Annales de la Fondation Louis de Broglie
Vol. 6, n° 1, 1981

MESURES PHYSIQUES ET THEORIE DE L'INFORMATION

par J. DAYANTIS

Centre de Recherches sur les Macromolécules
6, rue Boussingault
67083 STRASBOURG CEDEX

(Manuscrit reçu le 8 avril 1980)

Résumé : On montre d'abord que la formule de Brillouin pour l'information obtenue lors de la mesure d'une observation X ne peut être acceptée que si X est une quantité bien définie. Si X est une variable aléatoire (v.a.), la formule de Brillouin n'est pas acceptable et l'on montre qu'une quantification adéquate de l'information obtenue par une mesure approximative peut se faire dans ce cas en utilisant la théorie de Shannon pour le taux de transmission de l'information dans un canal bruité. On ne considère que des mesures macroscopiques, de sorte qu'il n'y a pas lieu de prendre en considération le principe d'incertitude de la mécanique quantique. On souligne, par la suite, l'analogie existant entre un appareil de mesure conventionnel et l'observateur humain faisant usage de ses sens. On montre qu'une même matrice de transition Q peut être utilisée dans les deux cas pour relier la distribution originale p (x) de la v.a. X avec la distribution modifiée par l'appareil de mesure, ce qui conduit à une distribution P (y) de la v.a. Y. La connaissance de l'une des distributions ci-haut et de la matrice de transition Q permet dans tous les cas de procéder à une quantification adéquate de l'information obtenue. Cependant, dans certains cas particuliers, cette quantification peut se faire par l'intermédiaire des seules distributions p (x) et P (y) sans connaissance de la matrice Q.

I. Introduction

Cet article traite de l'information obtenue lorsqu'une observable physique est connue de manière approximative. Cette connaissance approximative peut venir soit de mesures effectuées au moyen d'un instrument de mesure nécessairement imparfait, soit des estimations d'un observateur humain dénué d'instruments de mesure adéquats et faisant usage pour ses estimations uniquement de ses sens.

Les instruments de mesure sont en général imparfaits parce qu'ils sont d'une précision et d'une exactitude limitées. La signification donnée aux mots "*précis*" et "*exact*" est la suivante : un instrument de mesure est dit précis si un grand nombre de déterminations successives d'une même observable conduisent à une distribution serrée autour de quelque valeur moyenne ; l'instrument est de plus exact si ladite valeur moyenne coïncide ou est très proche de la valeur réelle de l'observable.

On ne considère ici que des mesures macroscopiques, de sorte que les problèmes qui font intervenir le principe d'incertitude de la mécanique quantique sortent du cadre de l'article. Par ailleurs, pour éviter toute confusion, nous tenons à préciser que nous ne serons pas concernés par l'estimation des différentes moyennes du calcul statistique, à déterminer à partir d'un ensemble de mesures de précision donnée. Ce problème, d'importance capitale en pratique, a été traité par le passé par certains grands mathématiciens et c'est durant les années 20 que R.A. Fisher ([1]) a donné à la théorie de l'estimation statistique, la forme définitive sous laquelle nous la connaissons aujourd'hui.

Notre propos restreint est ici de traiter du problème spécifique de l'information obtenue par l'intermédiaire de mesures approximatives, que celles-ci soient dues à des mesures appropriées ou à des estimations humaines.

Le mot "*information*" doit être compris dans son sens mathématique précis, dû à Shannon ([2-6]). Soulignons encore que le mot "*estimation*" n'a jamais dans ce qui suit le sens d'une moyenne statistique ; il s'agira toujours d'une estimation au sens courant du terme. Par exemple, un observateur pourra estimer, faisant usage de ses sens, la hauteur d'un arbre, la vitesse d'une voiture automobile sur la route, la hauteur d'un son musical et ainsi de suite. Soit donc une certaine propriété x d'un objet physique s. Le plus souvent d'ailleurs s fera partie d'un ensemble S d'objets de même nature. x sera alors, par exemple, la hauteur d'un arbre particulier dans une forêt, ou bien la vitesse d'une voiture particulière sur une autoroute et ainsi de suite. La propriété x étendue à l'ensemble de la population S devient alors une variable aléatoire (v.a.) X. Brillouin ([7]) pose que si L est le champ d'observation de la v.a. X et ΔX la précision de l'instrument de mesure, l'information obtenue lors d'une mesure particulière est donnée par

$$I = k \ln \frac{L}{\Delta X} \qquad (1)$$

où k est une certaine constante, prise par Brillouin, de manière plausible mais sans démonstration rigoureuse, égale à la constante de Boltzmann. On ne peut accepter la relation (1) comme étant une définition adéquate de l'information obtenue que dans le cas où X est une quantité bien définie, telle la charge ou la masse d'un électron et non pas une variable aléatoire. L'ensemble précédent S ne contient alors qu'un seul élément s et il n'existe pas une distribution de probabilités objectives p (x) pour la distribution de X, de sorte que l'on est en droit de poser que toutes les valeurs à priori possibles de X sont équiprobables, ce qui conduit à la relation de Brillouin. L'absence d'une distribution de probabilités p (x) pour X fait cependant que la formule (1) contient nécessairement un élément arbitraire, à savoir l'étendue du champ d'observation. Aussi, nombre de physiciens pourraient-ils être tentés de la rejeter même dans ce cas. Si par ailleurs X est une v.a., la relation (1) ne peut être admise comme étant une mesure adéquate de l'information pour la raison suivante : admettons que ΔX soit suffisamment petit pour qu'en pratique on puisse admettre que l'instrument de mesure permet de déterminer la distribution des valeurs prises par X de manière quasi-parfaite. La relation (1) nous dit alors que l'information obtenue lorsque l'on effectue une mesure particulière dépend de la précision ΔX et du champ d'observation L mais nullement de la distribution de la v.a. X dans la population S. Or, l'essence même de la théorie de Shannon [2,6] tient à ce que l'information est justement fonction de la distribution de X. On sait que c'est cette théorie de Shannon qui a ouvert l'ère mathématique de la théorie de l'information. Comme on peut le constater sur la figure 1, la formule de Brillouin n'est adéquate que dans le cas où les valeurs de X dans le champ d'observation L sont équiprobables. On examine plus en détail ce point dans la discussion (paragraphe VI). Dans ce qui suit, nous voudrions montrer qu'une quantification convenable peut être obtenue très simplement par une extension directe de la théorie de Shannon pour le taux de transmission de l'information au travers d'un canal bruité, théorie qui fait usage du concept "*d'équivocation*" [8]. Nous montrons schématiquement, ci-

bas, l'analogie, au moyen des cas A, B et C. Le cas A concerne le taux de transmission de l'information dans un canal bruité, d'une source à un observateur. Le cas B concerne la mesure d'une observable physique au moyen d'un instrument de mesure imparfait. Enfin, dans le cas C, c'est l'observateur lui-même qui fait fonction d'instrument de mesure.

A.	*source*	→ *signal*	*canal bruité* →	*signal équivoque*
B.	*observable* →	*mesure en utilisant un appareil imparfait*	*convulation de l'observable et de l'appareil de mesure* →	*mesure approximative*
C.	*observable* →	*estimation d'un observateur*	*convulation de l'observable et de l'observateur* →	*mesure approximative*

Les observateurs humains sont donc assimilés dans de qui suit à des instruments de mesure, lorsqu'ils effectuent des estimations. De fait, la différence entre les deux est plus quantitative que qualitative. La différence qualitative essentielle est la suivante : nous nous attendons à ce qu'un instrument de mesure approprié réponde, compte tenu de sa précision, de manière identique à des stimulations identiques. Par contre, l'on ne s'attend pas à ce qu'il en soit en général ainsi pour un observateur humain. La réponse de ce dernier peut, en effet, être influencée par différents facteurs externes ou internes à l'observateur, objectifs ou subjectifs. A titre d'exemple, on peut citer l'heure du jour, l'éclairement, la condition psychologique de l'observateur au moment de l'estimation. Toutefois, pour le propos de l'article, nous supposerons que les estimations humaines satisfont à certaines conditions de stationnarité, ce qui doit permettre d'assurer un minimum de reproductibilité des estimations de l'observateur. Ce point est examiné un peu plus en détail au paragraphe VI.

Le problème de l'information obtenue lorsque l'on effectue des mesures physiques a été examiné par le passé par de nombreux chercheurs, surtout des mathématiciens ([5], [13]). Le résultat essentiel, à savoir nos relations (23) et (23') tenant compte de l'équivocation de l'appareil de mesure remonte au moins à Lindley ([13]). Cependant, ce résultat semble moins connu des physiciens et surtout des expérimentateurs, qui pourtant sont en contact quotidien avec des instruments de mesure. La valeur de l'information obtenue par une mesure physique, telle qu'elle est donnée dans le livre populaire de Brillouin, est à notre sens inexacte ou, à tout le moins incomplète. Ailleurs, le problème n'est même pas abordé. Aussi, l'utilité du présent article pourrait surtout être de diffuser des choses connues, mais cependant insuffisamment connues des principaux intéressés, en utilisant une approche élémentaire et intuitive, qui s'appuie sur la théorie de Shannon pour le taux de transmission de l'information dans un canal bruité. Ceux qui voudraient compléter la présente approche par des approches plus générales et plus rigoureuses, devraient se reporter aux travaux des mathématiciens ([5], [13]).

II. Equivocation et taux de transmission de l'information

Afin de rappeler la théorie de Shannon sur le taux de transmission de l'information dans un canal bruité, il convient d'abord d'introduire les différentes "*entropies*" concernées. Nous donnons d'abord la définition dans le cas discret, suivie, après le symbole d'identité ≡ par la définition dans le cas continu. Pour ne pas être redondant, l'explication des symboles est donnée seulement dans le cas discret ceux du cas continu devenant de la sorte évidents.

Soit deux variables aléatoires X et Y où X peut prendre les valeurs discrètes 1, 2,...i,... et Y, les valeurs discrètes 1, 2,...j,... ; si p_{ij} est la probabilité pour que X prenne la valeur i et Y la valeur j, alors l'entropie de la distribution des deux v.a. X et Y est donnée, d'après Shannon ([9]) par :

$$H\,[X,Y] = -\sum_{ij} p_{ij} \ln p_{ij} \equiv -\iint p(x,y) \ln p(x,y)\,dx\,dy \quad (2)$$

Dans le premier membre de l'équation 2, des crochets sont mis à la place des parenthèses habituelles, pour souligner que H n'est pas une fonction de X et Y dans le sens usuel mais une quantité bien définie par les distributions de probabilité de ces variables aléatoires. L'entropie de l'une de ses distributions indépendemment de l'autre est donnée par :

$$H\,[X] = -\sum_{i} p_i \ln p_i \equiv -\int p(x) \ln p(x)\,dx \quad (3)$$

$$H\,[Y] = -\sum_{j} p_j \ln p_j \equiv -\int p(y) \ln p(y)\,dy \quad (4)$$

où

$$p_i = \sum_j p_{ij} \qquad p(x) = \int p(x,y)dy \qquad (5a)$$

$$p_j = \sum_i p_{ij} \qquad p(y) = \int p(x,y)dx \qquad (5b)$$

Si, comme cela est souvent le cas, X et Y sont interdépendants. Nous écrirons :

$$p_{ij} = p_i q_{ij} = p_j Q_{ji} \qquad p(x,y) = p(x)q_x(y) \qquad (6a)$$

$$= p(y)q_y(x) \qquad (6b)$$

ce qui revient à dire que la probabilité pour que X et Y prennent respectivement les valeurs i et j est égale à la probabilité pour que X prenne la valeur i multipliée par la probabilité conditionnelle q_{ij} pour que Y prenne la valeur j lorsque l'on sait que la valeur prise par X est i(et raisonnement analogue si on utilise la probabilité conditionnelle Q_{ji}). En portant (6) dans (2), on trouve facilement que :

$$H[X,Y] = H[X] + H_X[Y] = H[Y] + H_Y[X] \qquad (7)$$

$$H_X[Y] = -\sum_{ij} p_{ij} \ln q_{ij} = -\int p(x,y)\ln q_x(y)dxdy \qquad (8a)$$

$$H_Y[X] = -\sum_{ij} p_{ij} \ln Q_{ji} = -\int p(x,y)\ln q_y(x)dxdy \qquad (8b)$$

L'entropie conditionnelle $H_X[Y]$ mesure l'incertitude moyenne sur la valeur de Y quand la valeur prise par X est connue ; de même, l'entropie conditionnelle $H_Y[X]$ mesure l'incertitude moyenne sur la valeur de X quand la valeur prise par Y est connue.

La théorie de Shannon sur le taux de transmission de l'information dans un canal bruité stipule que si une source produit de la néguentropie ([10]) au taux de $H[X]$ par seconde et que cette néguentropie est transmise par un canal bruité, la néguentropie effectivement transmise par seconde est moindre que la néguentropie produite à la source et est donnée par :

$$R = H[X] - H_Y[X] = H[Y] - H_X[Y] \qquad (9a,9b)$$

L'entropie conditionnelle $H_Y[X]$ est appelée par Shannon "*l'équivocation*" et mesure notre incertitude sur la valeur de X à la source lorsque l'on spécifie la valeur de Y à l'arrivée.

Considérons maintenant, qu'au lieu de vouloir transmettre un signal par l'intermédiaire d'un canal bruité, nous désirons déterminer la distribution d'une certaine propriété X dans une population S. Il n'est pas nécessaire de préciser la nature de la population. Cela pourrait être, par exemple, notre précédente population d'arbres dans une forêt dont nous désirons connaître la distribution des hauteurs, ou encore la population d'automobiles à un moment donné sur une autoroute dont nous désirons connaître la distribution des vitesses. S'il était possible de déterminer très exactement la distribution vraie pour la hauteur des arbres dans la forêt (que nous supposerons continue) alors l'information obtenue lorsque l'on détermine la hauteur d'un arbre particulier serait donnée par la mesure de Shannon :

$$I[X] = -\int p(x)\ln p(x)dx \qquad (3')$$

Toutefois, en raison de l'imperfection des appareils de mesure et des erreurs d'observation, la distribution exacte normée p(x) est remplacée par une distribution approchée normée P(y) (voir fig. 2). On appellera cette distribution modifiée la distribution convoluée. En d'autres termes, dans un cas spécifique donné, la valeur mesurée pour X n'est pas trouvée être x (valeur exacte) mais (y) (valeur approchée). x et y peuvent différer de manière plus ou moins importante, d'après les caractéristiques de l'appareil de mesure. Une seconde mesure de X conduira à la valeur y' en général différente de y et ainsi de suite. La distribution normée globale P(y) pourra différer de manière plus ou moins importante de la distribution réelle p(x). L'information obtenue, à partir de la distribution P(y) ne peut être donnée par la relation (3') où l'on a simplement remplacé p(x) par P(y). Il est facile de montrer que telle procédure conduirait à des absurdités. De fait, le bruit dans un canal est remplacé ici par la nature approximative de toute mesure physique. Le message reçu comportant de l'équivocation est l'analogue de la distribution convoluée P(y). Comme conséquence de cette équivalence, l'information obtenue par des mesures approximatives peut être traitée de manière analogue à l'information obtenue en utilisant un canal bruité.

III. Distributions exactes et convoluées d'observables physiques - La matrice de transition

Considérons d'abord un système physique unique S dont nous désirons déterminer certaine propriété X. Soit x_0 dans ce cas particulier la valeur de X. Si la propriété est mesurée un très grand nombre de fois en utilisant un même instrument de mesure de caractéristiques données, la valeur x_0 conduira à une distribution $q_{x_0}(y)$ comme conséquence des fluctuations de l'appareil de mesure et des erreurs d'observation. $q_{x_0}(y)$ est la fonction convoluée de la fonction de Dirac $\delta(x-x_0)$ et de l'appareil de mesure ([11]). Supposons maintenant qu'au lieu d'avoir un système unique s, nous ayons une population s_i de systèmes formant un ensemble S, de sorte qu'à chaque s_i corresponde une valeur x_i de l'observable X. Si la distribution des x_i est continue ou peut être approchée par une distribution continue, nous pouvons la représenter par une distribution normée p(x). Pour chaque valeur de x la convolution avec l'instrument de mesure conduira à une distribution convoluée $q_x(y)$. Dans le cas particulier où l'on peut postuler l'invariance par translation de l'observable des caractéristiques de l'appareil de mesure, c'est-à-dire dans le cas où

$$q_{x_j}(y) = q_{x_k}(y + x_k - x_j) \tag{10}$$

la distribution globale P(y) est donnée par l'intégrale usuelle de convolution.

$$P(y) = p(x) * q(x) = \int p(x)q(y - x)dx \tag{11}$$

où dans q(x) l'indice x a été éliminé en tant que superflu.

Dans ce cas, en faisant usage d'un théorème classique, si T_p et T_q sont les transformées de Fourier de p(x) et de q(x), la transformée de Fourier de P(y) est donnée par :

$$T_P = T_p \cdot T_q \tag{12}$$

Cependant, dans le cas général, l'hypothèse ([11]) ne sera pas acceptable et nous nous attendons en particulier à ce qu'il en soit ainsi lorsque des observateurs humains se comportent en tant qu'instruments de mesure (voir paragraphe IV). Dans ce cas, la distribution $q_x(y)$ dépendra en général de la valeur particulière x considérée, de sorte que l'on peut poser $q_x(y) \equiv q(x,y)$ et la convolution de p(x) avec l'appareil de mesure sera donnée par :

$$P(y) = \int p(x)q(x,y)dx \tag{13}$$

La relation (13) est une sorte d'intégrale de convolution généralisée, pour laquelle la relation (12) n'est plus en général vérifiée.

Une autre définition, qui sera utile, de la distribution convoluée P(y) peut être donnée en utilisant une matrice de transition. Soit L le champ d'observation (que nous supposons fini) et divisons-le en N parties égales au moyen de N + 1 points (fig. 2). Soit alors q_{ij} la probabilité conditionnelle pour que Y prenne une valeur comprise entre j, j + 1 lorsque l'on sait que x a pris une valeur comprise entre i, i + 1. Les q_{ij} définissent une matrice carrée d'ordre N que nous supposerons régulière et qui est stochastique, c'est-à-dire que nous avons $\sum_j q_{ij} = 1$ avec $q_{ij} \geqslant 0$ quels que soient i et j. Définissons maintenant deux vecteurs N-dimensionnels $\vec{p}$ et $\vec{P}$, tels que les composantes $p_1, \ldots p_N, P_1, \ldots P_N$ soient proportionnelles aux aires délimitées respectivement par i, i + 1 et p(x) et j, j + 1 et P(y) respectivement (fig. 2). Appelons ces vecteurs, pour lesquels nous avons $\sum_i pi = 1$, $\sum_j P_j = 1$, avec p_i, $P_j \geqslant 0$, vecteurs stochastiques. Nous avons, de manière évidente :

$$\vec{P} = {}^t Q \, \vec{p} \tag{14}$$

où ${}^tQ$ est la matrice transposée de Q. Comme nous avons supposé Q régulière, nous avons aussi :

$$\vec{p} = ({}^tQ)^{-1}\vec{P} \tag{15}$$

où $({}^tQ)^{-1}$ est la matrice inverse de ${}^tQ$. Il est facile de montrer que si $\vec{p}$ est un vecteur stochastique et Q une matrice stochastique, $\vec{P}$ est nécessairement un vecteur stochastique. En effet, puisque $P_j = \sum_i p_i q_{ij}$, nous avons :

$$\sum_j P_j = \sum_{ji} p_i q_{ij} = \sum_i p_i \sum_j q_{ij} = \sum_i p_i = 1$$

On peut appeler $\vec{P}$ le vecteur convolué de $\vec{p}$ par Q. On admettra que lorsque N tend vers l'infini, les vecteurs $\vec{p}$ et $\vec{P}$ tendront à représenter exactement les distributions p(x) et P(y). Cependant, il ne sera pas toujours nécessaire de faire tendre N vers l'infini pour obtenir de manière exacte les distributions p(x) et P(y). En effet, selon le théorème de l'échantillonnage de Shannon ([12]), pour pouvoir reconstruire très exactement une fonction, il suffit de connaître sa valeur en M points équidistants, où M est donné par :

$$M = 2T\nu_{max} \tag{16}$$

Si la fonction est périodique, T est sa période et ν_{max} la fréquence maximale de sa décomposition en série de Fourier ; si elle ne l'est pas T sera son support et ν_{max} la fréquence maximale de son spectre de Fourier. Le corollaire de ceci est évidemment que si la fréquence maximale est infinie, la fonction n'est pas reconstructible de manière exacte à partir d'un nombre fini de ses points. Prenons donc pour ν_{max}

la plus grande des deux valeurs définies par p(x) et P(y), il suffit alors de diviser le champ L en :

$$N = \frac{L}{T} \cdot M = 2L\nu_{max} \qquad (17)$$

points équidistants. N sera aussi la dimension de la matrice de transition Q et des vecteurs stochastiques $\vec{p}$ et $\vec{P}$ associés.

Si ν_{max} est infini, on peut toujours travailler avec N points équidistants, qui vont définir des vecteurs $\vec{p}$ et $\vec{P}$ représentant les fonctions p(x) et P(y) de manière approchée. On peut ensuite chercher à définir de manière précise p(x) et P(y) en faisant tendre N vers l'infini. L'alternative à cette situation est d'utiliser l'intégrale de convolution, les relations (11) et (13). L'approfondissement de ce point sort du cadre de l'article.

Voici quelques propriétés de la matrice Q. Dans le cas le plus général, il n'y a pas de relations entre les N^2 éléments matriciels, sauf pour les N conditions stochastiques.

$$\sum_j q_{ij} = 1 \qquad q_{ij} \geqslant 0 \qquad (18)$$

Si l'on suppose qu'il n'y a pas d'éléments matriciels égaux entre eux (et qu'en particulier il n'y ait pas plus d'un élément nul), nous aurons au total $N^2 - N = N(N - 1)$ termes indépendants.

Si l'on se donne deux vecteurs stochastiques arbitraires $\vec{p}$ et $\vec{P}$, l'équation (15) ne peut être résolue pour déterminer les q_{ij}, ou plus exactement, les q_{ij} sont indéterminés. Ce cas le plus général correspond à l'équation (13) du cas continu, où la fonction

q(x,y) ne peut être déterminée à partir des distributions p(x) et P(y). Physiquement, cela signifie que la réponse de l'appareil de mesure à un très grand nombre de mesures portant sur une même valeur x_0 de la v.a. X ne peut être déterminée à partir des distributions globales p(x) et P(y). En théorie de la communication, si le signal envoyé par la source, parmi N possibles, est i, ce signal peut être reçu par l'observateur en tant que i, j, k... en raison de l'équivocation. La statistique globale des signaux émis à la source et reçus par l'observateur ne permet pas alors à ce dernier de définir, à priori, la distribution des signaux reçus à l'arrivée si un même signal i est envoyé un très grand nombre de fois par la source. Pour illustrer ce qui précède, prenons N juste égal à trois. La matrice de transition est :

$$\begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix}$$

où nous avons les trois conditions stochastiques.

$$\sum_j q_{1j} = 1 \qquad \sum_j q_{2j} = 1 \qquad \sum_j q_{3j} = 1 \qquad (18')$$

qui donnent trois relations entre les q_{ij}. Il reste six termes indéterminés. Si l'on se donne deux vecteurs stochastiques $\vec{p}$ et $\vec{P}$, on a trois relations supplémentaires, venant de l'équation (14) et reliant les q_{ij}. Restent trois q_{ij} dont les valeurs peuvent être choisies arbitrairement compte tenu des conditions stochastiques et le problème est indéterminé. Si maintenant on considère l'équation (11), l'équation de convolution usuelle, nous aurons les conditions supplémentaires que tous les q_{ij} sont égaux pour lesquels i - j est constant : tous les termes parallèles à la diagonale principale sont égaux. Nous avons :

$$N - 1 + 2\,[(N - 2) + (N - 3) + \ldots + 1] = (N - 1)^2$$

conditions additionnelles, lesquelles ajoutées aux N conditions stochastiques font un total de $N^2 - N + 1$ relations. S'il n'y a pas d'autres contraintes, le choix de deux vecteurs stochastiques reliés par une relation résoudra exactement le problème de la détermination des q_{ij}. Si l'on introduit des contraintes supplémentaires (sous la forme par exemple d'un certain nombre de termes nuls) alors les vecteurs $\vec{p}$ et $\vec{P}$ devront être reliés par un nombre de relations égal à n, si n est le nombre de contraintes en excès de N^2. Autrement, le problème sera surdéterminé et en général impossible. Considérons par exemple, la matrice tridimensionnelle précédente avec la condition supplémentaire que les q_{ij} pour lesquels i - j est constant sont égaux entre eux. Nous avons alors la matrice :

$$\begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{11} & q_{12} \\ q_{31} & q_{21} & q_{11} \end{pmatrix}$$

qui après utilisation des conditions stochastiques (18') se réduit à :

$$\begin{pmatrix} q_{11} & q_{12} & 1-(q_{11} + q_{12}) \\ 1-(q_{11} + q_{12}) & q_{11} & q_{12} \\ q_{12} & 1-(q_{11} + q_{12}) & q_{11} \end{pmatrix}$$

où il n'y a plus que deux inconnues q_{11} et q_{12}. La solution du problème si deux vecteurs stochastiques $\vec{p}$ et $\vec{P}$ sont donnés ne sera possible que si une relation existe entre ces derniers qui est donnée en annulant le déterminant.

$$D = \begin{Vmatrix} P_1 - p_2 & P_1 - P_2 & P_3 - P_2 \\ P_2 - p_3 & P_2 - P_3 & P_1 - P_3 \\ P_3 - p_1 & P_3 - P_1 & P_2 - P_1 \end{Vmatrix} = 0$$

Nous avons considéré jusqu'ici la possibilité de résoudre algébriquement l'équation (14) pour les éléments q_{ij} de la matrice Q soumis aux conditions stochastiques, si l'on se donne deux vecteurs stochastiques $\vec{p}$ et $\vec{P}$. Toutefois, même si l'on se place dans le cas où le problème est algébriquement soluble (égalité du nombre des contraintes et des inconnues), certaines de ces solutions, dépendant du choix des vecteurs $\vec{p}$ et $\vec{P}$ ne seront pas physiquement acceptables, pouvant conduire à des valeurs négatives pour certains des q_{ij}. Par conséquent, si l'on se choisit un vecteur stochastique $\vec{P}$, le choix d'un deuxième vecteur stochastique $\vec{p}$ ne peut être complétement arbitraire. $\vec{P}$ devra être choisi dans un sous-espace, s'il y en a, des P_j, de façon que tous les q_{ij} soient positifs ou nuls. Il sort du cadre de ce travail d'essayer de définir le découpage de l'espace des P_j résultant du choix d'un vecteur arbitraire $\vec{p}$. Disons simplement que puisqu'aussi bien $\vec{p}$ que $\vec{P}$ sont des quantités qui dérivent de l'expérience, nous nous attendons à ce que ceci soit une condition suffisante pour assurer la non-négativité des q_{ij}.

IV. La courbe d'estimation de l'observateur

Nous avons déjà souligné dans l'introduction que les estimations humaines d'observables physiques peuvent être assimilées à des mesures physiques obtenues grâce à des instruments de mesure si certaines conditions de stationnarité sont satisfaites. Un exemple peut illustrer comment les fonctions supposées stationnaires $q_x(y)$ et $P(y)$ d'un observateur humain peuvent être construites. Soit l'exemple précédent d'automobiles sur une autoroute. Nous pouvons préciser le temps qu'il fait, la saison, le jour de la semaine et l'heure. Avec ces restrictions, il est naturel d'admettre qu'il existe une distribution objective des vitesses $p(v)$ sensiblement stationnaire pour la durée de l'expérience. Soit maintenant, un observateur, posté aux abords de l'autoroute, auquel on demande d'estimer la vitesse des voitures qui passent devant lui. On peut lui demander, par exemple, de faire ses estimations par intervalles de 2 ou 5 kilomètres, (ceci définit le découpage du champ d'observation). Au même moment, la vitesse précise de chaque automobile est mesurée par d'autres observateurs, disposant d'instruments de mesure adéquats. Lorsqu'un nombre statistiquement significatif d'estimations et de mesures aura été obtenu, l'ensemble des estimations est divisé en sous-ensembles, chaque sous-ensemble groupant toutes les estimations correspondant à une vitesse mesurée donnée,(à l'intervalle de tolérance précédemment défini près). Ceci va donner la fonction de probabilité conditionnelle $q_v(y)$ de l'observateur, ou encore les éléments q_{ij} de la matrice de transition Q. La distribution estimée globale des vitesses $P(v)$ sera appelée la "*courbe d'estimation*" de l'observateur et sera le produit de convolution de la distribution réelle $p(v)$ par l'instrument de mesure qu'est en l'occurence l'observateur. L'introduction des distributions $q_v(y)$ et $P(v)$ est utile, car ceci permet de quantifier l'information d'excès obtenue par

un observateur avec un appareil adéquat, qui a procédé à une estimation préalable (voir section V).

Supposons par exemple, que la distribution réelle des vitesses est une gaussienne centrée sur la valeur v_0 et d'écart-type σ_0 et que la réponse de l'observateur $q_v(y)$ est une gaussienne d'écart-type σ_1 indépendant de la valeur de v considérée. Avec ces hypothèses, nous sommes dans le cas correspondant à l'équation (11), l'équation de convolution usuelle, qu'il est dans ce cas approprié d'utiliser directement. Nous avons en effet :

$$p(x) = (2\pi\sigma_0^2)^{-1/2}\exp(-x^2/2\sigma_0^2) \qquad (19a)$$

$$q_x(y) = (2\pi\sigma_1^2)^{-1/2}\exp\{-(x-y)^2/2\sigma_1^2\} \qquad (19b)$$

où l'on a posé $x = v - v_0$. T signifiant la transformée de Fourier, nous avons :

$$T\,\{(2\pi\sigma_0^2)^{-1/2}\exp(-x^2/2\sigma_0^2)\} = \exp(-2\pi\sigma_0^2K^2) \qquad (20)$$

et en utilisant (12) :

$$T_{P(y)} = \exp\{-2\pi K^2(\sigma_0^2 + \sigma_1^2)\}$$

de sorte que :

$$P(y) = \{2\pi(\sigma_0^2 + \sigma_1^2)\}^{-1/2}\exp\{-y^2/2(\sigma_0^2 + \sigma_1^2)\} \qquad (21)$$

La courbe d'estimation de l'observateur, qui est un cas particulier de distribution convoluée est une gaussienne centrée sur la valeur v_0 comme la distribution réelle mais d'écart-type $(\sigma_0^2 + \sigma_1^2)^{1/2}$ au lieu de σ_0. Observons que dans ce cas particulier si $p(x)$ et $P(y)$ sont donnés, l'équation (11) peut être déconvoluée de façon à obtenir

$q_x(y)$. Le cas discret équivalent est celui où tous les éléments q_{ij} de Q peuvent être déterminés sans ambiguïté par la donnée des deux vecteurs $\vec{p}$ et $\vec{P}$.

V. La quantification de l'information

On obtient la quantification désirée par l'intermédiaire des équations (9a) et (9b). En introduisant les vecteurs stochastiques $\vec{p}$ et $\vec{P}$ ainsi que la matrice Q, les entropies correspondantes s'écrivent :

$$H(X) = -\sum_i p_i \ln p_i \tag{22a}$$

$$H(Y) = -\sum_j P_j \ln P_j \tag{22b}$$

$$H_X(Y) = -\sum_{ij} p_i q_{ij} \ln q_{ij} \tag{22c}$$

$$H_Y(X) = -\sum_{ij} P_j Q_{ji} \ln Q_{ji} \tag{22d}$$

où les Q_{ji} sont les éléments de la matrice Q^{-1}. Nous obtenons ainsi, à partir de (9a) et (9b) :

$$I = -\sum_i p_i \ln p_i + \sum_{ij} P_j Q_{ji} \ln Q_{ji} = -\sum_j P_j \ln P_j + \sum_{ij} p_i q_{ij} \ln q_{ij} \tag{23}$$

Dans le cas continu, nous écrirons de manière analogue :

$$I = -\int p(x) \ln p(x) dx + \iint p(x,y) \ln Q_y(x) dx dy = -\int P(y) \ln P(y) dy + \iint p(x,y) \ln q_x(y) dx dy \tag{23'}$$

Rappelons que $q_x(y)$ est la distribution conditionnelle des y lorsque x est fixé, $Q_y(x)$ la distribution conditionnelle des x lorsque y est fixé.

L'information obtenue est au plus égale à l'entropie de la distribution p(x) de la v.a. X. L'égalité est acquise lorsque q_{ij}, $Q_{ij} = \delta_{ji}$, où δ_{ij} est le symbole de Kronecker égal à 1 si i = j et zéro si $i \neq j$. Physiquement ceci signifie que les mesures reproduisent exactement les valeurs prises par l'observable, de sorte que les distributions p(x) et P(y) (ou encore les vecteurs $\vec{p}$ et $\vec{P}$) se confondent. L'autre limite est celle où l'information obtenue est nulle et elle se produit lorsque deux v.a. X et Y ne sont pas corrélées, de sorte que les probabilités conditionnelles q_{ij} et Q_{ji} sont respectivement égales aux probabilités globales P_j et p_i respectivement. Dans ce cas, nous avons par exemple :

$$\sum_{ij} p_i q_{ij} \ln q_{ij} = \sum_i p_i \sum_j P_j \ln P_j = + \sum_j P_j \ln P_j$$

de sorte que I calculée à partir de (23) est nulle.

Observons que dans certains cas particuliers, l'information I peut être calculée à partir des deux distributions marginales p(x) et P(y), lorsque l'équation (13) peut être déconvoluée pour donner la distribution de probabilité conditionnelle $q_x(y)$ (ou, de manière alternative, si à partir des vecteurs $\vec{p}$ et $\vec{P}$, on peut obtenir sans ambiguïté les éléments de la matrice de transition Q). Si cela n'est pas possible, la quantification à partir des seules distributions p(x) et P(y) ne sera pas non plus possible.

Supposons maintenant que nous effectuons deux mesures successives de la même valeur d'une observable en utilisant deux instruments de mesure de précision et d'exactitude différentes. Soit I_1 et I_2, les informations respectives obtenues, considérées indépendemment l'une de l'autre. Nous avons :

$$I_1 = H^{(1)}[Y] - H_X^{(1)}[Y] \qquad (24a)$$

et

$$I_2 = H^{(2)}[Y] - H_X^{(2)}[Y] \qquad (24b)$$

Si l'instrument 2 est plus précis et exact que l'instrument 1, l'information d'excès obtenue lorsque l'on effectue une mesure avec l'instrument 2 après en avoir effectué une avec l'instrument 1 est donnée, en tenant compte de ce que $H^{(1)}[X] \equiv H^{(2)}[X]$, par

$$\Delta I = I_2 - I_1 = H^{(2)}[Y] - H^{(1)}[Y] - \{H_X^{(2)}[Y] - H_X^{(1)}[Y]\} = H_Y^{(1)}[X] - H_Y^{(2)}[X] \qquad (25)$$

En particulier, l'instrument de mesure 1 peut être simplement un observateur humain faisant usage de ses sens ou de sa connaissance antérieure de la situation pour procéder à une estimation approximative. L'équation (25) donne alors l'information d'excès obtenue lorsque l'observable est par la suite mesurée avec plus de précision grâce à un instrument de mesure adéquat.

VI. Discussion

Le but du présent article est de mettre l'accent sur deux analogies :

- la première, c'est l'analogie entre la théorie de Shannon pour le taux de transmission de l'information dans un canal bruité et l'information obtenue par des mesures effectuées avec des instruments de mesure d'exactitude et de précision nécessairement limitées. Le cas idéal de mesures effectuées avec des instruments parfaits est analogue au cas de la transmission de l'information dans un canal sans bruit ;
- la seconde analogie est celle qui existe entre un instrument de mesure proprement dit et un observateur humain procédant à des estimations, utilisant pour cela ses sens et éventuellement la connaissance préalable qu'il peut avoir d'une certaine situation.

Ces deux analogies conduisent à un traitement unifié des cas A, B et C considérés dans l'introduction.

Examinons maintenant certaines différences propres à ces analogies. En ce qui concerne la première analogie, la quantité R donnée par l'équation (9) est une quantité purement formelle, alors que la quantité I de l'équation (23) a toujours une signification sémantique. On peut en effet, à partir d'une source envoyer une série de signaux n'ayant aucune signification dans aucun langage connu. Malgré cela, l'information "*de l'ingénieur*", selon Shannon, sera toujours donnée par (9). Par ailleurs, toute série de mesures d'une observable X a toujours un contenu sémantique. La même remarque est valable lorsqu'un observateur humain se comporte en tant qu'instrument de mesure.

S'agissant de la deuxième analogie, on a déjà observé que lorsque l'on considère des observateurs humains, il convient d'introduire de manière explicite une hypothèse de stationnarité des distributions de probabilités, hypothèse que l'on fait de manière implicite pour les appareils de mesure usuels. En d'autres termes, il est nécessaire de faire l'hypothèse que les éléments de la matrice de transition Q d'un observateur humain sont constants, ce que l'on admet de manière quasi-automatique pour les appareils usuels.

On peut cependant faire un pas de plus et introduire dans le cas des observateurs humains, une matrice Q dont les éléments dépendraient d'un certain nombre de paramètres α_i de sorte que l'on aurait $q_{ij} = q_{ij}(\alpha_i)$. Parmi ces paramètres, le temps t à partir d'une origine à préciser dans chaque cas spécifique ainsi que l'intervalle de temps $\Delta\tau$ entre une mesure et la suivante jouent à coup sûr un rôle important. Il est possible que l'introduction de tels paramètres, que nous ne faisons que suggérer ici, puisse éliminer toute différence entre instruments de mesure conventionnels et observateurs humains, d'autant plus qu'au niveau des paramètres q la différence semble de nouveau de nature plus quantitative que qualitative.

Ce point ne peut évidemment être tranché qu'à partir d'une expérimentation adéquate, toutefois, à la connaissance de l'auteur la matrice Q (ainsi du reste que les vecteurs stochastiques $\vec{p}$ et $\vec{P}$) n'ont jamais été utilisés dans pareille perspective.

Revenons maintenant à l'équation (1) de Brillouin et montrons que cette équation est un cas particulier de la relation générale (23). Divisons en effet, le champ d'observation L en N parties égales, de sorte que le ΔX de la figure 1 couvre N' divisions. Supposons que la probabilité à priori pour la valeur que peut prendre X dans le champ d'observation est constante dans tout le champ et égale à 1/N (en d'autres termes p(x) se réduit à une

constante) ; supposons de surcroît que si ΔX couvre le domaine D, D + N', alors les probabilités conditionnelles Q_{ji} sont toutes égales à 1/N' si i et j sont à l'intérieur du domaine ΔX et nulles autrement. A partir de la première des équations (23), prenant en considération que $0 \ln 0 = 0$ et que $\sum_j P_j = 1$, on trouve que :

$$I = \ln N + \sum_j P_j \sum_i \frac{1}{N'} \ln N' = \ln N - N' \frac{1}{N'} \ln N' = \ln \frac{N}{N'}$$

Ceci n'est autre chose que la relation de Brillouin où la constante arbitraire k a été éliminée. Elle peut être utilisée comme suite à ce qui a été dit dans l'introduction, que X soit une v.a. ou non. Resserrons maintenant la discussion au cas où X est une v.a. Une ligne de pensée est de dire que si au départ nous ne savons rien de la distribution de X, alors il est permis de poser que toutes les valeurs de X sont équiprobables. Par exemple, l'observateur qui ne saurait rien de la distribution de la vitesse des véhicules sur l'autoroute aurait le droit de poser que l'information qu'il obtient lorsqu'il évalue la vitesse d'un véhicule donné est $I = -k \ln \frac{\Delta V}{V_{max}}$ où l'on admet seulement que l'observateur connaît la vitesse maximale possible (ce qui délimite le champ d'observation) ainsi que la précision de ses estimations. Par la suite, s'il acquière quelque meilleure connaissance de la distribution des vitesses, l'information obtenue devrait être modifiée en conséquence pour tenir compte de cette connaissance supplémentaire. Nous voudrions critiquer cette approche, d'abord parce qu'elle nous semble ambiguë, ensuite parce qu'à notre sens l'ignorance où l'on peut se trouver sur la distribution p(x) n'autorise pas de poser que p(x) est une constante. Pour se tirer d'affaire, on peut proposer la démarche suivante : l'ensemble des observateurs est divisé en deux classes distinctes 1 et 2. Les observateurs de la classe 1 sont supposés ignorer la distribution p(x) et sont soumis par les observateurs de la classe 2, qui eux sont supposés connaître p(x), à des tests préliminaires ayant pour objet de déterminer la matrice de transition Q de chaque observateur de la classe 1. Lorsque les tests préliminaires seront complétés, on pourra demander aux observateurs de la classe 1 de procéder à leurs estimations proprement dites. Les observateurs de la classe 2 sont alors en mesure de calculer, pour tout observateur de la classe 1, la quantité d'information que chacun de ces derniers acquiert lorsqu'il procède à une estimation (ce que ne peuvent faire les observateurs de la classe 1 eux-mêmes). Evidemment, cette information dépend pour chaque observateur de son "*équation personnelle*" (sa matrice Q) et est donc une quantité variable selon l'observateur, ce qui est conforme à la philosophie poursuivie dans cet article.

Cette démarche a pour avantage de rendre constante l'information qu'obtient un observateur donné qui effectue une estimation, sauf que le calcul réel ne peut être effectué que par un observateur de la classe 2, après que les tests préliminaires auront été effectués. Par ailleurs, les observateurs de la classe 1 peuvent être transformés en observateurs de la classe 2 (et être ainsi en mesure de calculer eux-mêmes l'information qu'ils obtiennent lors d'une estimation) si les observateurs de la classe 2 leur communiquent le résultat des tests et le vecteur $\vec{p}$. Le recours à la relation (1) de Brillouin est ainsi évité, ainsi que la modification graduelle de cette dernière à mesure que les connaissances s'accumulent sur le vecteur $\vec{p}$ et les matrices Q. L'approche est cohérente avec le fait que si les observateurs de la classe 1 sont tout simplement des instruments de mesure usuels (lesquels évidemment ne peuvent être transformés en observateurs de la classe 2) alors une fois leurs matrices caractéristiques Q déterminées, l'information obtenue par les observateurs de la classe 2 (qui sont maintenant des observateurs usuels) est celle donnée par les relations (23) et (23'), c'est-à-dire la théorie de l'équivocation de Shannon.

REFERENCES

(1) R.A. Fisher, Proc. Cambridge Phil. Soc., 22 700, (1925)

(2) C.E. Shannon and W. Weaver, "*The Mathematical Theory of Communication*", The University of Illinois Press, Urbana, (1949)

(3) A.I. Khinchin, "*Mathematical Foundations of Information Theory*", Dover, (1957)

(4) P.M. Woodward, "*Probability and Information Theory with Application to Radar*", Pergamon, London, (1953)

(5) S. Kullback, "*Information Theory in Statistics*", Wiley, (1959)

(6) L. Brillouin, "*Sciences and Information Theory*", Second Ed., Academic Press N.Y., (1962)

(7) Réf. 6, chap. 15

(8) Réf. 2, Partie II, pp 66-ff

(9) Réf. 2, Partie III, p.87

(10) Nous utilisons ici le terme introduit par Brillouin (voir la réf. 6, chap. 12). La néguentropie ou information obtenue lorsqu'une v.a. X prend la valeur x est égale à l'entropie de la distribution p (x) changée de signe.

(11) En électronique $q_{x_0}(y)$ est la réponse en fonction du temps de l'instrument de mesure à l'impulsion unique $\delta(x-x_0)$ à l'instant x_0. Ici, $q_{x_0}(y)$ est une distribu-

tion qui résulte d'un très grand nombre de mesures successives portant sur le même objet s_0 d'un ensemble S, x_0 étant la valeur de X correspondant à s_0. $q_{x_0}(y)$ peut encore être la distribution qui résulte d'un grand nombre de mesures effectuées sur des objets différents de l'ensemble précédent mais correspondant à la même valeur x_0 de la v.a. X. On suppose que chaque mesure conduit à un résultat bien défini. Bien que les deux définitions soient différentes, elles conduisent cependant au même formalisme. Par ailleurs, la notation $q_x(y)$ pour la probabilité conditionnelle de Y lorsque la valeur de X est donnée qui est utilisée tout le long de l'article a la même signification que q(x,y) utilisée dans la relation (13). L'on a préféré la notation avec x subscrit car elle est plus suggestive que les v.a. X et Y n'ont pas des rôles symétriques.

(12) C.E. Shannon, Proc. Inst. Radio Eng., 37, 10 (1949)

(13) D.V. Lindley, Ann. Math. Statis., 27, 986 (1956)

-=-=-=-=-=-=-=-=-=-=-=-=-=-

LEGENDE DES FIGURES

Fig. 1 Représentation schématique de deux distributions de probabilités $p_1(x)$ et $p_2(x)$ de la variable aléatoire X. Si L est le champ d'observation et ΔX la précision de l'instrument de mesure, la relation $I = k \ln \frac{L}{\Delta X}$ stipule que l'information obtenue quand on effectue une mesure particulière dépend de L et ΔX mais non pas de la forme des distributions $p_1(x)$ et $p_2(x)$. Ceci est contraire à la mesure de l'information de Shannon, notion essentiellement statistique, pour laquelle l'information est plus grande pour $p_2(x)$, la plus large des deux distributions. Par ailleurs, si p(x) se réduit à une constante, l'information selon Shannon se réduit à la formule ci-haut de Brillouin.

Fig. 2 Obtention de la distribution de probabilités convoluée P(y) à partir de la distribution "*réelle*" p(x). Si x prend une valeur comprise entre i et i + 1, il existe une probabilité conditionnelle q_{ij} pour que Y prenne une valeur comprise entre j et j + 1. Si le champ d'observation est divisé en N parties égales, les N^2 éléments q_{ij} définissent une matrice carrée N-dimensionnelle, qui permet la détermination d'un vecteur "*convolué*" $\vec{P}$ à partir d'un vecteur $\vec{p}$ où $\vec{P}$ et $\vec{p}$ sont étroitement reliés respectivement à P(y) et p(x).

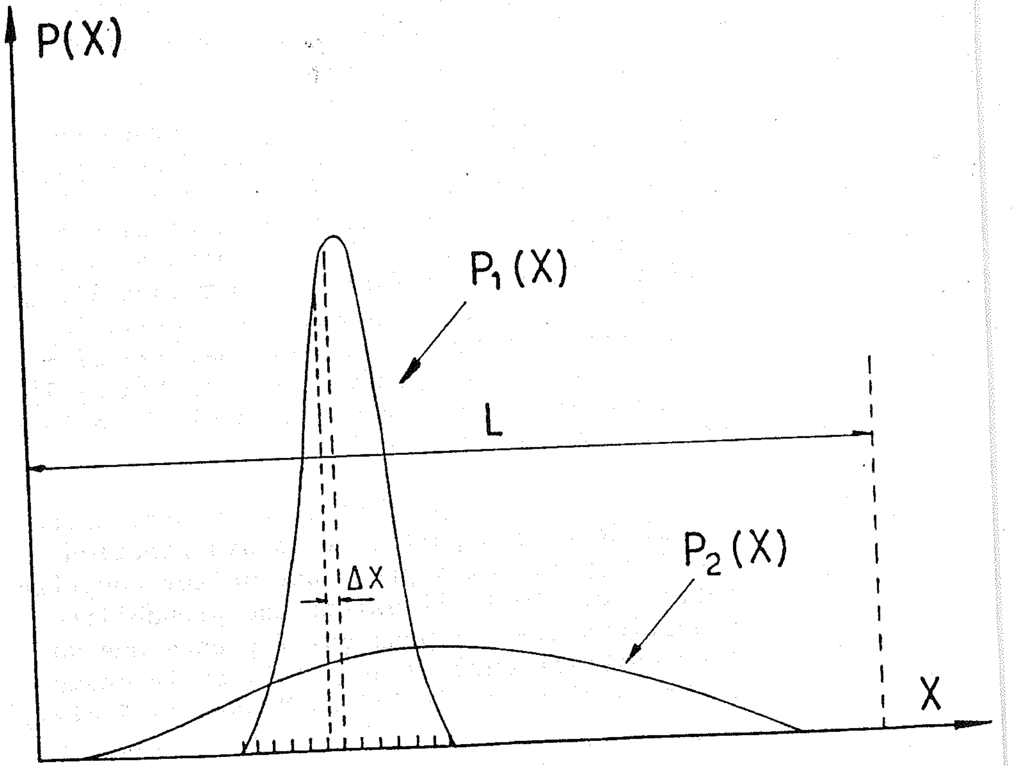

Figure 1

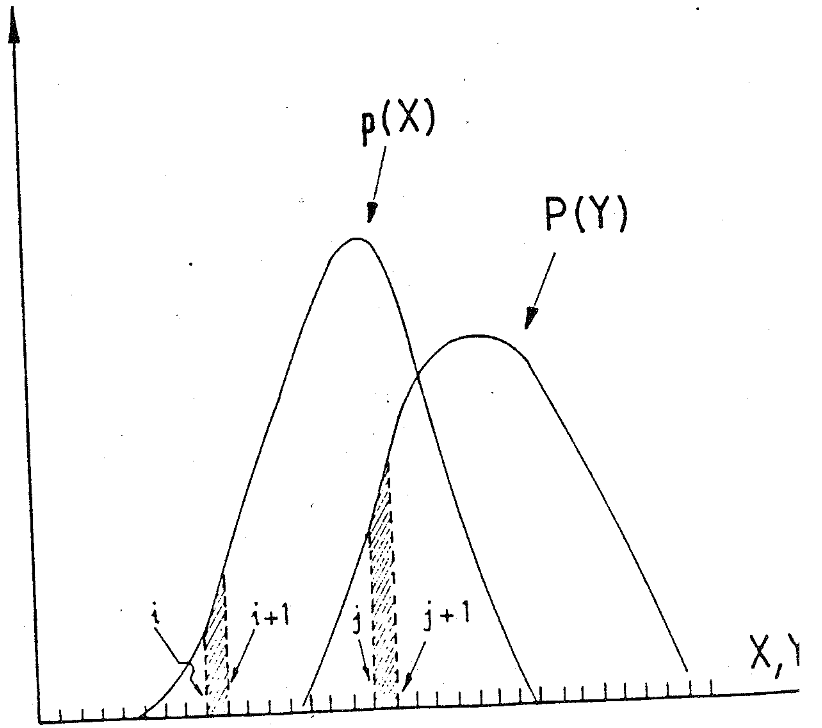

Figure 2