langage de la logique modale on devrait écrire $Bel(A) = P(\Box A)$ où $\Box$ représente la modalité *l'agent croit que*. De la même façon, $Pl(A)$ mesure le degré de cohérence logique de A avec m.

Les fonctions de croyance Bel sont k-monotones pour tout entier k positif :

$$Bel(\cup_{i=1,\ldots,k}A_i) \geq \sum_{i=1}^{k}(-1)^{i+1}\sum_{I:|I|=i}Bel(\cap_{j\in I}A_j). \tag{3.20}$$

Les fonctions de plausibilité vérifient une propriété similaire, en renversant le sens de l'inégalité ci-dessus.

Une fonction de commonalité $Q(A) = \sum_{E\supseteq A} m(E)$ a été aussi introduite (Shafer, 1976) pour des raisons essentiellement calculatoires, même si il est apparu ensuite que c'était ici la contrepartie de la fonction de possibilité garantie de la théorie des possibilités (Dubois *et al.*, 2001b) (cf. Section 3.4.1). Inversement, connaissant la fonction Bel (ou Pl, ou Q), on peut recalculer de façon unique la fonction de masse m par la transformation de Möbius : $m(E) = \sum_{A\subseteq E}(-1)^{|E\setminus A|}Bel(A)$. Voir (Kennes, 1992) pour des algorithmes efficaces pour faire ce calcul.

Les fonctions de croyance sont le plus souvent définies sur des référentiels finis. Cependant, l'analogie entre fonction de croyance et ensemble aléatoire permet de définir simplement des fonctions de croyance sur les réels (Dempster, 1968; Strat, 1984; Smets, 2005a; Denœux, 2009) ou même sur des espaces topologiques plus généraux (Nguyen, 1978, 2006). Par ailleurs, on peut étendre les fonctions de croyance et de plausibilité à des événements flous (Smets, 1981) sous la forme de ce qui s'avère être des intégrales de Choquet

$$Bel(F) = \sum_{E} m(E)\cdot \min_{s\in E} F(s) \tag{3.21}$$

et

$$Pl(F) = \sum_{E} m(E)\cdot \max_{s\in E} F(s), \tag{3.22}$$

pour le cas fini. On peut également « fuzzifier » la théorie des fonctions de croyance en autorisant soit les éléments focaux à être des ensembles flous (Zadeh, 1979; Yen, 1990), soit les masses de croyance à être des intervalles ou des nombres flous (Denœux, 1999, 2000a).

Deux cas particuliers

Deux cas particuliers remarquables de fonctions de croyance sont à signaler :

1. Les fonctions de probabilité sont obtenues en supposant que les ensembles focaux sont des singletons. Il est clair que si $m(A) > 0$ implique $\exists s \in S, A = \{s\}$, alors $Bel(A) = Pl(A) = P(A)$ pour la fonction de probabilité telle que $P(\{s\}) = m(\{s\}), \forall s \in S$. Inversement, Bel est une fonction de probabilité si et seulement si $Bel(A) = Pl(A), \forall A \subseteq S$.
2. Les fonctions de plausibilité sont des mesures de possibilité (ou par dualité, les fonctions de croyance sont des mesures de nécessité) si et seulement si les ensembles focaux sont emboîtés, c'est-à-dire, $\forall A \neq B \in \mathcal{F}, A \subset B$ ou $B \subset A$. Dans ce cas, $Pl(A \cup B) = \max(Pl(A), Pl(B))$ et $Bel(A \cap B) = \min(Bel(A), Bel(B))$.

On peut associer à m une distribution de possibilité π, en posant $\pi(s)$ = Pl({s}) (plausibilité des singletons), soit :

$$\forall s \in S, \pi(s) = \sum_{s\in E} m(E). \tag{3.23}$$

Il est facile de voir que π est à valeurs sur [0, 1], normalisée ($\pi(s) = 1$ pour un état $s \in S$) dès que les ensembles focaux s'intersectent tous (a fortiori s'ils sont emboîtés). La reconstruction de m à partir de π n'est possible que lorsque les ensembles focaux sont emboîtés ou disjoints. Supposons, de fait, que les ensembles focaux sont emboîtés et figurent dans la suite croissante $E_1 \subset E_2 \subset, \dots, \subset E_n$, où $E_i = \{s_1, \dots s_i\}$, alors

$$\pi(s_i) = \sum_{j=i}^{n} m(E_j).$$

Les mesures de possibilité Π et de nécessité N, définies à partir de π coïncident respectivement avec les fonctions de plausibilité et de croyance induites par m. La fonction de masse se recalcule à partir de π comme suit (en posant $\pi(s_{n+1}) = 0$) :

$$m_\pi(E_i) = \pi(s_i) - \pi(s_{i-1}), i = 1, \dots, n. \tag{3.24}$$

3.5.3 Règles de combinaison

La combinaison d'informations issues de différentes sources (éléments d'évidence) joue un rôle fondamental dans la théorie des fonctions de croyance (voir aussi chapitre **??**). Le mécanisme de base est la règle de Dempster (Shafer, 1976), qui permet la combinaison d'informations d'informations indépendantes. Ce mécanisme, ainsi que la définition très précise de la notion d'indépendance dans ce contexte, peuvent être introduits en reprenant la métaphore des codes aléatoires introduite précédemment.

Règle de Dempster

Soient m_1 et m_2 deux fonctions de masse sur S auxquelles correspondent des ensembles aléatoires (C_1, P_1, Γ_1) et (C, P_2, Γ_2), où C_1 et C_2 sont comme précédemment interprétés comme des ensembles de codes. Supposons que les deux codes soient tirés au hasard indépendamment. Pour chaque couple $c_1 \in C_1$ et $c_2 \in C_2$, la probabilité qu'ils soient tirés conjointement est $P_1(\{c_1\}) \cdot P_2(\{c_2\})$; on en déduit alors que $X \in \Gamma_1(c_1) \cap \Gamma_2(c_2)$. Cependant, si $\Gamma_1(c_1) \cap \Gamma_2(c_2) = \emptyset$, on peut en déduire que la paire (c_1, c_2) n'a pas pu être tirée : par conséquent, la probabilité jointe sur $C_1 \times C_2$ doit être conditionnée par l'ensemble $\{(c_1, c_2) \in C_1 \times C_2 | \Gamma_1(c_1) \cap \Gamma_2(c_2) \neq \emptyset\}$. Ce raisonnement conduit à la règle suivante, appelée règle de Dempster :

$$(m_1 \oplus m_2)(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) m_2(C) \tag{3.25}$$

pour tout $A \subseteq S$, $A \neq \emptyset$ et $(m_1 \oplus m_2)(\emptyset) = 0$, où

$$\kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{3.26}$$

est appelé *degré de conflit* entre m_1 et m_2. Si $\kappa = 1$, les deux éléments d'évidence sont logiquement contradictoires et ne peuvent donc pas être combinés. Une version non normalisée de cette règle a été introduite par Smets (1990a).

La règle de Dempster est commutative, associative et elle admet la fonction de masse vide comme élément neutre. Elle s'exprime facilement à l'aide de la fonction de commonalité ; en notant Q_1, Q_2 et $Q_1 \oplus Q_2$ les fonctions de commonalité associées respectivement à m_1, m_2 et $m_1 \oplus m_2$, on a la relation suivante :

$$Q_1 \oplus Q_2 = \frac{1}{1-\kappa} Q_1 \cdot Q_2. \tag{3.27}$$

Règle disjonctive

La règle de Dempster tend à concentrer les masses sur des ensembles focaux plus petits : elle a donc un comportement conjonctif. On peut en définir un équivalent disjonctif (Dubois et Prade, 1986; Smets, 1993) :

$$\forall A \subseteq S, \quad (m_1 \cup m_2)(A) = \sum_{B \cup C = A} m_1(B) m_2(C). \tag{3.28}$$

Cette règle suppose que l'une au moins des sources d'informations est fiable, contrairement à la règle de Dempster qui suppose que les deux sources d'information sont fiables. La règle $\cup$ est commutative et associative ; elle admet comme élément neutre la fonction de masse m telle que $m(\emptyset) = 1$. Elle s'exprime à l'aide de la fonction Bel par l'équation

$$Bel_1 \cup Bel_2 = Bel_1 \cdot Bel_2, \tag{3.29}$$

qui est à mettre en parallèle à l'équation (3.27).

Approximations

L'application de la règle de Dempster ou de la règle duale disjonctive a pour effet d'augmenter le nombre d'éléments focaux. Pour éviter une explosion combinatoire, il peut être utile d'approcher une fonction de croyance par une autre plus simple, en regroupant des éléments focaux similaires ou de masse faible (Lowrance *et al.*, 1986; Tessem, 1993). Un algorithme efficace pour opérer un tel regroupement en minimisant la perte d'information, basé sur le principe de la classification hiérarchique ascendante, a été proposé par (Denœux, 2001).

Lorsque les équations (3.27) ou (3.29) sont utilisées, la complexité devient fonction non plus du nombre d'éléments focaux, mais de la cardinalité du cadre de discernement S. Un algorithme d'approximation efficace basé sur la recherche d'un grossissement (regroupement d'éléments focaux) minimisant la perte d'information a été proposé par (Denœux et Ben Yaghlane, 2002).

Gestion du conflit

La gestion du conflit entre les sources d'information est un problème important en pratique, qui a fait l'objet de nombreux travaux en théorie des fonctions de croyance (Smets, 2007). Lorsqu'on détecte un conflit important entre des éléments d'information, deux stratégies sont possibles : on peut revenir sur la modélisation de l'information, ou utiliser des règles *robustes*, produisant un résultat cohérent en cas de conflit.

Un exemple d'une telle règle est celle de Dubois et Prade (1988) définie de la manière suivante :

$$(m_1 *_{DP} m_2)(A) = \sum_{B\cap C=A} m_1(B)m_2(C) + \sum_{\{B\cap C=\emptyset, B\cup C=A\}} m_1(B)m_2(C), \qquad (3.30)$$

pour tout $A \subseteq \Omega$, $A \neq \emptyset$, et $(m_1 *_{DP} m_2)(\emptyset) = 0$. Lorsque le degré de conflit entre m_1 et m_2 est nul, on a $m_1 *_{DP} m_2 = m_1 \oplus m_2$: en l'absence totale de conflit, la règle de Dubois et Prade est équivalente à la règle de Dempster. En revanche, lorsque le degré de conflit est égal à 1, on a $m_1 *_{DP} m_2 = m_1 \cup m_2$: dans ce cas, la règle de Dubois et Prade est équivalente à la règle disjonctive. Dans tous les autres cas, le comportement de la règle $*_{DP}$ est intermédiaire entre les modes conjonctif et disjonctif : il s'agit d'une règle de combinaison adaptative. Notons encore que cette règle est commutative mais qu'elle n'est pas associative.

Combinaison d'informations dépendantes

La règle de Dempster et son équivalent disjonctif supposent tous deux l'indépendance des informations combinées. S'il est souvent possible de décomposer un corpus d'évidence en éléments indépendants, ce n'est pas toujours le cas, notamment dans des applications de fusion multi-capteurs par exemple. Il est alors utile de disposer de règles permettant la combinaison d'informations non indépendantes.

Une telle règle, appelé règle prudente, a été proposée par (Denœux, 2008). Cette règle, justifiée par le principe de moindre engagement (cf. infra), est commutative, associative et idempotente. Elle admet une règle duale disjonctive appelée règle hardie. Avec la règle de Dempster et la règle disjonctive définie par l'équation (3.28), les règles prudente et hardie peuvent être vues comme des cas particuliers de familles infinies de règles basées sur des normes triangulaires et des uninormes (Pichon et Denœux, 2010). D'autres règles idempotentes, mais non associatives, ont récemment été proposés par (Destercke et Dubois, 2011) et (Cattaneo, 2011).

Prise en compte de métaconnaissances sur les sources

Lorsque l'on fusionne des informations issues de plusieurs sources, il est souvent utile de prendre en compte non seulement les informations fournies par les sources, mais aussi des métaconnaissances sur leurs propriétés (fiabilité, sincérité, etc.). L'opération d'affaiblissement, en prélevant une proportion de la masse affectée aux éléments focaux pour la transférer sur le cadre de discernement S, permet de prendre en compte la fiabilité d'une source d'information (Shafer, 1976; Smets, 1993). Cette opération a été généralisée par (Mercier *et al.*, 2008) afin de prendre en compte la fiabilité d'une source dans différents contextes. (Pichon *et al.*, 2012) ont proposé un mécanisme très général pour la correction et la combinaison d'informations prenant en compte à la fois la pertinence et la sincérité des sources, et ont montré que les connecteurs de la logique booléenne pouvaient être interprétés en fonction de ces deux propriétés.

3.5.4 Imprécision, spécialisation et mesures d'information

On peut vouloir comparer les fonctions de croyance par leur contenu informatif. Cela permet notamment d'appliquer un principe de maximum d'incertitude ou de « moindre engagement » (Smets, 1993), qui remplit la même fonction que le principe du maximum d'entropie en théorie des probabilités. Ce principe consiste à poser que, lorsque plusieurs fonctions de croyance sont compatibles avec un ensemble de contraintes, la moins informative doit être choisie. L'application de ce principe suppose donc la définition d'un ordre sur les fonctions de croyance. Pour cela, on peut chercher à mesurer le degré d'imprécision et/ou d'incertitude d'une fonction de croyance , ou l'on peut adopter une approche plus qualitative et définir directement une relation d'ordre dans l'ensemble des fonctions de croyance.

Approche quantitative

Les fonctions de croyance modélisant à la fois de l'information imprécise et incertaine, on peut vouloir mesurer leur imprécision et leur incertitude séparément. Un indice d'imprécision naturel est la cardinalité espérée d'une fonction de masse :

$$Imp(m) = \sum_{E \subseteq S} m(E) \cdot Card(E). \tag{3.31}$$

Il est clair que $Imp(m^?) = Card(S)$ (où la fonction de masse $m^?(S) = 1$ code l'ignorance totale), et $Imp(m) = 1$ si la fonction de masse est une probabilité. On vérifie que $Imp(m) = \sum_{s \in S} Pl(s)$.

L'incertitude d'une fonction de croyance peut se mesurer en généralisant la mesure d'entropie probabiliste

$$H(p) = -\sum_{i=1}^{card(S)} p_i \cdot \ln p_i. \tag{3.32}$$

Plusieurs extensions ont été proposées de la forme

$$D(m) = -\sum_{E \subseteq S} m(E) \cdot \ln g(E), \tag{3.33}$$

où g peut être, par exemple, Pl ou Bel (Dubois et Prade, 1987a; Klir et Wierman, 1999). Pour $g = Pl$, on obtient une mesure de dissonance, maximale pour les mesures de probabilité uniforme, et minimale (nulle) dès que tous les ensembles focaux E s'intersectent : $\bigcap\{E|m(E) > 0\} \neq \emptyset$. Pour $g = Bel$, on a plutôt une mesure de confusion, minimale (nulle) dès que $m(E) = 1$ pour un seul ensemble focal (information imprécise mais certaine), mais élevée pour les fonctions de masse uniformes sur tous les ensembles de cardinalité $card(S)/2$. Voir aussi (Ramer et Klir, 1993; Klir et Wierman, 1999).

La règle de Dempster étant le mécanisme fondamental en théorie des fonctions de croyance pour combiner des informations indépendantes, une autre approche consiste à exiger d'une mesure d'incertitude I l'additivité par rapport à cette règle : $I(m_1 \oplus m_2) = I(m1) + I(m_2)$. Cette exigence associée à d'autres conditions assez naturelles conduit Smets (1983) à proposer la mesure suivante :

$$I(m) = \sum_{E \subseteq S} \ln Q(E). \tag{3.34}$$

D'autres critères quantitatifs tentent de mesurer simultanément l'imprécision et l'incertitude. Par exemple, la mesure d'incertitude agrégée $AU(m)$ (Klir et Wierman, 1999) se définit de la façon suivante, pour une fonction de masse m normalisée :

$$AU(m) = \max_{P \in \mathcal{P}(m)} H(P), \tag{3.35}$$

où $\mathcal{P}(m)$ est l'ensemble des mesures de probabilité compatibles avec m :

$$\mathcal{P}(m) = \{P, P(A) \leq Pl(A), \forall A \subseteq S\}. \tag{3.36}$$

Il est clair que $AU(m)$ est maximale à la fois pour $m = m^?$ et pour m telle que $m(\{s\}) = 1/card(S)$ pour tout $s \in S$, ces deux fonctions de masse correspondant, respectivement, à l'imprécision et à l'incertitude maximales.

Approche comparative

La seconde approche pour comparer le contenu informationnel de fonctions de croyance consiste à définir directement une relation d'ordre partiel sur l'ensemble des fonctions de croyance. Etant données deux fonctions de masse m_1 et m_2 normalisées (c'est-à-dire telles que $m_1(\emptyset) = 0$ et $m_2(\emptyset) = 0$), on dit que m_1 est plus précise que m_2 (ce que l'on note $m_1 \sqsubseteq_{Pl} m_2$) si et seulement si, pour tout événement A, l'intervalle $[Bel_1(A), Pl_1(A)]$ est inclus dans l'intervalle $[Bel_2(A), Pl_2(A)]$. A cause de la propriété de dualité entre Pl et Bl, il suffit que l'inégalité $\forall A, Pl_1(A) \leq Pl_2(A)$ soit vérifiée. En termes de probabilités imprécises, si m_1 est plus précise que m_2, cela veut dire que $\mathcal{P}(m_1)$ est un sous-ensemble de $\mathcal{P}(m_2)$ (Dubois et Prade, 1986; Yager, 1986). La fonction m est donc maximalement précise quand elle coïncide avec une probabilité unique, et minimalement précise si $m = m^?$.

Une fonction de masse m_1 est une *spécialisation* d'une fonction de masse m_2 (ce que l'on note $m_1 \sqsubseteq_s m_2$) si et seulement si les trois conditions suivantes sont vérifiées :

1. Tout ensemble focal de m_2 contient au moins un ensemble focal de m_1 ;
2. Tout ensemble focal de m_1 est inclus dans au moins un ensemble focal de m_2 ;
3. Il existe une matrice stochastique W dont le terme w_{ij} est la proportion de la masse $m_1(E_i)$ de l'ensemble focal E_i de m_1 qu'on peut réattribuer à l'ensemble focal F_j de m_2 pour reconstruire la masse $m_2(F_j)$, au sens où $m_2(F_j) = \sum_i w_{ij} \cdot m_1(E_i)$, avec la contrainte $w_{ij} > 0$ seulement si $E_i \subseteq F_j$.

Cette relation est plus exigeante que la précédente : si m_1 est une spécialisation de m_2, alors m_1 est aussi plus précise que m_2 (et non l'inverse, voir (Dubois et Prade, 1986)). Il est aussi évident que si m_1 est une spécialisation de m_2, alors $Imp(m_1) \leq Imp(m_2)$.

On vérifie donc que, dans le cas consonant, m_π (défini par 3.24) et π contiennent la même information, et dans ce cas $Pl = \Pi$ et $Bel = N$. Pour les mesures de possibilité, les ordres informationnels de précision et de spécialisation coïncident avec l'ordre de spécificité des distributions de possibilité sur les singletons : m_{π_1} est une spécialisation de m_{π_2} si et seulement si $\Pi_1(A) \leq \Pi_2(A), \forall A \subseteq S$ si et seulement si $\pi_1(s) \leq \pi_2(s), \forall s \in S$ (Dubois et Prade, 1986).

D'autres relations d'ordre informationnel ont été proposées. Par exemple, m_1 est dite plus informative que m_2 au sens des commonalités (ce que l'on note $m_1 \sqsubseteq_Q m_2$) si $Q_1 \leq Q_2$ (Dubois et Prade, 1986; Yager, 1986). Cette relation peut s'interpréter à partir de l'équation (3.27) : plus les nombres $Q_1(A)$ sont proches de 1, plus faible sera l'influence de m_1 par combinaison de Dempster avec une autre fonction de masse m_2, et donc plus faible est le contenu informationnel de m_1. La relation $\sqsubseteq_Q$ est plus faible que $\sqsubseteq_s$, mais elle n'est pas comparable avec $\sqsubseteq_{Pl}$. D'autres relations d'ordre, proposées par (Denœux, 2008), reposent sur la décomposition canonique d'une fonction de croyance (Smets, 1995). Elles permettent de dériver les règles de combinaison prudentes et hardies du principe de moindre engagement.

3.5.5 Probabilité pignistique et décision

Smets (1990b) a tenté de réconcilier la théorie du pari échangeable (qui justifie les probabilités subjectives) et l'hypothèse que les connaissances d'un agent puissent être représentées par des fonctions de croyance. Une des objections majeures à la théorie des probabilités subjective est son incapacité à distinguer entre une situation de hasard parfait (dé non pipé) et une situation d'ignorance (dé non testé), comme on l'a indiqué plus haut. La théorie des fonctions de croyance permet de capturer cette différence : le cas d'ignorance totale se représente par la fonction de masse $m^?(S) = 1$, qui code la situation où $Bel(A) = 0, Pl(A) = 1, \forall A \neq S, \emptyset$ (ce qui correspond à la distribution de possibilité $\pi^?$ uniformément égale à 1). En revanche, la probabilité uniforme exprime bien l'idée que toutes les réalisations de la variable x sont connues comme équiprobables.

Si un agent ignore tout sur la variable x, il sera amené à proposer une probabilité uniforme sur S, selon le principe d'indifférence de Laplace ; si l'agent a une connaissance partielle descriptible par une fonction de croyance de fonction de masse m, Smets (1990b) suggère que l'agent devra parier avec la distribution de probabilité définie en remplaçant chaque ensemble focal E par une probabilité uniforme sur E, puis en effectuant le mélange convexe de ces probabilités, pondéré par les masses $m(E)$. C'est la probabilité pignistique définie par la distribution $pari$:

$$pari(s) = \sum_{E:s \in E} \frac{m(E)}{card(E)}. \tag{3.37}$$

Cette transformation d'une fonction de croyance en probabilité a été proposée par Dubois et Prade (1982) dans l'esprit d'une généralisation du principe de Laplace. Smets (1990b) l'a justifiée axiomatiquement, en cherchant une probabilité qui obéisse à une propriété de linéarité (la probabilité pignistique d'une somme convexe de fonctions de croyance est la somme convexe de leurs probabilités) et à une propriété d'anonymité (la probabilité pignistique d'un événement ne doit pas changer si on permute les réalisations de cet événement). Une autre tentative de justification a plus tard été proposée par Smets (2005b). En fait, la probabilité pignistique était déjà connue en théorie des jeux coopératifs depuis les années 50 sous le nom de valeur de Shapley et les axiomes de Smets sont mathématiquement les mêmes que ceux proposés par Shapley (1953) dans un contexte très différent.

On peut chercher la fonction de croyance la moins informative, au sens de la spécialisation définie plus haut, correspondant à une probabilité pignistique fixée. On peut montrer qu'elle est unique et consonante, et qu'elle induit donc une distribution de possibilité (Dubois *et al.*, 2008).

Dans le modèle des croyances tranférables de (Smets et Kennes, 1994), la transformation pignistique est vue comme le passage du niveau crédal, où sont combinées les informations, au niveau pignistique où sont prises les décisions. Notons qu'il existe d'autres règles de décision en théorie des fonctions de croyance, comme celle du maximum de plausibilité (Appriou, 1991, 1998; Cobb et Shenoy, 2006) et celle du maximum de croyance correspondant, respectivement, à la minimisation des risques inférieur et supérieur (Denœux, 1997).

3.5.6 Deux types de conditionnement

La plupart du temps, l'information codée par une distribution de probabilité se réfère à une population (l'ensemble des situations qui correspondent aux résultats de tests statistiques).

C'est de l'information générique, typiquement fréquentiste. On utilise cette information pour inférer des croyances sur une situation particulière sur laquelle on a fait des observations incomplètes mais claires. C'est ce qu'on appelle la prédiction. Si $P(A \mid C)$ est la probabilité (fréquentiste) d'apparition de A dans le contexte C, on mesure la confiance $g(A \mid C)$ de l'agent dans la proposition A, lorsqu'il ne connaît que l'information C, à l'aide de la quantité $P(A \mid C)$, en supposant que la situation courante est typique de l'environnement C. La croyance de l'agent relative à la proposition A dans la situation particulière passe de $g(A) = P(A)$ à $g(A \mid C) = P(A \mid C)$ quand on observe que C est vrai dans la situation courante et rien d'autre. Le conditionnement sert alors à mettre à jour les croyances contingentes de l'agent sur la situation courante en exploitant l'information générique. Par exemple, la probabilité P représente la connaissance médicale (souvent compilée sous la forme d'un réseau bayésien). L'information contingente C représente alors les résultats de tests d'un patient. $P(A \mid C)$ est la probabilité d'apparition de la maladie A pour les patients chez qui on observe C ; cette valeur évalue aussi la probabilité (contingente) que ce patient ait cette maladie. Notons que dans ce type d'inférence, la mesure de probabilité P n'évolue pas. On se contente d'appliquer les connaissances génériques à la classe de référence C, ce qu'on peut appeler *focalisation*.

L'information en théorie des fonctions de croyance est supposée représentée par l'affectation de poids positifs $m(E)$ à des sous-ensembles E de S. Dans l'optique de la modélisation de la connaissance générique, $m(E)$ sera, par exemple, la proportion de résultats imprécis, de la forme $x \in E$, d'un test statistique sur une variable aléatoire x. Dans ce cadre, l'inférence plausible dans un contexte C consiste à évaluer les fonctions de poids $m(\cdot \mid C)$ induites par la fonction de masse m sur l'ensemble d'états C. Trois cas sont à envisager (de Campos *et al.*, 1990) :

1. $E \subseteq C$: dans ce cas, $m(E)$ reste attribué à E ;
2. $E \cap C = \emptyset$: dans ce cas, $m(E)$ ne compte plus et est éliminé ;
3. $E \cap C \neq \emptyset$ et $\overline{E} \cap C \neq \emptyset$: dans ce cas, une fraction $\alpha_E \cdot m(E)$ de $m(E)$ reste attribuée à $E \cap C$ et le reste, soit $(1 - \alpha_E) \cdot m(E)$, est attribué à $\overline{E} \cap C$. Mais cette répartition est inconnue.

Le troisième cas correspond aux observations incomplètes E qui ne confirment ni n'infirment C. On n'a pas l'information suffisante pour savoir si, dans chacune des situations correspondant à ces observations, C est vrai ou non. Supposons connues les valeurs $\{\alpha_E, E \subseteq S\}$. Elles valent toujours $\alpha_E = 1$ et 0 dans les premier et second cas respectivement. On construit donc une fonction de masse $m_\alpha(\cdot \mid C)$. Notons qu'une renormalisation de la fonction de masse résultante est nécessaire, en général, dès que $Pl(C) < 1$ (on divise la masse obtenue ci-dessus par $Pl(C)$). Si on note $Bel_\alpha(A \mid C)$ et $Pl_\alpha(A \mid C)$ les fonctions de croyance et de plausibilité obtenues par focalisation sur C avec le vecteur de partage α, on définit les degrés de croyance et de plausibilité conditionnelle sur C par

$$Bel(A \mid C) = \inf_\alpha Bel_\alpha(A \mid C), \tag{3.38}$$

et

$$Pl(A \mid C) = \sup_\alpha Pl_\alpha(A \mid C). \tag{3.39}$$

On obtient encore des fonctions de croyance et de plausibilité [1] (Jaffray, 1992) et les résultats suivants montrent qu'on obtient bien une généralisation du conditionnement bayésien (de Campos *et al.*, 1990) :

$$Bel(A \mid C) = \inf\{P(A \mid C),\ \text{t.q. } P(C) > 0, P \geq Cr\} = \frac{Bel(A \cap C)}{Bel(A \cap C) + Pl(\overline{A} \cap C)}, \quad (3.40)$$

$$Pl(A \mid C) = \sup\{P(A \mid C),\ \text{t.q. } P(C) > 0, P \geq Cr\} = \frac{Pl(A \cap C)}{Pl(A \cap C) + Bel(\overline{A} \cap C)}. \quad (3.41)$$

On voit facilement que $Pl(A \mid C) = 1 - Bel(\overline{A} \mid C)$, et que ces formules généralisent le conditionnement probabiliste au sens où : $Bel(A \mid C)$ est bien une fonction de $Bel(A \cap C)$ et de $Bel(\overline{C} \cup A)$ (et de même pour $Pl(A \mid C)$). Notons que si $Bel(C) = 0$ et $Pl(C) = 1$ (ignorance totale quant à C) alors tous les éléments focaux de m chevauchent C sans que C ne les contienne. Dans ce cas, $Bel(A \mid C) = 0$ et $Pl(A \mid C) = 1, \forall A \neq S, \emptyset$: on ne sait plus rien inférer dans le contexte C.

Un autre conditionnement, dit de Dempster, proposé par Shafer (1976) et Smets et Kennes (1994), suppose systématiquement $\alpha_E = 1$ dès que $E \cap C \neq \emptyset$. On suppose un transfert intégral de la masse de chaque élément focal de E sur $E \cap C \neq \emptyset$ (suivi d'une renormalisation). Cela signifie qu'on interprète la nouvelle information C comme venant modifier la fonction de croyance initiale de sorte que $Pl(\overline{C}) = 0$: les situations où C est faux sont considérées comme impossibles. Si on note $Pl(A \mid\mid C)$ la fonction de plausibilité après révision, on a :

$$Pl(A \mid\mid C) = \frac{Pl(A \cap C)}{Pl(C)}. \quad (3.42)$$

Cela constitue clairement une autre généralisation du conditionnement probabiliste. La croyance conditionnelle est obtenue alors par dualité $Bel(A \mid\mid C) = 1 - Pl(\overline{A} \mid\mid C)$. Notons qu'avec ce conditionnement, la taille des éléments focaux diminue, donc l'information devient plus précise, et les intervalles $[Bel, Pl]$ deviennent plus étroits (toujours plus étroits que ceux obtenus par focalisation). Le conditionnement de Dempster correspond bien à un processus d'enrichissement de l'information, contrairement à la focalisation. Il est un cas particulier de la règle de combinaison de Dempster (cf. Section 3.5.3), la fonction de masse m étant combinée avec une fonction de masse m_C telle que $m_C(C) = 1$. Si $Bel(C) = 0$ et $Pl(C) = 1$ (ignorance totale quant à C), conditionner sur C au sens de Dempster augmente beaucoup la précision des croyances résultantes.

Ce mode de conditionnement n'a rien à voir avec l'approche décrite auparavant, parce que dans l'optique de Shafer et Smets, la fonction de masse m ne représente pas la connaissance générique mais plutôt de l'information contingente incertaine (témoignages non totalement fiables, indices plus ou moins sûrs) recueillie sur une situation précise.

1. Ces deux formules appliquées au cas particulier consonant des mesures de nécessité et de possibilité préservent cette consonance (Dubois et Prade, 1997a). Ce type de conditionnement existe donc aussi pour les possibilités numériques.

3.5.7 Applications en classification

En intelligence artificielle, la théorie des fonctions de croyance a été principalement utilisée, jusqu'au début des années 1990, pour modéliser l'incertitude dans les systèmes experts (Shafer, 1987; Shenoy, 1989). Un autre domaine d'application qui s'est considérablement développé depuis les années 1990 est celui de l'apprentissage statistique. La théorie des fonctions de croyance s'est notamment révélée être un formalisme efficace pour combiner des classifieurs, construire des règles de décision à partir de données d'apprentissage imparfaites ou mettre en évidence de nouvelles structures de classification.

Combinaison de classifieurs

Une première approche pour appliquer la théorie des fonctions de croyance en classification consiste à envisager les sorties de classifieurs comme des éléments d'évidence relativement à la classe de l'objet considéré, et à les combiner par la règle de Dempster ou par une autre règle. Etant donnée la généralité des fonctions de croyance, cette approche peut être appliquée pour combiner des classifieurs de types très divers dont les sorties peuvent être converties en fonctions de croyance.

Par exemple, (Xu *et al.*, 1992) ont proposé d'utiliser une matrice de confusion pour convertir la décision d'un classifieur en fonction de masse et ont obtenu de bons résultats sur une problème de reconnaissance d'écriture manuscrite. Plus récemment, Bi *et col.* (2008) ont proposé une représentation des scores de classifieurs sous forme de fonctions de masses « triplets » avec trois éléments focaux. (Bi, 2012) a étudié l'influence de la diversité des classifieurs combinés par différentes règles, tandis que Quost *et col.* (2011) ont proposé d'optimiser la règle de combinaison au sein d'une famille paramétrée de règles basées sur des normes triangulaires.

Dans une optique différente, Quost *et col.* (2007) se sont intéressés au problème de la combinaison de classifieurs à deux classes, pour résoudre des problèmes multi-classes. Par exemple, dans le schéma dit « un-contre-un », la sortie d'un classifieur peut être vue comme une fonction de masse conditionnelle. Le problème consiste alors à construire une fonction de masse non conditionnelle sur l'ensemble des classes, maximalement compatible avec les fonctions de masse conditionnelles issues des différents classifieurs binaires.

Classifieurs évidentiels

Un classifieur évidentiel est un classifieur dont la sortie est une fonction de masse sur l'ensemble de classes $\Omega = \{\omega_1, \ldots, \omega_c\}$. Deux approches différentes ont été proposées pour construire un tel classifieur à partir d'observations.

La première approche, introduite et justifiée axiomatiquement par (Appriou, 1991), consiste à construire une fonction de masse m sur Ω à partir des vraisemblances $p(x|\omega_k)$, x désignant le vecteur d'attributs. L'une des deux méthodes proposées par Appriou n'est autre que l'application du théorème de Bayes généralisé introduit par Smets (1993). La fonction de masse m a l'expression suivante :

$$m = \bigoplus_{k=1}^{c} \overline{\{\omega_k\}}^{\alpha_k p(x|\omega_k)}, \tag{3.43}$$

α_k étant un coefficient tel que $\alpha_k p(x|\omega_k) \leq 1$, et la notation A^w désignant la fonction de masse simple μ telle que $\mu(A) = 1 - w$ et $\mu(\Omega) = 1 - w$. Un intérêt majeur de cette méthode est qu'elle est applicable en l'absence d'a priori, ou avec un a priori faible sur les classes. En revanche, lorsqu'un a priori probabiliste existe, on retrouve le même résultat que dans l'approche bayésienne. (Appriou, 1991) a bien montré la robustesse de cette méthode, en particulier lorsque la distribution des observations n'est pas la même dans les ensembles d'apprentissage et de test (du fait, par exemple, de conditions d'acquisition différentes ou du dysfonctionnement d'un capteur).

La seconde approche, introduite par (Denœux, 1995), consiste à envisager chaque exemple d'apprentissage (ou chacun des K plus proches voisins de x dans l'ensemble d'apprentissage) comme un élément d'évidence relativement à la classe de l'objet considéré. Les différents éléments d'évidence sont modélisés par des fonctions de masse, qui sont ensuite combinées par la règle de Dempster. Dans la forme la plus générale de cette méthode, on considère un ensemble d'apprentissage $\{(x^{(1)}, m^{(1)}), \ldots, (x^{(N)}, m^{(N)})\}$ où $x^{(i)}$ est le vecteur d'attributs pour l'exemple i et $m^{(i)}$ est une fonction de masse sur Ω caractérisant une connaissance partielle sur la classe de cet exemple. Dans le cas particulier où cette connaissance est certaine, on a $m^{(i)}(\{\omega_j\}) = 1$ pour un élément ω_j de Ω. Dans le cas général, il s'agit d'un problème d'apprentissage partiellement supervisé, l'information partielle sur la classe des exemples d'apprentissage pouvant être, par exemple, apportée par un expert. On suppose par ailleurs connue une mesure de dissimilarité δ caractérisant de manière pertinente le degré de dissemblance entre deux vecteurs.

Considérons maintenant un nouvel objet décrit par le vecteur d'attributs x connu et de classe y inconnue. La fonction de masse représentant l'information apportée par l'exemple d'apprentissage $(x^{(i)}, m^{(i)})$ est définie par l'expression suivante :

$$m(A|e^{(i)}) = \varphi\left(\delta(x, x^{(i)})\right) m^{(i)}(A), \quad \forall A \subset \Omega \tag{3.44}$$

$$m(\Omega|e^{(i)}) = 1 - \sum_{A \subset \Omega} m(A|e^{(i)}) \tag{3.45}$$

où φ est une fonction décroissante vérifiant $\varphi(0) \leq 1$, et $\lim_{d\to\infty} \varphi(d) = 0$. La fonction de masse $m(\cdot|e^{(i)})$ s'obtient donc par *affaiblissement* (Shafer, 1976) de $m^{(i)}$, le facteur d'affaiblissement étant fonction croissante de la dissimilarité entre les vecteurs x et $x^{(i)}$. La condition $\lim_{d\to\infty} \varphi(d) = 0$ traduit le fait que la fonction de masse $m(\cdot|e^{(i)})$ tend vers la fonction de masse vide $m^?$ lorsque la dissimilarité entre les vecteurs x et $x^{(i)}$ tend vers l'infini. La structure vide étant l'élément neutre de l'opération de somme orthogonale, une observation $x^{(i)}$ très éloignée de x n'aura qu'une influence négligeable sur la croyance en la classe de l'objet décrit par x.

Zouhal et Denoeux (1998) ont proposé une méthode permettant de choisir la fonction φ parmi une famille paramétrée en minimisant un critère d'erreur. Denoeux (2000b) a introduit une variante de cette méthode basée sur des prototypes. Le classifieur obtenu peut être vu comme un modèle connexionniste présentant certaines particularités avec les réseaux de neurones à fonctions de base radiales. Zouhal et Denoeux (2001) ont étudié différentes variantes de la méthode dans le cas où l'information sur les classes est de nature possibiliste, et (Petit-Renaud et Denœux, 2004) ont étendu l'approche aux problèmes de régression, la variable y étant quantitative. Plus récemment, Denoeux *et col.* (2010) ont proposé une autre extension aux problèmes de classification multi-label, dans lesquels chaque exemple peut appartenir simultanément à plusieurs classes. Cette extension est basée sur la définition de fonctions de croyance dans un treillis de sous-ensembles de Ω (les intervalles pour l'ordre $\subseteq$). Pour une présentation très générale de cette approche, voir (Denœux et Masson, 2012).

Les deux méthodes de classification résumées ci-dessus, basées respectivement sur les vraisemblances et sur les distances, ont été comparées expérimentalement par Fabre *et col.* (2001) et d'un point de vue théorique par Denœux et Smets (2006), qui ont montré qu'elles pouvaient toutes les deux être dérivées du théorème de Bayes généralisé.

Classification non supervisée

La théorie des fonctions de croyance a également été appliquée au problème de la classification non supervisée (clustering). Il s'agit ici de mettre en évidence automatiquement des groupes d'objets relativement homogènes parmi un ensemble donné. Etant donné un ensemble de classes Ω, Denœux et Masson (2004) ont défini une partition crédale de n objets comme une famille $(m_1, \ldots, m_n)$ de fonctions de masse sur Ω, m_i décrivant une information incertaine sur la classe de l'objet i. Les notions de partitions nette, floue et possibiliste sont des cas particuliers de ce formalisme très général.

Un algorithme permettant de construire une partition crédale à partir de données de proximité a été proposé par Denœux et Masson (2004) et étendu par Masson et Denœux (2004) dans le cas où les proximités sont mal connues et décrites par des intervalles. Masson et Denœux (2008) ont proposé un algorithme d'optimisation alternée de type *fuzzy c-means* générant une partition crédale à partir d'un tableau individus-variables, en optimisant une fonction de coût. Cet algorithme a été étendu par Masson et Denœux (2009) à des données de proximité.

Dans une optique différente, Masson et Denœux (2011) se sont intéressés au problème de la combinaison de partitions obtenues par différents algorithmes de classification (et/ou à partir de tableaux de distances différents). Ils ont proposé de modéliser le résultat de chaque algorithme de classification par une fonction de masse définie sur le treillis des partitions de l'ensemble des n objets, puis de combiner les différentes fonctions de masse par la règle de Dempster. On obtient ainsi une fonction de croyance dans l'ensemble des partitions, qui peut être résumée sous différentes formes. Cet exemple montre qu'il est parfaitement possible de définir et de combiner des fonctions de croyance dans de très grands espaces, à condition d'exploiter la structure de ces espaces (structure de treillis par exemple) et de se limiter à des fonctions de masse de forme simple (Denœux et Masson, 2012).

3.6 Théorie des probabilités imprécises

La théorie des probabilités imprécises (Walley, 1991) procède en quelque sorte d'une démarche inverse de celle des fonctions de croyance. Au lieu de probabiliser l'approche ensembliste de l'imprécis, on introduit l'imprécision ensembliste dans l'approche probabiliste. Avec une vision fréquentiste, l'incertitude épistémique se superpose à la variabilité aléatoire. Dans une vision subjectiviste, Gilboa et Schmeidler (1989) ont montré qu'en relaxant de façon appropriée les axiomes de Savage, on peut justifier par la théorie de la décision l'hypothèse qu'un agent utilise une famille de probabilités pour choisir ses décisions. Pour se prémunir contre l'incertain, il sélectionne, pour évaluer chaque décision, la mesure de probabilité qui assure une utilité espérée minimale. On suppose alors que les connaissances de l'agent sont représentées par une famille convexe de mesures de probabilité sur S.

La spécification d'un modèle de l'incertain de type probabilité imprécise peut s'effectuer de deux façons

– Soit en définissant directement une famille $\mathcal{P}$ de probabilités.
– Soit en affectant à certains événements une borne inférieure et /ou supérieure de probabilités.

La première approche est compatible avec l'idée d'un modèle probabiliste imprécis. Cette imprécision peut avoir diverses origines :

– Dans un cadre fréquentiste, on peut renoncer à l'hypothèse que les fréquences d'observation convergent. A la limite, on sait seulement que la fréquence de chaque événement appartient à un intervalle (Walley et Fine, 1982).
– On peut disposer de plusieurs modèles stochastiques sans savoir lequel choisir. Par exemple, disposer d'un modèle paramétré sans bien connaître la valeur des paramètres. Dans ce cas les bayésiens supposent l'existence d'une distribution de probabilité a priori sur les modèles possibles ou les valeurs des paramètres. C'est précisément ce qui n'est pas supposé ici.
– Des informations incomplètes sont fournies par un agent sur une distribution de probabilité (support, moyenne, mode, médiane, certains fractiles) dans un cadre non-paramétrique ; ou encore l'agent donne la probabilité de certaines propositions (des quantiles par exemple).

Si on part directement d'une famille $\mathcal{P}$ de mesures de probabilité, on peut calculer les bornes (Smith, 1961)

$$P_*(A) = \inf_{P \in \mathcal{P}} P(A); \;\; P^*(A) = \sup_{P \in \mathcal{P}} P(A). \tag{3.46}$$

Les fonctions P_* et P^* sont monotones et vérifient la propriété de dualité $P^*(A) = 1 - P_*(A^c)$. P_* et P^* sont respectivement appelées *enveloppes inférieures et supérieures* (Walley, 1991). L'additivité de P impose les conditions suivantes (Good, 1962) : $\forall A, B \subseteq S$, t. q. $A \cap B = \emptyset$,

$$P_*(A) + P_*(B) \leq P_*(A \cup B) \leq P_*(A) + P^*(B) \leq P^*(A \cup B) \leq P^*(A) + P^*(B). \tag{3.47}$$

La largeur de l'intervalle $[P_*(A), P^*(A)]$ représente en quelque sorte le degré d'ignorance de l'agent quant à la proposition A. Quand cet intervalle occupe tout l'intervalle unité, l'agent n'a aucune connaissance sur A. Quand cet intervalle se réduit à un point, la connaissance probabiliste est maximale. En général la seule connaissance de ces intervalles de probabilité ne suffit pas à reconstituer la famille $\mathcal{P}$. Si on construit à partir de la fonction P_* l'ensemble de mesures de probabilité $\mathcal{P}(P_*) = \{P | \forall A \subseteq S, P(A) \geq P_*(A)\}$, il est facile de voir que $\mathcal{P}(P_*)$ est convexe (si $P_1 \in \mathcal{P}(P_*)$ et $P_2 \in \mathcal{P}(P_*)$ alors, $\forall \lambda \in [0,1], \lambda \cdot P_1 + (1-\lambda) \cdot P_2 \in \mathcal{P}(P_*)$) et contient la fermeture convexe de l'ensemble $\mathcal{P}$ de départ. Mais $\mathcal{P}$ et $\mathcal{P}(P_*)$ produisent les mêmes enveloppes inférieures et supérieures.

Une propriété caractéristique d'une enveloppe supérieure (engendrée par un ensemble non-vide de probabilités) a été établie par Giles (1982). Dans la suite, on confond A et sa fonction caractéristique à valeur dans $\{0, 1\}$ ($A(s) = 1$ si $s \in A$ et 0 sinon). Une fonction d'ensemble g est une enveloppe supérieure si et seulement si pour toute famille $A_0, A_1, \ldots, A_k$ de sous-ensembles de S, et toute paire d'entiers (r, s) tels que $\sum_{i=1}^{k} A_i(\cdot) \geq r + s \cdot A_0(\cdot)$, on a

$$\sum_{i=1}^{k} g(A_i) \geq r + s \cdot g(A_0).$$

Dans l'autre approche, un agent propose des prix d'achat $\bar{p}(f), f \in \mathcal{F}$ pour un ensemble $\mathcal{F}$ d'actes risqués. Un acte risqué est une fonction f de S dans les réels indiquant les états résultants en une perte ($f(s) < 0$) ou un gain ($f(s) > 0$). Un événement A est vu comme un acte risqué coïncidant avec sa fonction caractéristique. On ne suppose plus des paris échangeables : l'agent propose un prix d'achat pour l'acte risqué mais peut proposer un prix de vente plus haut (Walley (1991)). Dans ce cas il peut donc y avoir un écart entre le prix d'achat maximal qu'un agent accepte de proposer pour parier sur un événement A et le prix de vente minimal qu'il utiliserait s'il échangeait sa place avec le banquier. Dans ce cas, $P_*(A)$ est le prix d'achat maximal pour l'acte risqué associé ̀l'événement A, et $P^*(A)$ en est le prix de vente minimal pour l'agent.

La condition de Giles s'interprète en termes de pari et correspond à la cohérence d'un ensemble de bornes supérieures de probabilité (prix de vente minimaux) affectées aux sous-ensembles de S prémunissant un agent qui achète $k+1$ billets de loterie correspondant aux événements $A_0, A_1, \ldots, A_k$ contre une perte sûre d'argent. En effet, l'ensemble $\mathcal{P} = \{P | P(A_i) \geq \bar{p}(A_i), i = 0, ...k\}$ peut être vide (ce qui indique une perte sûre d'argent avec les prix d'achats trop hauts $\bar{p}(A_i)$) ; et si $\mathcal{P} \neq \emptyset$, ces prix peuvent parfois être augmentés sans altérer l'information : en général,

$$P_*(A_i) = \inf\{P(A_j) | P(A_i) \geq \bar{p}(A_i), i = 0, ..., k\} \geq \bar{p}(A_i)$$

Si on a l'égalité, à savoir si une fonction d'ensemble monotone g est telle que $g(A) = \inf\{P(A) | P \geq g\}$, elle est dite *Walley-cohérente*. C'est alors l'enveloppe inférieure d'une famille convexe de probabilités (souvent appelée ensemble crédal). De plus, la Walley-cohérence peut être vue comme la généralisation d'une fermeture déductive : si $\mathcal{K}$ est une famille logiquement cohérente de propositions $A_0, A_1, \ldots, A_k$, et qu'on pose $\bar{p}(A_i) = 1, i = 0, ...k$, alors $P_*(A) = 1$ si et seulement si $\mathcal{K} \models A$.

Deux types de conditionnement

Dans le cas où la connaissance générique de l'agent est représentée par des probabilités imprécises, l'inférence plausible bayésienne se généralise en effectuant une analyse de sensibilité sur la probabilité conditionnelle (Walley, 1991). Soit $\mathcal{P}$ une famille de mesures de probabilité sur S. Pour chaque proposition A on connaît une borne inférieure $P_*(A)$ et une borne supérieure $P^*(A)$ du degré de probabilité de A. En présence d'observations contingentes résumées par un contexte C, la croyance de l'agent relative à la proposition A est représentée par l'intervalle $[P_*(A \mid C), P^*(A \mid C)]$ défini par

$$P_*(A \mid C) = \inf\{P(A \mid C), P(C) > 0, P \in \mathcal{P}\}$$

$$P^*(A \mid C) = \sup\{P(A \mid C), P(C) > 0, P \in \mathcal{P}\}.$$

Il peut arriver que l'intervalle $[P_*(A \mid C), P^*(A \mid C)]$ soit plus large que $[P_*(A), P^*(A)]$ ce qui correspond à une perte d'information dans les contextes plus spécifiques par rapport aux contextes généraux. Cela correspond à l'idée que plus on connaît d'informations singulières caractérisant une situation, moins on sait lui appliquer l'information générique (car le nombre d'observations statistiques lui correspondant sera très faible). On voit que ce conditionnement ne correspond pas du tout à l'idée d'enrichissement de l'information générique.

Dans le cadre plus général des probabilités imprécises, une application brutale de la révision par une information C consiste à imposer la contrainte supplémentaire $P(C) = 1$ à la famille $\mathcal{P}$, soit

$$P_*(A \mid\mid C) = \inf\{P(A \mid C), P(C) = 1, P \in \mathcal{P}\};$$

$$P^*(A \mid\mid C) = \sup\{P(A \mid C), P(C) = 1, P \in \mathcal{P}\}.$$

Mais il se peut que l'ensemble $\{P \in \mathcal{P}, P(C) = 1\}$ soit vide (c'est toujours le cas dans le cadre classique bayésien car $\mathcal{P}$ est un singleton). On applique alors le principe de maximum de vraisemblance (Gilboa et Schmeidler, 1992) et on remplace la condition $P(C) = 1$ par $P(C) = P^*(C)$ dans l'équation ci-dessus. On généralise ainsi la règle de Dempster (qui est retrouvée si P^* est une fonction de plausibilité.)

En tant que cadre générique de représentation de l'incertain numérique, susceptible de multiples interpétations, et incluant les autres théories comme des cas particuliers, les probabilités imprécises connaissent un intérêt croissant, et suscite de nombreux travaux théoriques (ainsi, on a pu par exemple établir des liens (De Cooman et Hermans, 2008) entre la théorie des probabilités imprecises au sens de Walley, et une vue des probabilités en termes de jeux (Shafer et Vovk, 2001)), mais aussi algorithmiques comme par exemple la version imprécise des réseaux bayésiens (Cozman, 2000; de Campos et Cozman, 2005).

3.7 Conclusion

L'intelligence artificielle, en s'intéressant à la représentation et à l'exploitation d'informations imparfaites a été naturellement conduite à reconnaître les insuffisances des cadres de la logique classique et de la théorie des probabilités pour le traitement de telles informations, et à introduire et développer de nouveaux cadres de représentation plus adaptés, que ce chapitre a passé en revue. Ces nouveaux cadres sont multiples et complémentaires, plutôt que concurrents, même si la recherche dans ce domaine demeure trop compartimentée. Ces nouvelles approches offrent, comme on a pu le voir une grande richesse de représentation, qu'il faut correctement appréhender pour les utiliser à bon escient. Ils peuvent être quantitatifs, et permettre en particulier l'expression de probabilités imprécises, ou être qualitatifs. Il est clair aussi que malgré les nombreux travaux concernant ces différents cadres depuis 40 ans, il reste encore à faire pour parvenir à une unification des formalismes de l'incertain et à en explorer les liens avec les statistiques, et plus généralement à en maîriser complètement l'usage.

Références

ABADI, M. et HALPERN, J. Y. (1994). Decidability and expressiveness for first-order logics of probability. *Inf. Comput.*, 112(1):1–36.

APPRIOU, A. (1991). Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, (11):27–40.

APPRIOU, A. (1998). Uncertain data aggregation in classification and tracking processes. *In* BOUCHON-MEUNIER, B., éditeur : *Aggregation and Fusion of imperfect information*, pages 231–260. Physica-Verlag, Heidelberg.

ASSAGHIR, Z., KAYTOUE, M. et PRADE, H. (2010). A possibility theory-oriented discussion of conceptual pattern structures. *In* DESHPANDE, A. et HUNTER, A., éditeurs : *Proc. Int. Conf. on Scalable Uncertainty Management (SUM'10), Toulouse, Sept. 27-29*, numéro 6379 de LNCS, pages 70–83. Springer.

BACCHUS, F. (1991). *Representing and Reasoning With Probabilistic Knowledge : A Logical Approach to Probabilities*. MIT Press, Cambridge, Ma.

BANERJEE, M. et DUBOIS, D. (2009). A simple modal logic for reasoning about revealed beliefs. *In* SOSSAI, C. et CHEMELLO, G., éditeurs : *Proc. 10th Europ. Conf. Symb. and Quantit. Approaches to Reasoning with Uncertainty (ECSQARU'09), Verona, July 1-3*, volume 5590 de *LNCS*, pages 805–816. Springer.

BARBUT, M. et MONTJARDET, B. (1970). *Ordre et Classification : Algèbre et Combinatoire*. Hachette.

BELOHLAVEK, R. (2002). *Fuzzy Relational Systems. Foundations and Principles*. Kluwer.

BEN AMOR, N., BENFERHAT, S., DUBOIS, D., MELLOULI, K. et PRADE, H. (2002). A theoretical framework for possibilistic independence in a weakly ordered setting. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(2):117–155.

BENFERHAT, S., DUBOIS, D., GARCIA, L. et PRADE, H. (2002). On the transformation between possibilistic logic bases and possibilistic causal networks. *Int. J. of Approximate Reasoning*, 29:135–173.

BENFERHAT, S., DUBOIS, D., KACI, S. et PRADE, H. (2008). Modeling positive and negative information in possibility theory. *Inter. J. of Intelligent Systems*, 23:1094–1118.

BENFERHAT, S., DUBOIS, D. et PRADE, H. (1997). Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92:259–276.

BENFERHAT, S., DUBOIS, D. et PRADE, H. (1998). Practical handling of exception-tainted rules and independence information in possibilistic logic. *Applied Intelligence*, 9:101–127.

BENFERHAT, S., DUBOIS, D. et PRADE, H. (1999a). An overview of inconsistency-tolerant inferences in prioritized knowledge bases. *In* DUBOIS, D., PRADE, H. et KLEMENT, E., éditeurs : *Fuzzy Sets, Logic and Reasoning about Knowledge*, volume 15 de *Applied Logic Series*, pages 395–417. Kluwer, Dordrecht.

BENFERHAT, S., DUBOIS, D. et PRADE, H. (1999b). Possibilistic and standard probabilistic semantics of conditional knowledge bases. *J. of Logic and Computation*, 9:873–895.

BENFERHAT, S., DUBOIS, D. et PRADE, H. (2000). Kalman-like filtering in a possibilistic setting. *In* HORN, W., éditeur : *Proc.14th Europ.Conf. on Artificial Intelligence (ECAI'00), Berlin, Aug. 20-25*, pages 8–12.

BENFERHAT, S. et KACI, S. (2003). Logical representation and fusion of prioritized information based on guaranteed possibility measures : Application to the distance-based merging of classical bases. *Artificial Intelligence*, 148:291–333.

BI, Y. (2012). The impact of diversity on the accuracy of evidential classifier ensembles. *Int. J. of Approximate Reasoning*, 53(4):584–607.

BI, Y., GUAN, J. et BELL, D. (2008). The combination of multiple classifiers using an evidential reasoning approach. *Artificial Intelligence*, 172(15):1731–1751.

BIAZZO, V., GILIO, A., LUKASIEWICZ, T. et SANFILIPPO, G. (2002). Probabilistic logic under coherence, model-theoretic probabilistic logic, and default reasoning in system p. *J.*

of Applied Non-Classical Logics, 12(2):189–213.

BOLT, J. H., VAN DER GAAG, L. C. et RENOOIJ, S. (2005). Introducing situational signs in qualitative probabilistic networks. *Int. J. Approx. Reasoning*, 38:333–354.

BONNEFON, J.-F., DA SILVA NEVES, R., D.DUBOIS et PRADE, H. (2008). Predicting causality ascriptions from background knowledge : model and experimental validation. *Int. J. Approx. Reasoning*, 48:752–765.

BONNEFON, J.-F., DA SILVA NEVES, R., D.DUBOIS et PRADE, H. (2012). Qualitative and quantitative conditions for the transitivity of perceived causation - theoretical and experimental results. *Annals of Mathematics and Artificial Intelligence*, 64:311–333.

BUCHANAN, B. G. et SHORTLIFFE, E. H. (eds.) (1984). *Rule-Based Expert Systems*. Addison-Wesley, Reading, Mass.

BURMEISTER, P. et HOLZER, R. (2005). Treating incomplete knowledge in formal concepts analysis. *In* GANTER, B., éditeur : *Formal Concept Analysis : Foundations and Applications*, volume 3626 de *LNCS*, pages 114–126. Springer.

CATTANEO, M. E. G. V. (2011). Belief functions combination without the assumption of independence of the information sources. *Int. J. of Approximate Reasoning*, 52(3):299–315.

CAYRAC, D., DUBOIS, D. et PRADE, H. (1996). Handling uncertainty with possibility theory and fuzzy sets in a satellite fault diagnosis application. *IEEE Trans. on Fuzzy Systems*, 4(3):251–269.

CAYROL, M., FARRENY, H. et PRADE, H. (1982). Fuzzy pattern matching. *Kybernetes*, 11(2):103–116.

CHEESEMAN, P. (1988). An inquiry into computer understanding. *Computational Intelligence*, 4:58–66. with comments by R. Aleliunas, A. Bundy, N. C. Dalkey, A. P. Dempster, D. Dubois and H. Prade, M. L. Ginsberg, R. Greiner, P. J. Hayes, D. Israel, L. Kanal and D. Perlis, H. Kyburg, D. McDermott, D. L. McLeish, C. G. Morgan, E. Neufeld and D. Poole, J. Pearl, L. Rendell, E. H. Ruspini, L.K. Schubert, G. Shafer, D. J. Spiegelhalter, R. R. Yager, L. A. Zadeh (67–128), and a reply by P. Cheeseman (129–142).

CHOQUET, G. (1953). Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295.

COBB, B. R. et SHENOY, P. P. (2006). On the plausibility transformation method for translating belief function models to probability models. *Int. J. of Approximate Reasoning*, 41(3):314–330.

COLETTI, G. et SCOZZAFAVA, R. (2002). *Probabilistic Logic in a Coherent Setting*. Kluwer Acad. Publ.

COLETTI, G. et SCOZZAFAVA, R. (2003). Coherent conditional probability as a measure of uncertainty of the relevant conditioning events. *In* NIELSEN, T. D. et ZHANG, N. L., éditeurs : *Proc. 7th Europ. Conf. Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'03), Aalborg, July 2-5*, volume 2711 de *LNCS*, pages 407–418. Springer.

COX, R. T. (1946). Probability, frequency, and reasonable expectation. *Am. Jour. Phys.*, 14:1–13.

COZMAN, F. G. (2000). Credal networks. *Artificial Intelligence*, 120:199–233.

de CAMPOS, C. P. et COZMAN, F. G. (2005). The inferential complexity of Bayesian and credal networks. *In* KAELBLING, L. P. et SAFFIOTTI, A., éditeurs : *Proc. 19th Int. Joint*

Conf. on Artificial Intelligence (IJCAI'05), Edinburgh, July 30-Aug. 5, pages 1313–1318.

DE CAMPOS, L. M., LAMATA, M. T. et MORAL, S. (1990). The concept of conditional fuzzy measure. *Int. J. of Intelligent Systems*, 5:237–246.

DE COOMAN, G. (1997). Possibility theory. Part I : Measure- and integral-theoretic groundwork ; Part II : Conditional possibility ; Part III : Possibilistic independence. *Int. J. of General Syst.*, 25:291–371.

DE COOMAN, G. et AEYELS, D. (1999). Supremum preserving upper probabilities. *Information Sciences*, 118(1-4):173–212.

DE COOMAN, G. et HERMANS, F. (2008). Imprecise probability trees : Bridging two theories of imprecise probability. *Artificial Intelligence*, 172:1400–1427.

DE FINETTI, B. (1936). La logique des probabilités. *In Congrès International de Philosophie Scientifique*, pages 1–9, Paris. Hermann et Cie.

DE FINETTI, B. (1974). *Theory of Probability*. Wiley and Sons, New York.

DEMPSTER, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Statistics*, 28:325–339.

DEMPSTER, A. P. (1968). Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics*, 39(3):957–966.

DENŒUX, T. (1995). A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics*, 25(05):804–813.

DENŒUX, T. (1997). Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7):1095–1107.

DENŒUX, T. (1999). Reasoning with imprecise belief structures. *Int. J. of Approximate Reasoning*, 20:79–111.

DENŒUX, T. (2000a). Modeling vague beliefs using fuzzy-valued belief structures. *Fuzzy Sets and Systems*, 116(2):167–199.

DENŒUX, T. (2000b). A neural network classifier based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics A*, 30(2):131–150.

DENŒUX, T. (2001). Inner and outer approximation of belief structures using a hierarchical clustering approach. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(4): 437–460.

DENŒUX, T. (2008). Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence. *Artificial Intelligence*, 172:234–264.

DENŒUX, T. (2009). Extending stochastic ordering to belief functions on the real line. *Information Sciences*, 179:1362–1376.

DENŒUX, T. et BEN YAGHLANE, A. (2002). Approximating the combination of belief functions using the fast Mœbius transform in a coarsened frame. *Int. J. of Approximate Reasoning*, 31(1-2):77–101.

DENŒUX, T. et MASSON, M.-H. (2004). EVCLUS : Evidential clustering of proximity data. *IEEE Trans. on Systems, Man and Cybernetics B*, 34(1):95–109.

DENŒUX, T. et MASSON, M.-H. (2012). Evidential reasoning in large partially ordered sets. application to multi-label classification, ensemble clustering and preference aggregation. *Annals of Operations Research*, 195(1):135–161.

DENŒUX, T. et SMETS, P. (2006). Classification using belief functions : the relationship between the case-based and model-based approaches. *IEEE Transactions on Systems, Man and Cybernetics B*, 36(6):1395–1406.

DENOEUX, T., YOUNES, Z. et ABDALLAH, F. (2010). Representing uncertainty on set-valued variables using belief functions. *Artificial Intelligence*, 174(7-8):479–499.

DENŒUX, T. et ZOUHAL, L. M. (2001). Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, 122(3):47–62.

DESTERCKE, S. et DUBOIS, D. (2011). Idempotent conjunctive combination of belief functions : Extending the minimum rule of possibility theory. *Information Sciences*, 181(18): 3925–3945.

DESTERCKE, S., DUBOIS, D. et CHOJNACKI, E. (2008). Unifying practical uncertainty representations : I. Generalized p-boxes. II. Clouds. *Int. J. of Approximate Reasoning*, 49:649–663, 664–677.

DJOUADI, Y., DUBOIS, D. et PRADE, H. (2011). Graduality, uncertainty and typicality in formal concept analysis. *In* CORNELIS, C., DESCHRIJVER, G., NACHTEGAEL, M., SCHOCKAERT, S. et SHI, Y., éditeurs : *35 Years of Fuzzy Set Theory - Celebratory Volume Dedicated to the Retirement of Etienne E. Kerre*, pages 127–147. Springer.

DOMOTOR, Z. (1985). Probability kinematics - Conditional and entropy principles. *Synthese*, 63:74–115.

DUBOIS, D. (1986). Belief structures, possibility theory and decomposable confidence measures on finite sets. *Computers and Artificial Intelligence*, 5(5):403–416.

DUBOIS, D., FARIÑAS DEL CERRO, L., HERZIG, A. et PRADE, H. (1999). A roadmap of qualitative independence. *In* DUBOIS, D., PRADE, H. et KLEMENT, E., éditeurs : *Fuzzy Sets, Logics and Reasoning about Knowledge*, volume 15 de *Applied Logic series*, pages 325–350. Kluwer Acad. Publ., Dordrecht.

DUBOIS, D., FODOR, J. et PRADE, H. (2010). Conditional measures : An alternative to Cox functional equation. *In* CINTULA, P., KLEMENT, E. P. et STOUT, L. N., éditeurs : *Proc. 31st Linz Seminar on Fuzzy Set Theory, Linz, Austria, Feb. 9-13)*, pages 43–46.

DUBOIS, D., FOULLOY, L., MAURIS, G. et PRADE, H. (2004). Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10:273–297.

DUBOIS, D., GRABISCH, M., DE MOUZON, O. et PRADE, H. (2001a). A possibilistic framework for single-fault causal diagnosis under uncertainty. *Int. J. of General Systems*, 30(2):167–192.

DUBOIS, D., HAJEK, P. et PRADE, H. (2000a). Knowledge-driven versus data-driven logics. *J. of Logic, Language, and Information*, 9:65–89.

DUBOIS, D. et HÜLLERMEIER, E. (2007). Comparing probability measures using possibility theory : A notion of relative peakedness. *Int. J. of Approximate Reasoning*, 45:364–385.

DUBOIS, D., LANG, J. et PRADE, H. (1994). Possibilistic logic. *In* GABBAY, D., HOGGER, C., ROBINSON, J. et NUTE, D., éditeurs : *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3*, pages 439–513. Oxford Univ. Press.

DUBOIS, D., MORAL, S. et PRADE, H. (1997). A semantics for possibility theory based on likelihoods. *J. Math. Anal. Appl.*, 205:359–380.

DUBOIS, D., PAP, E. et PRADE, H. (2000b). Hybrid probabilistic-possibilistic mixtures and utility functions. *In* FODOR, J., BAETS, B. D. et PERNY, P., éditeurs : *Preferences and Decisions under Incomplete Knowledge*, pages 51–73. Physica-Verlag.

DUBOIS, D. et PRADE, H. (1980). *Fuzzy Sets and Systems - Theory and Applications*. Academic Press, New York.

DUBOIS, D. et PRADE, H. (1982). A class of fuzzy measures based on triangular norms. a general framework for the combination of uncertain information. *Int. J. of General Systems*, 8(1):43–61.

DUBOIS, D. et PRADE, H. (1985 ; 2e édit.1987b). *Théorie des Possibilités. Applications à la Représentation des Connaissances en Informatique*. (avec la collaboration de H. Farreny, R. Martin-Clouaire, C. Testemale), Masson, Paris.

DUBOIS, D. et PRADE, H. (1986). A set-theoretic view of belief functions : Logical operations and approximations by fuzzy sets. *Int. J. General Systems*, 12:193–226.

DUBOIS, D. et PRADE, H. (1987a). Properties of information measures in evidence and possibility theories. *Fuzzy Sets and Systems*, 24:161–182.

DUBOIS, D. et PRADE, H. (1988). Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4:244–264.

DUBOIS, D. et PRADE, H. (1989). Handling uncertainty in expert systems : pitfalls, difficulties, remedies. *In* HOLLNAGEL, E., éditeur : *The Reliability of Expert Systems*, pages 64–118. Ellis Horwood, Chichester, U.K.

DUBOIS, D. et PRADE, H. (1991). Epistemic entrenchment and possibilistic logic. *Artificial Intelligence*, 50:223–239.

DUBOIS, D. et PRADE, H. (1992). Putting rough sets and fuzzy sets together. *In* SLOWINSKI, R., éditeur : *Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory*, pages 203–232. Kluwer Acad. Publ.

DUBOIS, D. et PRADE, H. (1994). Conditional objects as nonmonotonic consequence relationships. *IEEE Trans. on Systems, Man and Cybernetics*, 24(12):1724–1740.

DUBOIS, D. et PRADE, H. (1996). What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84:169–185.

DUBOIS, D. et PRADE, H. (1997a). Bayesian conditioning in possibility theory. *Fuzzy Sets and Systems*, 92:223–240.

DUBOIS, D. et PRADE, H. (1997b). Focusing vs. belief revision : A fundamental distinction when dealing with generic knowledge. *In* GABBAY, D. M., KRUSE, R., NONNENGART, A. et OHLBACH, H. J., éditeurs : *Proc. 1st Int. Joint Conf. on Qualitative and Quantitative Practical Reasoning (ECSQARU-FAPR'97), Bad Honnef, June 9-12*, volume 1244 de *LNCS*, pages 96–107. Springer.

DUBOIS, D. et PRADE, H. (1998). Possibility theory : Qualitative and quantitative aspects. *In* GABBAY, D. M. et SMETS, P., éditeurs : *Quantified Representation of Uncertainty and Imprecision*, volume 1 de *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 169–226. Kluwer Acad. Publ.

DUBOIS, D. et PRADE, H. (2001). Possibility theory, probability theory and multiple-valued logics : A clarification. *Annals of Mathematics and Artificial Intelligence*, 32:35–66.

DUBOIS, D. et PRADE, H. (2004). Possibilistic logic : a retrospective and prospective view.

Fuzzy Sets and Systems, 144:3–23.

DUBOIS, D. et PRADE, H. (2006). Représentations formelles de l'incertain et de l'imprécis. *In* BOUYSSOU, D., DUBOIS, D., PIRLOT, M. et PRADE, H., éditeurs : *Outils de Modélisation*, volume 1 de *Concepts et Méthodes pour l'Aide à la Décision*, chapitre 3, pages 111–171. Traité IC2, Hermes, Lavoisier.

DUBOIS, D. et PRADE, H. (2012). Possibility theory and formal concept analysis : Characterizing independent sub-contexts. *Fuzzy Sets and Systems*, 196:4–16.

DUBOIS, D. et PRADE, H. (eds.) (2008). *Bipolar Representations of Information and Preference. Part 1A & Part 1B : Cognition and Decision ; Part 2 : Reasoning and Learning*. Special issue, Int. J. of Intelligent Systems, 23 (8,9,10), Wiley.

DUBOIS, D., PRADE, H. et RICO, A. (2013). Qualitative capacities as imprecise possibilities. *In* van der GAAG, L. C., éditeur : *Proc. 12th Europ. Conf. Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'13), Utrecht, July 7-10*, volume 7958 de *LNCS*, pages 169–180. Springer.

DUBOIS, D., PRADE, H. et SCHOCKAERT, S. (2012). Stable models in generalized possibilistic logic. *In* BREWKA, G., EITER, T. et MCILRAITH, S. A., éditeurs : *Proc.13th Int. Conf. Principles of Knowledge Representation and Reasoning (KR'12), Rome, June 10-14*, pages 519–529.

DUBOIS, D., PRADE, H. et SMETS, P. (2001b). "Not impossible" vs. "guaranteed possible" in fusion and revision. *In* BENFERHAT, S. et BESNARD, P., éditeurs : *Proc. 6th Europ. Conf. Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'01), Toulouse, Sept. 19-21*, volume 2143 de *LNCS*, pages 522–531. Springer.

DUBOIS, D., PRADE, H. et SMETS, P. (2008). A definition of subjective possibility. *Int. J. Approx. Reasoning*, 48:352–364.

DUPIN DE SAINT-CYR, F. et PRADE, H. (2008). Handling uncertainty and defeasibility in a possibilistic logic setting. *Int. J. Approx. Reasoning*, 49:67–82.

FABRE, S., APPRIOU, A. et BRIOTTET, X. (2001). Presentation and description of two classification methods using data fusion based on sensor management. *Information Fusion*, 2(1):49–71.

FERRÉ, S. et RIDOUX, O. (2004). Introduction to logical information systems. *Inf. Process. Manage.*, 40(3):383–419.

FINE, T. (1983). *Theories of Probability*. Academic Press, New York.

GAIFMAN, H. et SNIR, M. (1982). Probabilities over rich languages, testing and randomness. *J. Symbolic Logic*, 47(3):495–548.

GANTER, B. et KUZNETSOV, S. O. (2001). Pattern structures and their projections. *In* DELUGACH, H. S. et STUMME, G., éditeurs : *Proc. 9th Int. Conf. on Conceptual Structures (ICCS'01), Stanford, Jul. 30-Aug. 3*, volume 2120 de *LNCS*, pages 129–142. Springer.

GANTER, B. et WILLE, R. (1999). *Formal Concept Analysis. Mathematical Foundations*. Springer-Verlag.

GÄRDENFORS, P. (2nd ed., College Publications, 2008). *Knowledge in Flux. Modeling the Dynamics of Epistemic States*. 1st ed., MIT Press, 1988.

GILBOA, I. et SCHMEIDLER, D. (1989). Maxmin expected utility with a non-unique prior. *J. of Mathematical Economics*, 18:141–153.

GILBOA, I. et SCHMEIDLER, D. (1992). Updating ambiguous beliefs. *In* MOSES, Y., éditeur : *Proc. of the 4th Conf. on Theoretical Aspects of Reasoning about Knowledge (TARK'92), Monterey*, pages 143–162. Morgan Kaufmann.

GILES, R. (1982). Foundations for a theory of possibility. *In* GUPTA, M. M. et SANCHEZ, E., éditeurs : *Fuzzy Information and Decision Processes*, pages 183–195. North-Holland.

GINSBERG, M. L. (1990). Bilattices and modal operators. *J. of Logic and Computation*, 1:1–41.

GOLDSZMIDT, M. et PEARL, J. (1991). System Z^+ : A formalism for reasoning with variable-strength defaults. *In Proc. 9th National Conf. on Artificial Intelligence (AAAI'91), Anaheim, July 14-19*, volume 1, pages 339–404.

GOOD, I. J. (1962). Subjective probability as the measure of a non-measurable set. *In* NAGEL, E., SUPPES, P. et TARSKI, A., éditeurs : *Logic, Methodology, and Philosophy of Science*, pages 319–329. Stanford University Press.

GUAN, J. W. et BELL, D. A. (1991). *Evidential Reasoning and its Applications. Vol 1*. North-Holland, Amsterdam.

HALPERN, J. Y. (1990). An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350.

HALPERN, J. Y. (1999a). A counterexample to theorems of Cox and Fine. *J. Artif. Intell. Res. (JAIR)*, 10:67–85.

HALPERN, J. Y. (1999b). Technical addendum, Cox's theorem revisited. *J. Artif. Intell. Res. (JAIR)*, 11:429–435.

HALPERN, J. Y. (2001). Plausibility measures : A general approach for representing uncertainty. *In* NEBEL, B., éditeur : *Proc. 17th Int. Joint Conf. on Artificial Intelligence (IJCAI'01), Seattle, Aug.4-10, 2001*, pages 1474–1483. Morgan Kaufmann.

HALPERN, J. Y. (2003). *Reasoning About Uncertainty*. MIT Press, Cambridge, Ma.

HALPERN, J. Y. et PUCELLA, R. (2002). A logic for reasoning about upper probabilities. *J. Artif. Intell. Res. (JAIR)*, 17:57–81.

HALPERN, J. Y. et PUCELLA, R. (2006). A logic for reasoning about evidence. *J. Artif. Intell. Res. (JAIR)*, 26:1–34.

HIGASHI, M. et KLIR, G. J. (1982). Measures of uncertainty and information based on possibility distributions. *Int. J. General Systems*, 8:43–58.

HORVITZ, E., HECKERMAN, D. et LANGLOTZ, C. (1986). A framework for comparing alternative formalisms for plausible reasoning. *In* KEHLER, T., éditeur : *Proc. 5th Nat. Conf. on Artificial Intelligence. Philadelphia, Aug.11-15, 1986. Vol. 1*, pages 210–214. Morgan Kaufmann.

JAEGER, M. (2001). Automatic derivation of probabilistic inference rules. *Int. J. Approx. Reasoning*, 28:1–22.

JAEGER, M. (2006). Probabilistic role models and the guarded fragment. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 14(1):43–60.

JAFFRAY, J.-Y. (1992). Bayesian updating and belief functions. *IEEE Trans. on Systems, Man, and Cybernetics*, 22:1144–1152.

JAYNES, E. T. (1979). Where do we stand on maximum entropy. *In* LEVINE, I. et TRIBUS, M., éditeurs : *The Maximum Entropy Formalism*, pages 15–118. MIT Press.

JAYNES, E. T. (2003). *Probability Theory : The Logic of Science*. Cambridge Univ. Press. preprint version, 1996.

KENNES, R. (1992). Computational aspects of the Möbius transformation of graphs. *IEEE Trans. on Systems, Man, and Cybernetics*, 22:201–223.

KLIR, G. J. et WIERMAN, M. J. (1999). *Uncertainty-Based Information. Elements of Generalized Information Theory*. Springer-Verlag, New-York.

KRAUS, S., LEHMANN, D. et MAGIDOR, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207.

KYBURG, JR., H. E. et TENG, C. M. (2012). The logic of risky knowledge, reprised. *Int. J. Approx. Reasoning*, 53:274–285.

LEHMANN, D. J. et MAGIDOR, M. (1992). What does a conditional knowledge base entail ? *Artificial Intelligence*, 55:1–60.

LINDLEY, D. V. (1982). Scoring rules and the inevitability of probability. *Int. Statistics Rev.*, 50:1–26.

LIU, W. (2001). *Propositional, Probabilistic and Evidential Reasoning : Integrating Numerical and Symbolic Approaches*. Physica Verlag, Springer.

LOWRANCE, J. D., GARVEY, T. D. et STRAT, T. M. (1986). A framework for evidential-reasoning systems. *In* et AL., T. K., éditeur : *Proceedings of AAAI'86*, volume 2, pages 896–903, Philadelphia. AAAI.

LUCAS, P. et VAN DER GAAG, L. (1991). *Principles of Expert Systems*. Addison-Wesley.

MARCHIONI, E. et GODO, L. (2004). A logic for reasoning about coherent conditional probability : A modal fuzzy logic approach. *In* ALFERES, J. J. et LEITE, J. A., éditeurs : *Proc. 9th Europ. Conf. on Logics in Artificial Intelligence (JELIA'04), Lisbon, Sept. 27-30*, volume 3229 de *LNCS*, pages 213–225. Springer.

MARTIN, T. (2006). Logique du probable de Jacques Bernoulli à J.-H. Lambert. *Journ@l Electronique d'Histoire des Probabilités et de la Statistique*, 2(1b): http ://www.jehps.net/Novembre2006/Martin3.pdf.

MASSON, M.-H. et DENŒUX, T. (2004). Clustering interval-valued data using belief functions. *Pattern Recognition Letters*, 25(2):163–171.

MASSON, M.-H. et DENŒUX, T. (2008). ECM : an evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41(4):1384ñ–1397.

MASSON, M.-H. et DENŒUX, T. (2009). RECM : relational evidential c-means algorithm. *Pattern Recognition Letters*, 30:1015–1026.

MASSON, M.-H. et DENŒUX, T. (2011). Ensemble clustering in the belief functions framework. *Int. J. of Approximate Reasoning*, 52(1):92–109.

MERCIER, D., QUOST, B. et DENŒUX, T. (2008). Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, 9(2):246–258.

MILCH, B. et RUSSELL, S. J. (2007). First-order probabilistic languages : Into the unknown. *In* MUGGLETON, S., OTERO, R. P. et TAMADDONI-NEZHAD, A., éditeurs : *Revised Selected Papers of the16th Int. Conf. on Inductive Logic Programming (ILP'06), Santiago de Compostela, Aug. 24-27*, volume 4455 de *LNCS*, pages 10–24. Springer.

MONGIN, P. (1994). Some connections between epistemic logic and the theory of nonadditive probability. *In* HUMPHREYS, P., éditeur : *Patrick Suppes : Scientific Philosopher. Vol.1 : Probability and Probabilistic Causality*, Synthese Library Vol. 234, pages 135–171. Springer.

NGUYEN, H. (2006). *An Introduction to Random Sets*. Chapman and Hall/CRC Press, Boca Raton, Florida.

NGUYEN, H. T. (1978). On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65:531–542.

NILSSON, N. J. (1993). Probabilistic logic revisited. *Artificial Intelligence*, 59:39–42.

PARIS, J. (1994). *The Uncertain Reasoner' Companion*. Cambridge University Press.

PARSONS, S. (2001). *Qualitative Approaches for Reasoning Under Uncertainty*. MIT Press.

PASQUIER, N., BASTIDE, Y., TAOUIL, R. et LAKHAL, L. (1999). Efficient mining of association rules using closed itemset lattices. *Inf. Syst.*, 24:25–46.

PAWLAK, Z. (1991). *Rough Sets. Theoretical Aspects of. Reasoning about Data*. Kluwer Acad. Publ., Dordrecht.

PAWLAK, Z. et SKOWRON, A. (2007a). 1. Rudiments of rough sets. *Inf. Sci.*, 177(1):3–27.

PAWLAK, Z. et SKOWRON, A. (2007b). 2. Rough sets : Some extensions. *Inf. Sci.*, 177(1):28–40.

PAWLAK, Z. et SKOWRON, A. (2007c). 3. Rough sets and Boolean reasoning. *Inf. Sci.*, 177(1): 41–73.

PEARCE, D. (2006). Equilibrium logic. *Annals of Mathematics and Artificial Intelligence*, 47:3–41.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufmann Publ.

PEARL, J. (1990). System Z : A natural ordering of defaults with tractable applications for default reasoning. *In Proc. of theoretical aspects of reasoning about knowledge*, pages 121–135.

PETIT-RENAUD, S. et DENŒUX, T. (2004). Nonparametric regression analysis of uncertain and imprecise data using belief functions. *Int. J. of Approximate Reasoning*, 35(1):1–28.

PICHON, F. et DENŒUX, T. (2010). The unnormalized Dempster's rule of combination : a new justification from the least commitment principle and some extensions. *Journal of Automated Reasoning*, 45(1):61–87.

PICHON, F., DENŒUX, T. et DUBOIS, D. (2012). Relevance and truthfulness in information correction and fusion. *Int. J. of Approximate Reasoning*, 53(2):159–175.

QUOST, B., DENŒUX, T. et MASSON, M.-H. (2007). Pairwise classifier combination using belief functions. *Pattern Recognition Letters*, 28(5):644–653.

QUOST, B., MASSON, M.-H. et DENŒUX, T. (2011). Ensemble clustering in the belief functions framework. *Int. J. of Approximate Reasoning*, 52(3):353–374.

RAMER, A. et KLIR, G. J. (1993). Measures of discord in the Dempster-Shafer theory. *Information Sci.*, 67:35–50.

RENOOIJ, S. et VAN DER GAAG, L. (1999). Enhancing QPNs for trade-off resolution. *In* LASKEY, K. B. et PRADE, H., éditeurs : *Proc. 15th Conf. on Uncertainty in Artificial Intelligence*

(UAI '99), Stockholm, July 30 - Aug. 1, pages 559–566. Morgan Kaufmann.

RENOOIJ, S. et VAN DER GAAG, L. C. (2008). Enhanced qualitative probabilistic networks for resolving trade-offs. *Artif. Intell.*, 172:1470–1494.

SCHWEIZER, B. et SKLAR, A. (1963). Associative functions and abstract semi-groups. *Publ. Math. Debrecen*, 10:69–180.

SHACKLE, G. L. S. (1961). *Decision, Order and Time in Human Affairs*. (2nd edition), Cambridge University Press, UK.

SHAFER, G. (1976). *A Mathematical Theory of Evidence*. Princeton Univ. Press.

SHAFER, G. (1978). Non-additive probabilities in the work of Bernoulli and Lambert. *Arch. for History of Exact Sciences*, 19 (4):309–370.

SHAFER, G. (1981). Constructive probability. *Synthese*, 48(1):1–60.

SHAFER, G. (1987). Probability judgment in artificial intelligence and expert systems. *Statistical Science*, 2(1):3–44.

SHAFER, G. et VOVK, V. (2001). *Probability and Finance : It's Only a Game !* Wiley, New York.

SHAPLEY, L. S. (1953). A value for n-person games. *In* KUHN, H. W. et TUCKER, A. W., éditeurs : *Contributions to the Theory of Games, volume II*, volume 28 de *Annals of Mathematical Studies series*, pages 307–317. Princeton University Press.

SHENOY, P. P. (1989). A valuation-based language for expert systems. *Int. J. of Approximate Reasoning*, 3:383–411.

SHORTLIFFE, E. H. (1976). *Computer-based Medical Consultations MYCIN*. Elsevier.

SMETS, P. (1981). The degree of belief in a fuzzy event. *Information Sci.*, 25:1–19.

SMETS, P. (1982). Possibilistic inference from statistical data. *In Proc. 2nd World Conf. on Mathematics at the Service of Man*, pages 611–613, Las Palmas.

SMETS, P. (1983). Information content of an evidence. *Int. J. of Man-Machine Studies*, 19:33–43.

SMETS, P. (1990a). The combination of evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458.

SMETS, P. (1990b). Constructing the pignistic probability function in a context of uncertainty. *In* HENRION, M., SHACHTER, R. D., KANAL, L. N. et LEMMER, J., éditeurs : *Uncertainty in Artificial Intelligence 5*, pages 29–39. Elsevier Science Publ.

SMETS, P. (1993). Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem. *Int. J. of Approximate Reasoning*, 9:1–35.

SMETS, P. (1995). The canonical decomposition of a weighted belief. *In Int. Joint Conf. on Artificial Intelligence*, pages 1896–1901, San Mateo, Ca. Morgan Kaufman.

SMETS, P. (2005a). Belief functions on real numbers. *Int. J. Approx. Reasoning*, 40:181–223.

SMETS, P. (2005b). Decision making in the TBM : the necessity of the pignistic transformation. *Int. J. of Approximate Reasoning*, 38:133–147.

SMETS, P. (2007). Analyzing the combination of conflicting belief functions. *Information Fusion*, 8(4):387–412.

SMETS, P. et KENNES, R. (1994). The transferable belief model. *Artificial Intelligence*, 66: 191–234.

SMITH, C. A. B. (1961). Consistency in statistical inference and decision. *J. of the Royal Statistical Society*, B-23:1–37.

SNOW, P. (1999). Diverse confidence levels in a probabilistic semantics for conditional logics. *Artificial Intelligence*, 113:269–279.

SPOHN, W. (1988). Ordinal conditional functions : a dynamic theory of epistemic states. *In* HARPER, W. L. et SKYRMS, B., éditeurs : *Causation in Decision, Belief Change, and Statistics*, volume 2, pages 105–134. Kluwer.

SPOHN, W. (2012). *The Laws of Belief : Ranking Theory and Its Philosophical Applications*. Oxford Univ. Press.

STRAT, T. M. (1984). Continuous belief functions for evidential reasoning. *In* BRACHMAN, R. J., éditeur : *Proc. National Conf. on Artificial Intelligence (AAAI'84), Austin, Aug. 6-10*, pages 308–313.

SUGENO, M. (1977). Fuzzy measures and fuzzy integrals - A survey. *In* GUPTA, M. M., SARIDIS, G. N. et GAINES, B. R., éditeurs : *Fuzzy Automata and Decision Processes*, pages 89–102. North Holland, Amsterdam.

TESSEM, B. (1993). Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61:315–329.

WALLEY, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.

WALLEY, P. et FINE, T. (1982). Towards a frequentist theory of upper and lower probability. *The Annals of Statistics*, 10:741–761.

XU, L., KRZYZAK, A. et SUEN, C. Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3):418–435.

YAGER, R. R. (1986). The entailment principle for Dempster-Shafer granules. *Int. J. of Intelligent Systems*, 1:247–262.

YAGER, R. R. et LIU, L. P. (Eds.), (2008). *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Springer Verlag, Heidelberg.

YEN, J. (1990). Generalizing the Dempster-Shafer theory to fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3):559–569.

ZADEH, L. A. (1965). Fuzzy sets. *Information Control*, 8:338–353.

ZADEH, L. A. (1968). Probability measures of fuzzy events. *J. Math. Anal. Appl.*, 23(2):421–427.

ZADEH, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, 8:199–249.

ZADEH, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28.

ZADEH, L. A. (1979). Fuzzy sets and information granularity. *In* M. M. GUPTA, R. K. R. et YAGER, R. R., éditeurs : *Advances in Fuzzy Sets Theory and Applications*, pages 3–18. North-Holland, Amsterdam.

ZOUHAL, L. M. et DENŒUX, T. (1998). An evidence-theoretic k-NN rule with parameter optimization. *IEEE Trans. on Systems, Man and Cybernetics C*, 28(2):263–271.

Index