Chapitre 2

Mesures de confiance

Dans ce chapitre, nous présentons différentes approches élaborées en vue de mesurer la confiance que l'on peut accorder à un fait. La théorie des probabilités, s'inscrit en effet maintenant au sein d'un éventail de théories dont la plupart sont moins restrictives. Par ailleurs les probabilités (issues dans la pratique de mesures statistiques) peuvent rendre compte d'un degré de vérité (un degré de prédiction) mais pas de l'incertitude. La théorie de Dempster-Shafer, issue d'une autre approche, rentre également dans ce cadre.

2.1. Mesures de confiance et théorie des possibilités

MESURE DE CONFIANCE

Afin de donner un cadre général aux différentes approches, la première définition de confiance dans un évenement, encore appelée mesure non additive ou mesure floue, ne comporte que trois axiomes. C'est la notion la plus générale définie sur un ensemble Ω d'éventualités. En général, nous parlerons d'ensembles finis auquel cas les «évenements» sont toutes les parties de Ω, mais au cas où Ω n'est pas fini, les évenements doivent former une tribu de parties, c'est à dire un ensemble de parties comprenant la partie pleine (Ω en totalité), stable par complémentation, intersections et unions finies ou dénombrables. Nous verrons plus loin une extension de cette définition à des sous-ensembles flous de l'univers Ω. On peut alors nommer mesure de confiance, toute fonction c d'une tribu de Ω vers [0, 1] ayant (au minimum) les propriétés $c(\emptyset) = 0$ et $c(\Omega) = 1$ et vérifiant :

$A \subset B \Rightarrow c(A) \leq c(B)$ et pour toute suite croissante ou décroissante :

$$A_1 \subset A_2 \subset A_3 \subset ..... \subset A_n..... \Rightarrow \lim c(A_n) = c(\lim(A_n)) \quad \text{(continuité)}$$

En conséquence on aura :

$$c(A \cup B) \geq \max(c(A), c(B)) \quad \text{et} \quad c(A \cap B) \leq \min(c(A), c(B))$$

INVERSE DE MÖBIUS D'UNE MESURE

Si c est une mesure sur un univers fini Ω pour chaque partie A de Ω, en désignant par |A| le cardinal de A, on pose m, l'inverse ou «masse» de Möbius de c, la fonction définie par :

$$m(A) = \sum_{B \subset A} (-1)^{|A - B|} c(B)$$

On aura alors $c(A) = \sum_{B \subset A} m(B)$, ce qui signifie que la confiance en A est la somme de toutes les masses de ses parties disjointes ou non.
Démonstration : pour une partie quelconque A, on a :

$$\sum_{B \subset A} m(B) = \sum_{B \subset A} \sum_{C \subset B} (-1)^{|B - C|} c(C) = \sum_{(C \subset B \subset A)} (-1)^{|B - C|} c(C)$$

La mesure c(C) est comptée positivement dans cette somme chaque fois qu'on peut trouver un sur-ensemble de C (dans un B lui même inclu dans A) formé en rajoutant un nombre pair d'éléments. Or, si la différence A - C comporte k éléments, le nombre de parties de A - C est 2^k dont la moitié sont des parties paires et l'autre moitié des parties impaires, il y a donc annulation du terme en c(C) dans la dernière expression sauf au cas où C = A où alors le seul B que l'on puisse intercaler entre C et A, est B = A, et donc $\sum_{B \subset A} m(B) = (-1)^0 c(A) = c(A)$.
A l'origine, le résultat de Möbius fut démontré pour une suite entière c qui s'exprime à partir de la suite m par $c(n) = \sum_{d \mid n} m(d)$, où | désigne la relation de divisibilité.
En posant μ(k) = [si (k = 1) alors 1 sinon si k se décompose en p nombres premiers distincts alors $(-1)^p$ sinon 0], m s'exprime à son tour par $m(n) = \sum_{d \mid n} \mu(n / d).c(d)$.

MESURE DE POSSIBILITÉ

C'est la solution la plus pessimiste ou la plus prudente [Zadeh 78], pour une confiance, à savoir que l'on prend comme définition supplémentaire :

$$ps(A \cup B) = \max(ps(A), ps(B))$$

donc en notant ¬A l'événement contraire, c'est à dire le complémentaire, et on a en particulier :

$$\max(ps(A), ps(\neg A)) = 1$$

MESURE DE NÉCESSITÉ

C'est une mesure de confiance vérifiant $nc(A \cap B) = \min(nc(a), nc(B))$ sans qu'il soit question d'aucune notion d'indépendance. Chaque fois qu'une mesure de possibilité ps est définie, alors la nouvelle mesure définie par : $nc(A) = 1 - ps(\neg A)$ est une nécessité, et réciproquement si nc est une mesure de nécessité, alors ps est définie par $ps(A) = 1 - nc(\neg A)$ qui est à son tour une possibilité. Ces deux notions sont donc duales.
Ayant l'une ou l'autre, on a donc $\max(ps(A), ps(\neg A)) = 1$, ce qui entraîne les deux propriétés :

$$ps(A) < 1 \Rightarrow nc(A) = 0 \qquad 0 < nc(A) \Rightarrow ps(A) = 1$$

Ces deux propriétés sont très importantes, car elles signifient que pour le couple (nc, ps) situé dans [0, 1], l'un est toujours à une extrémité de l'intervalle (0 ou 1) avec bien sûr nc ≤ ps, et que l'on n'a absolument pas tous les couples possibles de cet intervalle comme cela est le cas pour les mesures suivantes. La théorie des possibilités (et donc nécessités) a été initialement prévue pour tenir compte des informations incomplètes à propos de propositions booléennes c'est à dire soit vraie, soit fausse. Le degré de possibilité n'est donc en rien un degré de vérité mais plutôt un degré de préférence alors que la nécéssité mesurera plutôt une priorité.

MESURE DE CRÉDIBILITÉ (FONCTION DE CROYANCE)

C'est une mesure de confiance qui a la propriété supplémentaire :

$$g(A_1 \cup ... \cup A_n) \geq \Sigma g(A_i) - \Sigma_{i<j}\, g(A_i \cap A_j) + ... + (-1)^{n+1} g(A_1 \cap ... \cap A_n)$$

(sur-additivité générale ou encore appelée ∞-monotonie).
Alors si $A_1 = A$ et $A_2 = \neg A$ on a en particulier : $g(A) + g(\neg A) \leq g(\Omega) \leq 1$
Voir à ce sujet [Smets, Mamdani, Dubois, Prade 88].
Une crédibilité est par ailleurs une mesure de confiance dont la masse de Möbius est non-négative [Chateauneuf, Jaffray 89].

PLAUSIBILITÉ

C'est la notion duale, si cr est une crédibilité, alors $pl(A) = 1 - cr(\neg A)$ est une «plausibilité», mesure sous-additive. Nous verrons comment à partir d'une distribution de masse m, on peut définir pour tout $A \neq \emptyset$ la «croyance» en A (c'est une crédibilité) et la plausibilité par :

$$cr(A) = \Sigma_{B \subset A}\, m(B) \text{ et } pl(A) = \Sigma_{A \cap B \neq \emptyset}\, m(B)$$

On obtient $pl(A) + cr(\neg A) = 1 - m(\emptyset)$, donc 1 si m est normale.

MESURE DE PROBABILITÉ

On retrouve les probabilités comme mesure de confiance pr ayant l'axiome supplémentaire

$$A \cap B = \emptyset \Rightarrow pr(A \cup B) = pr(A) + pr(B)$$

C'est alors une crédibilité avec l'égalité au lieu d'une inégalité, soit la formule de Poincaré :

$$g(\cup_{1\leq i\leq n} A_i) = \Sigma_{I \subset \{1, .., n\}\ I \neq \emptyset} (-1)^{|I|+1} g(\cap_{j \in I} A_j)$$

Les probabilités, en accordant 0 à ∅, ne distinguent pas entre l'ignorance et la contradiction, qui cependant, dans le raisonnement de sens commun, ne recèlent pas la même quantité de savoir.

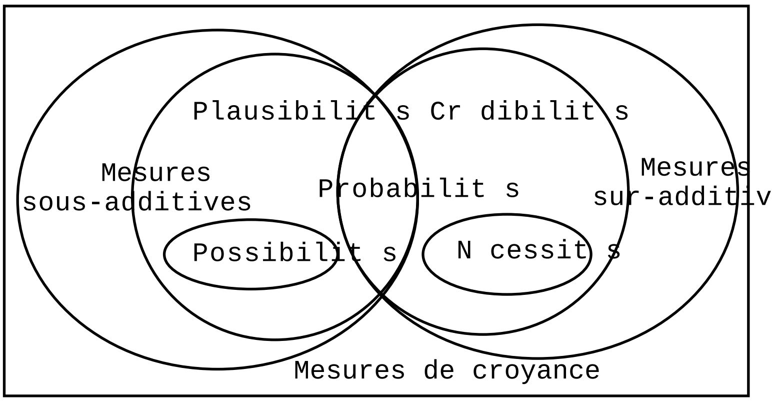

Figure 2.1 Illustration de la hiérarchie des types de mesures.

MESURES DE CROYANCE ET D'INCRÉDULITÉ, L'EXEMPLE DE MYCIN [BUCHANAN 84]

Ces mesures sont introduites empiriquement : MB (belief) qui est une possibilité et MD(A) = MB(¬A) (disbelief) qui est le complément à 1 d'une possibilité. Toute deux vérifiant respectivement max et min pour les disjonctions et conjonctions.

On a $0 < MD(A) \Rightarrow MB(A) = 1$ et $MB(A) < 1 \Rightarrow MD(A) = 1$

Le coefficient de certitude d'un fait est défini par : cf(A) = MB(A) - MD(A) qui est donc dans [-1, 1] alors cf(A) + cf(¬A) = 0, cf est une sorte de degré d'ignorance qui fait correspondre en particulier le vrai à 1, le faux à -1 et l'incertain à 0.

Si cf = 1 alors MB = 1 et MD = 0, et si cf = -1 c'est le contraire.

Il revient au même de parler de v = (1 + cf)/2 comme un degré de vérité, ou du coefficient λ = pr(évidence / A) / pr(évidence / ¬A) avec cf(A) = (λ - 1) / (λ + 1), bijection de [0, +∞[vers [-1, 1[.

Inférences : pour une règle le «cf» de l'ensemble des prémisses est le min des cf s'ils sont tous supérieurs à 0.2, le max s'ils sont tous inférieurs à -0.2 et 0 sinon, et le cf de la conclusion est le produit du cf des prémisses par le cf de la règle. Cette option min-produit est la plus courante dans les sytèmes experts.

Agrégation : si deux règles aboutissent à la même conclusion avec les coefficients x et y dans [-1, 1], alors :

cf(conclusion) = si x, y > 0 alors x + y - xy
(on reconnait la disjonction probabiliste)
si x, y < 0 alors x + y + xy
sinon (x+y) / (1 - min(|x|, |y|))

DANS LE SYSTÈME PROSPECTOR, connaissant les probabilités conditionnelles, on définit pour P ⇒ Q, un degré de «vraisemblance ou suffisance» ds = pr(P / Q) / pr(P / ¬Q), un «degré de nécessité» dn = pr(¬P / Q) / pr(¬P / ¬Q), et aussi un degré de «prédiction» v(Q / P) = pr(Q / P) / pr(¬Q / P). Enfin les «probabilités» a priori ne respectent pas la théorie des probabilités car Prospector utilisent le min pour la conjonction et le max pour la disjonction, et il définit en outre une «priorité» ou «cote» ou encore «degré de diagnostic» o(Q) = pr(Q) / pr(¬Q). Le degré de prédiction est infini si P ⇒ Q, égal à 1 s'il y a une indifférence supposée, et nul si P ⇒ ¬Q.

DISTRIBUTION DE POSSIBILITÉ, MESURE DE POSSIBILITÉ ET ENSEMBLE FLOU

Une distribution de possibilité sur un univers U est une application π de U dans l'intervalle [0, 1] qui est normalisée, c'est à dire vérifiant $\sup(\pi) = 1$.
Le rapport évident avec les possibilités sur U, est que toute distribution de possibilité permet de définir une mesure de possibilité pour les parties de U simplement par $ps(A) = \sup\{\pi(x) / x \in A\}$, de plus $nc(A) = \inf \{1 - \pi(x) / x \notin A\}$ définit la nécessité duale.

Réciproquement, une mesure de possibilité ps sur U permet de définir une distribution de possibilité $\pi(x) = ps(\{x\})$ pour $x \in U$. Une distribution de possibilité équivaut donc à la donnée d'un ensemble flou normalisé sur U.

REMARQUE : une distribution de possibilité n'a rien à voir avec une distribution de probabilité car la somme de π (somme discrète si U est fini, intégrale de π sur U sinon) n'est pas 1 en général.
Exemple, sur U = [0, 100] on pourrait définir «trentaine» par un intervalle flou A de noyau [30, 38] et de support [28, 42]. En ce cas, si X est une variable (linguistique telle que «âge») prenant ses valeurs dans U, «X est A», par exemple «âge est trentaine» est une proposition atomique.
On peut définir la restriction associée à X notée $ps(X = x / A) = \mu_A(x)$.
Alors si X = 29, ps(X est A) = 0.5 et si X = 39, ps(X est A) = 0.75.
Si A est normalisé, et c'est le cas de notre exemple, cela induit une mesure de possibilité conditionnelle pour toute partie B de U par :
$ps(B) = ps(X \in B / A) = \sup \{\mu_A(x) / x \in B\}$ qu'il faut lire comme la possibilité de l'évenement B, connaissant A.

Concrètement, les systèmes à base de connaissances sont confrontés au problème suivant, étant donné un ensemble flou (normalisé) A (ou donc, ce qui revient au même, à une distribution de possibilité π), et une variable X dont la valeur est elle-même imprécise, mal connue (connue seulement par un ensemble flou A'), comment mesurer la confiance dans le fait «X est A» ? On peut voir par exemple A comme un prédicat «grand» et X assigné à une valeur floue A' (ou une distribution de possibilité π'). On posera la définition de [Zadeh 78] :
$ps(X \text{ est } A) = ps(A / A') = \sup\{\min(\pi'(x), \mu_A(x)) / x \in U\}$ noté sup min $(\mu_{A'}, \mu_A)$ borne supérieure d'une fonction.
On lira «la possibilité d'avoir A sachant que X est A'».
Cette définition a pour conséquences (pour le même X) que $A \subset B \Rightarrow ps(A) \leq ps(B)$ et (max et min étant des lois binaires mutuellement distributives) que la possibilité de la disjonction est $ps(A \cup B) = \max(ps(A), ps(B))$.
Ce «ps» conditionné par le fait que X est assigné à A' est donc bien une mesure de possibilité sur U.

En résumé, si on a des représentations $\mu_{A'}$ pour X et un prédicat A, la possibilité que X soit A est donnée par la hauteur d'une intersection d'ensemble flous. On peut voir ps comme un degré d'intersection de A' avec A. Pour la nécessité que X soit A, on va utiliser le complément de la possibilité du contraire, soit :

$$nc(X \text{ est } A) = 1 - ps(X \text{ est } \neg A) = 1 - \sup \min(\mu_{A'}, 1 - \mu_A) = \inf \max(1 - \mu_{A'}, \mu_A)$$

Cette nécessité est aussi notée de façon conditionnelle nc (A / A'). C'est un degré d'inclusion de X (plus précisément de sa valeur floue A') dans A car :

«X satisfait entièrement A» $\Leftrightarrow$ nc(X est A) = 1 $\Leftrightarrow$ supp(A') $\subseteq$ supp(A)

$$\mathbf{ps(X\ est\ A) = sup\ min(\mu_{A'}, \mu_A) = ps(A, A')}$$
$$\mathbf{nc(X\ est\ A) = inf\ max(1 - \mu_{A'}, \mu_A)}$$

REMARQUE

[Dubois, Prade 85] et d'autres auteurs définissent la nécessité nc(A / A') au contraire par : 1 - ps(¬A' est A) = 1 - sup min(1 - $\mu_{A'}$, μ_A) = inf max($\mu_{A'}$, 1 - μ_A)
C'est alors un degré d'inclusion de A dans A'.
L'une ou l'autre de ces deux options sont constamment utilisées dans les travaux réalisés sur des domaines concrets et sont connues sous le nom de filtrage flou.
Dans les deux cas, dès que X est assigné à un A' non exact, on sort de la théorie des possibilités car on peut avoir 0 < nc < ps < 1 pour le fait «X est A» et non l'inverse comme ci-dessus.
Par contre, si X est exactement connu par une valeur précise, la seconde option vérifie bien les deux axiomes 0 < nc $\Rightarrow$ ps = 1 et ps < 1 $\Rightarrow$ nc = 0 et donc cette deuxième définition est plus conforme à la théorie des possibilités. C'est cependant, la plupart du temps, la première option qui est adoptée dans les systèmes concrets, c'est pourquoi nous l'avons présenté en premier lieu.
Pour cette première définition, si X a une valeur précise, on vérifiera que ps = nc redonne un degré de vérité.
Cette remarque montre qu'il est plus général d'envisager des systèmes où la confiance accordée à un fait est un couple (x, y) pouvant prendre toutes les valeurs en accord avec $0 \le x \le y \le 1$, c'est le cas des crédibilités-plausibilités.
Une autre notion est la compatibilité définie elle-même comme un ensemble flou :
$\mu_{cp\ (X,\ A)}(y)$ = si $\mu_X^{-1}(y) = 0$ alors 0 sinon sup{μA(x) / y = $\mu_X(x)$} [Zadeh 79].
Une «véritable logique floue» n'aurait pas seulement la présence de prédicats vagues, mais les valeurs de vérité seraient elles mêmes des ensembles flous dans [0, 1]. Pour un degré de vérité perçu comme un ensemble flou, voir aussi [Godo 89].

VARIABLE LINGUISTIQUE

En logique floue les propositions sont structurées à partir des propositions atomiques du type «X est A» où la sémantique leur donne la valeur $\mu_A(x)$ pour chaque valeur précise de la variable X. B.Bouchon a introduit la notion de variable linguistique comme un triplet (X, E, P) où X est une varable prenant ses valeurs dans l'univers U, et P une famille de prédicats flous sur U (petit, grand, ...). Les sytème de contrôle flou du chapitre 4 mettent en oeuvre par exemple trois variables linguistiques s'ils ont deux entrées et une sortie.
La variable linguistique X en elle-même est un attribut tel que âge, température, etc recevant des valeurs : jeune, vieux, chaud, ... définies par des sous-ensembles flous.

EXTENSION DE LA NOTION DE MESURE FLOUE

Soit F(Ω) l'ensemble de toutes les parties floues de Ω. Une mesure (de confiance) floue peut être définie comme une application c de F(Ω) vers [0, 1], croissante au regard de l'inclusion et vérifiant c(∅) = 0. Il est toujours possible de se ramener à c(Ω) = 1 si c est bornée. Par exemple, la hauteur, le cardinal, l'intégrale ou même le cardinal du support peuvent constituer des mesures floues définies pour chaque sous-ensemble flou de Ω.

Dans le but de comparer des sous-ensembles flous [Rifqi 95] introduit :

Différence sur $F(\Omega)$: une opération de différence doit vérifier pour toutes parties floues A et B :

$A \subset B \Rightarrow A - B = \emptyset$, $A - B = A - (A \cap B)$, $(B \subset B') \Rightarrow (B - A \subset B' - A)$

La différence entre ensembles exacts est généralisée, et par exemple les deux définitions $\mu_{A-B} = \max(0, \mu_A - \mu_B)$ et μ_{A-B} = [si ($\mu_B = 0$) alors μ_A sinon 0], peuvent convenir, mais pas celle de $\min(\mu_A, 1 - \mu_B)$.

Mesure de comparaison s entre ensembles flous : une mesure de comparaison s est définie pour toutes parties floues A, B par $s(A, B) = f(c(A \cap B), c(B - A), c(A - B))$ où f est une fonction de $[0, 1]^3$ dans [0, 1] croissante au sens large pour son premier argument. Nous avons un cadre général pour définir des similitudes, et la définition usuelle de «possibilité» sup min (μ_A, μ_B) en est un exemple.

Mais ce qui est intéressant, est que l'on peut définir une «satisfiabilité» (A satisfait B) dès lors que f ne dépend pas de son troisième argument et que de plus $f(0, v, w) = 0$, et $f(u, 0, w) = 1$ pour $u \neq 0$.

On peut également définir une mesure d'inclusion dès lors que f est indépendante de son second argument et que $f(0, v, w) = 0$ et $f(1, v, 0) = 1$ de sorte que $s(A, A) = 1$.

Ainsi par exemple inclu $(A, B) = |A \cap B| / |A|$ convient à ces contraintes ou bien : inclu $(A, B) = \inf(\min(1 - \mu_A + \mu_B, 1))$.

A noter encore l'exemple de $s(A, B) = (2/\pi) \operatorname{Arctan}[(c(A \cap B) / c(A - B)]$ qui est décroissante en $c(A - B)$ et qui ne vaut 1 que si A = B.

Une dissimilarité sera au contraire une mesure déterminée par f indépendante de son premier argument, croissante au sens large vis-à-vis de ses autres arguments et vérifiant $f(u, 0, 0) = 0$, par exemple $ds(A, B) = \sup |\mu_A - \mu_B|$.

2.2. La théorie de Dempster-Shafer

Cette théorie [Shafer 76] se fonde sur un partage de l'univers de référence Ω dans lequel on considère une famille F de parties (un corps d'évidence normal, des événements «focaux») dont on est sûr d'une «évidence» m (ce n'est pas une mesure de confiance définie plus haut).

DISTRIBUTION DE MASSE

C'est une fonction m de $P(\Omega)$ dans [0, 1] telle que $[m(A) > 0 \Leftrightarrow A \in F]$ et telle que l'on ait la somme $\sum_{A \subset \Omega} m(A) = 1$

Il faut bien noter que m ne vérifie pas la croissance vis à vis de l'inclusion.

L'ignorance totale serait traduite par une masse vérifiant $m(\Omega) = 1$ et $m(A) = 0$ pour toute partie A non pleine de Ω. Si par exemple, on ne considère que deux éventualités $\Omega = \{a, b\}$ avec $m(\{a\})$ et $m(\{b\})$, alors $m(\{a, b\})$ symbolise l'indifférence à a ou b, et $m(\emptyset)$ symbolise la part de ce qui pourrait être autre chose que a ou b. On ne requiert donc pas $m(\emptyset) = 0$, mais si c'est le cas, m est dite normale ou complète, et alors, on pourra montrer qu'une telle masse définit une famille de probabilités.

Exemple, en termes d'évenements, avec deux évenements A et B, si les seuls éléments focaux sont B de masse q et $\neg B$ de masse 1 - q, cette masse représente toutes les probabilités vérifiant pr(A et B) = x, pr(A et ¬B) = y, pr(¬A et B) = z, pr(¬A et ¬B) = t avec x + z = q et y + t = 1 - q et donc x + y + z + t = 1.

COMPLÉMENT D'UNE MASSE Si m est définie sur E par $m(A_i) = m_i$, les éléments focaux du complément m' sont les compléments des A_i et $m'(\neg A_i) = m_i$.

RELATION DE RESTRICTION ENTRE DEUX MASSES

On dit que m' est une restriction de m si, sur un élément focal F de m, il existe deux parties distinctes A, B de F, focales pour m et telles que m'(A) = m(A) + x, m'(B) = m(B) + y, m'(F) = m(F) - x - y, m' étant confondue avec m par ailleurs.
En définissant m' ≤ m comme la clôture transitive de cette relation, on peut montrer que cela signifie que la famille des probabilités issues de m' est inclue dans celle issue de m [Baldwin, Martin 94].
Deux masses m_1 et m_2 sont dites orthogonales (incomparables et noté $m_1 \perp m_2$) si, ni m_1 n'est restriction de m_2, ni l'inverse.

MASSE ET PROBABILITÉS ASSOCIÉES À UN ENSEMBLE FLOU

Etant donné un ensemble flou discret défini par les valeurs $\mu_i = \mu(x_i)$ que l'on va supposer ordonnées de façon décroissante :
$1 \geq \mu_1 \geq \mu_2 \geq \mu_3 \geq ... \geq \mu_n$ en posant alors $m(\varnothing) = 1 - \mu_1$ et $m(\{x_1, ... , x_i\}) = \mu_i - \mu_{i+1}$ convenant que $\mu_{n+1} = 0$, on définit une masse sur l'ensemble des parties décrites.
Si l'ensemble flou est normalisé, alors la masse est normale, sinon une normalisation est toujours possible en remplaçant μ par $\mu^* = \mu / \mu_1$.
Exemple, si on dispose de l'ensemble flou défini par : $\mu(a) = 1$, $\mu(b) = 0.7$, $\mu(c) = 0.5$, $\mu(d) = 0.1$, on aura m{a} = 0.3, m{a, b} = 0.2, m{a, b, c} = 0.4 et m{a, b, c, d} = 0.1.
Toutes les probabilités pr associées à cette masse vérifieront 0.3 ≤ pr(a) ≤ 1, 0 ≤ pr(b) ≤ 0.7, 0 ≤ pr(c) ≤ 0.5, 0 ≤ pr(d) ≤ 0.1 car pr(a) + pr(b) + pr(c) + pr(d) = 1, enfin on a pr(b) + pr(c) + pr(d) ≤ 0.7, pr(c) + pr(d) ≤ 0.5 et pr(d) ≤ 0.1.

BORNE INFÉRIEURE DE DEUX MASSES

Pour deux masse m_1 et m_2, la borne inférieure $m_1 \wedge m_2$ est la masse associée à l'intersection des deux familles de probabilités. Si les éléments focaux de m_1 sont les M_i et et ceux de m_2 les Nj pour un nombre fini d'indices i et j, alors les éléments focaux de $m_1 \wedge m_2$ sont toutes les intersections $L_{ij} = M_i \cap N_j$ avec les masses l_{ij} de telle sorte que $\Sigma_j l_{ij} = m_1(M_i)$ et $\Sigma_i l_{ij} = m_2(N_j)$. En ce cas si L est un élément focal de $m_1 \wedge m_2$, il peut avoir été obtenu de plusieurs façons et :
$m_1 \wedge m_2 (L) = \Sigma_{Lij = L}\, m(L_{ij})$.
Exemple 1 : soient deux masses m et m' issues des ensembles flous μ défini par {a/1, b/0.8, c/0.3} et μ' défini par {a/1, b/0.5, c/0.4}, les deux masses m et m' ont ainsi leurs valeurs 0.2, 0.5, 0.3 et 0.5, 0.1, 0.4 respectivement sur {a}, {a, b}, {a, b, c} comme indiqués à l'extérieur du tableau de gauche. Pour calculer la distribution de masse de $m_1 \wedge m_2$ on commence par l'élément focal {a, b, c} qui ne peut recevoir que la valeur 0.3 (dernière case en bas à droite). Puis, pour {a, b}, ce sont les 3 cases encadrant la précédente. En essayant de maximiser le nombre de zéros, à partir des sommes marginales, on peut remplir le tableau et donc donner 0.2 pour {a, b} et enfin 0.5 pour les 5 cases restantes correspondant à {a}.

De $m_1 \wedge m_2$, on peut déduire ici un ensemble flou {a/1, b/0.5, c/0.3} qui n'est autre que $\mu \cap \mu'$.

	{a}	{a, b}	{a, b, c}		
{a}	0.2	0	0	0.2	
{a, b}	0.3	0.1	0.1	0.5	m
{a, b, c}	0	0	0.3	0.3	
	0.5	0.1	0.4		
		m'			

	{a}	{a, c}	{a, b, c}		
{a}	0.2	0	0	0.2	
{a, b}	0.1	0.2	0.2	0.5	m
{a, b, c}	0	0.3	0	0.3	
	0.3	0.5	0.2		
		m'			

Exemple 2 : Soient les deux ensembles flous μ et μ' définis sur {a, b, c} par {a/1, b/0.8, c/0.3} et {a/1, b/0.2, c/0.7}.
Cette fois les masses m et m' qui en sont déduites et dont les valeurs figurent comme sommes marginales du tableau de droite sont m{a} = 0.2, m{a, b} = 0.5, m{a, b, c} = 0.3, et m'{a} = 0.3, m'{a, c} = 0.5, m'{a, b, c} = 0.2.

Les valeurs de $m_1 \wedge m_2$ sont calculées à partir du tableau et sont {a, b, c} 0, {a, b} 0.2, {a, c} 0.3, et sur {a} les 6 autres cases soit 0.5, mais cette masse ne correspond à aucun ensemble flou. En revanche la solution m_0{a} = 0.7, m_0{a, c} = 0.1, m_0{a, b, c} = 0.2 convient aux sommes et correspond à l'ensemble flou {a/1, c/0.3, b/0.2}, qui est $\mu \cap \mu'$, mais $m_0 < m_1 \wedge m_2$.

CONFIANCE ACCORDÉE AU FAIT «X EST A»

Lorsque A est un ensemble flou (un prédicat tel que «grand» ...) et X une variable (telle que «untel») assignée à un ensemble flou A' (une donnée imprécise sur sa taille par exemple), une solution étudiée dans [Baldwin, Martin 94], [Baldwin, Lawry, Martin 96] est de calculer la crédibilité et la plausibilité de «X est A» à partir d'une masse définie sur l'ensemble L = {f, i, t} des trois valeurs de vérité faux, indéterminé et vrai.
Supposons X à valeurs dans l'univers U où A et A' conduisent aux distributions de masse m et m' dont les éléments focaux sont respectivement notés M_i et N_j. On construit la matrice à valeurs dans {t, i, f} en donnant la valeur vrai si $N_j \subseteq M_i$, faux si $N_j \cap M_i = \varnothing$ et indéterminé dans les autres cas.

	{b}	{b, c}	{a, b, c}	{a, b, c, d}		
{a}	f 0.09	f 0.15	i 0.03	i 0.03	0.3	
{a, b}	t 0.15	i 0.25	i 0.05	i 0.05	0.5	m pour A
{a, b, c}	t 0.06	t 0.1	t 0.02	i 0.02	0.2	
	0.3	0.5	0.1	0.1		
		valeurs de m' pour A'				

La crédibilité de «X est A» lorsque X est assigné à A', sera la masse totale attribuée à l'élément t, soit ici 0.33, la plausibilité de «X est A» sera 1 - cr(X est ¬A) = 1 - masse (f) = 1 - 0.24 = 0.76. La masse 0.43 de i est ce qui reste, elle mesure l'écart entre la plausibilité 0.76 et la crédibilité 0.33. Si A n'est pas normalisé, le calcul est conduit de la même façon, mais avec des $M_i \neq \varnothing$.

Remarque, en prenant les définitions «possibilistes» vues plus haut, pour le même exemple on obtient : ps(X est A) = sup min (μ_A, $\mu_{A'}$) = 0.7 (valeur voisine de 0.76), et nc(X est A) = 1 - ps(X est ¬A) = 0.3 qui est aussi voisine de 0.33, mais quel sens donner à ce couple ?

LIEN AVEC LES PROBABILITÉS

Pour la même proposition «X est A» où X est connu par un ensemble flou A' qu'il faut donc confronter à l'ensemble flou A, en posant :

$pr(X \text{ est } A) = pr(A / A') = \sum m_i.n_j.|M_i \cap N_j| / |N_j|$

La fraction figurant dans chaque terme est donc égale à 1 pour les $N_j \subseteq M_i$, donc cr ≤ pr, elle est nulle pour ceux qui satisfont $N_j \cap M_i = \emptyset$, pour les autres il s'agit d'une fraction de la masse de l'élément i, et on a bien pr ≤ pl, ce calcul donnant bien une probabilité compatible avec la crédibilité.

Dans le cas de l'exemple ci-dessus pr(X est A) = 0 + 0 + 0.03 / 3 + 0.03 / 4 + 0.15 + 0.25 / 2 + 0.05 (2 / 3) + 0.5 / 2 + 0.06 + 0.1 + 0.02 + 0.02 (3 / 4) = 0.53908

PLAUSIBILITÉ ET CRÉDIBILITÉ ASSOCIÉES À UNE MASSE

Pour tout $A \neq \emptyset$, on définit la «croyance» en A (c'est une crédibilité) par :

$cr(A) = \sum_{B \subseteq A} m(B)$ c'est naturellement la part d'évidence que contient A, c'est une crédibilité. En termes de logique ce serait $cr(q) = \sum_{p \Rightarrow q} m(p)$

On définit également :

$pl(A) = \sum_{A \cap B \neq \emptyset} m(B)$ donne l'ensemble des pièces d'évidence rendant une occurrence de A possible, on peut l'exprimer. $pl(q) = \sum\{m(p) / p \text{ n'entraine pas } \neg q\}$ et c'est une plausibilité.

$$\boldsymbol{cr(A) = \sum_{B \subseteq A} m(B) \quad \text{et} \quad pl(A) = \sum_{A \cap B \neq \emptyset} m(B)}$$

On obtient $pl(A) + cr(\neg A) = 1 - m(\emptyset)$, donc 1 si m est normale.

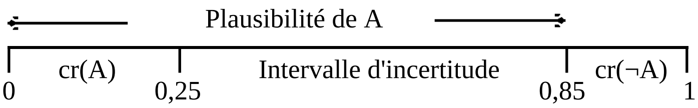

Figure 2.2 Illustration d'un ecxemple où cr(A) = 0.25 et cr(¬A) = 1.15.

Exercice 2.1

Montrer la sur-additivité $cr(A \text{ ou } B) \geq cr(A) + cr(B) - cr(A \text{ et } B)$ (faire un dessin) et la sub-additivité $pl(A \text{ et } B) \leq pl(A) + pl(B) - pl(A \text{ ou } B)$.

Ainsi, il est possible d'avoir $cr(A) = cr(\neg A) = 0$ et $pl(A) = pl(\neg A) = 1$, de plus on a clairement les inégalités : $0 \leq cr(A) \leq pl(A) \leq 1$

L'intervalle [cr(A), pl(A)] peut être vu comme l'encadrement d'une probabilité mal connue, c'est pourquoi ces bornes sont souvent appelées probabilités basse et haute.

Chaque fait focal $m(A) \neq 0$, est incompatible avec tout fait qu'il n'entraine pas.

Inversement, si cr est une crédibilité, et si on pose $m(A) = \sum_{B \subseteq A} (-1)^{card(A-B)} cr(B)$, on montre que cr est bien la crédibilité définie à partir de cette masse m.

Théorème Si cr = pl alors cette mesure commune est une probabilité

En effet si B est focal et coupe A, alors B ⊂ A d'après les définitions de cr et pl. On suppose alors que les événements ne sont pas toutes les parties de E mais une tribu quelconque, la sur-additivité et la sous-additivité entraînent la propriété supplémentaire des probabilités. On peut même montrer que ceci est équivalent au fait que tous les faits focaux sont «premiers» c'est à dire entraînés par rien d'autre qu'eux-mêmes et le faux.

Théorème Une crédibilité sur un ensemble fini, est une probabilité si et seulement si les éléments focaux sont seulement des singletons.

Si cr est une probabilité, alors $cr(\{x_1, x_2,... ,x_n\})= \sum cr(\{x_i\}) = \sum m(x_i)$ et d'après la définition de cr, toute autre partie qu'un singleton, inclue dans A ne peut avoir qu'une masse nulle.
Inversement $\sum m(x) = 1$, alors si A, B parties disjointes,
$cr(A) + cr(B) = \sum_{x \in A} m(x) + \sum_{x \in B} m(x) = \sum_{x \in A \cup B} m(x) = cr(A \cup B)$
C'est une probabilité.

Autre cas particulier avec m normale :

Théorème Si les propositions focales sont totalement ordonnées par l'implication ou, ce qui revient au même, que les événements focaux sont totalement ordonnés par l'inclusion, alors cr est une mesure de nécessité, et pl une possibilité.

Si F est le plus grand élément focal de A et G celui de B, on peut supposer F ⊂ G et donc F ⊂ A ∩ B. L'intersection A ∩ B n'en a donc pas de plus grand et :

$$cr(A \cap B) = cr(A) = \sum_{H \subset F} m(H) = cr(A) = \min (cr(A), cr(B))$$

Réciproquement supposons que pour toutes parties A et B, on ait $cr(A \cap B) = \min (cr(A), cr(B))$. Alors, pour deux éléments focaux F et G quelconques la propriété vaut, F et G ne peuvent donc être disjoints car m est normale. Si F et G se coupent donc, en supposant $cr(F) \leq cr(G)$, $cr(F \cap G) = cr(F) = \sum_{H \subset F \cap G} m(H) = \sum_{H \subset F} m(H)$ signifie que les éléments focaux H de F sont exactement les mêmes que ceux contenus à la fois dans F et dans G, en particulier pour H = F, on a F ⊂ G.

PROBABILITÉ PIGNISTIQUE

Si m est une masse sur Ω fini et si |F| désigne le cardinal de F, elle est définie comme la distribution $pp(x) = \sum_{x \in F \text{ focal}} m(F).|F|$, [Smets 82, Smets 90] on vérifie que la somme totale égale à 1.
On déduit donc pour une partie quelconque A de Ω que :

$$pp(A) = \sum_{F \cap A \neq \varnothing} m(F).|F \cap A| / |F| \leq pl(A)$$

On peut montrer qu'il s'agit de la moyenne de toutes les probabilités qui sont inférieures à cette plausibilité pl définie par m.

EXTENSION DE LA NOTION DE MASSE À DES ÉLÉMENTS FOCAUX FLOUS [Yager 95]

Si m est définie comme un poids non nul accordé à un nombre fini de sous-ensembles flous normalisés (dits focaux) de Ω, et si la masse totale est 1, il est possible de poser $cr(A) = \sum_{B \text{ focal}} m(B).Inclu(B, A)$ et $pl(A) = \sum_{B \text{ focal}} m(B).Inter(B, A)$ où les relations floues d'inclusion et d'intersection sont définies comme d'habitude par $Inter(B, A) = \sup(\min(\mu_A, \mu_B))$ et (en notant $\neg A$ le complémentaire flou de A), et $Inclu(B, A) = 1 - Inter(B, \neg A) = \inf(\max(\mu_A, 1 - \mu_B))$.

RÈGLE DE DEMPSTER

Si plusieurs sources d'informations sont disponibles pour une proposition P, par des mesures de crédibilité et plausibilité, chaque source fournit une pondération probabiliste, soit deux «masses» m_1 et m_2. La règle de Dempster (1968) encore appelée somme orthogonale que nous noterons par le symbole $\oplus$, généralise Bayes (1763) en donnant une combinaison commutative et associative de m_1 et m_2 :

$$m_1 \oplus m_2 (P) = \frac{\sum \{m_1(Q)*m_2(R) \ / \ Q \wedge R = P\}}{1 - \sum \{m_1(Q)*m_2(R) \ / \ Q \wedge R = \varnothing\}}$$

Le principe de cette opération est de considérer, pour P, la somme des masses pour les couples Q et R dont la conjonction est P, c'est à dire :
$\alpha\Sigma\{m_1(Q)m_2(R) \ / \ Q \wedge R = P\}$ où $\alpha = 1 / (1 - c)$ est un coefficient normalisateur mesurant de 1 à l'infini le conflit.
$c = \Sigma \{m_1(Q)m_2(R) \ / \ Q, R \text{ incompatibles}\}$ entre 0 et 1, est appelé le degré de conflit entre m_1 et m_2. Lorsque c devient proche de 1, la normalisation opérée en divisant par 1-c devient évidemment discutable, lorsqu'il est égal à 1, la combinaison n'est pas définie.

Exercice 2.2
En posant $(Q(m))(A) = \Sigma_{A \subseteq C}\, m(C)$ appelée «commonalité» de m, montrer un résultat inverse donnant la crédibilité $cr(A) = \Sigma_{A \cap B = \varnothing} (-1)^{|B|} Q(B)$ ainsi que :
$Q(m_1 \oplus m_2) = \alpha Q(m_1).Q(m_2)$.

Pour un ensemble de propositions $\{0, P, \neg P, 1\}$, si $m(P) = L$ et $m(\neg P) = 1-U$, une partie de la masse n'est pas attribuée, c'est $1 - \sum m(q) = 1 - [L+ (1-U)] = U - L$
L'application de la règle à partir de deux couples (L_1, U_1) et (L_2, U_2) c'est à dire $m_1(P) = L_1$ et $m_2(P) = L_2$, $m_1(\neg P) = 1-U_1$, $m_2(\neg P) = 1-U_2$ (L pour «low» et U pour «up») peut recevoir plusieurs légitimations :
En fait le point de vue probabiliste est le suivant, pour chaque source d'information nous avons trois événements, en premier lieu P connu avec au moins la probabilité L, puis $\neg P$ avec au moins 1-U, et enfin «l'ignorance sur P» que nous notons ?P dont la probabilité est le complément à 1 soit U-L. En confrontant deux sources nous avons le tableau ci-dessous tel que la somme des lignes et des colonnes donne ces trois valeurs pour P_1 et P_2.

On considère alors que si P_1 et P_2 sont les deux conclusions pour P relatives aux sources m_1, m_2, P est obtenu avec dans le numérateur de la formule, P_1 et P_2, P_1 et Ignorance sur P_2, et enfin, Ignorance sur P_1 et P_2.
Les situations de conflit sont les cases grisées dont la probabilité est degré de conflit

$$c = L_1(1-U_2) + L_2(1-U_1).$$

	Preuve 1	P	¬P	?P
Preuve 2		L1	1 - U1	U1 - L1
P	L2	L1*L2	L2(1 - U2)	L2(U1 - L1)
¬P	1 - U2	L1.(1 - U2)	(1 - U1).(1 - U2)	(1 - U2).(U1 - L1)
?P	U2 - L2	L1.(U2 - L2)	(1 - U1).(U2 - L2)	(U1 - L1).(U2 - L2)

Figure 2.3 Confrontation de deux preuves de P établies avec les supports respectifs (L_1, U_1) et (L_2, U_2)

La probabilité basse de P est obtenue par l'événement «au moins une fois P» (cases entourées en gras) qui est donc :

$L' = [L_1*L_2 + L_1*(U_2-L_2) + (U_1-L_1)*L_2] = L_1*U_2 + L_2*U_1 - L_1*L_2$

$= L_1 + L_2 - L_1*L_2 - c$ mais en divisant par 1-c (les situations acceptables), on fait une normalisation soit :

$$L = (L_1 + L_2 - L_1L_2 - c) / (1 - c)$$

De même la probabilité basse de ¬P est (zones blanches) :

$$L(\neg P) = [(1-U_1)(1-U_2) + (1-U_1)(U_2-L_2) + (U_1-L_1)(1-U_2)] = 1 - c - U_1*U_2$$

puis, en normalisant (en divisant par par 1-c) et en prenant le complément à 1, on a

$$U = U_1U_2 / (1 - c)$$

On vérifiera que la zone d'incertitude (pointillée) $(U_1 - L_1)(U_2 - L_2)$ normalisée par (1 - c) est bien quantifiée par U - L.
En définitive, cette opération d'agrégation dans $S = \{(x, y) / 0 \le x \le y \le 1\}$ s'exprime par :

$$\boxed{(x, y) \oplus (x', y') = \left(\frac{x + x' - xx' - c}{1 - c}, \frac{yy'}{1 - c}\right) = \left(\frac{xy' + x'y - xx'}{1 - x - x' + xy' + x'y}, \frac{yy'}{1 - x - x' + xy' + x'y}\right)}$$

Cette opération est commutative, associative, mais non idempotente.
Le vrai (1, 1) et le faux (0, 0) sont absorbants, l'incertain (0, 1) est neutre, et elle est partout définie sauf pour le couple de supports le plus antagoniste (0, 0) avec (1, 1).
Plus c est grand, plus cette règle devient comme on l'a déjà dit, discutable. Ces formules d'agrégation ont aussi été employées dans Mycin (voir aussi l'annexe langage FRIL).
Pour un point de vue algébrique, voir [Gacôgne 93] et surtout [Daniel 94]. En cherchant les fonctions # de [0, 1] dans lui-même respectant les trois contraintes : deux points fixes $f(1) = 1$ et $f(0) = 0$, (1/2, 1/2) centre de symétrie (on souhaite l'auto-dualité $\#\neg x = \neg\# x$) et $f'(x) = 0$ pour x = 0 ou 1 (T et F sont atteints plus vite par #P que par P), la fonction polynômiale de plus petit degré est $\#(x) = x^2(3 - 2x)$ et la fonction rationnelle de plus petits degrés étant $\#x = x^2 / (2x^2 - 2x + 1)$.

C'est cette dernière qui, étant choisie comme «connecteur de renforcement», se généralise naturellement à $([0, 1]^2)^2$ sous l'opération de Dempster. D'autres renforcements sont possibles [Pearl 90], ainsi $\#x = 1 - (1 - x^a)^a$, mais ils ne sont pas auto-duaux.

EXEMPLE [Klir 88]

Face à un tableau, on pose 3 questions R = peinture de Raphaël, D = oeuvre de l'un de ses disciples, C = contrefaçon, deux experts donnent leurs «poids» à 7 événements. On calcule c = 0,03 et m est le résultat de l'agrégation de m1, m2.

	m1	cr1	m2	cr2	m	cr
R	0.05	0.05	0.15	0.15	0.21	0.21
D	0	0	0	0	0.01	0.01
C	0.05	0.05	0.05	0.05	0.09	0.09
R ou D	0.15	0.20	0.05	0.20	0.12	0.34
R ou C	0.1	1	0.5	1	0.31	1
R ou D ou C	0.6	1	0.5	1	0.31	1

APPLICATION EN RECHERCHE PÉTROLIÈRE [Aminzadeh 94]

Dans ce système expert, plusieurs connaissances d'ordre géochimiques ou géologiques sont formalisées en vue d'arriver à déterminer un type de saturation dans une couche géologique. Les règles utilisent des prédicats flous comme «plutôt supérieur à 2» ou «plutôt inférieur à 1». L'intérêt de cette application réside dans le fait que plusieurs conclusions s'excluent mutuellement (autres qu'une connaissance et son contraire) bien qu'il soit difficile de séparer certaines conclusions (eau ou pétrole...).
Par exemple, si les règles de la géologie concluent à «gaz ou pétrole» avec une probabilité minimale de 0.4, une première agrégation avec les conclusions de la géochimie va se présenter :

Géol. \ Géochimie	Eau 0.5	Pétrole 0.3	Indétermination 0.2
Gaz ou Pétrole 0.4	Conflit 0.2	Pétrole 0.12	Gaz ou Pétr. 0.08
Indét. 0.6	Eau 0.3	Pétrole 0.18	Indéterminé 0.12

La normalisation s'opère en multipliant par 1 / (1 - c) = 1.25 ce qui donne la première colonne du tableau ci-dessous (par exemple le pétole obtient 1.25*(0.12 + 0.18)). A présent une agrégation est effectuée avec les conclusions d'une autre série de règles issues de la géophysique (première ligne) :

Précéd. \ Géoph.	Gaz 0.03	Eau ou Pétrole 0.4	Indéterminé 0.3
Gaz ou Pétrole 0.1	Gaz 0.03	Pétrole 0.04	Gaz ou Pétrole 0.03
Eau 0.375	Conflit 0.1125	Eau 0.15	Eau 0.1125
Pétrole 0.375	Conflit 0.1125	Pétrole 0.15	Pétole 0.1125
Indéterminé 0.15	Gaz 0.045	Eau ou Pétrole 0.06	Indéterminé 0.045

La normalisation s'effectue en multipliant par 1 / (1 - 0.1125 - 0.1125) = 1.29, ce qui donne les crédibilités 0.097 pour le Gaz, 0.039 pour Gaz ou Pétrole, 0.339 pour l'eau, 0.077 pour Eau ou Pétrole, 0.39 pour le pétrole seul et enfin 0.058 pour l'indétermination.
Pour les trois conclusions intéressantes, on a donc :
pl (gaz) = 1 - cr(¬gaz) = 1 - cr(eau) - cr(eau ou pétrole) - cr(pétrole)
= 1 - 0.339 - 0.077 - 0.39 = 0.194 et pl(pétrole) = 1 - cr(gaz) - cr(eau)
= 1 - 0.097 - 0.339 = 0.564
D'où les supports (L - U) pour le gaz (0.097 - 0.194), pour l'eau (0.339 - 0.474) et le pétrole (0.39 - 0.564) qui est donc le plus probable.

UN POINT DE VUE UNIFICATEUR ENTRE 3 TYPES DE THÉORIE

On peut représenter la logique booléenne par les deux points T(vrai) et F (faux), les logiques multivaluées par le segment FT, la théorie des possibilité par l'ensemble des couples des segments FI et IT, (en notant I l'indéterminé) et enfin la théorie des probabilités haute et basse (notées ici obligation et éventualité) par la zone hachurée. La négation étant dans tous les cas la symétrie par rapport à IM.

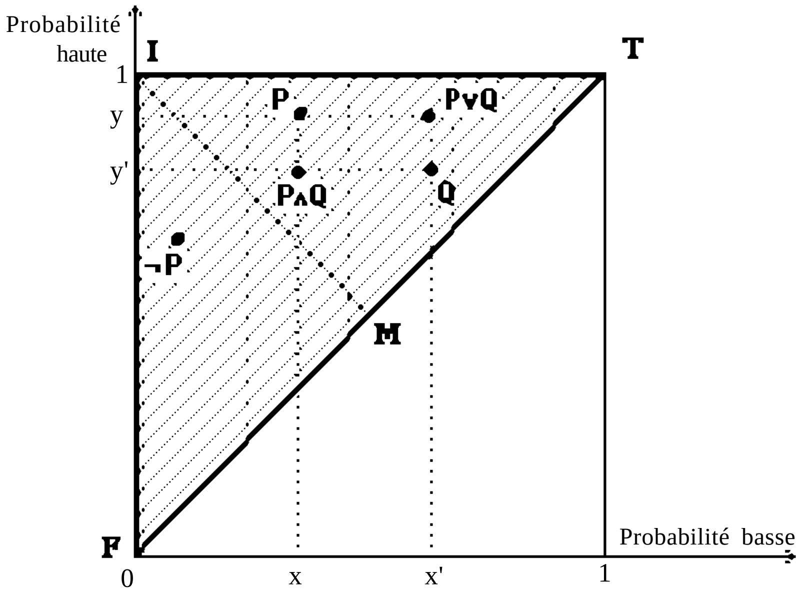

Figure 2.4 L'ensemble $\{(x, y) / 0 \leq x \leq y \leq 1\}$ des supports.

Ajoutons que l'axe MI est celui des incertitudes croissantes et IM celui, au contraire, de la connaissance croissante, son maximum en (1, 0) serait la «contradiction» notée TF.

Voir l'annexe sur le système MVL pour le système d'axes formés par (FT) et (IM).
Un autre système d'axes formé IT et IF où les coordonnées deviennent I (0, 0), T (1, 0), F (0, 1), M (1/2, 1/2), c'est à dire par a = x et b = 1 - y, peut être envisagé, la négation est alors ¬(a, b) = (b, a) le symétrique par rapport à (IM) et la règle de Dempster s'écrit :

$(a, b) \otimes (c, d) = [1 - (1 - a)(1 - c) / (1 - (ad + bc), 1 - (1 - b)(1 - d) / (1- (ad + bc)]$

Pour tout élément (a, b), son image par la fonction définie comme $h(a, b) = (a, b) \otimes M = ((1 - b) / (2 - a - b), (1 - a) / (2 - a - b))$ est le point de [F, T], donc sans incertitude, intersection avec le segment joignant (a, b) et le sommet (1, 1) qui représente la «contradiction». Pour une telle structure (un dempsteroïde) [Daniel 94, 96] montre que [F, T] est un groupe pour $\otimes$ et que tout automorphisme préservant I et M est entièrement déterminé comme extension d'un couple d'automorphismes sur FT et sur FI $\cup$ IT.

REMARQUES

En observant les itérés par la règle de Dempster (non idempotente) d'un point dans l'ensemble des supports (x, y), on s'aperçoit que suivant sa position, le résultat consiste en une «accentuation» vers le faux (0, 0) ou bien vers le vrai (1, 1). Les exceptions sont le cas de l'incertain (0, 1) dont le carré pour cette coopération est lui-même et les autres éléments de IM dont on peut montrer que les itérés se rapprochent de M.
L'opération de Dempster est donc bien un renforcement de la certitude.

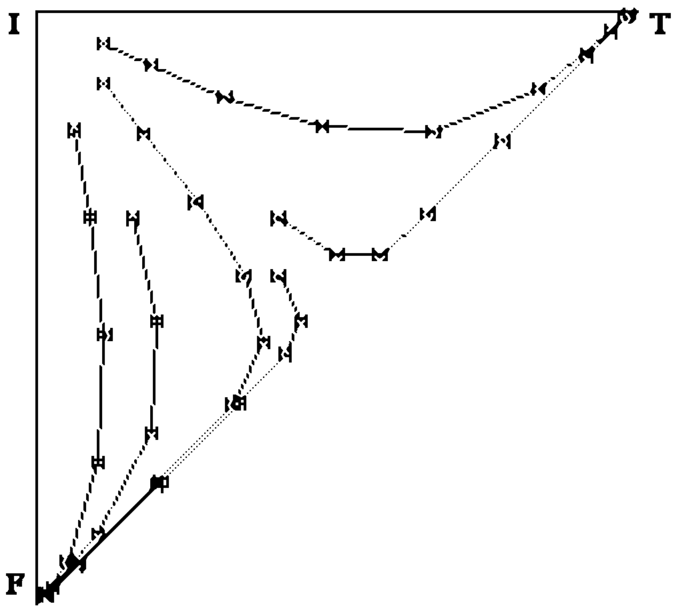

Figure 2.5 Six exemples d'itération spar Dempster convergent vers T ou F.

En disposant régulièrement des points suivant un quadrillage, les carrés au sens de l'opération de Dempster montrent un mouvement vers moins d'incertitude en se rapprochant du segment FT, tout en se rapprochant également des deux valeurs booléennes F et T.

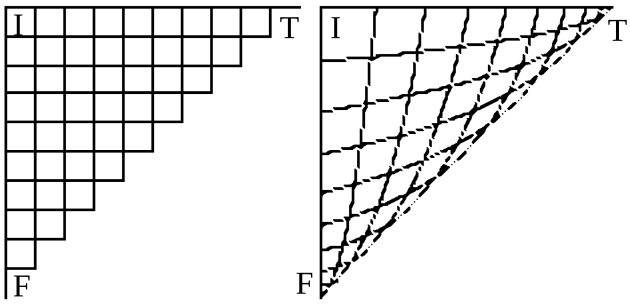

Figure 2.6 Déformation d'un quadrillage par Dempster

APPLICATION À LA FUSION DE DONNÉES MULTI-CAPTEURS

Le problème général est de déterminer une hypothèse (la meilleure) parmi n objets ou n classes, lorsque p sources d'information (des capteurs par exemple) envoient leurs «avis». Ce problème se présente notamment pour le traitement des images prises par satellites.
On appelle «cadre de discernement» un ensemble de n hypothèses (une classification en fait) non nécessairement exclusives, par exemple E = {ami, ennemi, neutre} ou encore E = {avion-de-chasse, transporteur, bombardier, hélicoptère-léger, hélicoptère-lourd} avec n = 5. On dispose de m capteurs ayant des sensibilités différentes d'une cible à l'autre et le problème est de fusionner leurs conclusions.

On nomme matrice de confusion, la matrice (à trois dimensions) formée par les :
p(i, k, j) = proba(«le capteur j émette l'avis k en présence de l'objet i»)
Si H_i est une hypothèse et Dj une déclaration (de capteur) :

Dans l'approche bayésienne, $p(H_i) = \prod_{j=1}^{p} \frac{p(D_j / H_i)\, p(H_i)}{p(D_j)}$

En fait, le capteur j renvoie simplement une «grandeur de discrimination» e(k, j) mesurant la vraisemblance de l'hypothèse k et p(i, k, j) est la probabilité que ce niveau e(k, j) soit le plus grand de tous les e(h, j) pour toutes les hypothèses h, sachant que la cible présentée est i.
Dans le cadre restreint $E = \{H_i, \neg H_i\}$, seuls $H_i \cap H_i = H_i$ et $H_i \cap E = H_i$, alors la règle de Demspter donne $m_i(H_i) = \sum m_{i1}(A_1).m_{i2}(A_2)....m_{im}(A_m) / (1 - c)$ d'où :
$(1\text{-}c)m_i(H_i) = \prod_{1\leq j\leq m} (m_{ij}(H_i) + m_{ij}(E)) - \prod m_{ij}(E)$
On en déduit pour ce même E une plausibilité pl_i que l'on va confondre avec celle relative au cadre de discernement $E = \{H_1, H_2, ... ,H_n\}$ donc $pl(H_i) = pl_i(H_i) = m_i(H_i) + m_i(E) = \{\prod_{1\leq j\leq m} m_{ij}(H_i)\}/(1\text{-}c)$
On souhaite maintenant choisir l'hypothèse H_i telle que $C_{i1} = pl(\neg H_i) = 1\text{-} cr(H_i)$ soit minimum et que $C_{i2} = pl(H_i)$ soit maximum. La conservation des variations relatives exige le système différentiel $\partial Ci/Ci = -\partial C_{i1}/C_{i1}$ et $\partial C_i/C_i = \partial C_{i2}/C_{i2}$. On peut donc choisir de maximiser la solution $C_i = C_{i2}/C_{i1}$ soit :

$$C_i = \frac{pl(H_i)}{1 - cr(H_i)} = \frac{\prod_{j=1}^{m} m_{ij}(H_i)}{\prod_{j=1}^{m} 1 - m_{ij}(H_i)} = \prod_{j=1}^{m} \frac{e_{ij}^{max}}{1 - e_{ij}^{min}}$$

EXPÉRIENCE, si n = m = 3 et p_i = proba (déclaration i / hypothèse i) en supposant que les trois capteurs fonctionnent de la même façon. La matrice de confusion est :

Hypothèses :	H1	H2	H3
D1	p_1	$(1 - p_2) / 2$	$(1 - p_3) / 2$
Déclarations : D2	$(1 - p_1) / 2$	p_2	$(1 - p_3) / 2$
D3	$(1 - p_1) / 2$	$(1 - p_2) / 2$	p_3

A partir des p_i, sont calculés les autres éléments de la matrice par $p_{ij} = (1-p_i) / (n-1)$, la cible choisie i sera celle qui va maximiser le critère Ci défini par :
Solutions de la théorie de l'évidence :

$C_i = \prod_{1\leq k\leq n} \prod_{1\leq j\leq m} (1 - p(i, k, j)*(1 - e(k, j)))$.

puis $C_i = \prod_{1\leq k\leq n} \prod_{1\leq j\leq m} [(1 - p(i, k, j)*(1 - e(k, j))) / (1 - p(i, k, j)*e(k, j))]$ (maximisations de la plausibilité et de la crédibilité, courbe «ev»)
De façon analogue dans [Appriou 88] d'autres méthodes sont comparées :
Solutions Bayesienne $C_i = \prod_{1\leq j\leq m} e(i, j)$ pour «prob1» et $C_i = \prod_{1\leq j\leq m} p(i, k, j)$ où p(i, k, j) est le coefficient de la matrice de confusion du capteur j pour la cible i réalisant le plus grand e(h, j) pour toutes les hypothèses h (courbes «prob2»).
Solution possibiliste : max des $C_i = \min_{1\leq k\leq n,\ 1\leq j\leq m} (1 - p(i, k, j)*(1 - e(k, j)))$ (courbe «poss»).

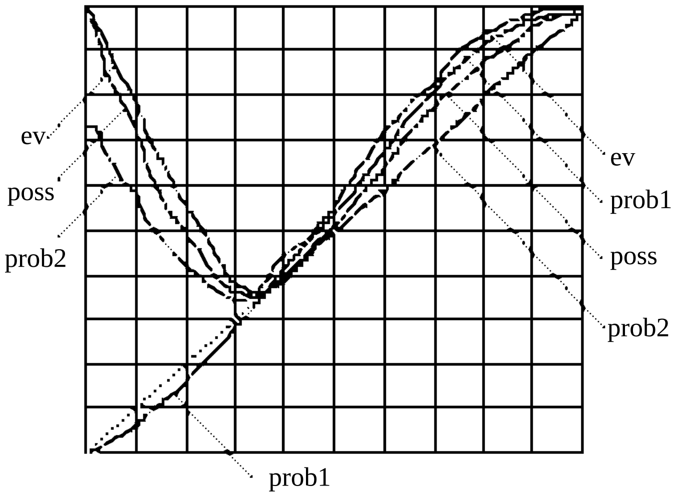

Figure 2.7 Taux de reconnaissance en fonction de a.

On effectue des tirages aléatoires des coefficients e(k,j) un grand nombre de fois (plus de 5000), et trace les courbes du taux de reconnaissance par fusion : (nombre de succès) / (nombre d'essais) en fonction de la valeur commune a = p_i.
Si a < 1/3 (très mauvais capteurs) les résultats sont meilleurs à ceux d'un capteur unique, mais cela signifie que les capteurs ont plus de chance de proposer une cible différente de celle qui est présentée. Un capteur complètement aléatoire signifie a = 1/3, en ce cas la fusion de plusieurs de ces capteurs doit également fournir un taux de reconnaissance 1/3, ce qui est le cas. Les cas intéressants se réalisent pour a > 1/3, alors les résultats de ev sont meilleurs que ceux d'un capteur unique, et supérieurs à ceux de poss.

En faisant varier le nombre de capteurs (second schéma), l'expérience montre que l'on peut classer prob1, prob2, pos, ev dans l'ordre croissant, mais qu'à partir de 8 capteurs les performances de «poss» sont comparables à celles de «ev».

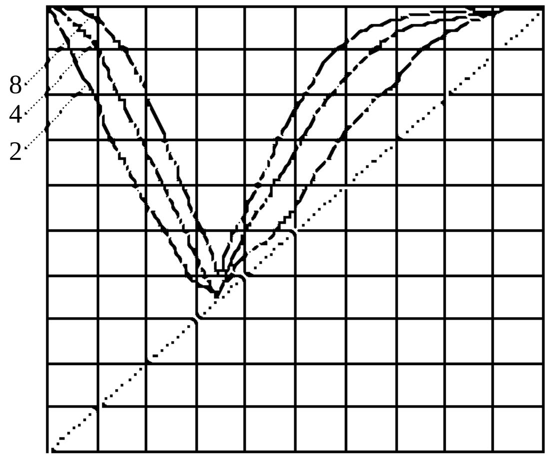

Figure 2.8 (a varie de 0 à 1 en abscisse et C_i de 0 à 1 en ordonnée) En augmentant la taille n du discernement, les performances augmentent en convergeant vers une courbe proche de celle de n = 10**.**