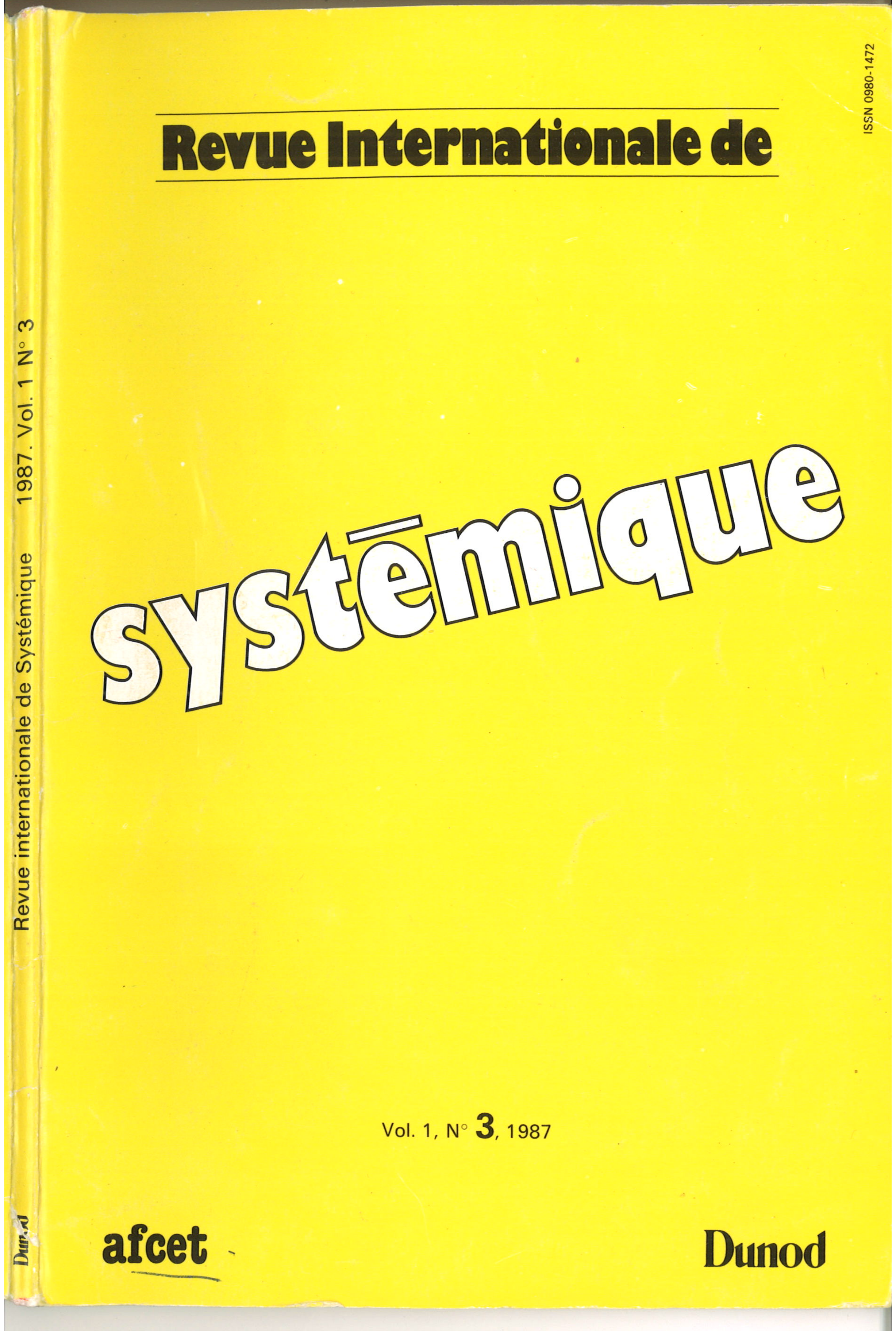


AFSCET

Revue Internationale de systémique

# Revue Internationale de Sytémique

volume 01, numéro 3, pages 275 - 294, 1987

# Approche en compréhension dans la gestion d'incertitudes en vue de l'estimation des dépendances souhaitées dans un système

Albert Perez

Numérisation Afscet, décembre 2015.

maintenant un maëlström qui nous amène à tout confondre. On prend le représenter (machine) pour l'exprimer (organisme) et vice-versa. Représentation et expression s'effacent au profit de la confusion Frankenstein. Les deux métaphores jadis isolées l'une de l'autre et raisonnables, s'emballent jusqu'à délirer.

Désastre ou voyage dans un Nouveau Monde ? Ce n'est pas le lieu d'en juger. Remarquons seulement la place de Turing et sa responsabilité dans cette étrange affaire où il fut plus qu'un agent-charnière : un véritable embrayeur.

## Références

HOFSTADTER D.R., DENNETT D.C., *the Mind I*, Basic Books Inc., New-York, 1981.

MINSKI, «Society of minds», ronéoté, 1985.

SEARLE J., «Minds brains and programs», in *The Behavioral and Brain Sciences*, Vol. 3, Cambridge University, Prss., 1980.

SFEZ L., *Critique de la communication*, Le Seuil, Paris, 1988.

TURING, «Computing Machinery and intelligence», *Mind*, Vol. 49, n. 236, 1950.

REVUE INTERNATIONALE DE SYSTEMIQUE
Vol. 1, N° 3, pp. 275 à 294

# APPROCHE EN COMPREHENSION DANS LA GESTION D'INCERTITUDES EN VUE DE L'ESTIMATION DES DEPENDANCES SOUHAITEES DANS UN SYSTEME

Albert PEREZ

Institut de Théorie de l'Information et d'Automatique
de l'Académie Tchécoslovaque des Sciences [1]

Résumé

Disposant d'un ensemble de connaissances partielles concernant la structure de dépendance mutuelle entre les nœuds d'un système incertain, nœuds susceptibles de prendre différents états, on cherche à estimer globalement cette structure dans le but de déduire de bonnes approximations des dépendances souhaitées. On étudie des systèmes à structure probabiliste.

Abstract

Having at our disposal a set of pieces of knowledge concerning the structure of mutual dependence between the nodes of the uncertain system considered, nodes able to take different states, we try to estimate this structure, in an overall way, with the aim to derive from it good approximations of the required dependencies. We study namely systems with probability structure.

Ayant à sa disposition un ensemble de connaissances concernant la structure de dépendance mutuelle entre les nœuds du système incertain considéré, nœuds susceptibles de prendre différents états (par exemple, système à base de connaissances), on cherche à estimer globalement cette structure en vue d'en déduire de bonnes approximations des dépendances souhaitées. Par opposition à une approche en extension dans la gestion d'incertitudes, caractérisée par des règles rigides (a priori) pour la composition (synthèse) de tout sous-ensemble de connaissances, règles qui, une fois choisies, ne sont pas influencées par le reste des connaissances (contexte), l'approche en compréhension

1. rue Pod Vodàrenskou vezi 4, 18208 Prague 8, Tchécoslovaquie

Revue internationale de systémique. 0980-1472 Vol. 1/87/03/275/20/$ 4.00/

que nous adoptons ici prend en considération la totalité des connaissances dont nous disposons en vue d'établir (automatiquement), pour chaque sous-ensemble particulier de connaissances, la démarche spécifique appropriée à leur synthèse. Dans l'hypothèse d'une structure probabiliste du système, par exemple, l'approche en compréhension revient à approximer au mieux la distribution multidimensionnelle sous-jacente par une gestion appropriée de l'ensemble des connaissances dont on dispose, ce que nous ferons par la suite.

Plan de l'article

1. Approche en compréhension versus approche en extension
2. Systèmes incertains à structure probabiliste
3. Le Principe d'Entropie Maximale dans l'intégration des connaissances
4. Le concept de barycentre dans l'intégration des connaissances

## 1. Approche en compréhension versus approche en extension

Rappelons qu'en logique on distingue, sur une structure, un calcul *en extension* d'un calcul *en compréhension* par le fait que dans le premier cas les valeurs (non nécessairement numériques) des objets composés de la structure considérée dépendent uniquement des valeurs de leurs composantes dans cette structure, tandis que dans le second cas les valeurs des objets composés dépendent en plus d'autres paramètres («état de l'univers»).

Par analogie, nous avons introduit dans Perez (1983) une distinction fondamentale entre systèmes à base de connaissances (systèmes experts) en extension et en compréhension (en anglais : intensional). A cette époque tous les systèmes experts connus étaient de type extensionnel. L'idée de construire par contre un système expert en compréhension (en anglais : intensional expert system), présentée dans le travail ci-dessus, a été réalisée dans le système expert INES, cf. Perez and Jirousek (1985), Perez (1985), Jirousek and Perez (1986), Studeny (1987).

Le point critique permettant de distinguer si un système expert est de type en extension ou en compréhension réside dans la gestion d'incertitudes qu'il met en œuvre, c'est-à-dire, dans la façon d'estimer le «poids», non nécessairement numérique (degré de certitude ou de validité, probabilité objective, subjective ou comparative, etc.) d'une certaine conclusion (c'est-à-dire, d'une certaine combinaison d'états pris par un sous-ensemble de nœuds du système) à partir des «poids» d'un ensemble d'antécédents (c'est-à-dire, d'états pris par certains autres nœuds) en exploitant les connaissances dans la base de connaissances dont chacune représente, en règle générale, quelque dépendance



mutuelle plus ou moins forte (donc aussi pondérée) entre un certain nombre de variables parmi celles attribuées aux nœuds du système afin d'exprimer leurs états. Dans un système expert de type extensionnel le poids d'une conclusion est estimé par la valeur prise par une fonction (de combinaison cf. Hajek (1985) des poids des antécédents et des poids affectés à la dépendance de la conclusion par rapport à chacun d'eux, fonction choisie a priori, une fois pour toutes, et qui ne tient pas compte, en particulier, d'une dépendance mutuelle possible entre les antécédents eux-mêmes qui pourrait être estimée à partir de l'ensemble des connaissances dans la base de connaissances. En fait, on ne tient même pas compte d'une telle dépendance déduite dans une échelle «locale» et à plus forte raison dans une échelle globale. Par contre, dans un système expert en compréhension le poids d'une conclusion est estimé à partir des poids des antécédents et des poids de dépendance de la conclusion par rapport à chacun d'eux, à l'aide d'une fonction, a priori inconnue, qui apparait automatiquement, a posteriori, et qui dépend de l'ensemble des nœuds correspondants aux antécédents et à la conclusion considérée.

Par une généralisation naturelle, cette distinction entre approche en extension et approche en compréhension dans la gestion d'incertitudes, en vue d'estimer les dépendances souhaitées, peut concerner l'étude de tout système incertain (stochastique) au sujet duquel on dispose d'un ensemble de connaissances (dépendances partielles ou autres contraintes auxquelles il est soumis) insuffisant pour son identification «complète» mais peut-être suffisant pour obtenir une bonne qualité de décision sous réserve d'une gestion appropriée.

Un exemple illustratif très important est celui d'un système à structure probabiliste, caractérisé par une distribution multidimensionnelle simultanée des variables incertaines (stochastiques) attribuées aux nœuds du système susceptibles de prendre différents états suivant les valeurs de ces variables. Indépendamment de son interprétation, cette distribution étant inconnue on cherche à l'approximer aussi bien que possible à partir de l'ensemble des connaissances, de nature probabiliste, dont on dispose concernant le système. L'estimation de toute dépendance souhaitée dans le système (probabilités marginales ou conditionnelles) s'obtient alors comme projection de l'approximation globale, résultat de l'intégration de l'ensemble des connaissances. C'est ce qui distingue l'approche en compréhension d'une approche en extension qui, devant la tâche difficile (mais réalisable dans une certaine mesure, comme nous le verrons) de cette intégration de l'ensemble des connaissances, utilise prématurément des méthodes trop simplifiées pour être capables de saisir toute la richesse des types de dépendances qui peuvent se présenter dans un système stochastique complexe.

Parmi les méthodes extensionnelles largement utilisées se trouve celle des ensembles flous. Bien entendu, le hasard et l'incertain ont des aspects multiples et ne sont pas toujours probabilisables en ce qui concerne leur structure (signalons que des concepts de même structure peuvent avoir des interprétations différentes). On dit que la théorie des sous-ensembles flous a apporté une meilleure description de ce qui n'est pas mesurable au sens probabiliste. En principe, ceci parait vrai compte tenu du fait que l'axiomatique de la théorie des probabilités se situe dans un treillis distributif et complémenté tandis que celle de la théorie des sous-ensembles flous se situe dans un treillis distributif, de sorte que la structure de «valuation» est plus générale que celle de «probabilité», cette dernière étant une valuation particulière puisque «additive».

Examinons cependant la situation du point de vue des applications. Plaçons-nous notamment dans le cadre d'un problème d'estimation de la structure de dépendance caractérisant un système incertain complexe conformément à notre tâche. La première question qui se pose est de savoir quel type de valuation est approprié à la description de l'incertitude qui y règne et pourquoi, éventuellement, la valuation probabiliste est à rejeter. Ce n'est pas une question dictée par un «entêtement probabiliste». La raison en réside plutôt dans le fait qu'une valuation à structure probabiliste a l'avantage de se laisser appliquer d'une façon *conséquente* puisqu'elle est fondée sur un appareil mathématique bien développé offrant des possibilités de modélisation *opérationnelle* beaucoup plus riches que d'autres appareils connus en ce qui concerne surtout la notion-clé de *dépendance multilatérale* entre les variables incertaines (stochastiques) attribuées aux nœuds du système.

Même si on peut introduire des «valuations conditionnelles», semblables dans le cas général aux probabilités conditionnelles, afin de pouvoir exprimer cette dépendance, on s'aperçoit que pour cela on doit avoir à sa disposition toutes les valuations pertinentes correspondant aux variables incertaines, intervenant dans le problème, prises individuellement ou par deux ou par trois, etc. Appelons-les valuations unidimensionnelles, bidimensionnelles, tridimensionnelles..., N-dimensionnelles, où N est le nombre de toutes les variables. Mais qui est capable de nous fournir ces valuations si N dépasse quelques unités ? Il s'agit, évidemment, d'un système de valuations cohérent. Si la valuation est une probabilité la réponse se trouve dans la relation qui existe entre une distribution simultanée et ses différentes distributions marginales, par exemple. Si on n'est pas capable de fournir la distribution simultanée mais seulement quelques de ses marginales (le plus souvent de basse dimension), la théorie des probabilités fournit de moyens raisonnables permettant d'approximer la distribution simultanée à partir de ces quelques marginales et cela en un sens convenable eu égard au problème de décision considéré.

Mais comment surmonter ces difficultés dans le cas d'une valuation non-probabiliste ? On nous propose en règle générale (avec M. Zadeh en tête) une approche extensionnelle, c'est-à-dire, on se donne des méthodes a priori pour composer les fonctions floues ou les valuations en passant du cas unidimensionnel au cas bidimensionnel, tridimensionnel, etc..., sans pouvoir aisni tenir compte de toute la richesse des types de dépendance qui peuvent se présenter entre les variables incertaines d'un système donné. En fait, cette façon de procéder est contraire à l'esprit même d'une théorie générale des ensembles flous laquelle, pourvu quelle soit bien développée comme la théorie des probabilités, rendrait possible comme elle une approche en compréhension pour intégrer les connaissances partielles, leur composition étant dictée dans une large mesure par l'*ensemble* de toutes les valuations de plus basse dimension que l'on dispose au début.



Pour conclure, on peut dire qu'il ne s'agit pas tant de disposer d'autant de représentations de l'incertain qu'on le veut mais plutôt d'en avoir telles qui, fondées sur un appareil mathématique bien développé permettant une approche en compréhension dans la gestion des connaissances et valuations partielles concernant un système incertain complexe, permettent de faire des inférences aussi efficaces que possibles sans se livrer prématurément à des méthodes extensionnelles lesquelles, bien que plus simples, sont mal fondées et par suite moins puissantes.

## 2. Systèmes incertains à structure de dépendance probabiliste

Soient $X_1, X_2, \ldots X_N$ les variables incertaines attribuées aux N nœuds du système incertain considéré, nœuds susceptibles de prendre différents états suivant les valeurs prises par les variables respectives. Par la suite nous admettrons que la structure de dépendance entre les nœuds du système est de nature probabiliste, c'est-à-dire, que cette structure se laisse décrire par une distribution de probabilités simultanée $P_A$ de l'ensemble A des variables (stochastiques) $X_1, X_2, \ldots, X_N$. Cette distribution pourrait être conçue comme dépendant de l'état du système évoluant dans le «temps» mais nous n'adopterons pas ce point de vue ici.

Supposons que nous ayons à notre disposition, au lieu de $P_A$, seulement un ensemble de connaissances K concernant le système, par exemple quelques probabilités conditionnelles ou marginales, en règle générale, de dimension beaucoup plus faible que N que nous supposerons d'autant plus grand que le système considéré est plus complexe (dans un système à base de connaissances ou système expert, par exemple, N peut être de l'ordre de 100). Parmi les connaissances on pourrait avoir des données concernant la valeur moyenne de certaines fonctions $h_i$ ($i = 1, 2, \ldots, L$) de variables $X_1, X_2, \ldots, X_N$ ou même quelque famille de distributions à laquelle $P_A$ appartient. D'une façon générale, chaque connaissance partielle a pour effet de réduire l'ensemble $\mathcal{P}$ de toutes les distributions N-dimensionnelles possibles à un sous-ensemble de $\mathcal{P}$ contenant $P_A$. L'intersection $E_k$ de tous ces sous-ensembles correspondant à l'ensemble de connaissances K, explicitement ou implicitement, contiendra d'ordinaire beaucoup plus que le seul élément inconnu $P_A$.

L'approche en compréhension, à laquelle nous avons fait allusion à la section 1, nous amène à la recherche dans $E_K$ d'une approximation convenable $\overline{P}_A$ de $P_A$. C'est ce qu'on appelle *intégration* de l'ensemble des connaissances K.

Qu'entend on par approximation convenable ? En adaptant nos procédures de décision, concernant le système, à $\overline{P}_A$ au lieu du $P_A$ réel on obtient quelques conséquences négatives concernant la qualité de la décision dues à la mauvaise adaptation. Il est donc souhaitable de choisir $\overline{P}_A$ de façon à réduire autant que possible ces conséquences négatives. On connaît (voir, par exemple, Perez (1965), (1967)), à ce sujet des bornes concernant l'augmentation du risque, due à la mauvaise adaptation en question, exprimées à l'aide de certaines mesures de divergence, entre $P_A$ et $\overline{P}_A$, comme par exemple l'entropie relative de Shannon ou la divergence de Kullback-Leibler $H(P_A, \overline{P}_A)$ de $P_A$ par rapport à $\overline{P}_A$. Grosso modo, plus cette divergence est petite, plus les conséquences négatives de la mauvaise adaptation sont réduites. Dans la suite on va utiliser $H(P_A, \overline{P}_A)$ comme mesure de divergence. Rappelons sa définition :

$$(1)\quad H(P,Q) \begin{cases} := \int dP \log \dfrac{dP}{dQ}, \text{ pour } P \ll Q, \\ := +\infty, \text{ dans les autres cas} \end{cases}$$

Si, en particulier, on prend dans (1) $P = P_A$ et $Q = P_{X_1} \dots P_{X_N}$, c'est-à-dire, le produit des marginales unidimensionnelles de $P_A$, on obtient ce que nous avons appelé dans Perez (1977) *degré de dépendance*, $I(P_A)$, entre les variables aléatoires $X_1, X_2, \dots X_N$ avec distribution simultanée $P_A$. Dans le cas particulier où $N = 2$, $I(P_A)$ coincide avec l'information mutuelle de Shannon. Par extension, on peut appeler aussi $I(P_A)$, pour N quelconque, *multiinformation* (cf. Perez (1985)) d'autant plus qu'elle a des propriétés très semblables à celles de l'information de Shannon. De plus, si $G_1, \dots, G_k$ est une partition de A, on peut écrire la relation d'additivité suivante

$$(2)\quad I(P_A) = I(A) = I(G_1) + I(G_2) + \dots + I(G_k) + I(G_1, \dots, G_k),$$



c'est-à-dire que le degré de dépendance (multiinformation) du système stochastique A (représenté par $P_A$) est égal à la somme des degrés de dépendance des sous-sytèmes $G_1, G_2, \dots G_k$ (représentés par $P_{G_1}, \dots, P_{G_k}$ marginales de $P_A$ correspondant à $G_1, \dots, G_k$) en lesquels le système A est décomposé, *plus* le degré de dépendance entre sous-systèmes (cf. Perez (1977), (1980)).

Remarquons que pour toute permutation $G_{i_1}, \dots, G_{i_k}$ de $G_1, \dots, G_k$

$$(3)\quad I(G_1,\dots,G_k) = I(G_{i_1}, G_{i_2}) + I(G_{i_1} \cup G_{i_2}, G_{i_3}) + \dots + I(G_{i_1} \cup \dots \cup G_{i_{k-1}}, G_{i_k}).$$

Ceci résulte immédiatement du fait que, symboliquement,

$$(4)\quad P_A = P_{G_{i_1}} P_{G_{i_2}/G_{i_1}} P_{G_{i_3}/G_{i_1} \cup G_{i_2}} \cdots P_{G_{i_k}/G_{i_1} \cup \dots \cup G_{i_{k-1}}},$$

où par exemple nous désignons par $P_{G_{i_2}/G_{i_1}}$ la probabilité conditionnelle du vecteur aléatoire, ayant ses composantes dans $G_{i_2}$, le vecteur aléatoire ayant ses composantes dans $G_{i_1}$ étant donné. En fait, $P_{G_{i_2}/G_{i_1}} = P_{G_{i_1} \cup G_{i_2}} / P_{G_{i_1}}$, etc.

En prenant dans (4), à la place de $G_{i_1}$ dans le second facteur un sous-ensemble $F_1$ de $G_{i_1}$, à la place de $G_{i_1} \cup G_{i_2}$ dans le troisième facteur un sous-ensemble $F_2$ de $G_{i_1} \cup G_{i_2}, \dots$ à la place de $G_{i_1} \cup \dots \cup G_{i_{k-1}}$ dans le k-ème facteur un sous-ensemble $F_{k-1}$ de ce dernier ensemble, on obtient une nouvelle distribution simultanée $\overline{P}_A$ que j'appelle une *simplification de la structure de dépendance* (s.s.d.) de l'ensemble A de variables aléatoires (du système stochastique) initialement représenté par $P_A$. On obtient alors pour $H(P_A, \overline{P}_A)$.

$$(5)\quad H(P_A, \overline{P}_A) = I(P_A) - I_{P_A}(\overline{P}_A),$$

où par $I_{P_A}(\overline{P}_A)$(noté abusivement par $I(\overline{P}_A)$ dans les citations précédentes) nous désignons le degré de dépendance de $\overline{P}_A$ relativement à $P_A$ :

$$(6)\quad I_{P_A}(\overline{P}_A) = I(G_1) + \dots + I(G_k) + I(F_1, G_{i_2}) + \dots + I(F_{k-1}, G_{i_k}) = I(G_{i_1}) + I(F_1 \cup G_{i_2}) - I(F_1) + \dots + I(F_{k-1} \cup G_{i_k}) - I(F_{k-1})$$

Si l'ensemble de connaissances K se réduit à un ensemble M de de marginales $P_{S_1}, \dots, P_{S_k}$ de $P_A$, correspondant aux sous-ensembles $S_1, \dots, S_k$ de A, avec $\bigcup_{i=1}^{k} S_i = A$, nous pouvons, pour chaque permutation $i_1, \dots, i_k$ des indices $1, \dots, k$, construire une s.s.d. $\bar{P}_A^{i_1 \dots i_k}$ de $P_A$, en prenant : $G_{i_1} = S_{i_1}$, $G_{i_2} = S_{i_2} - S_{i_1}, \dots$, $G_{i_k} = S_{i_k} - \bigcup_{j=1}^{k-1} S_{i_j}$ ; $F_1 = S_{i_2} \cap S_{i_1}$, $F_2 = S_{i_3} \cap \bigcup_{j=1}^{2} S_{i_j}, \dots$, $F_{k-1} = S_{i_k} \cap \bigcup_{j=1}^{k-1} S_{i_j}$. On a donc, symboliquement,

$$\bar{P}_A^{i_1 \dots i_k} = P_{S_{i_1}} \cdot P_{S_{i_2}} / P_{F_1} \cdot P_{S_{i_3}} / P_{F_2} \dots P_{S_{i_k}} / P_{F_{k-1}}, \tag{7}$$

de sorte que (6) nous donne

$$I_{P_A}(\bar{P}_A^{i_1 \dots i_k}) = \sum_{i=1}^{k} I(P_{S_i}) - \sum_{j=1}^{k-1} I(P_{S_{i_{j+1}}} \cap \bigcup_{s=1}^{j} S_{i_s}) =: I_M(\bar{P}_A^{i_1 \dots i_k}), \tag{8}$$

puisque pour calculer cette quantité, comme pour construire $\bar{P}_A^{i_1 \dots i_k}$, il nous suffit de connaître l'ensemble M de marginales. Si quelque $F_i$ ci-dessus est vide, le facteur correspondant dans (7) est pris égal à 1 et $I(F_i)$ égal à 0.

Soit $E_M$ l'ensemble des distributions $Q_A \in \mathcal{P}$ dont les marginales en $S_1, \dots, S_k$ coïncident avec celles en M, c'est-à-dire, $Q_{S_i} = P_{S_i}$, $i = 1 \dots, k$.

Soit $L_M$ l'ensemble des distributions $\bar{P}_A^{i_1 \dots i_k} \in \mathcal{P}$ engendrées, comme en (7), à partir de M pour toutes les permutations $i_1, \dots, i_k$.

Evidemment, pour que $E_M \neq \emptyset$ il est nécessaire et suffisant que l'ensemble des distributions M soit *compatibles*, c'est-à-dire, qu'il existe au moins une $Q_A \in \mathcal{P}$ ayant comme marginales toutes ces distributions. Pour les conditions de compatibilité voir Kellerer (1964). Pour que $L_M \neq \emptyset$ il suffit (et c'est aussi nécessaire) que les $P_{S_i} \in M$ soient *faiblement compatibles*, c'est-à-dire, que, pour chaque couple ($P_{S_i}$, $P_{S_j}$) d'éléments de M, leurs restrictions à $S_i \cap S_j$ coïncident. Si, de plus, pour quelque permutation $i_1 \dots, i_k$, les $S_{i_1}, \dots, S_{i_k}$ sont tels que

$$S_{i_{j+1}} \cap \bigcup_{s=1}^{j} S_{i_s} \subset S_{i_r}, \text{ pour quelque } r \leqslant j,\ j = 1, \dots, k\ , \tag{*}$$



alors les $P_{S_1}, \dots, P_{S_k}$ sont compatibles. On peut alors voir que $E_M \cap L_M \neq \emptyset$ et que cette intersection contient un seul point, la distribution $\bar{P}_A^{i_1 \dots i_k}$ pour $i_1, \dots, i_k$ satisfaisant (*). Pour cette distribution on obtient (cf. (8) )

$$I_M(\bar{P}_A^{i_1 \dots i_k}) = I(\bar{P}_A^{i_1 \dots i_k}) = \tag{9}$$

$$= \min_{Q_A \in E_M} I(Q_A) = \max_{Q_A \in L_M} I_M(Q_A).$$

cf. Jirousek and Perez (1986). En comparant (5) et (9) on obtient

$$H(P_A, \bar{P}_A^{i_1 \dots i_k}) = \min_{Q_A \in L_M} H(P_A, Q_A), \text{ pour tout } P_A \in E_M, \tag{10}$$

c'est-à-dire que la distribution $\bar{P}_A^{i_1 \dots i_k}$, pour $i_1, \dots, i_k$ satisfaisant (*), est la meilleure approximation de la distribution $P_A$, réelle mais inconnue, pourvu que $P_A \in E_M$.

En général, la compatibilité des distributions dans M n'entraîne pas obligatoirement $E_M \cap L_M \neq \emptyset$ mais si cela a lieu alors cette intersection contient nécessairement une seule distribution et cette distribution satisfait (9) et, par conséquent, aussi (10). Dans Studeny (1987), par exemple, on montre que cette intersection est non-vide en supposant, au lieu de (*), que chaque ensemble $S_i$ dans M peut avoir une intersection non-vide avec au plus trois autres ensembles dans M.

Même si $E_M \cap L_M = \emptyset$ , on a quelquefois intérêt à choisir notre approximation de $P_A$ dans $L_M$ (au lieu de $E_M$), on prend alors celle qui satisfait (10), c'est-à-dire, celle qui par le choix de la permutation $i_1,\dots,i_k$, *maximise* $I_M(\overline{P}_A^{\,i_1\dots i_k})$ ou, ce qui est équivalent (cf. (8) ), *minimise* la seconde somme figurant dans (8). Cet intérêt peut résulter par exemple, du besoin de pouvoir effectivement approximer $P_A$ (à l'aide d'un ordinateur, bien entendu) même pour une dimension élevée (N de l'ordre de 100 variables aléatoires binaires). On peut voir comment en tirer ensuite, d'une façon effective, les dépendances souhaitées, notamment en estimant les probabilités conditionnelles (poids) de certaines conclusions à partir d'un nombre quelconque d'antécédents dans un système expert à compréhension, par exemple, dans Jirousek and Perez (1986).

Toutes les fois que ce sera possible, il est, bien entendu, préférable de choisir l'approximation de $P_A$ dans $E_M$ et, plus généralement, dans le sous-ensemble $E_K \subseteq \mathcal{P}$ correspondant à l'ensemble de connaissances K. Comme nous l'avons vu, cet ensemble K ne se réduit pas toujours à un ensemble M de marginales, bien que, par ce qu'on appelle «petite intégration» de certaines connaissances plus élémentaires, on pourrait obtenir parfois des estimations de telles marginales. On pourrait obtenir aussi de bonnes estimations de telles marginales sur la base de données statistiques suffisamment riches par rapport à la dimension des marginales.

Très souvent on applique le Principe d'Entropie Maximale (PEM), plutôt intuitif, pour choisir une approximation dans $E_K$. Malgré ses nombreux succès, le PEM n'est pas bien fondé et il y a même des cas importants où il échoue, comme il arrive pour le «canal donné». Ces désavantages sont supprimés par l'application, à sa place, d'une approche minimax basée sur le concept de *barycentre* que nous avons introduit dans Perez (1984) et développé dans Perez (1986). L'idée d'appliquer ce concept à l'intégration de connaissances de nature probabiliste a été présentée pour la première fois dans Perez (1987).

## 3. Le Principe d'Entropie Maximale dans l'intégration de connaissances

Le Principe d'Entropie Maximale (PEM) a été introduit par E.T. Jaymes (1957) comme critère très naturel pour trouver des distributions de probabilités d'états microscopiques, en mécanique statistique, compatibles avec les valeurs données de quelques caractéristiques macroscopiques comme l'énergie moyenne, etc. Voir, par exemple, Guiasu, Shenitzer (1985). Dans l'ensemble E de toutes les mesures de probabilité sur un espace mesurable $(X, \mathcal{X})$, compatibles avec les valeurs



données $c_1,\dots,c_m$ de certaines variables aléatoires $h_1(x),\dots, h_m(x)$, on prend celle qui maximise l'entropie de Shannon (cas discret) dont les propriétés fondamentales comme mesure d'incertitude sont bien connues. A la suite des succès du PEM permettant de retrouver quelques distributions classiques, son usage a été étendu à beaucoup d'autres domaines d'application comme la statistique, la théorie statistique de la décision, la classification automatique, la théorie de l'information, l'analyse des suites temporelles et l'analyse des données en général. Il est ainsi naturel, du point de vue intuitif, d'essayer d'appliquer la PEM aux problèmes d'intégration de connaissances, tels que ceux mentionnés à la section précédente, ou tout-au moins, aux problèmes de «petite intégration» si les trop hautes dimensions empêchent son application. Comme nous l'avons remarqué plus haut, c'est justement pour cette raison que, dans les travaux mentionnés concernant l'approximation des distributions multidimensionnelles de l'ordre de 100 variables aléatoires $P_A \in E_M$ nous avons cherché les approximations $\overline{P}_A$ dans l'ensemble correspondant $L_M$ de s.s.d.

Le PEM est étendu au cas continu dans la littérature d'une façon moins satisfaisante en remplaçant automatiquement les probabilités dans l'expression de l'entropie (cas discret) par les densités correspondantes (prises d'ordinaire par rapport à la mesure de Lebesgue) et en intégrant au lieu de sommer (par rapport à la même mesure L). En fait, au signe près, l'expression ainsi obtenue est du type d'une entropie relative de Shannon. Dans Perez (1956) l'entropie de Shannon (cas fini) était aussi considérée comme une entropie généralisée, du type précédent, prise par rapport à la mesure d'équiprobabilité. Dans cet ordre d'idées, nous pouvons, d'une façon plus satisfaisante, donner une présentation unifiée du PEM, valable tant dans le cas discret que dans le cas continu.

Dans l'ensemble E ci-dessus (plus généralement, $E_K$ de la section précédente) il faut prendre la mesure de probabilité qui minimise l'entropie relative de Shannon, ou la divergence de Kullback-Leibler, par rapport à la mesure L «d'incertitude maximale».

En appliquant la méthode des multiplicateurs de Lagrange on trouve facilement que la solution est une mesure de probabilité $R_{PEM} \in E$ de la forme

$$(11)\qquad dR_{PEM} = dL \,.\, \exp\,(b_0 + \sum_{i=1}^{m} b_i h_i),$$

où les constantes $b_0, b_1,\dots, b_m$ sont obtenues comme solution (pourvu qu'elle existe) du système suivant d'équations (contraintes)

$$(12)\qquad \int dR = 1\ ; \quad \int h_i(x)\, dR = c_i, \qquad i = 1,2,\dots,m.$$

Remarquons que des contraintes (connaissances partielles) du type : «quelques probabilités (conditionnelles) ou marginales sont données», peuvent aussi s'exprimer en donnant les valeurs moyennes de quelques fonctions h(x) convenables.

Ainsi, pour E : = ensemble de toutes les distributions simultanées d'un couple de variables aléatoires X et Y prenant les valeurs 0 ou 1, avec $\Pr(X=0) = p$ et $\Pr(Y = 1 \mid X = 1) = q(1/1)$ données, (c'est-à-dire, avec valeur moyenne, p, de la fonction $h_{0.}(z)$ définie sur $Z = \{0,1\}^2$ par $h_{0.}(0,0) = h_{0.}(0.1) = 1$ et par $h_{0.}(z) = 0$ dans les autres cas ; et avec valeur moyenne 0 de la fonction $h_{1/1}(z) := h_{11}(z) - q(1/1) . h_{1.}(z)$, où $h_{1.}(1,1) = 1$ et $h_{11}(z) = 0$, dans les autre cas et $h_{1.}(1,0) = h_{1.}(1,1) = 1$ et $h_{1.}(z) = 0$, dans les autres cas), on trouve que $R_{PEM}(0,0) = R_{PEM}(0,1) = p/2$ ; $R_{PEM}(1,0) = (1-p).(1-q(1/1))$ ; $R_{PEM}(1,1) = (1-p).q(1/1)$.

Ce résultat pourrait s'obtenir plus simplement à partir de la relation bien connue

$$H(X,Y) = H(X) + p.H(Y|X=0) + (1-p).H(Y|X=1),$$

puisque max H(X,Y) correspond ici à max $H(Y|X=0)$, c'est-à-dire, à $\Pr(Y=0|X=0) = \Pr(Y=1|X=0) = 1/2$.

Cependant si, dans la définition de E ci-dessus, la contrainte $\Pr(X=0) = p$ est remplacée par la contrainte $\Pr(Y=0) = q$, (c'est-à-dire si au lieu de la valeur moyenne p de $h_{0.}(z)$ on donne la valeur moyenne q de $h_0(z)$ défini par $h_0(0,0) = h_0(1,0) = 1$ et $h_0(z) = 0$, dans les autres cas), la situation n'est pas tellement simple ; c'est pourquoi la méthode générale ci-dessus est recommandable.

De la même façon on trouve, par exemple, que, pour E := ensemble de toutes les distributions de probabilités «continues» sur la droite avec valeur moyenne et dispersion données, $R_{PEM}$ coincide avec la distribution de Laplace correspondante. Ce résultat se généralise d'une façon naturelle dans le cas multidimensionnelle.

Si, d'autre part, E := ensemble de toutes les distributions de probabilités «continues» sur la demi-droite positive avec valeur moyenne donnée, $R_{PEM}$ est la distribution exponentielle bien connue.

Comme exemple simple du cas avec «canal donné» prenons E : = ensemble de toutes les distributions simultanées d'un couple de variables aléatoires X et Y binaires avec canal donné :

$$\Pr(Y|X=0) = Q_0 \text{ et } \Pr(Y|X=1) = Q_1.$$

Pour tout $P_{XY} \in E$ on a pour l'entropie



$$H(P_{XY}) = H(P_X) + p\,H(Q_0) + (1-p)\,H(Q_1),$$

où $P_X(0) = p$ et $P_X(1) = 1-p$ est la distribution marginale correspondante de X. On trouve facilement que le maximum de $H(P_{XY})$ est obtenu pour

$$p = p_{PEM} = 1/(\exp(H(Q_1) - H(Q_0)) + 1), \tag{13}$$

de sorte que $R_{PEM}$ est engendré par le canal $(Q_0, Q_1)$ et la distribution d'entrée $P_X(0) = p_{PEM}$ et $P_X(1) = 1-p_{PEM}$. Celle-ci préfère, comme on devait s'attendre, l'hypothèse statistique $Q_i$ avec une entropie $H(Q_i)$ plus grande, en donnant la même probabilité (1/2, 1/2) aux deux hypothèses si, et seulement si, $H(Q_0) = H(Q_1)$. Comme nous allons le voir dans la section suivante, on a ici affaire à un exemple clair d'insuccès du PEM, tout-au-moins, en ce qui concerne les conséquences négatives de l'adaptation de nos procédures de décision à la distribution de probabilités qu'il propose.

Nous ne voulons pas allonger ici la liste des exemples où le PEM connait des succès ou des échecs du point de vue ci-dessus. Ajoutons seulement que dans le premier exemple de notre liste (p et q(1/1) donnés) le PEM donne la bonne solution tandis que dans le second exemple (q et q(1/1) donnés) il échoue comme dans le cas du «canal donné», mais pas d'une façon aussi claire. Le but que nous visons est d'attirer l'attention sur le fait que, si le PEM est bien justifié pour exprimer certaines lois fondamentales de la nature, cela est loin de signifier qu'il ne puisse pas conduire, dans certaines situations de décision, à des solutions inadéquates puisque, de ce point de vue, le PEM est plutôt intuitif et mal fondé. Il est donc recommandable avant d'appliquer une solution $R_{PEM}$, obtenue par le PEM, de vérifier si ses propriétés sont convenables du point de vue de la décision considérée. En particulier, il est utile de voir si $R_{PEM}$ est un barycentre au sens de la section suivante.

## 4. Le concept de barycentre dans l'intégration de connaissances

Le concept de barycentre d'un ensemble de mesures de probabilité (m.p.) a été introduit dans Perez (1984) et développé davantage dans Perez (1986). On se borne ici au cas d'ensembles de m.p. non-pondérées.

Soit, ainsi, $\mathcal{P}$ l'ensemble de toutes les m.p. sur un espace mesurable $(X, \mathcal{X})$ donné et soient S et T deux sous-ensembles de $\mathcal{P}$.

Soit encore D(P,Q) une mesure de divergence définie sur $\mathcal{P}$ .

Par D–*barycentre* de S pris en T nous entendons (pourvu qu'elle existe) une m.p. ${}_T R_S^D \in T$ satisfaisant la relation

$$\sup_{P \in S} D({}_T R_S^D, P) = \min_{R \in T} \sup_{P \in S} D(R,P). \tag{14}$$

Par la suite, $D(R,P) := H(P,R) = \int dP \log \frac{dP}{dR}$ (pour $P \ll R$) et $H(P,R) = +\infty$, dans les autres cas. Pour des raisons de compatibilité avec les notations utilisées dans les articles précédents, nous allons écrire, à la place de D, H' au lieu de H où $H'(R,P) := H(P,R)$. La position de R par rapport à P est, en effet, importante.

En confondant par la suite l'ensemble S ci-dessus avec l'ensemble E (respectivement, l'ensemble $E_k$, ou $E_M$ introduit dans la section 2), de toutes les m.p. de $\mathcal{P}$ qui sont compatibles avec les contraintes (connaissances partielles) données, notre idée est de choisir, pour intégrer les connaissances, le barycentre de E comme approximation de la m.p. réelle $P \in E$, en principe, inconnue. Le plus naturel serait d'identifier avec E aussi l'ensemble T où l'on prend le barycentre. Cependant, ceci n'est pas obligatoire si d'autres considérations interviennent. Un exemple de cette nature a été mentionné dans la section 2 à propos de l'introduction des approximations $\bar{P}_A$ de $P_A \in E_M$ prises pour des raisons de réalisabilité dans l'ensemble $L_M$ de toutes les s.s.d., lesquelles peuvent se construire à partir de l'ensemble donné M de marginales. Dans ce cas-là, $S = E_M$ et $T = L_M$ et on peut se demander si $\bar{P}_A := \arg\max_{Q_A \in L_M} I_M(Q_A) = \arg\min_{Q_A \in L_M} H(P_A, Q_A)$ est égal au barycentre ${}_{L_M} R_{E_M}^{H'}$. La réponse est positive puisque

$$\sup_{P_A \in E_M} H(P_A, \bar{P}_A) = \sup_{P_A \in E_M} I(P_A) - I_M(\bar{P}_A) = \sup_{P_A \in E_M} I(P_A) - \max_{Q_A \in L_M} I_M(Q_A) =$$

$$= \min_{Q_A \in L_M} \sup_{P_A \in E_M} H(P_A, Q_A),$$

ce qui montre que notre méthode d'approximation en question est en accord avec l'approche du barycentre. Autrement dit, toutes les fois qu'on obtient une approximation $\bar{P}_A := \arg\max_{Q_A \in L_M} I_M(Q_A)$, on obtient du même coup le barycentre correspondant.



La solution de ce problème, dans le cas particulier où M contient seulement des marginales de dimension ne dépassant pas 2, $|S_i| \leqslant 2$, s'obtient facilement en appliquant l'algorithme «greedy» de complexité linéaire, bien connu en théorie de graphes (cf. Perez (1983) et Perez and Jirousek (1985) ).

La solution du même problème dans le cas particulier où les ensembles $S_i$ dans $M = \{P_{S_1}, ..., P_{S_k}\}$ ont la propriété (*), mentionnée dans la section 2, s'obtient à partir de (9) qui indique qu'en même temps, dans ce cas-là, $\bar{P}_A = \arg\min_{Q_A \in E_M} I(Q_A)$, ou, ce qui revient au même, que $\bar{P}_A = \arg\max_{Q_A \in E_M} H(Q_A)$ puisque (cas discret) $I(Q_A) = -H(Q_A) + H(Q_{X_1}) + ... + H(Q_{X_N})$ et pour $Q_A \in E_M$ ses marginales unidimensionnelles sont connues comme résultant de M. Autrement dit, l'approximation $\bar{P}_A$ cherchée peut aussi s'obtenir ici soit par l'algorithme de Deming and Stephan (1940), précisé par Csissar (1975), (cf. procédure IPFP dans Perez and Jirousek (1985) ), soit par l'application du PEM (section 3). Malheureusement, la complexité peut augmenter dans ce cas plus rapidement que N.

Une situation analogue se présente pour la solution du problème ci-dessus si, au lieu de la propriété (*), on suppose avec Studeny (1987) que chaque ensemble $S_i$ dans M peut avoir une intersection non-vide avec au plus trois autres ensembles dans M, comme nous l'avons mentionné dans la section 2.

Comme nous l'avons constaté dans cette section, la solution $\bar{P}_A$, correspondant au second et au troisième cas ci-dessus, coïncide avec le seul point de l'intersection $E_M \cap L_M$ (laquelle est dans ces deux cas non-vide). Donc ce point donne du même coup le barucentre ${}_{L_M} R_{E_M}^{H'}$.

D'après ce qui précède on peut donc dire que ce barycentre est obtenu par l'application du PEM à $E_M$. Il faut cependant distinguer en général ${}_{L_M} R_{E_M}^{H'}$ de ${}_{E_M} R_{E_M}^{H'}$.

Pour E : = ensemble de toutes les distributions n-nomiales on a

$$(15)\qquad {}_E R_E^{H'} = (1/n,\dots,1/n).$$

En effet, pour tout $P = (p_1,\dots,p_n)$ et $R = (r_1,\dots,r_n)$ de E on a

$$H(P,R) = \sum_{i=1}^{n} p_i \log(p_i/r_i) = -H(P) + \sum_{i=1}^{n} p_i \log(1/r_i).$$

Soit $r_{i_0} := \min_{1 \leqslant i \leqslant n} r_i = n^{-s}$. On a $s \geqslant 1$ avec "=" si, et seulement si, $r_i = 1/n$ pour $i = 1,\dots,n$, ce que nous notons par $\bar{R}$.

Soit $P' = (0,\dots,0,p'_{i_0} = 1,0,\dots,0)$ de E. On obtient

$$\max_{P \in E} H(P,R) \geqslant H(P',R) = 0 + \log(1/r_{i_0}) = s \log n \geqslant \log n$$

$$= \max_{P \in E} H(P,\bar{R}) \ (= H(P',\bar{R}))$$

avec "=" si et seulement si $R = \bar{R}$. Autrement dit, (cf. (14) ),

$$\bar{R} = \arg \min_{R \in E} \max_{P \in E} H(P,R) = {}_E R_E^{H'} \quad \text{C.Q.F.D.}$$

Remarquons qu'on obtient le même résultat en prenant par exemple comme mesure de divergence, au lieu de H', la quantité

$$D(R,P) = \sum_{i=1}^{n} (r_i - p_i)^2 .$$

Ici donc l'approche du barycentre (AB) donne une solution $R_{AB}$ laquelle coincide avec celle, $R_{PEM}$, donnée par le Principe d'Entropie Maximale (PEM).

Pour E : = ensemble de toutes les distributions Q n–nomiales avec les probabilités en $i_1,\dots,i_m$ $(m \leqslant n-1)$ données, de somme totale w,

on a

(16) ${}_E R_E^{H'}$ = élément de E avec probabilité de chacun de n–m points restants égale à $(1-w)/(n-m)$

La méthode de démonstration est tout à fait analogue à celle du cas précédent. La même remarque est valable. Ici aussi $R_{AB} = R_{PEM}$.



Nous ne voulons pas allonger ici la liste des cas où $R_{AB} = R_{PEM}$. Dans tous ces cas, l'intégration de l'ensemble des connaissances K, effectuée par $R_{PEM}$, peut être considérée comme convenable (cf. section 2) dans un sens minimax puisqu'alors, en adaptant nos procédures de décision à $R_{PEM}$, on minimise en quelque sorte les conséquences négatives de la mauvaise adaptation par rapport à la distribution réelle la plus défavorable, compatible avec K.

Dans la section 3 nous avons cependant signalé certains cas, notamment le cas avec q et q(1/1) donnés, et le cas avec «canal donné», où le PEM échoue. En effet, tandis que, par exemple, dans le cas avec p et q (1/1) donnés on obtient $R_{PEM} = R_{AB}$, au contraire, dans les deux cas précédents cette égalité n'a pas lieu.

En nous réservant de donner dans une publication ultérieure certains résultats généraux permettant de distinguer entre ces deux possibilités, nous nous bornons ici à étudier brièvement le cas du «canal donné».

On peut, notamment, montrer aisément le résultat suivant :

Soit $E_X$ un ensemble de mesures de probabilité sur un espace mesurable $(X, \mathfrak{X})$ et soit C un canal avec espace mesurable d'entrée $(X, \mathfrak{X})$ et espace mesurable de sortie $(Y, \mathfrak{Y})$, c'est-à-dire, un système de mesures de probabilité $\{P_{Y|x},\ x \in X\}$ sur $(Y, \mathfrak{X})$ tel que, pour chaque ensemble $F \in \mathfrak{Y}$, la fonction $P_{Y|x}(F)$ de $x \in X$ soit $\mathfrak{X}$ - mesurable.

Soit E l'ensemble des mesures de probabilité $P_{XY}$ sur $\mathfrak{X} \times \mathfrak{Y}$ engendré par tout $P_X \in E_X$ et le canal C.

Alors le barycentre ${}_E R_E^{H'}$ de E (cf. (14) ) est donné par la mesure de probabilité sur $\mathfrak{X} \times \mathfrak{Y}$ engendrée par le barycentre ${}_{E_X} R_{E_X}^{H'}$ de $E_X$ et le canal C.

Appliquons ce résultat à l'exemple simple du «canal donné» de la section 3, où le canal C était donné par $Q_0$ et $Q_1$ et où $E_X$ était l'ensemble de toutes les distributions binomiales. On obtient, puisque ${}_{E_X} R_{E_X}^{H'}$ est ici donnée par (1/2, 1/2) conformément au résultat (15) de la présente section, que le barycentre ${}_E R_E^{H'}$ de E correspondant est engendré par une distribution $P_X$ et le canal $(Q_0, Q_1)$ avec $P_X(0) = p_{AB} = 1/2 = P_X(1) = 1 - p_{AB}$, tandis que la solution obtenue par le PEM était (cf. (13) ) $P_X(0) = p_{PEM} = 1/(\exp(H(Q_1) - H(Q_0)) + 1) \neq 1/2$, sauf si $H(Q_0) = H(Q_1)$.

Remarquons à ce propos que la méthode du maximum de vraisemblance très souvent appliquée en décision statistique en l'absence de toute connaissance concernant une distribution de probabilités a priori, est en accord avec l'approche du barycentre puisque dans une situation où on connaît seulement le canal (donné par le système des hypothèses statistiques) cette procédure opère essentiellement sur la base d'une mesure de probabilité globale laquelle est le barycentre de l'ensemble de toutes les mesures de probabilité compatibles avec le canal donné. Par contre, la méthode du maximum de vraisemblance est en désaccord avec le PEM sauf dans le cas exceptionnel où toutes les hypothèses statistiques ont la même entropie.

Poursuivons l'étude de l'exemple simple précédent et posons $Q_0 = (q_0, 1-q_0)$, $Q_1 = (q_1, 1-q_1)$. En supposant $q_0 > q_1 > 1/2$, on obtient $H(Q_1) > H(Q_0)$, de sorte que $p_{PEM} < 1/2$ tandis que $p_{AB} = 1/2$. Supposons en plus que $(1-p_{PEM}).q_1 > p_{PEM}\, q_0$. On peut voir que cette inégalité a lieu si, et seulement si,

$$(17) \quad (1-q_0).\log(q_0/1-q_0) < (1-q_1).\log(q_1/1-q_1)$$

Elle est, par exemple, satisfaite pour $q_0 = 0{,}9$ et $q_1 = 0{,}75$ puisque $0{,}1 \times \log 9 < 0{,}25 \times \log 3$.

En prenant, maintenant, comme critérium pour la discrimination des hypothèses statistiques $Q_0$ et $Q_1$ la probabilité moyenne d'erreur, la fonction de décision $d_{AB}$ adaptée au barycentre est $d_{AB}(Y=0) = Q_0$, $d_{AB}(Y=1) = Q_1$, tandis que la fonction de décision $d_{PEM}$ adaptée au PEM est : $d_{PEM}(Y=0) = d_{PEM}(Y=1) = Q_1$. La distribution réelle de probabilités a priori $(p,1-p)$ est cependant inconnue et p peut prendre toute valeur entre 0 et 1, 0 et 1 inclus. En notant par $e_{AB}^{(p)}$ et $e_{PEM}^{(p)}$ les probabilités moyennes d'erreur correspondant à la distribution a priori $(p,1-p)$ et aux fonctions de décisions $d_{AB}$ et $d_{PEM}$, respectivement, on obtient

$$e_{AB}^{(p)} = (1-p).q_1 + p.(1-q_0)\,; \quad e_{PEM}^{(p)} = p.q_0 + p.(1-q_0) = p.$$

$$\text{Ainsi,} \quad \max_{p \in [0,1]} e_{AB}^{(p)} = e_{AB}^{(o)} = q_1\,; \quad \max_{p \in [0,1]} e_{PEM}^{(p)} = e_{PEM}^{(1)} = 1.$$



Ce résultat confirme l'avantage de l'AB par rapport au PEM. En effet, si (17) n'est pas valable alors $d_{AB} = d_{PEM}$.

**Références**

W.E. DEMING and F.F. STEPHAN (1940) : On a least square adjustement of a sampled frequency table when the expected marginal totals are known. Ann. Math. Statist. 11 (1940), 427-444.

A. PEREZ (1956) : Notions généralisées d'incertitude, d'entropie et d'information du point de vue de la théorie de martingales. Trans. First Prague Conf. on Information Theory (1956), Academia (1957), 183-208.

E.T. JAYNES (1957) : Information Theory and Statistical Mechanics, Phys. Rev. 106 (1957), 620-630 ; 108 (1957), 171-182.

H.G. KELLERER (1964) : Verteilungsfunktionen mit gegeben Marginalverteilungen. Z. Wahrscheinlichkeitstheorie 3, 247-270.

A. PEREZ (1965) : Information, $\epsilon$-Sufficiency and Data Reduction Problems. Kybernetika 1 (1965), 297-323.

A. PEREZ (1967) : Information-Theoretic Risk Estimates in Statistical Decision. Kybernetika 3 (1967), 1-21.

I. CSISZAR (1975) : I-divergence geometry of probability distributions and minimization problems. The Annals of Probability 3 (1975), 146-158.

A. PEREZ (1977) : Simplifications $\epsilon$-admissibles de la structure de dépendance d'un ensemble de variables aléatoires. In : Actes du Colloque International no. 276 (1978) «Les développements récents de la théorie de l'information et leurs applications», organisé dans le cadre des colloques internationaux du CNRS, à Cachan, juillet 1977, par Claude Picard, 35-46.

A.PEREZ (1980) : Approximation of a stochastic system by a set of weakly mutually dependent sybsystems. In : Abhandlungen der Akademie der Wissenschaften der DDR No. 6, 1981 ; Statistische Nachrichtentheorie und ihre Anwendungen. 2. Internationalen Seminar – SHT 80 – in Schnett, 1980, 9-17.

A. PEREZ (1983) : Probability approach in integrating pieces of knowledge for medical decision making (in czech). In : Transactions BMI'83 – Conference on biomedical engineering, Marianske, Lazné, october 1983, 221-226.

A. PEREZ (1984) : Barycenter of a Set of Probability Measures and its Application in Statistical Decision. Proceedings COMPSTAT 1984, Physica-Verlag, Wien, 154-159.

A.PEREZ (1985) : Information theoretic approach to expert systems. In : Open Problems of the Second Joint Swedish-Soviet International Workshop on Information Theory, Gränna 1985, 19-22.

A. PEREZ, and R. JIROUSEK (1985) : Constructing an Intensional Expert System (INES). In : Medical Decision Making : Diagnostic Strategies and Expert Systems, Elsevier Science Publisher (North-Holland), IFIP-IMIA, 1985, 307-315.

P. HAJEK (1985) : Combining functions for certainty degrees in consulting systems. Inter. J. Man-Machine Studies, 22 (1985), 59-76.

A. PEREZ (1986) : The Barycenter Concept of a Set of Probability Measures as a Tool in Statistical Decision. In : Proceedings of the IV. Vilnius Conference on Probability Theory and Mathematical Statistics Vilnius 1985, VNU Science Press, Utrecht, 1986, Vol. II, 437-450.

R. JIROUSEK and A. PEREZ (1986) : A Partial Solution of the Marginal Problem. In : Trans. 10th Prague Conference on Information Theory (1986), Academia 1987.

M. STUDENY (1987) : The concept of multiinformation in probabilistic decision making. Candidate's dissertation, UTIA CSAV, Prague.

A. PEREZ (1987) : The Maximal Entropy Principle and the Barycenter Approach in Knowledge Integration. Presented at the Third Joint USSR-Swedish International Workshop on Information Theory, May 1987, Sochi (USSR).

REVUE INTERNATIONALE DE SYSTEMIQUE
Vol. 1, N° 3, pp. 295 à 318

# LES NOUVELLES SCIENCES sont bien des Sciences Repères historiques et épistémologiques

Jean-Louis LE MOIGNE

Université d'Aix-Marseille III *

Résumé

A partir de 1948, foisonnent de nombreuses nouvelles disciplines scientifiques, telles que la Cybernétique, l'Informatique, les Sciences de la Communication, de l'Organisation, de la Décision, de la Cognition, la Systémique, et plus récemment les sciences de l'Autonomie et de la Complexité. Ces nouvelles sciences qui ne sont pas toutes encore pleinement acceptées par les Institutions Scientifiques contemporaines, peuvent-elles légitimement prétendre au même statut scientifique que les sciences classiques (qui furent, elles aussi nouvelles autrefois) ? Ne risquent-elles pas de devenir bientôt des pseudo-sciences ? On propose de repérer les conditions d'émergence et d'institutionnalisation de ces nouvelles sciences contemporaines en s'efforçant de reconnaître leur mode de connaissance critique et leur capacité à produire et à valider des énoncés enseignables.

Cette discussion historique permet de mettre en évidence la progressive construction d'un discours épistémologique fondateur explicitement différent du discours positiviste et post-néo-positiviste qui prévalait jusqu'ici pour fonder les sciences classiques : l'épistémologie constructiviste et la nouvelle Rhétorique.

On argumente succinctement sa pertinence, sa généralité et son urgence pour la Science et donc pour toutes les disciplines scientifiques contemporains. On souligne l'effet catalytique de l'émergence de la Systémique à partir de 1970 dans ce développement épistémologique contemporain. On conclut en soulignant l'actualité des sciences fondamentales de l'ingéniérie, sciences de

* Faculté d'Economie Appliquée, GRASCE, (CNRS 935), 3 Avenue R. Schuman, 13100 Aix-en-Provence, France.

Revue internationale de systémique. 0980-1472 Vol. 1/87/03/295/24/$ 4.40/