

ANNALES DE L'I. H. P., SECTION A

COLETTE PADET

**Sur une théorie informationnelle de l'observation
pour des événements de nature probabiliste**

Annales de l'I. H. P., section A, tome 33, n° 3 (1980), p. 319-343

[<http://www.numdam.org/item?id=AIHPA_1980__33_3_319_0>](http://www.numdam.org/item?id=AIHPA_1980__33_3_319_0)

© Gauthier-Villars, 1980, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P., section A » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

Sur une théorie informationnelle de l'observation pour des événements de nature probabiliste

par

Colette PADET

Université de Reims, Laboratoire de Mécanique Quantique

RÉSUMÉ. — On se propose de relier certains concepts fondamentaux dans le domaine de l'observation avec certains autres concepts de base de la théorie de l'information, en particulier celui d'entropie.

Une axiomatique définissant un processus d'observation est d'abord établie. Pour des phénomènes de nature probabiliste, on peut alors définir et calculer, en les mettant clairement en relation avec les caractéristiques du procédé d'observation, l'information associée à un résultat d'observation et l'entropie d'une variable aléatoire relativement à un procédé d'observation. La démarche suivie, applicable aussi bien au cas des variables aléatoires continues qu'à celui des variables aléatoires discrètes, permet d'apporter une solution au problème très controversé de l'entropie des variables aléatoires continues.

En faisant la relation avec les caractéristiques du procédé utilisé, on est amené à introduire quatre familles d'entropies (entropie utile, entropie mesurable, entropie absolue et entropie brute) dont on met en évidence les principales propriétés ainsi que les relations avec d'autres familles d'entropie.

SUMMARY. — The present paper aims at linking some fundamental concepts of the observation with some other basic concepts in the information theory, particularly the concept of entropy.

An axiomatic formulation in order to describe a process of observation is stated at first. For a statistical set of events, we define and calculate, with reference to the parameters of the observation, the information associated with a result of observation and the entropy of a random variable for a proceed of observation. The method can be applied as well for the

continuous random variables as for the discrete random variables, allow to give a solution to the much debated problem of the entropy of a continuous random variable.

Taking into account characteristic parameters of the proceed of observation, we are lead up to introduce four classes of entropies, for which we examine the main properties just as the relations with others classes of entropies.

Il est généralement fructueux, dans de nombreux domaines, de réaliser une synthèse entre deux groupes de phénomènes ou entre deux théories complémentaires qui ont connu pendant un certain temps des évolutions autonomes. On peut ainsi mettre en évidence des propriétés nouvelles, voire élaborer des concepts nouveaux. Pour ce faire, la démarche la plus intéressante consiste souvent à réunir les bases conceptuelles des deux théories plutôt qu'à regrouper leurs développements et leurs résultats. Une telle tentative d'unification entre la physique statistique et la théorie de l'information a ainsi été à l'origine de la construction d'un nouveau concept probabiliste, l'opacité d'une complexion statistique [1].

La période actuelle nous offre un autre exemple de deux domaines importants visiblement complémentaires entre lesquels aucune synthèse véritable ne semble avoir été réalisée : celui de l'observation et celui de l'information. On ne s'informe pourtant qu'en observant mais cette évidence ne s'est pas encore traduite dans la théorie.

Le travail que nous présentons ici est une tentative pour relier certains concepts fondamentaux dans le domaine de l'observation avec certains autres concepts de base de la théorie de l'information, en particulier celui d'entropie. Nous avons d'abord été amenés à approfondir l'analyse d'un processus d'observation et à en proposer une axiomatique. Pour des phénomènes de nature probabiliste, cela nous a permis de définir et de calculer, en les mettant clairement en relation avec les caractéristiques du procédé d'observation, l'information associée à un résultat d'observation et l'entropie d'une variable aléatoire relativement à un procédé d'observation. La démarche suivie, applicable aussi bien au cas des variables aléatoires continues qu'à celui des variables aléatoires discrètes, a le mérite d'apporter une solution au problème si controversé de l'entropie des variables aléatoires continues.

1. CONCEPTS DE BASE EN THÉORIE DE L'INFORMATION

Considérons un espace de probabilité constitué par le triplet $\{\Omega, \mathcal{C}, p\}$ où Ω est l'ensemble de toutes les réalisations possibles de l'expérience

étudiée, $\mathcal{C}$ une tribu de parties mesurables de Ω et p une mesure de probabilité définie sur $\mathcal{C}$ et telle que $p(\Omega) = 1$, les éléments de $\mathcal{C}$ étant appelés événements [2] [3].

Soit $I(\alpha_i) = -K \log p(\alpha_i)$ la mesure d'information relative à l'événement α_i . Elle caractérise le degré d'incertitude *a priori* lié à cet événement [4] [5]. La valeur de la constante K dépend uniquement de l'unité d'information choisie. Nous considérerons par la suite l'unité appelée bit, correspondant à $K = \frac{1}{\log 2}$ d'où l'expression :

$$I(\alpha_i) = -\log_2 p(\alpha_i)$$

Si l'on réitère un grand nombre de fois une expérience dans des conditions identiques, divers événements $\alpha_1 \dots \alpha_n$ de probabilités

$$p(\alpha_1) = p_1, \dots, p(\alpha_i) = p_i, \dots, p_n$$

peuvent être observés. Par suite, l'information moyenne par événement, introduite pour la première fois par Shannon en 1948, s'écrit [6] :

$$H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log_2 p_i$$

En raison de son analogie formelle avec l'entropie physique, cette expression est souvent appelée « entropie informationnelle ». De nombreux auteurs la désignent également par « entropie informationnelle $H(A)$ de la variable aléatoire A ». Ceci s'explique par le fait qu'en physique on associe presque toujours à un événement un nombre réel, ce qui nécessite la considération d'une variable aléatoire vectorielle. C'est cette dénomination que nous utiliserons désormais.

L'entropie de Shannon dont nous venons de rappeler la définition est l'entropie d'une variable aléatoire discrète [7 à 9]. Il apparaît naturel d'envisager son extension au cas continu. Mais on se heurte alors à une difficulté liée à la divergence de la fonction $H(A)$. De nombreux auteurs ont cependant souhaité associer une fonction finie aux variables aléatoires continues. La plus communément utilisée est celle qui a été proposée par Shannon [6] :

$$H(A) = - \int_{\mathbb{R}^n} \rho(\alpha) \log_2 \rho(\alpha) d\alpha$$

où A est une variable aléatoire continue de densité $\rho(\alpha)$.

Or, il faut remarquer que, si les conventions adoptées pour définir l'entropie dans le cas de distributions discrètes sont cohérentes et relativement satisfaisantes, il en va tout autrement dans le cas des distributions

continues. La fonction proposée par Shannon présente des difficultés d'interprétation provenant des propriétés suivantes [10 à 13] :

- $H(A)$ peut être négative,
- la formule définissant $H(A)$ n'est pas homogène puisque la grandeur figurant sous le logarithme possède une dimension qui est celle d'une densité,
- la fonction $H(A)$ n'est pas invariante dans une transformation des coordonnées.

Ces difficultés ont souvent été sous-estimées, voire ignorées, en raison du fait qu'elles disparaissent en général dans la plupart des applications, qui sont basées sur des recherches de maxima d'entropies. Ceci ne diminue en aucun cas cependant les faiblesses de $H(A)$ et quelques auteurs [14 à 17] ont cherché une définition de l'entropie plus satisfaisante que celle de Shannon. Tous pratiquement (Gibbs, Kolmogorov, Sempì, etc.) ont été amenés à intégrer la notion de précision d'une observation à la théorie de l'information. Malheureusement, celle-ci se voit attribuer dans ces différentes tentatives des significations assez différentes et quelquefois imprécises. Une voie semble cependant se dégager qui mérite d'être explorée plus avant, à savoir la prise en compte des paramètres caractéristiques du procédé d'observation et c'est elle que nous nous proposons de suivre maintenant. Auparavant, il est toutefois indispensable d'élaborer une analyse théorique des propriétés habituellement reconnues à tout système expérimental, puis du processus d'observation lui-même, dans le cas le plus général où les événements observés ne sont pas obligatoirement de nature probabiliste.

II. ÉLÉMENTS D'UNE THÉORIE MACROSCOPIQUE DE L'OBSERVATION

II.1. Caractéristiques d'un procédé d'observation.

Il est connu depuis longtemps que tout procédé d'observation que nous noterons $\mathcal{P}$ est caractérisé par trois qualités : sensibilité, fidélité et précision. En particulier, l'échelle de lecture de l'appareil est nécessairement discrète, l'écart entre les « graduations » ne pouvant matériellement pas être rendu infiniment petit. Mais ces concepts classiques bien connus des expérimentateurs sont rarement formalisés de façon satisfaisante au niveau théorique. Souvent même ils ne sont pas dissociés les uns des autres.

Notre premier souci sera donc d'en exécuter une analyse détaillée pour parfaire et généraliser ensuite leur formalisation. Remarquons au préalable que, du fait de l'imperfection de l'appareil et de la présence de l'échelle de

lecture, trois ensembles d'événements interviennent dans toute observation, à savoir l'ensemble $\mathcal{A}$ des événements observés α , l'ensemble $\mathcal{A}_e$ des événements enregistrés α_e et l'ensemble $\mathcal{A}_l$ des événements lus α_l . Les deux premiers ensembles seraient identiques si l'appareil était idéal. Les deux derniers le seraient si l'étalonnage et l'affichage étaient parfaits [18].

II.1.1. SENSIBILITÉ

La limite de sensibilité caractérise une sorte d'inertie du procédé d'observation vis-à-vis d'une variation de l'événement observé. Pour tout événement α_0 , l'événement enregistré α_{e0} est identique à α_0 . Si des fluctuations apparaissent autour de α_0 , l'appareil continue à marquer α_{e0} . Ainsi, l'enregistrement d'une valeur α_{e0} provient soit d'un événement $\alpha_0 \equiv \alpha_{e0}$, soit d'un ensemble d'événements α appartenant au borélien $\Delta_s(\alpha_{e0})$ (fig. 1) qui, d'un point de vue physique, peut être considéré comme centré sur $\alpha_0 \equiv \alpha_{e0}$. Le borélien $\Delta_s(\alpha_{e0})$ est appelé « limite de sensibilité » de l'appareil.

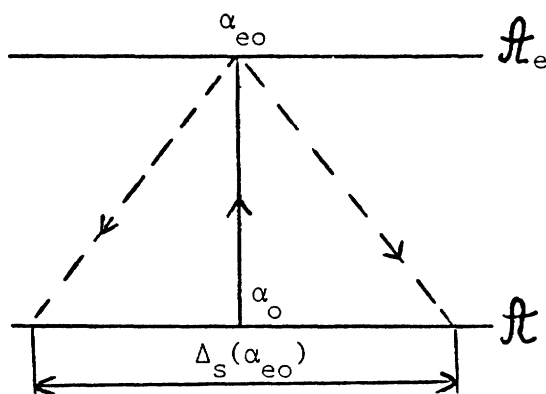


FIG. 1.

II.1.2. FIDÉLITÉ

La fidélité d'un procédé d'observation est une aptitude à la reproductibilité de l'observation pour des événements identiques; elle est affectée par les fluctuations aléatoires qui surviennent dans l'appareil d'observation. La conséquence en est que, à chaque valeur α à mesurer, correspond tout un ensemble de valeurs enregistrées α_e possibles formant un segment $\Delta_{ef}(\alpha)$ (fig. 2). Ce segment est appelé « limite de fidélité » de l'appareil pour la valeur α . Notons au passage que la fidélité est plus volontiers appelée depuis quelques années fiabilité ou encore sûreté de fonctionnement.

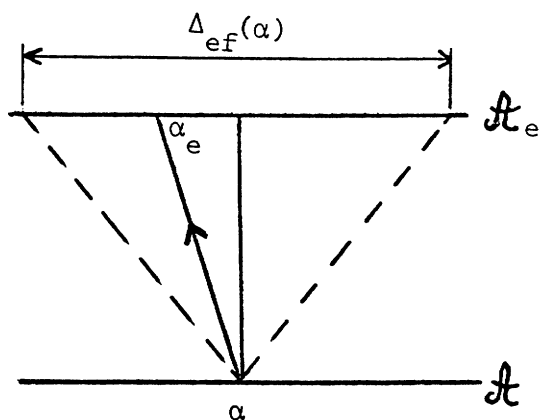


FIG. 2.

II.1.3. PRÉCISION

Le manque de précision d'un appareil de mesure caractérise un écart systématique entre la valeur enregistrée α_e et la valeur lue α_l qui procède d'une part du fait que l'ensemble des valeurs lues est toujours discret et dénombrable, d'autre part d'un défaut d'étalonnage ou d'une détérioration de l'appareil (introduction d'une distorsion) (fig. 3 et 4).

Le caractère discret de l'échelle de lecture implique qu'à chaque valeur α_l susceptible d'être affichée corresponde tout un ensemble $\Delta_{ep}(\alpha_l)$ de valeurs enregistrées. Quant à la distorsion, elle introduit un décalage entre valeurs enregistrées et valeurs affichées.

Compte tenu de ce qui précède, nous noterons tout procédé d'observation $\mathcal{P}\{\Delta_s, \Delta_{ef}, \Delta_{ep}, \Delta_l\}$, Δ_l étant l'intervalle $[\alpha_l, \alpha_{l+1}[$.

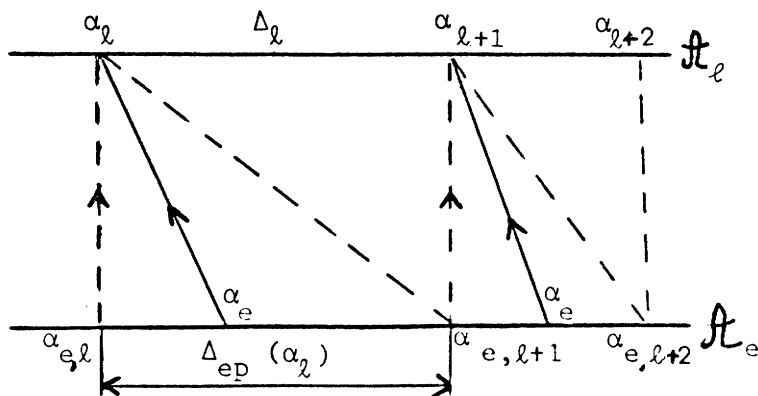


FIG. 3.

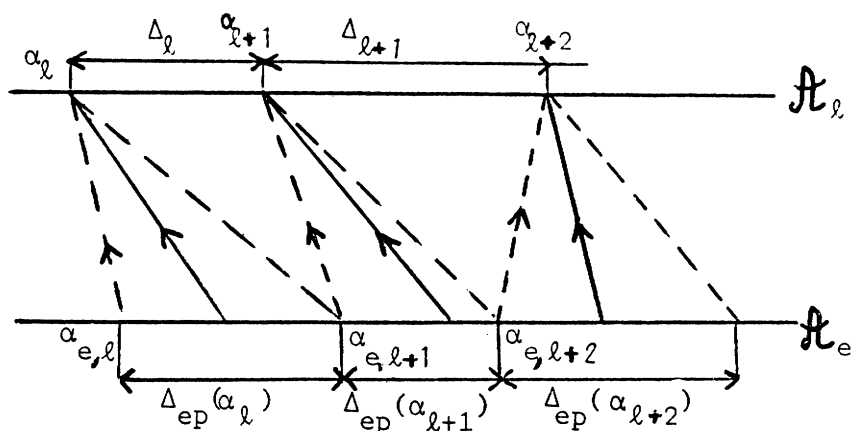


FIG. 4.

II.2. Structure d'un processus de mesure.

Il est nécessaire de poursuivre l'analyse en considérant non plus le détail du mécanisme d'observation mais l'aspect global de l'acte de mesure. Un processus de mesure se décompose en deux phases : l'observation proprement dite et l'interprétation de cette observation [18].

Le processus d'observation conduit d'un événement α à observer à la valeur lue α_i selon le schéma représenté figure 5.

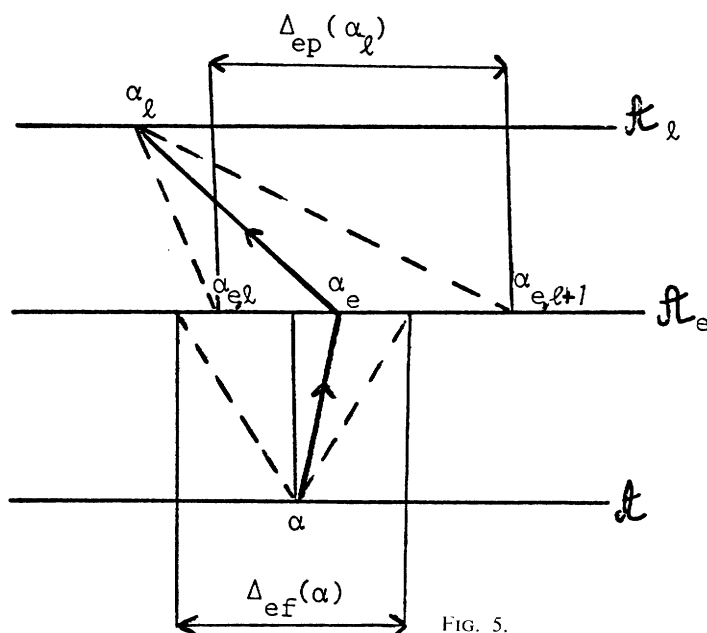


FIG. 5.

Le processus d'interprétation permet ensuite, connaissant la valeur lue α_l , de remonter à une valeur possible de α . Étant donnée la limitation des qualités d'un appareil de mesure, nous pouvons seulement conclure que la lecture de la valeur α_l implique l'appartenance de l'événement α observé à un certain borélien de $\mathcal{A}$ qui constitue en fait le « domaine total d'incertitude associé au résultat d'observation α_l ». Celui-ci peut être déterminé de la façon suivante : la lecture de α_l nous apprend que $\alpha_e \in \Delta_{ep}(\alpha_l)$. Étant donnée la limite de fidélité, à chaque α_e correspond tout un ensemble de $\alpha \in \Delta_f(\alpha_e)$. Ceci permet donc de déduire, en se référant aux notations de la figure 6, que :

$$\alpha \in [\alpha', \alpha''']$$

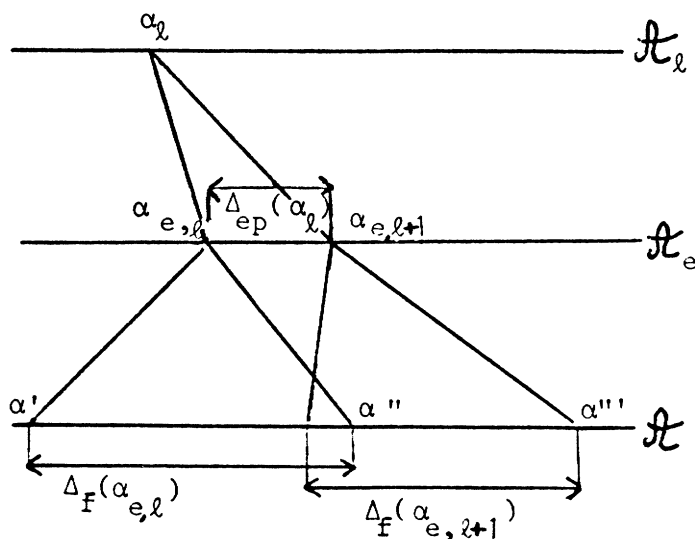


FIG. 6.

Dans le cas le plus général où la limite de sensibilité Δ_s n'est pas nulle, on déduit de la lecture de α_l que l'événement α appartient à l'intervalle (fig 7) :

$$\Delta_t(\alpha_l) = \left[\alpha' - \frac{\Delta_s(\alpha')}{2}, \alpha''' + \frac{\Delta_s(\alpha''')}{2} \right]$$

Le renseignement fourni par la lecture de α_l est donc en définitive :

$$\alpha \in \Delta_t(\alpha_l)$$

où $\Delta_t(\alpha_l)$ représente le domaine total d'incertitude associé au résultat d'observation α_l .

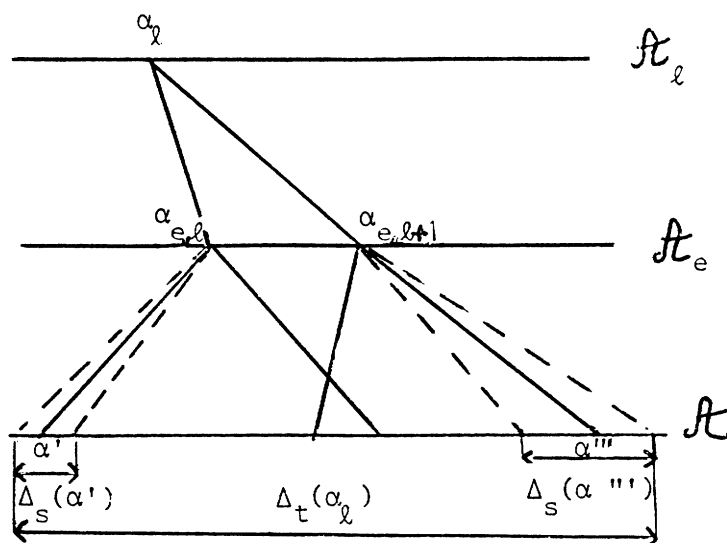


FIG. 7.

II.3. Axiomatisation.

Le véritable processus de mesure étant la succession des processus d'observation et d'interprétation, il apparaît maintenant indispensable d'en établir une axiomatique permettant de formaliser pour des cas très généraux le couple observation-interprétation [19]. Dans cette démarche, nous nous sommes inspirés du modèle mathématique de l'observation basé sur le formalisme des « opérateurs d'observation » [20 à 22].

Nous pouvons ainsi définir une classe de situations expérimentales correspondant à un procédé $\mathcal{P} \{ \Delta_s, \Delta_{ef}, \Delta_{ep}, \Delta_l \}$ par les axiomes suivants :

$\mathcal{C}_I : \exists$ une application Δ_{ef} de l'ensemble $\mathcal{A}$ dans l'ensemble $\mathcal{A}_e$, caractéristique du procédé d'observation $\mathcal{P}$, qui, à tout point de $\mathcal{A}$, fait correspondre un segment $\Delta_{ef}(\alpha)$ de $\mathcal{A}_e$. L'opérateur d'enregistrement σ_{ef} est une application définie sur l'ensemble $\mathcal{A}$ qui, à tout point $\alpha \in \mathcal{A}$ fait correspondre un point $\alpha_e \in \mathcal{A}_e$ avec une loi de probabilité $f(\alpha_e/\alpha)$ nulle en dehors de $\Delta_{ef}(\alpha)$.

$$\alpha \xrightarrow{\sigma_{ef}} \alpha_e = \sigma_{ef}(\alpha) \in \Delta_{ef}(\alpha)$$

$\mathcal{C}_{II} : \exists$ une partition dénombrable de $\mathcal{A}_e$ en segments totalement disjoints Δ_{ep} dont la réunion est $\mathcal{A}_e$. L'opérateur de lecture σ_{lp} est une application définie sur $\mathcal{A}_e$ qui, à tout $\alpha_e \in \Delta_{ep}(\alpha_l)$, fait correspondre le point α_l de $\mathcal{A}_l$.

$$\alpha_e \in \Delta_{ep}(\alpha_l) \xrightarrow{\sigma_{lp}} \alpha_l = \sigma_{lp}(\alpha_e)$$

Le produit σ_l des opérateurs d'enregistrement et de lecture est appelé opérateur d'observation.

$$\sigma_l = \sigma_{lp} \circ \sigma_{ef}$$

$\mathcal{C}_{III}$: L'application image réciproque σ de σ_l , ou opérateur d'interprétation, est une application de $\mathcal{A}_l$ dans $\mathcal{A}$ qui fait correspondre à chaque α_l de $\mathcal{A}_l$ un borélien $\Delta_l(\alpha_l)$ de $\mathcal{A}$.

$$\alpha_l \xrightarrow{\sigma} \sigma(\alpha_l) = \Delta_l(\alpha_l)$$

$\Delta_l(\alpha_l)$ est le domaine total d'incertitude associé à α_l , contenant l'ensemble de toutes les valeurs possibles pour α .

L'opérateur de mesure σ_M associé à un procédé général $\mathcal{P} \{ \Delta_s, \Delta_{ef}, \Delta_{ep}, \Delta_l \}$ a pour expression :

$$\sigma_M = \sigma \circ \sigma_l$$

Les axiomes précédents caractérisent la mesure (au sens expérimental du terme) d'événements par un procédé à échelle de lecture discrète, soumis à des fluctuations aléatoires et dont la limite de sensibilité n'est pas nulle. Ils vont maintenant permettre de traiter l'aspect informationnel de l'observation.

III. ENTROPIE D'UNE VARIABLE ALÉATOIRE RELATIVEMENT A UN PROCÉDÉ D'OBSERVATION

La connaissance acquise par mesure étant maintenant clairement caractérisée, il reste à en effectuer une évaluation quantitative à partir de la définition conventionnelle de l'information pour un ensemble statistique d'événements [18] [19]. Lors de cette étude, nous allons être amenés à introduire trois familles d'entropies.

III.1. Entropie utile.

La lecture d'une valeur α_l nous apprend que $\alpha \in \Delta_l(\alpha_l)$. Nous appellerons « information utile relative à l'événement lu α_l » l'information associée à ce résultat α_l et la définirons par la fonction :

$$I_u(\alpha_l) = -\log_2 \{ p[\alpha \in \Delta_l(\alpha_l)] \} \quad (1)$$

L'information moyenne que l'on peut recueillir en utilisant le procédé d'observation $\mathcal{P}$ est évidemment l'espérance mathématique de la fonction $I_u(\alpha_l)$. Nous appellerons cette expression « entropie utile de la variable aléatoire A relativement au procédé d'observation $\mathcal{P}$ » et la noterons $S(A, \mathcal{P})$.

$$S(A, \mathcal{P}) = \sum_l p(\alpha_l) I_u(\alpha_l) \quad (2)$$

Pour une variable aléatoire A continue de densité de probabilité $\rho(\alpha)$, nous avons d'une part :

$$I_u(\alpha_l) = -\log_2 \left[\int_{\Delta_t(\alpha_l)} \rho(\alpha) d\mu(\alpha) \right]$$

en notant μ la mesure de Lebesgue sur R .

D'autre part, $p(\alpha_l)$, étant la probabilité de lire la valeur α_l lors d'une mesure, est aussi la probabilité d'obtenir un événement enregistré α_e qui appartienne au segment $\Delta_{ep}(\alpha_l)$. Si $f(\alpha_e)$ est la densité de probabilité de α_e sur l'ensemble $\mathcal{A}_e$, alors :

$$p(\alpha_l) = p[\alpha_e \in \Delta_{ep}(\alpha_l)] = \int_{\Delta_{ep}} f(\alpha_e) d\mu(\alpha_e).$$

Une difficulté apparaît en ce qui concerne l'évaluation de la densité $f(\alpha_e)$. Nous avons vu que l'opérateur d'enregistrement σ_{ef} fait correspondre à tout point $\alpha \in \mathcal{A}$ une loi de probabilité de densité $f(\alpha_e/\alpha)$ telle que $f(\alpha_e/\alpha) = 0$ si $\alpha_e \notin \Delta_{ef}(\alpha)$. Par suite, la densité de probabilité $f(\alpha_e)$ de α_e , qui est en fait une densité marginale $f_{\alpha}(\alpha_e)$ de α_e pour tout α , s'écrit :

$$f(\alpha_e) = \int_{\mathcal{A}} f(\alpha_e/\alpha) \rho(\alpha) d\mu(\alpha)$$

Or la condition $\alpha_e \in \Delta_{ef}(\alpha)$ est équivalente à la condition $\alpha \in \Delta_f(\alpha_e)$, $\Delta_f(\alpha_e)$ étant le borélien de $\mathcal{A}$ contenant l'ensemble des événements de $\mathcal{A}$ susceptibles de donner α_e à l'observation. Il vient alors :

$$f(\alpha_e) = \int_{\Delta_f(\alpha_e)} f(\alpha_e/\alpha) \rho(\alpha) d\mu(\alpha)$$

et

$$p(\alpha_l) = \int_{\Delta_{ep}} d\mu(\alpha_e) \int_{\Delta_f(\alpha_e)} f(\alpha_e/\alpha) \rho(\alpha) d\mu(\alpha) \quad (3)$$

D'où l'expression générale de l'entropie utile $S(A, \mathcal{P})$:

$$S(A, \mathcal{P}) = -\sum \left[\int_{\Delta_{ep}} d\mu(\alpha_e) \int_{\Delta_f(\alpha_e)} f(\alpha_e/\alpha) \rho(\alpha) d\mu(\alpha) \right] \log_2 \left[\int_{\Delta_t(\alpha_l)} \rho(\alpha) d\mu(\alpha) \right]$$

En regroupant les formules (1), (2), (3), on voit que l'entropie utile $S(A, \mathcal{P})$ a la forme :

$$S(A, \mathcal{P}) = -\sum_l p(\alpha_l) \log_2 [p(\alpha \in \Delta_t(\alpha_l))]$$

Cette présentation permet de mettre en évidence trois propriétés essentielles :

- la formule définissant S est homogène puisqu'il n'y apparaît que des probabilités,
- les probabilités, étant des grandeurs sans dimension, sont invariantes

dans toute transformation des coordonnées. Il en est donc de même pour $S(A, \mathcal{P})$,

— l'entropie $S(A, \mathcal{P})$ est non négative car le logarithme d'une probabilité est toujours ≤ 0 .

Ces trois propriétés fondamentales de $S(A, \mathcal{P})$ assurent sa cohérence avec l'entropie de Wiener-Shannon définie pour les variables aléatoires discrètes hors de toute référence à un procédé d'observation, et la rendent exempte des défauts attachés à l'entropie de Shannon

$$H(A) = - \int \rho(\alpha) \log_2 \rho(\alpha) d\alpha$$

pour les variables aléatoires continues.

Nous devons également signaler qu'il existe une sous-classe de situations expérimentales très usuelles où $\Delta_{ef}(\alpha)$ est indépendante de α , et où les fluctuations ont une densité de probabilité uniforme sur chaque segment $\Delta_{ef}(\alpha)$. On a alors :

$$\mu \{ \Delta_{ef}(\alpha) \} = \mu \{ \Delta_f(\alpha_e) \} = \text{cte}$$

(ce que nous écrirons pour simplifier $\mu(\Delta_{ef}) = \mu(\Delta_f) = \text{cte}$) et

$$f(\alpha_e/\alpha) = \text{cte} \quad \text{sur } \Delta_f.$$

D'où

$$f(\alpha_e) = \frac{1}{\Delta_f} \int_{\Delta_f} \rho(\alpha) d\mu(\alpha)$$

et

$$S(A, \mathcal{P}) = - \sum_i \int_{\Delta_{ep}} \overline{\rho_f(\alpha_e)} d\mu(\alpha_e) \log_2 \left[\int_{\Delta_i(\alpha_i)} \rho(\alpha) d\mu(\alpha) \right]$$

où $\overline{\rho_f(\alpha_e)}$ représente la valeur moyenne de la fonction $\rho(\alpha)$ sur le borélien $\Delta_f(\alpha_e)$ de l'ensemble $\mathcal{A}$.

III.2. Entropie mesurable.

Étant donnée l'analyse que nous avons faite de la relation observation-observé, une autre famille d'entropies peut être définie.

En effet, il est toujours possible de dissocier le système d'affichage d'un moyen d'observation donné et de le traiter indépendamment. Si nous considérons alors un ensemble d'événements observés par un procédé $\mathcal{P} \{ \Delta_s, \Delta_{ef}, 0, 0 \}$ de précision parfaite, nous savons, après obtention du résultat d'observation α_e , qu'à l'issue du processus interprétatif, les valeurs possibles de α sont contenues dans un segment Δ_i de $\mathcal{A}$ que nous noterons $\Delta_{sf}(\alpha_e)$.

Par définition, nous appellerons « information mesurable apportée par l'enregistrement de la valeur α_e » l'expression :

$$I_m(\alpha_e) = - \log_2 [p(\alpha \in \Delta_{sf}(\alpha_e))]$$

Corrélativement à cette notion, nous pouvons définir une « entropie mesurable » $S_m(A, \mathcal{P})$ de la variable aléatoire A relativement à un procédé d'observation $\mathcal{P}$. C'est l'espérance mathématique de la fonction $I_m(\alpha_e)$ sur son ensemble de définition $\mathcal{A}_e$. C'est également la valeur que prend l'entropie utile $S(A, \mathcal{P})$ lorsque $\mathcal{P} = \mathcal{P} \{ \Delta_s, \Delta_{ef}, 0, 0 \}$. On a donc :

$$S_m(A, \mathcal{P}) = \int_{\mathcal{A}_e} f(\alpha_e) I_m(\alpha_e) d\mu(\alpha_e)$$

soit encore :

$$S_m(A, \mathcal{P}) = - \int_{\mathcal{A}_e} \left[\int_{\Delta_{sf}(\alpha_e)} f(\alpha_e/\alpha) \rho(\alpha) d\mu(\alpha) \right] \log_2 [\overline{\rho_{sf}(\alpha_e)} \mu \{ \Delta_{sf}(\alpha_e) \}] d\mu(\alpha_e)$$

où $\overline{\rho_{sf}(\alpha_e)}$ représente la valeur moyenne de la fonction $\rho(\alpha)$ sur le borélien $\Delta_{sf}(\alpha_e)$ de l'ensemble $\mathcal{A}$.

Le qualificatif « mesurable » est justifié par le fait que cette information serait accessible avec un moyen d'observation donné à condition de rendre parfait le système d'affichage.

Si nous considérons la sous-classe de situations expérimentales définie au paragraphe précédent, l'entropie mesurable $S_m(A, \mathcal{P})$ prend alors la forme :

$$S_m(A, \mathcal{P}) = - \int_{\mathcal{A}_e} \overline{\rho_f(\alpha_e)} \log_2 [\overline{\rho_{sf}(\alpha_e)} \mu \{ \Delta_{sf}(\alpha_e) \}] d\mu(\alpha_e)$$

III.3. Entropie absolue.

Nous devons enfin considérer l'information relative à un procédé idéal $\mathcal{P} \{ 0, 0, 0, 0 \}$. Elle ne fait en vérité aucune référence au procédé d'observation qui sert à recueillir l'information (lequel est implicitement considéré comme parfait) et ne concerne donc que l'objet d'observation en soi.

Soit α un événement à observer de probabilité $p(\alpha)$. On peut appeler « information absolue relative à l'événement α » l'expression :

$$I(\alpha) = - \log_2 p(\alpha)$$

C'est encore l'information associée à un opérateur σ_M identité.

Parallèlement à cette notion, nous pouvons définir une « entropie absolue $S(A)$ de la variable aléatoire A » qui est l'espérance mathématique des informations absolues relatives à chaque événement. Si la variable aléatoire A est continue, l'information $I(\alpha)$ et par suite l'entropie absolue $S(A)$ sont infinies.

III.4. Entropie brute.

Le processus de mesure considéré jusqu'à présent est constitué, comme nous l'avons vu, de la succession du processus d'observation et du processus interprétatif.

Or il est courant dans la pratique de s'en tenir à l'observation seule, qui représente en fait une « mesure brute » de l'événement. Aussi est-il très intéressant d'introduire les notions d'information brute et d'entropie brute [19].

L'événement final du processus d'observation est la lecture d'une valeur α_i de l'ensemble des résultats d'observation $\mathcal{A}_i$. Par suite, $p(\alpha_i)$ étant la probabilité de lire la valeur α_i à l'issue du processus d'observation, nous proposons d'appeler « information brute associée à l'événement α_i » l'expression :

$$I_b(\alpha_i) = -\log_2 p(\alpha_i)$$

« L'entropie brute $S_b(A, \mathcal{P})$ de la variable aléatoire A relativement à un procédé d'observation $\mathcal{P}$ » est alors l'espérance mathématique de $I_b(\alpha_i)$.

$$S_b(A, \mathcal{P}) = \sum_i p(\alpha_i) I_b(\alpha_i)$$

$$S_b(A, \mathcal{P}) = - \sum_i p(\alpha_i) \log_2 p(\alpha_i)$$

En remplaçant $p(\alpha_i)$ par sa valeur donnée dans l'expression (3), on obtient l'expression générale de l'entropie brute $S_b(A, \mathcal{P})$:

$$S_b(A, \mathcal{P}) = - \sum_i \left[\int_{\Delta_{ep}} d\mu(\alpha_e) \int_{\Delta_f(\alpha_e)} f(\alpha_e/\alpha) \rho(\alpha) d\mu(\alpha) \right] \log_2 \left[\int_{\Delta_{ep}} d\mu(\alpha_e) \int_{\Delta_f(\alpha_e)} f(\alpha_e/\alpha) \rho(\alpha) d\mu(\alpha) \right]$$

Une étude comparative entre l'entropie utile et l'entropie brute a permis de mettre en évidence, dans tous les cas où une conclusion a pu être obtenue, la relation d'ordre suivante :

$$S_b(A, \mathcal{P}) \geq S(A, \mathcal{P})$$

Nous pouvons en déduire qu'il y a diminution d'entropie lorsque l'observation d'un phénomène est complétée par l'interprétation des résultats de mesure. Cette diminution d'entropie traduit une augmentation de notre connaissance du système étudié provenant de l'interprétation.

IV. EXTENSION AU CAS DE VARIABLES ALÉATOIRES DISCRÈTES : ÉVÉNEMENTS DISCERNABLES PAR UN PROCÉDÉ DE MESURE

Nous avons défini plus haut l'entropie utile d'une variable aléatoire continue relativement à un procédé d'observation $\mathcal{P} \{ \Delta_s, \Delta_{ef}, \Delta_{ep}, \Delta_l \}$.

Nous allons maintenant étendre cette étude au cas de variables aléatoires discrètes et montrer qu'il est généralement nécessaire d'intégrer les caractéristiques du procédé d'observation dans l'expression de l'entropie.

IV.1. Évaluation de l'entropie utile d'une variable aléatoire discrète.

Nous considérons toujours les trois ensembles de valeurs réelles $\mathcal{A}$, $\mathcal{A}_e$ et $\mathcal{A}_l$. Les ensembles $\mathcal{A}$ et $\mathcal{A}_l$ sont discrets ; nous noterons α_v et α_l leurs éléments respectifs. L'ensemble $\mathcal{A}_e$ est continu, l'appareil pouvant *a priori* enregistrer n'importe quelle valeur α_e .

L'entropie utile de la variable aléatoire discrète A relativement à un procédé d'observation $\mathcal{P} \{ \Delta_s, \Delta_{ef}, \Delta_{ep}, \Delta_l \}$ a pour expression :

$$S(A, \mathcal{P}) = \sum_l p(\alpha_l) I_u(\alpha_l) \quad (4)$$

α_l étant le résultat d'observation et $I_u(\alpha_l)$ l'information utile relative à l'événement α_l .

Or l'information $I_u(\alpha_l)$ pour un procédé $\mathcal{P}$ est égale à :

$$I_u(\alpha_l) = -\log_2 [p(\alpha \in \Delta_l(\alpha_l))]$$

d'où

$$S(A, \mathcal{P}) = - \sum_l p(\alpha_l) \log_2 [p(\alpha \in \Delta_l(\alpha_l))]$$

Nous allons, pour la suite de cet exposé, introduire le symbole $\sum_{v, \Delta x}$ qui représente la sommation sur tous les α_v appartenant à un intervalle Δx . L'expression de $I_u(\alpha_l)$ peut alors s'écrire, en prenant $\Delta x = \Delta_l(\alpha_l)$:

$$I_u(\alpha_l) = -\log_2 \sum_{v, \Delta_l(\alpha_l)} p(\alpha_v) \quad (5)$$

Il reste à calculer la probabilité $p(\alpha_l)$ d'obtenir le résultat α_l lors d'une observation. Cette probabilité est aussi celle d'obtenir un événement enregistré α_e qui appartienne au segment $\Delta_{ep}(\alpha_l)$ associé à α_l .

$$p(\alpha_l) = p(\alpha_e \in \Delta_{ep}(\alpha_l)) = \int_{\Delta_{ep}(\alpha_l)} f(\alpha_e) d\mu(\alpha_e)$$

$f(\alpha_e)$ étant la densité de probabilité de α_e dans $\mathcal{A}_e$. Or, à tout point α_v de $\mathcal{A}$, par l'opérateur d'observation σ_{ef} , correspond sur $\mathcal{A}_e$ une loi de

probabilité de densité $f(\alpha_e/\alpha_v)$ telle que $f(\alpha_e/\alpha_v) = 0$ si $\alpha_e \notin \Delta_{ef}(\alpha_v)$. Par suite, la densité $f(\alpha_e)$ s'écrit :

$$f(\alpha_e) = \sum_v f(\alpha_e/\alpha_v)p(\alpha_v)$$

La condition $\alpha_e \in \Delta_{ef}(\alpha_v)$ étant équivalente à la condition $\alpha_v \in \Delta_f(\alpha_e)$, il vient alors :

$$f(\alpha_e) = \sum_{v, \Delta_f(\alpha_e)} f(\alpha_e/\alpha_v)p(\alpha_v)$$

et

$$p(\alpha_e) = \int_{\Delta_{ep}(\alpha_e)} \sum_{v, \Delta_f(\alpha_e)} f(\alpha_e/\alpha_v)p(\alpha_v)d\mu(\alpha_e) \quad (6)$$

En reportant (5) et (6) dans (4), l'entropie $S(A, \mathcal{P})$ s'écrit :

$$S(A, \mathcal{P}) = - \sum_l \left[\int_{\Delta_{ep}(\alpha_l)} \sum_{v, \Delta_f(\alpha_e)} f(\alpha_e/\alpha_v)p(\alpha_v)d\mu(\alpha_e) \right] \log_2 \left[\sum_{v, \Delta_t(\alpha_l)} p(\alpha_v) \right]$$

IV.2. Entropie mesurable et entropie absolue.

Comme nous l'avons fait pour une variable aléatoire continue, nous allons maintenant exprimer l'entropie mesurable d'une variable aléatoire discrète puis son entropie absolue.

Nous avons appelé entropie mesurable la valeur que prend l'entropie utile $S(A, \mathcal{P})$ lorsque $\mathcal{P} = \mathcal{P} \{ \Delta_s, \Delta_{ef}, 0, 0 \}$. C'est également l'espérance mathématique des informations mesurables $I_m(\alpha_e)$.

$$S_m(A, \mathcal{P}) = \int_{\mathcal{A}_e} f(\alpha_e)I_m(\alpha_e)d\mu(\alpha_e)$$

Dans le cas d'une variable aléatoire discrète, nous avons :

$$I_m(\alpha_e) = - \log_2 [p(\alpha \in \Delta_{sf}(\alpha_e))] = - \log_2 \sum_{v, \Delta_{sf}(\alpha_e)} p(\alpha_v)$$

Comme $f(\alpha_e) = \sum_{v, \Delta_f(\alpha_e)} f(\alpha_e/\alpha_v)p(\alpha_v)$, l'entropie mesurable s'écrit :

$$S_m(A, \mathcal{P}) = - \int_{\mathcal{A}_e} \left[\sum_{v, \Delta_f(\alpha_e)} f(\alpha_e/\alpha_v)p(\alpha_v) \right] \log_2 \left[\sum_{v, \Delta_{sf}(\alpha_e)} p(\alpha_v) \right] d\mu(\alpha_e)$$

On notera que, dans cette dernière intégrale, les sommes $\sum_{v, \Delta_f(\alpha_e)}$ et $\sum_{v, \Delta_{sf}(\alpha_e)}$

sont bien des fonctions de α_e puisque les segments $\Delta_f(\alpha_e)$ et $\Delta_{sf}(\alpha_e)$ sur lesquels sont faites les sommations sont eux-mêmes des fonctions de α_e .

Évaluons maintenant l'entropie absolue $S(A)$ de la variable aléatoire A , qui est la valeur de l'entropie utile $S(A, \mathcal{P})$ lorsque le procédé $\mathcal{P}$ devient idéal. Dans ce cas particulier, chaque valeur mesurée α_l correspond à une seule valeur observée α_v et lui est identique (fig. 8) de telle sorte que $p(\alpha_l) = p(\alpha_v)$ pour $l = v$ et $I_u(\alpha_l) = -\log_2 p(\alpha_v)$ d'où :

$$S(A, \mathcal{P}) = \sum_l p(\alpha_l) I_u(\alpha_l) = - \sum_v p(\alpha_v) \log_2 p(\alpha_v) = S(A)$$

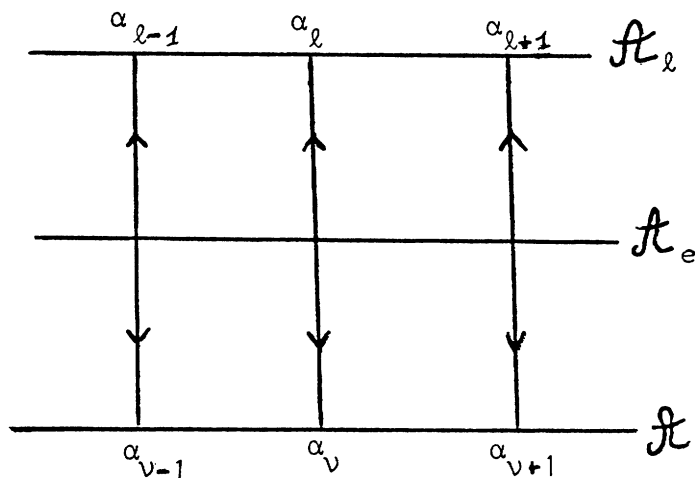


FIG. 8.

Ainsi, pour un procédé idéal, l'entropie utile d'une variable aléatoire discrète s'identifie à l'entropie absolue, elle-même identique à l'entropie de Shannon.

IV.3. Événements discernables par un procédé $\mathcal{P}$.

La capacité d'un procédé d'observation à distinguer plusieurs événements entre eux est une qualité importante et correspond à la notion de « pouvoir séparateur ». Nous examinons ici cette propriété du point de vue informationnel.

Les événements d'une distribution discrète sont discernables par un procédé d'observation $\mathcal{P} \{ \Delta_s, \Delta_{ef}, \Delta_{ep}, \Delta_l \}$ si :

— à chaque événement α_v est associé un résultat d'observation α_l et un seul,

— à chaque résultat d'observation α_l est attribué au plus un événement α_v .

Ceci signifie donc que le processus de mesure est un processus identique : le processus interprétatif est l'inverse du processus d'observation.

Calculons alors l'entropie utile de la variable aléatoire discrète A relativement à $\mathcal{P}$:

$$\begin{aligned} S(A, \mathcal{P}) &= \sum_l p(\alpha_l) I_u(\alpha_l) \\ &= - \sum_l p(\alpha_l) \log_2 [p(\alpha \in \Delta_l(\alpha_l))] \end{aligned}$$

Or, d'après la définition précédente des événements discernables, nous pouvons écrire :

$$\begin{aligned} p(\alpha_l) &= p(\alpha_v) \quad \text{pour} \quad l = v \\ \text{et} \quad p(\alpha \in \Delta_l(\alpha_l)) &= p(\alpha_v) \end{aligned}$$

On obtient par suite :

$$S(A, \mathcal{P}) = - \sum_v p(\alpha_v) \log_2 p(\alpha_v) \quad (7)$$

L'entropie utile d'événements discernables par un procédé d'observation $\mathcal{P}$ est donc indépendante des paramètres qui caractérisent $\mathcal{P}$.

Nous pouvons remarquer de plus que l'expression (7) est identique à l'entropie absolue définie précédemment ainsi qu'à l'entropie de Shannon. Ainsi sont mises en évidence les importantes propriétés suivantes qui spécifient les limites de validité de l'entropie de Shannon.

PROPRIÉTÉ 1. — « L'entropie absolue $S(A)$ d'une variable aléatoire discrète est l'entropie utile $S(A, \mathcal{P})$ de cette variable lorsque les événements sont discernables par le procédé d'observation ».

PROPRIÉTÉ 2. — « L'entropie absolue $S(A)$ d'une variable aléatoire discrète A est la valeur maximale de l'entropie utile $S(A, \mathcal{P})$ ».

Il est à signaler une autre situation fréquente en physique : celle des variables aléatoires dont la densité $\rho(\alpha)$ est nulle sauf sur un ensemble de domaines disjoints, la mesure de chaque domaine étant petite par rapport à la mesure du domaine sur lequel $\rho(\alpha)$ est définie. Nous appelons « distributions catastrophiques » (*) de telles distributions de probabilités. On trouvera dans la référence [18] une analyse de leurs propriétés.

En appliquant notre théorie de l'observation au cas des distributions

(*) L'expression est empruntée à THOM [23] et se réfère au fait que $\rho(\alpha)$ varie très rapidement à l'intérieur des domaines C_v .

discrètes, nous avons donc établi que, dès lors que les événements observés ne sont plus discernables par le procédé d'observation, l'expression de l'entropie doit inclure les caractéristiques de ce procédé d'observation.

V. ÉTUDE COMPARATIVE

Nous avons rappelé, au début de cet article, que plusieurs auteurs ont déjà tenté d'introduire directement la précision de l'observation dans l'expression de l'entropie d'une variable aléatoire continue, mais en se limitant à une analyse relativement sommaire du processus d'observation. Nous allons maintenant comparer l'expression de l'entropie $S(A, \mathcal{P})$ que nous avons proposée pour une variable aléatoire continue aux entropies déjà introduites, à savoir celles de Gibbs, de Kolmogorov, de Sempi et de Majernik. Il apparaît que ces entropies se présentent comme des limites ou des cas particuliers de l'entropie $S(A, \mathcal{P})$. Leur signification physique s'en trouve ainsi précisée

V.1. Entropie absolue et entropie de Shannon.

Rappelons tout d'abord la propriété concernant la relation qui relie l'entropie $S(A, \mathcal{P})$, l'entropie absolue $S(A)$ et la fonction de Shannon $H(A)$ pour une variable aléatoire continue A et qui s'exprime dans le théorème suivant [19] :

THÉORÈME. — La différence des entropies de deux variables aléatoires relativement à un même procédé $\mathcal{P} \{ \Delta_s, \Delta_{ef}, \Delta_{ep}, \Delta_l \}$ est égale à la différence de leurs fonctions de Shannon lorsque $\Delta_s, \Delta_{ef}, \Delta_{ep}$ et Δ_l tendent simultanément vers zéro :

$$\lim_{\mathcal{P} \text{ idéal}} \{ S(A, \mathcal{P}) - S(A', \mathcal{P}) \} = S(A) - S(A') = H(A) - H(A')$$

La fonction empirique de Shannon se voit ainsi attribuer un sens physique beaucoup plus précis.

De même, la fonction $\mathcal{H}$ de Boltzmann doit être replacée dans ce contexte où elle prend la signification suivante :

« La variation au cours du temps de l'entropie absolue d'un système thermodynamique est égale à la variation changée de signe de sa fonction $\mathcal{H}$ ».

V.2. Entropie utile et ε -entropie.

Examinons maintenant le rapport qui peut exister entre l'entropie utile et l' ε -entropie de Kolmogorov [15] dont on rappelle la définition proposée par Renyi [14]. « Si $\mathcal{A}$ est un espace métrique précompact,

pour tout $\varepsilon > 0$, $\mathcal{A}$ peut être divisé en un nombre fini d'ensembles disjoints ayant chacun une mesure inférieure ou égale à ε . On appelle ε -partition de $\mathcal{A}$ un tel système d'ensembles $\mathcal{A}_1(\varepsilon), \dots, \mathcal{A}_{n(\varepsilon)}(\varepsilon)$ dont la réunion est $\mathcal{A}$. A étant une variable aléatoire définie sur $\mathcal{A}$ et à valeurs réelles, l' ε -entropie de A par rapport à l' ε -partition $\mathcal{A}_1(\varepsilon), \dots, \mathcal{A}_{n(\varepsilon)}(\varepsilon)$ de $\mathcal{A}$ est, par définition :

$$H(\varepsilon, A) = - \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log_2 p_k(\varepsilon)$$

$p_k(\varepsilon)$ étant la probabilité d'observer l'événement α dans la partie $\mathcal{A}_k(\varepsilon)$ ».

Si l'on considère l'expression de l'entropie utile dans le cas d'un procédé $\mathcal{P} \{ 0, 0, \Delta_{ep}, \Delta_l \}$:

$$S(A, \mathcal{P}) = - \sum_i p(\alpha_i) \log_2 p(\alpha_i)$$

il apparaît qu'elle est identique à $H(\varepsilon, A)$.

Cette comparaison apporte toutefois une précision essentielle concernant le critère de construction de l' ε -partition, qui se trouve défini par la donnée de l'ensemble $\mathcal{A}_l$ des résultats d'observation (« échelle de lecture ») et par les caractéristiques de l'étalonnage (segments $\Delta_{ep}(\alpha_l)$).

L'entropie utile constitue en outre une généralisation de l' ε -entropie puisque le caractère précompact de l'ensemble $\mathcal{A}$ n'a pas eu à être postulé.

Cette constatation est particulièrement intéressante en raison de l'importance historique du concept d' ε -entropie en théorie de l'information.

V.3. Entropie utile et entropie grossière de Gibbs.

Les travaux de Gibbs [14] [15], antérieurs à ceux de Shannon, suscitent depuis quelques années un regain d'intérêt.

Gibbs divise l'espace des phases sur lequel il travaille en cellules dont la position, les dimensions et la forme dépendent des connaissances initiales que nous avons sur le système étudié. Ces cellules, dont l'extension et la position restent fixes dans le temps, ont une extension d'autant plus petite que nos connaissances sont plus précises. Gibbs définit alors la densité moyenne ρ_g de remplissage de ces cellules. Cette densité ρ_g dépend de l'extension des cellules, est uniforme à l'intérieur de chaque cellule, et varie d'une cellule à une autre. Il établit ensuite une entropie $\bar{H}$ appelée « grossière » qui a la même forme que celle de Shannon mais qui s'exprime en fonction de ρ_g .

$$\bar{H} = - \int \dots \int \rho_g \log_2 \rho_g d\tau = - \sum_k \rho_{gk} \delta\tau_k \log_2 \rho_{gk}$$

où ρ_{gk} représente la densité moyenne de remplissage de la cellule k dans l'espace des phases et $\delta\tau_k$ son extension.

Nous devons remarquer tout d'abord que cette expression présente un défaut évident, à savoir qu'elle comporte le logarithme d'une densité de probabilité, ce qui est contraire à l'essence même du concept de quantité d'information.

En conséquence, cette fonction de Gibbs se présente comme un cas à part pouvant difficilement apparaître comme une limite ou un cas particulier de l'entropie utile $S(A, \mathcal{P})$.

Elle possède cependant un lien avec $S(A, \mathcal{P})$ que nous avons établi dans la référence [19]. En effet, on démontre la relation suivante :

$$\bar{H} = \frac{S(A, \mathcal{P})}{\text{Log } 2} + \sum_k p_k \text{Log } \delta\tau_k$$

où $S(A, \mathcal{P})$ représente l'entropie relative au procédé $\mathcal{P} \{0, 0, \delta\tau_k, \delta\tau_k\}$ [26].

Ainsi, l'entropie de Gibbs se rattache à la précision proprement dite et non au pouvoir de résolution de $\mathcal{P}$ comme on pourrait le penser à première vue. La maille $\delta\tau_k$ n'est pas assimilable à une limite de résolution (ou limite de sensibilité).

V.4. Entropie mesurable et entropies avec erreur de Sempi.

Comme chez Kolmogorov, le but de Sempi [16] [17] est d'introduire des mesures d'information qui dépendent explicitement de l'erreur. Pour cela, il considère que, dans l'analyse d'une expérience, deux variables aléatoires entrent en jeu : la variable aléatoire A dont nous désirons déterminer la distribution et une variable aléatoire B qui représente l'erreur. La présence de l'erreur expérimentale inhérente à toute mesure ne nous permet pas de dire si la valeur prise par la variable aléatoire A , après réalisation de l'expérience, coïncide avec la valeur a lue sur l'appareil de mesure. On peut seulement conclure que la valeur mesurée est contenue dans l'intervalle $(a - b, a + b)$ où b représente l'erreur associée à la mesure.

Si F est la fonction de répartition de la variable aléatoire A ($F: \mathbb{R} \rightarrow [0, 1]$), la différence $F(a + b) - F(a - b)$ représente la probabilité de trouver la valeur mesurée dans l'intervalle $(a - b, a + b)$. L'information I fournie par l'événement « lire la valeur dans l'intervalle $(a - b, a + b)$ » est donnée par :

$$I = -\text{Log } [F(a + b) - F(a - b)]$$

Si nous considérons I comme une grandeur aléatoire, le calcul de sa valeur moyenne fournit l'expression de l'« entropie avec erreur ».

Deux cas se présentent alors :

— il est possible de prendre la moyenne par rapport à la mesure de

probabilité P sur l'espace-produit $R \times [0, \infty)$ (I prenant ses valeurs dans $[0, +\infty)$).

On obtient :

$$S(F, \phi) = - \iint_{R \times [0, \infty)} \text{Log } [F(a+b) - F(a-b)] d\tilde{P}$$

où ϕ est la fonction de répartition de l'erreur avec $\phi(b) = 0$ pour $b < 0$,

— il est souvent utile de considérer l'erreur comme un paramètre positif.

On a alors :

$$S(F, b) = - \int_R \text{Log } [F(a+b) - F(a-b)] d\mu_F(a)$$

où μ_F représente la mesure de Lebesgue-Stieltjes définie par F .

Cette entropie avec erreur présente une grande analogie avec l'entropie utile dans le cas d'un procédé $\mathcal{P} \{ \Delta_s, 0, 0, 0 \}$ [27]

$$S(A, \mathcal{P}) = - \int_{\mathcal{A}_e} \rho(\alpha_e) \log_2 [\overline{\rho_s(\alpha_e)} \cdot \mu \{ \Delta_s(\alpha_e) \}] d\mu(\alpha_e)$$

Si nous faisons intervenir la fonction de répartition F et la mesure de Lebesgues-Stieltjes μ_F définie par F , nous pouvons en effet écrire :

$$\overline{\rho_s(\alpha_e)} \cdot \mu \{ \Delta_s(\alpha_e) \} = \int_{\alpha_e - \frac{\Delta_s(\alpha_e)}{2}}^{\alpha_e + \frac{\Delta_s(\alpha_e)}{2}} \rho(\alpha) d\mu(\alpha) = F\left(\alpha_e + \frac{\Delta_s(\alpha_e)}{2}\right) - F\left(\alpha_e - \frac{\Delta_s(\alpha_e)}{2}\right)$$

$$d\mu_F(\alpha_e) = \rho(\alpha_e) d\mu(\alpha_e)$$

d'où

$$S(A, \mathcal{P}) = - \int_{\mathcal{A}_e} \log_2 \left[F\left(\alpha_e + \frac{\Delta_s(\alpha_e)}{2}\right) - F\left(\alpha_e - \frac{\Delta_s(\alpha_e)}{2}\right) \right] d\mu_F(\alpha_e)$$

On relève néanmoins que Sempì considère que l'« erreur de mesure » (jouant le rôle de notre terme Δ_s) peut aussi être traitée comme une variable aléatoire dans $\mathcal{A}_e$. Ce point de vue concerne plutôt un phénomène de fluctuation qui devrait alors se rattacher à la fidélité. En effet, il existe généralement une dépendance fonctionnelle entre α et $\Delta_s(\alpha)$ qui ne présente aucun aspect fondamentalement aléatoire.

L'argument suivi par Sempì pour justifier la forme de l'« entropie avec erreur » mérite d'ailleurs d'être analysé. Il peut se résumer ainsi : lorsqu'on obtient le résultat a , on peut seulement affirmer que la valeur à mesurer était comprise dans l'intervalle $[a-b, a+b]$. Ce raisonnement concerne donc l'interprétation de la mesure, mais il est quelque peu ambigu car il peut aussi bien s'appliquer au cas de la sensibilité (procédé $\mathcal{P} \{ \Delta_s, 0, 0, 0 \}$) qu'à celui de la fidélité (procédé $\mathcal{P} \{ 0, \Delta_{ef}, 0, 0 \}$); en effet pour les deux procédés, le processus interprétatif présente une analogie formelle. Cet

argument apparaît insuffisant pour justifier complètement l'expression proposée; il faut en outre recourir à l'analyse du processus d'observation, où la différence entre les deux procédés cités apparaît clairement.

V.5. Entropie fine.

D'autres auteurs [28] [29] signalent également que l'entropie d'une variable aléatoire continue doit inclure la finesse de la mesure effectuée sur cette variable et considèrent alors l'expression de l'entropie suivante :

$$H(X) = - \int \omega(x) \log_2 [\omega(x)\Delta x] dx \quad (9)$$

où $\omega(x)$ est la densité de probabilité de la variable X et Δx la finesse de la mesure. L'introduction de Δx répond ici à une intuition afin d'assurer l'homogénéité de l'expression $H(X)$ mais n'est nullement justifiée du point de vue physique. On ne voit pas très bien ce que représente le produit $\omega(x) \cdot \Delta x$ sauf dans le cas où $\omega(x)$ est constant sur l'intervalle Δx car alors $\omega(x) \cdot \Delta x$ est égal à la probabilité de trouver l'événement observé à l'intérieur de Δx .

En fait, cette expression de $H(X)$ représente la limite de l'entropie mesurable $S_m(A, \mathcal{P})$ lorsque Δ_s , Δ_f et Δ_{sf} tendent vers zéro.

En effet, on a :

$$S_m(A, \mathcal{P}) = - \int_{\mathcal{A}_e} \overline{\rho_f(\alpha_e)} \log_2 [\overline{\rho_{sf}(\alpha_e)} \cdot \mu \{ \Delta_{sf}(\alpha_e) \}] d\mu(\alpha_e)$$

Pour Δ_f et Δ_{sf} petits, nous avons :

$$\overline{\rho_f(\alpha_e)} \rightarrow \rho(\alpha_e) \quad \text{et} \quad \overline{\rho_{sf}(\alpha_e)} \rightarrow \rho(\alpha_e)$$

ce qui permet d'écrire :

$$S_m(A, \mathcal{P}) = - \int_{\mathcal{A}_e} \rho(\alpha_e) \cdot \log_2 [\rho(\alpha_e) \cdot \mu \{ \Delta_{sf}(\alpha_e) \}] d\mu(\alpha_e)$$

Cette expression est analogue à la formule (9). Notre concept d'entropie mesurable constitue donc une généralisation aux observations de faible qualité de l'expression proposée notamment par Majernik pour des mesures fines.

CONCLUSION

Le travail que nous avons présenté ici comporte deux aspects complémentaires, centrés respectivement sur l'observation et sur l'information.

Le premier aspect concerne la formalisation mathématique de l'observation scientifique en général. En partant de concepts propres à la physique expérimentale, nous avons approfondi et formalisé les concepts

classiques de sensibilité, fidélité et précision ; nous avons ainsi construit une ébauche de théorie de l'observation, qui intègre de manière organique à la fois les caractéristiques du procédé utilisé et un élément essentiel du processus de mesure : l'interprétation d'une observation. Cela nous a conduit à la définition et à l'évaluation de la « limite totale d'incertitude associée à un résultat d'observation α_i ».

Cette première étape nous a permis d'aborder ensuite le second aspect considéré : l'aspect informationnel. Nous avons proposé des définitions pour l'information associée à un résultat d'observation, et pour l'entropie $S(A, \mathcal{P})$ d'une variable aléatoire A relativement à un procédé d'observation.

L'entropie $S(A, \mathcal{P})$ apparaît tout particulièrement intéressante dans le cas où A est continue, car elle est exempte des défauts inhérents à l'entropie la plus utilisée en théorie de l'information, celle de Shannon : elle présente en outre des points de rapprochement avec plusieurs autres entropies (entropie grossière de Gibbs, ε -entropie de Kolmogorov, entropie de Shannon, entropie avec erreur de Sempi, entropie fine) qui permettent d'éclairer leurs significations physiques respectives et de préciser leurs limites de validité. En ce sens, notre concept, par sa généralité, tient le rôle d'un élément de référence.

Mais son intérêt débord largement le cadre des distributions continues puisqu'il s'applique également au cas de distributions discrètes, pour lesquelles nous avons établi une propriété qui pourrait s'avérer importante : l'entropie $S(A, \mathcal{P})$ d'une variable aléatoire discrète A diffère de l'entropie

classique $S(A) = - \sum_i p_i \log p_i$ dès lors que les événements observés α_i

(de probabilité p_i) ne sont plus discernables par le procédé $\mathcal{P}$.

Notre étude pourrait trouver de nombreuses applications, tant en ce qui concerne l'analyse de certains procédés d'observation, que l'évaluation de l'information acquise par l'observation de phénomènes aléatoires, particulièrement en transmission de l'information. Enfin, il serait certainement fructueux, en se basant sur les travaux de Kampé de Fériet, Picard et leurs collaborateurs, de l'étendre au cas où les événements observés ne sont pas de nature statistique, et ne peuvent pas être répétés. Il s'agit là en effet d'une classe de phénomènes couramment observés en pratique, mais pour lesquels l'approche informationnelle que nous avons utilisée n'est pas adaptée.

BIBLIOGRAPHIE

- [1] M. MÜGÜR-SCHÄCHTER, La relation entre l'entropie informationnelle et l'entropie statistique décrite par le concept probabiliste nouveau de fonctionnelle d'opacité d'une statistique. *C. R. Acad. Sci.*, t. **288 A**, 1979, p. 771.
- [2] J. BASS, *Éléments de calcul des probabilités*. Masson, 1974.
- [3] C. M. MARLE, *Mesures et probabilités*. Hermann, Paris, 1974.

- [4] A. RENYI, *Dimension, entropy and information*. Trans. Prague Conference on information theory, statistical decision functions, random processes, Prague, 1960, p. 545-556.
- [5] S. GUIASU, R. THEODORESCU, *Incertitude et information*. Les Presses de l'Université, Laval, Québec, 1971.
- [6] C. E. SHANNON, The mathematical Theory of Communication, I, II. *Bell Syst. Techn. J.*, t. 27, 1948, p. 379, p. 623.
- [7] C. E. SHANNON, W. WEAVER, *The mathematical theory of communication*. Univ. of Illinois Press, Urbana, 1949.
- [8] A. KOLMOGOROV, Logical basis for information theory. I. E. E. Trans. on Information Theory, II, t. 14, 5, 1968, p. 662-664.
- [9] A. I. KHINCHIN, *The entropy concept in probability theory* dans *Mathematical foundations of information theory*. Dover edition, 1967.
- [10] S. GUIASU, R. THEODORESCU, *La théorie mathématique de l'information*. Collection Sigma, Dunod, 1968.
- [11] B. PICINBONO, *Éléments sur les théories du signal, de la détection et de l'information*. Cours D. E. A. de mathématiques statistiques, Faculté des Sciences de Paris.
- [12] A. BLANC-LAPIERRE, *Mécanique statistique*. Masson, 1967.
- [13] E. ROUBINE, *Introduction à la théorie de la communication*. Masson, 1970.
- [14] J. W. GIBBS, *Elementary principles in statistical mechanics*. Yale Univ. Press, New Haven, 1902.
- [15] N. KOLMOGOROV, On the Shannon theory of information transmission in the case of continuous signals. I. R. E. Trans. on Information Theory, t. 12, 1956, p. 102.
- [16] C. SEMPI, The entropies with error. *Rendiconti di Matematica*, t. 9, 1, série VI, 1976, p. 1.
- [17] C. SEMPI, *On the limit of the entropies with error*. Communication personnelle, 1976.
- [18] J. PADET, C. PADET, A. M. CAMES, M. MÜGÜR-SCHÄCHTER, *Observation, informations et entropies*, *Entropie*, n° 72-73, 1977.
- [19] C. PADET, Contribution à une étude de l'observation et de la mesure à l'aide de concepts informationnels. *Thèse Sci. Phys. Reims*, 1978.
- [20] R. VALLÉE, *Sur la formalisation mathématique en théorie de l'observation*. Actes du VIII^e Congrès international de cybernétique, Namur, Sept. 1973.
- [21] R. VALLÉE, Sur deux classes d'opérateurs d'observation. *C. R. Acad. Sci.*, t. 233, 1951, p. 1350.
- [22] R. VALLÉE, Opérateurs d'observation et théorie de l'information. *C. R. Acad. Sci.*, 1951, p. 1428.
- [23] R. THOM, *Modèles mathématiques de la morphogenèse*. Coll. 10/18, Paris, 1974.
- [24] A. RENYI, On measures of entropy and information. *Proc. 4th Berkeley Symp. Math. Statist. and Prob.* University of California Press, t. 1, 1960, p. 547.
- [25] P. CASTAING, *Thermodynamique Statistique*. Masson, 1970.
- [26] M. MÜGÜR-SCHÄCHTER, C. PADET, J. PADET, Le concept de quantité d'information accessible par mesure correspondant à une distribution de probabilités continue. *C. R. Acad. Sci.*, t. 282 A, 1976, p. 487.
- [27] C. PADET, M. MÜGÜR-SCHÄCHTER, J. PADET, Une nouvelle expression de l'entropie pour les distributions statistiques continues. *C. R. Acad. Sci.*, t. 281 A, 1975, p. 993.
- [28] V. MAJERNIK, Some applications of information theory in quantum mechanics. *Acta physica austriaca*, t. 25, 3, 1967, p. 243.
- [29] J. FRONTEAU, *Vers une description non conservative de l'évolution en physique* (comm. personnelle), 1973.

(Manuscrit reçu le 29 mai 1980)