TOME LXII Juillet-Août 1955 N^os^ 7-8

REVUE GÉNÉRALE DES SCIENCES

PURES ET APPLIQUÉES

PUBLIÉE AVEC LE CONCOURS DU CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

ET BULLETIN DE L'ASSOCIATION FRANÇAISE POUR L'AVANCEMENT DES SCIENCES

(REVUE MENSUELLE)

Le Numéro : 175 fr.

SOMMAIRE :

Publication de la Société d'Edition d'Enseignement Supérieur

5, place de la Sorbonne - PARIS V^e^

Les problèmes de diagnostic et l'axiomatique des informations [1]

par P. SCHÜTZENBERGER

Le problème de la définition rigoureuse de l'information s'est posé aux physiciens et aux géomètres depuis le jour où des considérations d'ordre technologique ont fourni une base concrète à l'évaluation simultanée de sa valeur et du coût de son acquisition. Ainsi, en statistique, comme dans le domaine des télécommunications est apparu le besoin de chiffrer l'efficacité relative des procédures ou des appareillages et pour cela d'introduire une certaine quantité nouvelle « l'information » jouant, dans ces problèmes, un rôle analogue à celui de l'énergie dans les sciences développées à partir de la construction des machines.

D'autre part, dans la logique, dans certains chapitres de la physique théorique, etc..., on avait cherché également à définir certaines « informations » et peut-être près d'une dizaine de concepts différents portant ce nom peuvent être recensés dans la littérature. Sans méconnaître les mérites propres de ces diverses approches, nous essayerons ici une synthèse provisoire qui rassemble certains d'entre eux sous une axiomatique unique, à la fois formelle et pragmatique : formelle, parce que nous chercherons selon un mouvement classique dans les sciences de la nature non point tant ce qu'est l'information, qu'une expression analytique susceptible d'en exprimer les variations tout en satisfaisant à des critères abstraits d'additivité ou de commutativité ; pragmatique en ce sens que nous postulerons ceci : l'information ne peut être définie que relativement à une décision à prendre, à un but à atteindre et que notre objectif sera en définitive « des informations ».

***

Fixons d'abord les conditions dans lesquelles il y aura pour nous *changement d'information* : essentiellement quand un observateur O, en présence d'un système physique dont l'état ξ lui est inconnu, effectue une observation qui élimine diverses possibilités qui existaient *a priori*.

---

(1) Les mots *information*, *diagnostic*, ont beaucoup dépassé le sens qui leur est attribué dans la vie courante : ce qu'indique, d'ailleurs, la première section du présent article, dont le thème est inséparable d'une étude de M. Benoit MANDELBROT, qui paraîtra bientôt.

Les connaissances que O possède sur ξ, nous supposerons qu'elles peuvent être réduites à une distribution de probabilités *a priori*, chacun des états possibles X_1, X_2, ... X_n ayant une probabilité $P(X_1)$, $P(X_2)$, $P(X_n)$ d'être l'état inconnu de ξ.

Naturellement, il serait tout à fait concevable *in abstracto* que ξ fut susceptible de prendre un état non pas dans un ensemble discret, mais sur un continuum uni ou multi-dimensionnel ; c'est ce qui paraît légitime quand ξ est une mesure. De fait, la restriction que nous nous imposons en quantifiant ξ est assez peu importante, car en définitive toute mesure n'est jamais effectuée qu'avec une précision finie et nous pouvons donc toujours supposer qu'il existe une partition discrète $(X_1 ; X_2 ; X_3 ; \ldots X_n)$ de l'intervalle de variation de la variable considérée.

La seconde restriction, que ξ soit une aléatoire, est plus fondamentale et écarte malheureusement le cas important où ξ s'obtiendrait non pas par une mesure, mais par un calcul ou par un raisonnement, à moins que par un artifice l'on ne parvienne à réintroduire les probabilités. Notons, cependant, que nous n'exigeons pas que les $P_r\,(X_i)$ soient entièrement spécifiés : il est parfaitement possible et c'est là d'ailleurs le cas le plus intéressant, que dans leur expression interviennent des paramètres non aléatoires encore qu'inconnus. Nous symboliserons ceux-ci par θ.

D'autre part, l'observateur a un but à atteindre pour lequel la détermination de ξ n'est qu'un moyen.

Par exemple, selon les trois cas classiques :

— O cherche à estimer la valeur de θ ;

— O veut savoir si θ est ou non contenu dans un certain intervalle (ξ_0, ξ_1) (problèmes de test d'hypothèse) ;

— O veut savoir si θ est ou non dans un certain sous-ensemble Ξ des (X) (problèmes de tri).

Dans tous les cas, il lui est donc indifférent, du point de vue du « gain » qu'il réalise en effectuant ses observations que ξ soit un état ou un autre du moment qu'une certaine fonction donnée *a priori* de ces états a la même valeur.

Pour chaque situation concrète, il existe une *relation d'équivalence* (exactement *une projection*) entre les états X_1.

A la base se trouve évidemment le problème plus simple où l'objectif à atteindre est simplement la détermination de ξ : c'est le cas limite commun à tous les autres et celui qui se rencontre en théorie des communications en l'absence de bruit : nous l'appelerons le *problème de diagnostic*, selon une terminologie que nous avons introduite avec M. VILLE.

Venons-en maintenant au mécanisme même des observations : *a priori*, on pourrait considérer des observations quelconques dont chacune aurait pour seul effet de changer les $Pr(X_1)$. Pour des raisons de simplicité, on se bornera cependant à considérer des suites d'observations constituant des « diagnostics séquentiels ».

En empruntant aux systématiciens — botanistes ou zoologistes — leur notation par clefs dichotomiques nous pourrions, par exemple, représenter ainsi deux procédures typiques pour le diagnostic séquentiel de ξ entre les cinq états X_1, X_2 ... X_5.

PROCÉDURE I

$$\begin{array}{l}1^{re}\text{ observation}\\ \xi\varepsilon\ X_1+X_2\\ \text{ou}\ \xi\varepsilon X_3+X_4+X_5\end{array}\left\{\begin{array}{ll}\begin{array}{l}\text{si } \xi\varepsilon X_1+X_2\\ 2^e\text{ observation}\\ \xi=X_1\text{ ou } \xi=X_1\end{array} & \left\{\begin{array}{l}\xi=X_1\\ \xi=X_2\end{array}\right.\\ \begin{array}{l}\text{si } \xi\varepsilon X_3+X_4+X_5\\ 2^e\text{ observation}\\ \xi=X_3\\ \text{ou } \xi\varepsilon\ X_3+X_4\end{array} & \left\{\begin{array}{ll}\xi=X_3 & \\ \begin{array}{l}\text{si } \xi\varepsilon X_3+X_4\\ 3^e\text{ observation}\end{array} & \left\{\begin{array}{l}\xi=X_4\\ \xi=X_5\end{array}\right.\end{array}\right.\end{array}\right.$$

PROCÉDURE II

$$1^{re}\text{ observation}\left\{\begin{array}{ll}\xi=X_1 & \\ \begin{array}{l}\text{si } \xi\neq X_1\\ 2^e\text{ observation}\end{array} & \left\{\begin{array}{ll}\xi=X_2 & \\ \begin{array}{l}\text{si } \xi\neq X_2\\ 3^e\text{ observation}\end{array} & \left\{\begin{array}{ll}\xi=X_3 & \\ \begin{array}{l}\text{si } \xi\neq X_3\\ 4^e\text{ observation}\end{array} & \left\{\begin{array}{l}\xi=X_4\\ \xi=X_5\end{array}\right.\end{array}\right.\end{array}\right.\end{array}\right.$$

Il est clair que si les probabilités *a priori* de X_1, X_2, ... X_5 étaient respectivement 1/2, 1/4, 1/8, 1/16 et 1/16, la procédure II serait préférable puisqu'elle n'entraînerait en moyenne que $1/2 + 2 \times 1/4 + 3 \times 1/8 + 4 \times (1/16 + 1/16) = 15/8$ observations contre $2 \times (1/2 + 1/4) + 2 \times 1/8 + 3 \times (1/16 + 1/16) = 17/8$ observations pour la première procédure.

Nous postulerons :

A chaque observation est attachée une partition de l'ensemble des X_i en plusieurs sous-ensembles disjoints et l'observation nous apprend quel est celui d'entre eux qui contient ξ (*). L'effet sur les $Pr(X_i)$ est donc particulièrement simple : $Pr(X_i)$ devient zéro ou est multiplié par un coefficient qui normalise les probabilités des états encore possibles après l'observation.

Nous sommes en mesure maintenant de définir ce que nous appellerons une « quantité d'information » : à une observation à laquelle

(*) Par exemple : la première observation de la procédure I correspond à la partition $(X_1 + X_2)$ $(X_3 + X_4 + X_5)$. La deuxième observation de la procédure II correspond à (X_2) $(X_3 + X_4 + X_5)$, etc...

est attachée une partition (X) (Y) ... (Z), nous ferons correspondre « l'information » H(X, Y, ... Z) qui est une fonctionnelle continue et symétrique des probabilités $Pr(X)$, $Pr(Y)$, ... $Pr(Z)$ satisfaisant à l'axiome unique d'associativité suivant :

Axiome : si $\xi \varepsilon X + Y + Z$ et si X, Y et Z sont disjoints, on a identiquement :

$$A : H(X, Y + Z) + Pr(Y + Z)\, H(Y, Z) = H(X, Y, Z)$$

La signification de cet axiome est immédiate si l'on interprète H comme un coût : il revient à postuler que le coût moyen du diagnostic de ξ entre X, Y et Z est exprimable en fonction du coût des deux observations dichotomiques ((X) (Y + Z)) et ((Y) (Z)) d'une façon qui — en raison de la symétrie — ne dépend pas du choix de ces dichotomies : ou si l'on veut, que les trois schémas suivants ont le même coût :

Procédure (1)	Procédure (2)	Procédure (3)
$\begin{cases} \xi = X \\ \xi \varepsilon Y + Z \begin{cases} \xi = Y \\ \xi = Z \end{cases} \end{cases}$	$\begin{cases} \xi = Y \\ \xi \varepsilon X + Z \begin{cases} \xi = X \\ \xi = Z \end{cases} \end{cases}$	$\begin{cases} \xi = Z \\ \xi \varepsilon X + Y \begin{cases} \xi = X \\ \xi = Y \end{cases} \end{cases}$

L'axiome est donc tout aussi bien un *axiome de commutativité* ou encore *de linéarité de la combinaison des coûts.*

Il ne saurait entrer dans notre propos de dériver de l'équation précédente la forme générale de H par un raisonnement rigoureux et nous nous bornerons à indiquer que celui-ci se ferait en deux étapes successives :

1. H est une somme de termes de la forme

$$h(X) + h(Y) + h(Z) - h(X + Y + Z)$$

où h est une certaine fonctionnelle qui

2. a la forme particulière :

$$Pr(X)\ D \operatorname{Log} Pr(X)$$

où D symbolise un opérateur linéaire quelconque.

On aura donc pour valeur de l'information attachée à la partition X_1, X_2, ... :

$$\Sigma\ P(X_i)\ D \operatorname{Log} P(X_i)$$

Considérons maintenant les trois cas classiques.

Si D est une constante, H, au signe près, est l'entropie attachée à la distribution *a priori* de ξ. Un théorème fondamental montre alors que pour effectuer le diagnostic complet de π, il faudra en moyenne un nombre d'observations dichotomiques qui ne saurait être inférieur à C^{te}/H, quelle que soit la procédure employée.

Si les P(X) dépendent d'un paramètre inconnu et que nous cherchions à estimer la valeur de celui-ci au moyen d'un échantillon contenant n valeurs indépendantes de ξ, la précision de notre estimation (exprimée par sa variance) ne peut pas être inférieure à C^{te}/F où F est l'information correspondant à l'opérateur $D = -\frac{\partial^2}{\partial\theta^2}$: c'est le théorème de FRECHET-DARMOIS.

Si, dans les mêmes conditions, nous voulons seulement savoir si le paramètre θ est ou non compris dans un certain intervalle (θ_0, θ_i) et ceci avec des probabilités d'erreur fixées *a priori*, il existe encore une information qui livre une borne inférieure au nombre minimum moyen d'observations qui doivent être effectuées.

Or, si les deux derniers problèmes, l'estimation d'un paramètre inconnu et le test d'une hypothèse, constituent en quelque sorte l'essentiel de la statistique mathématique, le premier problème, au contraire, appartient à un chapitre distinct du calcul des probabilités la théorie des communications. Ainsi donc, l'information telle qu'elle a été introduite ici semble destinée à jouer un rôle unificateur malgré — ou peut-être à cause — du caractère formel de sa définition.

P. SCHÜTZENBERGER

LES LIVRES REÇUS

BASSE DE MENORVAL (Eliane). — Les Fossiles (Coll. Que sais-je ? Presses Universitaires, Paris).

BISANTI (A.) — Etude des couplages et du décalage de phase des transformateurs (Eyrolles, Paris), 500 francs.

BRUN (Edmond-A.). — Introduction à l'étude de la couche limite (Gauthier-Villars, Paris), 1.500 francs.

BOUTROUX (Pierre). — L'idéal scientifique des mathématiciens (Presses Universitaires, Paris), 800 francs.

BURGESS (Eric). — Frontier to Space (Chapman & Hall, Londres), 21/—.

CONTURIE (L.). — L'acoustique dans les bâtiments (Eyrolles, Paris). 1.600 fr.

DAVID (P.). — Cours de radioélectricité générale : propagation des ondes (Eyrolles, Paris), 1.900 francs.

DALMAT (Herbert T.). — The black Flies (Diptera, Simuliidae) of Guatemala and their role as vectors of Onchocerciasis (Smithsonian Institution, Washington).

DOIGNON (Pierre). — Flore des mousses de la plaine française (Ass. des Naturalistes de la vallée du Loing).

DUCLAUX (J.) — Traité de Chimie-Physique. XVII. Centrifuges et ultracentrifuges (Hermann & Cie, Paris), 800 francs.

DUGAST (Idelette). — Monographie de la tribu des Ndiki (Banen du Cameroun) (Institut d'Ethnologie, Paris), 4.000 francs.

ESSIG (Jean). — Douze, notre futur dix (Dunod, Paris).

FEYLING-HANSSEN (Rolf-W.) — Statigraphy of the marine late-pleistocene of Billefjorden, Vestspitsbergen (Norsk Polarinstitutt, Oslo), 22 couronnes.

FLEURY (P.) et **MATHIEU (J.-P.).** — Vibrations, mécanique, acoustique (Eyrolles, Paris), 3.000 francs relié.

(Voir suite, p. 247).