



HAL
open science

Observation, informations et entropies. Valeur informationnelle de la mesure en physique statistique

Jacques Padet, Colette Padet, Anne-Marie Cames-Pintaux, Miora Mugur-Schachter

► To cite this version:

Jacques Padet, Colette Padet, Anne-Marie Cames-Pintaux, Miora Mugur-Schachter. Observation, informations et entropies. Valeur informationnelle de la mesure en physique statistique. Entropie : thermodynamique – énergie – environnement – économie, 1977, 72-73, pp.5-20. hal-03407944

HAL Id: hal-03407944

<https://hal.univ-reims.fr/hal-03407944>

Submitted on 28 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Observation, Informations et Entropies

Valeur informationnelle de la mesure en physique statistique

Jacques PADET, Colette PADET,
Anne-Marie CAMES-PINTAUX
et Mioara MUGUR-SCHACHTER (1)

The present paper aims at stating an expression physically satisfactory of the entropy for a continuous statistical distribution, which implies we had to take into account characteristic parameters of the proceeds of observation, which have been formalized in a suitable manner. To reach this aim, first of all we define three sets of experimental situations to be referred to, corresponding to three classical properties of sensitiveness, reliability and accuracy for measurement apparatus; then in the general case where the three situations above occur together we determine the information linked with the observation of an event. The analysis of the measurement process then leads to defining the entropy $S(A, \mathcal{P})$ of a random variable A for a proceed of observation \mathcal{P} . Some properties of the function $S(A, \mathcal{P})$ are studied and a certain number of particular cases examined. The expressions of entropy proposed by Boltzmann, Gibbs, Shannon, Kolmogorov and Renyi appear as limits of the entropy $S(A, \mathcal{P})$, and their physical meaning is given precisely.

Ziel der vorliegenden Studie ist es, einen physikalisch zufriedenstellenden Ausdruck des Wärmegewichts für eine fortlaufende statistische Reihe aufzustellen, was die Berücksichtigung der für ein Beobachtungsverfahren charakteristischen Parameter, in brauchbare Form gebracht, voraussetzt. Dazu bestimmt man zuerst drei Bezugsversuchssituationen, die den drei klassischen Eigenschaften Empfindlichkeit, Treue und Genauigkeit, bei Messgeräten, entsprechen. Dann bestimmt man, im allgemeinen Fall, wo sich die drei Bezugssituationen überlagern, die Information, die der Beobachtung eines Ereignisses zugeordnet ist. Die Analyse des Messprozesses führt dann dazu, das Wärmegewicht $S(A, \mathcal{P})$ zu definieren, mit einer stochastischen Variabel A , bezogen auf ein Beobachtungsverfahren \mathcal{P} . Man untersucht einige Eigenschaften der Funktion $S(A, \mathcal{P})$ und eine bestimmte Anzahl Einzelfälle. Die früher vorgeschlagenen Ausdrücke des Wärmegewichts (von Boltzmann, Gibbs, Shannon, Kolmogorov und Renyi) erscheinen als Grenzwerte des Wärmegewichts $S(A, \mathcal{P})$ und ihr physikalischer Sinn wird verdeutlicht.

El presente estudio tiende a establecer una expresión físicamente satisfactoria de la entropía para una distribución estadística continua, lo que implica que hayan de tenerse en cuenta los parámetros característicos de un procedimiento de observación debidamente formalizados. Con este fin, se definen primeramente tres clases de situaciones experimentales de referencia, las cuales corresponden a las tres propiedades clásicas de sensibilidad, fidelidad y precisión en los aparatos de medida; luego, en el caso general de que se superpongan las tres situaciones de referencia, se determina la información asociada a la observación de un hecho acontecido. El análisis del proceso de medida conduce entonces a la definición de la entropía $S(A, \mathcal{P})$ de una variable aleatoria A , relativamente a un procedimiento de observación \mathcal{P} . Se estudian algunas propiedades de la función $S(A, \mathcal{P})$, examinándose un cierto número de casos particulares. Las expresiones de la entropía precedentemente propuestas por Boltzmann, Gibbs, Shannon, Kolmogorov y Renyi asimismo se presentan con carácter de límites de la entropía $S(A, \mathcal{P})$ precisándose su significación física.

1 INTRODUCTION

Les fondements et les implications de la notion classique d'entropie sont depuis un certain temps déjà assez clairement dégagés [1 à 4], mais il subsiste cependant à propos de ce concept un certain nombre de lacunes, tant en ce qui concerne la physique statistique que la transmission de l'information. Nous en relèverons particulièrement deux : l'une est relative au problème de l'entropie des distributions continues, qui n'a reçu jusqu'à présent que des solutions relativement partielles ; l'autre concerne l'incidence sur l'entropie des limitations inhérentes au procédé d'observation, encore incomplètement évaluée.

(1) Laboratoire de Mécanique Quantique et Groupe de Thermomécanique associé. Faculté des Sciences - B.P. 347 - 51062 REIMS Cedex.

En fait, des deux questions sont liées, et la clé de la première - celle des distributions continues - est fournie par la solution de la seconde, ainsi que l'avait déjà remarqué Kolmogorov. Des progrès notables ont certes été faits au plan mathématique, mais sans que la jonction avec les problèmes rencontrés en physique expérimentale soit clairement établie.

Dans l'étude qui suit, nous essayons d'aborder globalement le problème de l'observation en physique. Cela nous conduit à définir en premier lieu trois classes de situations expérimentales de référence, puis à montrer comment elles se superposent généralement lors d'une observation. Après avoir établi une expression aussi générale que possible pour l'entropie $S(A, \mathcal{P})$ d'une variable aléatoire A lorsque l'observation est réalisée au moyen d'un procédé \mathcal{P} , nous étudions brièvement quelques cas particuliers. Nous consacrons enfin un paragraphe aux distributions « catastrophiques » (avec les

distributions discrètes comme cas limite) en raison de leur importance pratique.

Le pluriel dont sont affectés, dans le titre, les mots information en entropie est inhabituel ; nous le justifierons en faisant apparaître trois « niveaux d'information » auxquels correspondent trois expressions différentes des grandeurs information et entropie, qui contribuent à une généralisation et à un approfondissement de ces concepts.

On pourrait traiter le problème d'une manière un peu plus synthétique et rigoureuse avec un outillage mathématique plus élaboré. Nous y avons toutefois délibérément renoncé pour dégager autant que possible les aspects physiques et conceptuels de la théorie proposée, et pour fournir au lecteur non mathématicien des résultats aisément lisibles et utilisables.

2 PRINCIPALES NOTATIONS UTILISEES

- \mathcal{A} : ensemble des événements observés.
- \mathcal{A}_e : ensemble des événements enregistrés par l'appareil.
- \mathcal{A}_m : ensemble discret des valeurs mesurées.
- α, α_e : éléments ponctuels de \mathcal{A} et \mathcal{A}_e .
- α_i : valeur mesurée, élément de l'ensemble \mathcal{A}_m .
- $\Delta_s(\alpha)$: limite de résolution du procédé d'observation pour la valeur α .
- $\Delta_f(\alpha)$: limite de fidélité du procédé d'observation pour la valeur α .
- Δ_i : écart entre les valeurs α_i et α_{i+1} .
- Δ_{pi} : ensemble des valeurs α_e ayant la même valeur mesurée α_i .
- $\Delta_{t,k}$: limite d'incertitude pour la valeur mesurée α_k .
- $f(\alpha_e)$: densité de probabilité de α_e dans \mathcal{A}_e .
- $\rho(\alpha)$: densité de probabilité de α dans \mathcal{A} .
- $\overline{\rho_f(\alpha)}$: valeur moyenne de $\rho(\alpha)$ sur le segment $\Delta_f(\alpha)$.
- $\overline{\rho_s(\alpha)}$: valeur moyenne de $\rho(\alpha)$ sur le segment $\Delta_s(\alpha)$.

3 CONCEPTIONS DE L'ENTROPIE EN PHYSIQUE

3.1 Terminologie et fondements physiques

3.1.1 Événements et processus

Soit un ensemble \mathcal{A} constitué d'événements physiques observables. Ce terme d'événements recouvre en fait une très grande variété de données physiques ; ainsi en thermodynamique, il peut s'agir de la pression, de la température en un point d'un système, ou bien d'une cellule de l'espace des phases ; en optique l'événement peut être une couleur, une frange d'interférence, une intensité de rayonnement... ; en transmission de l'information, ce sera la fréquence ou l'intensité d'un signal, ou encore une lettre d'un alphabet etc ; la liste précédente n'est évidemment pas exhaustive. Les événements sont usuellement appelés « états » en thermodynamique et « résultats d'une expérience » en physique.

On définit sur \mathcal{A} une variable aléatoire A . Dans le langage de la physique, A est une expérience. Pour chaque événement, A prend une valeur donnée $a \in \mathbf{R}$. Si A est discrète, l'ensemble \mathcal{A} est dénombrable et ses éléments seront notés α_i . A chaque événement élémentaire α_i est associée une probabilité de réalisation p_i ($0 \leq p_i \leq 1$) telle que $\sum p_i = 1$. La donnée des p_i définit une distribution de probabilités discrète.

Si A est continue, nous admettons que l'ensemble \mathcal{A} est identifiable à \mathbf{R} , et nous noterons α ses points. On appellera $\rho(\alpha)$ la densité de probabilité de A , qui est une fonction réelle intégrable telle que $\int_{\mathbf{R}} \rho(\alpha) d\alpha = 1$. A chaque segment E de \mathcal{A} est associée la probabilité $P(E) = \int_E \rho(\alpha) d\alpha$, la donnée de la fonction $\rho(\alpha)$ définissant une distribution continue de probabilités.

Dans le cas d'une variable aléatoire continue, on réserve en théorie des probabilités le nom d'événements à des parties mesurables de \mathbf{R} , auxquelles on peut associer une probabilité. Un événement ponctuel a donc en toute rigueur une probabilité nulle. Aussi, lorsque nous parlerons plus loin d'un événement ponctuel (ou isolé) α , cela sous-entendra en fait un événement infinitésimal de mesure $d\alpha$ et de probabilité $\rho(\alpha) d\alpha$ très petite, mais non nulle. Enfin, dans toute la suite, nous userons par commodité de l'abus de langage qui consiste à confondre l'événement ponctuel α ou α_i avec la valeur a correspondante de la variable aléatoire A .

D'autre part, on appellera « processus » un couple ordonné d'événement (α, β) α et β n'appartenant pas obligatoirement au même ensemble \mathcal{A} (ce qui est moins restrictif qu'en thermodynamique [6] [7]). La succession, que nous noterons \oplus , de deux processus (α, β) et (β, γ) est égale au processus (α, γ) :

$$(\alpha, \beta) \oplus (\beta, \gamma) = (\alpha, \gamma)$$

Cette opération n'est définie que lorsque les deux processus présentent un événement commun (ici β). Enfin, l'inverse du processus (α, β) est (β, α) ; en effet, la succession d'un processus et de son inverse est égale au processus identique :

$$(\alpha, \beta) \oplus (\beta, \alpha) = (\alpha, \alpha)$$

3.1.2 Information

Dans le cas où l'ensemble \mathcal{A} est discret, on appelle par définition « information relative à l'événement α_i » la grandeur :

$$I(\alpha_i) = -\log_2 p_i \quad (1)$$

exprimée en « bit » (binary unit). Il est essentiel de bien remarquer que $I(\alpha_i)$, étant une fonction décroissante de p_i représente une mesure conventionnelle de l'information qui nous manque relativement à α_i : c'est l'information qui pourrait être fournie par un résultat d'expérience, et non l'information dont nous disposons a priori, cette dernière étant mesurée par p_i .

3.1.3 Information moyenne ou entropie

Il est intéressant de considérer, sur l'ensemble discret \mathcal{A} , la moyenne des informations relatives aux différents événements α_i , généralement notée comme une fonction de la variable aléatoire A :

$$S(A) = \sum_i p_i I(\alpha_i) = - \sum_i p_i \log_2 p_i \quad (2)$$

Cette fonction a été introduite et largement exploitée en théorie de l'information par Shannon [3] [4], qui l'a nommée « entropie » en raison de son analogie avec l'entropie de la thermodynamique statistique.

Considérons en effet un état thermodynamique caractérisé par N molécules réparties dans n éléments de volume, l'élément i contenant N_i molécules ($i = 1$ à n). Le nombre W de complexions qui réalise cet état est donné par la formule de Boltzmann :

$$W = \frac{N!}{\prod_{i=1}^n N_i!}$$

où le signe Π signifie « produit des $N_i!$ », et l'entropie du système est définie par la relation :

$$S_{th} = k \log W \quad (3)$$

dans laquelle k est la constante de Boltzmann. Cette définition, primitivement établie pour des états d'équilibre, peut être étendue sans difficulté aux états hors d'équilibre [9].

Les nombres N et N_i étant très grands, on peut utiliser la formule de Stirling pour calculer $\log W$, et l'on obtient :

$$\log W \approx N \log N - \sum_i N_i \log N_i$$

En appelant $p_i = N_i/N$ la probabilité qu'une molécule soit dans l'élément i , on voit que S peut se mettre sous la forme suivante :

$$\begin{aligned} S_{th} &= -k \sum_{i=1}^n N_i \log \frac{N_i}{N} = -k N \sum_{i=1}^n \frac{N_i}{N} \log \frac{N_i}{N} \\ &= -K \sum_{i=1}^n p_i \log p_i \end{aligned} \quad (4)$$

Les expressions (2) et (4) étant identiques au point de vue formel, il apparaît que l'entropie de la thermodynamique statistique et l'entropie de la théorie de l'information sont de la même essence. Elles diffèrent néanmoins en ceci que la constante de Boltzmann, de par sa dimension, confère à l'entropie thermodynamique un contenu énergétique que ne possède pas la fonction adimensionnelle $S(A)$.

3.1.4 Fonction \mathcal{H} de Boltzmann et fonction H de Shannon

Envisageons maintenant le cas d'une distribution continue. Partant de l'expression (2) qui définit $S(A)$ dans le cas discret, on montre [4] que l'entropie n'est pas bornée lorsque A devient continue : $S(A) \rightarrow \infty$. Nous

reviendrons au paragraphe 5 sur l'interprétation de cette propriété, qui engendre évidemment des difficultés pratiques importantes. Historiquement, l'évolution des travaux relatifs aux distributions continues peut se ramener à deux grandes étapes.

A partir de considérations mécaniques et statistiques, Boltzmann a établi en 1872 que, pour un gaz isolé dont les molécules ont une densité de distribution des vitesses $f(u,v,w)$, la fonction

$$\mathcal{H} = \iiint f(u,v,w) \log f(u,v,w) du dv dw \quad (5)$$

est une fonction non croissante du temps. Cette propriété célèbre mais controversée [10] [11] est connue sous le nom de « Théorème \mathcal{H} ». Quelques années plus tard, Boltzmann a postulé l'identité de la fonction \mathcal{H} avec l'entropie thermodynamique : $S = -k \mathcal{H}$. Il a enfin proposé une expression plus générale de l'entropie pour un système quelconque dont les éléments obéissent aux lois de la mécanique classique : $S = -k \int_{\Gamma} f(p,q) \log f(p,q) dp dq$, où q et p représentent respectivement la position et l'impulsion du système dans l'espace des phases Γ , et $f(p,q)$ la densité de répartition correspondante normalisée dans Γ [12].

Plus récemment, Shannon [4] a adopté une expression semblable en théorie de l'information. Etant donnée une variable aléatoire A continue, de densité $\rho(\alpha)$, il appelle « entropie » la fonction :

$$H(A) = - \int_{-\infty}^{+\infty} \rho(\alpha) \log_2 \rho(\alpha) d\alpha \quad (6)$$

construite « par analogie » avec le cas discret. C'est cette définition quelque peu arbitraire qui est actuellement utilisée dans le traitement des signaux continus, où elle a rendu d'importants services malgré ses défauts évidents.

3.1.5 Critique

Les conventions adoptées pour définir l'information et l'entropie dans le cas des distributions discrètes sont cohérentes et relativement satisfaisantes ; elles ne sont assurément pas les seules possibles, mais elles ont probablement le mérite de la plus grande commodité d'emploi.

Il en va tout autrement dans le cas des distributions continues car la fonction $H(A)$ (ou la fonction \mathcal{H}) n'est pas pleinement cohérente avec son homologue $S(A)$ pour les cas discrets, et présente pour cette raison des difficultés d'interprétation (mis à part le fait, évident, qu'elle est la valeur moyenne de $\log_2 \rho(\alpha)$). Depuis Shannon, les propriétés suivantes sont en effet bien connues :

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

Toutefois, ces propriétés négatives sont pour ainsi dire gommées si l'on compare les fonctions de Shannon

de deux variables aléatoires continues A_1 et A_2 en faisant leur différence $H(A_1) - H(A_2)$. On a puisé dans cette constatation empirique un argument essentiel pour justifier l'utilisation de la fonction $H(A)$, qui est à la base de nombreux travaux [13 à 22]. Notons enfin que les difficultés signalées ont été parfois sous-estimées, voire ignorées ; mais elles sont cependant mentionnées clairement dans plusieurs traités [9] [15], et Roubine [14] en particulier en a fourni une analyse très lucide.

3.2 Introduction du concept de mesure

Nos raisonnements n'ont concerné jusqu'à présent que des distributions de probabilités supposées données a priori. Mais la notion d'information est essentiellement associée à l'observation d'un événement, c'est-à-dire à l'idée de mesure au sens le plus large du terme. Il semble donc nécessaire d'introduire les paramètres caractéristiques d'une mesure dans l'évaluation des grandeurs information et entropie. Cette constatation n'est pas nouvelle [5] [23 à 26] mais n'a donné naissance jusqu'à ces derniers temps qu'à un nombre relativement restreint de travaux.

C'est à Gibbs que l'on doit la première tentative d'intégrer la précision d'une mesure à l'évaluation de l'entropie. Pour remédier aux difficultés soulevées par la fonction \mathcal{H} de Boltzmann, il proposa d'identifier l'entropie à l'expression suivante :

$$S = -k \int_{\Gamma} f(p, q) \log \left\{ C f(p, q) \right\} dp dq \quad (7)$$

dans laquelle C est une constante déterminant la finesse de la structure granulaire introduite dans l'espace des phases par les limites intrinsèques des moyens d'observation utilisés. Nous montrerons plus loin (5.3) que la validité de la relation (7) est cependant limitée au cas où $C \rightarrow 0$.

Dans le même esprit, Kolmogorov [5] a introduit en théorie de l'information la notion de précision d'une observation. Il écrit : « En accord avec le sens propre du terme, l'entropie d'un objet A avec une distribution continue est toujours infinie. Si les signaux continus peuvent néanmoins servir à transmettre une information, c'est seulement parce qu'ils sont toujours observés avec une précision limitée. En conséquence, il est naturel de définir une ϵ -entropie appropriée $H(\epsilon, A)$ de l'objet A en donnant la précision de l'observation ϵ ».

On doit à Renyi [27] [15] une formalisation simple de l' ϵ -entropie. L'ensemble \mathcal{A} des événements α étant muni d'une métrique, on le suppose totalement borné, c'est-à-dire que pour tout $\epsilon > 0$, il peut être divisé en un nombre fini d'ensembles disjoints ayant chacun un diamètre inférieur ou égal à ϵ . On appelle ϵ -partition de \mathcal{A} un tel système $\mathcal{A}_1(\epsilon), \dots, \mathcal{A}_n(\epsilon)$ d'ensembles dont la réunion est \mathcal{A} . Soit $p_k(\epsilon)$ la probabilité d'observer un événement α dans l'élément $\mathcal{A}_k(\epsilon)$. Alors, l' ϵ -entropie est par définition :

$$H(\epsilon, A) = - \sum_{k=1}^{n(\epsilon)} p_k(\epsilon) \log_2 p_k(\epsilon) \quad (8)$$

Moyennant certaines conditions supplémentaires, on démontre qu'il existe une relation à la limite entre $H(\epsilon, A)$ et la fonction de Shannon $H(A)$,

$$\lim_{\epsilon \rightarrow 0} \left\{ H(\epsilon, A) - \log_2 \frac{1}{\epsilon} \right\} = H(A) \quad (9)$$

La formule (8), applicable dans l'esprit de ses auteurs à la transmission de l'information, ne paraît pas avoir été beaucoup utilisée en pratique, probablement en raison du caractère physique assez vague de ce qui est appelé « précision » de l'observation, la construction de l' ϵ -partition n'étant reliée à aucun critère expérimental vraiment précis. On verra plus loin qu'elle s'intègre néanmoins dans la théorie que nous allons exposer.

Depuis peu de temps, les recherches relatives à la relation observation-entropie semblent se multiplier dans des directions assez variées, et il serait trop long de les résumer ici. Nous relèverons seulement le fait que dans ces travaux on a attribué à la notion d'« erreur de mesure » des significations assez différentes et quelquefois imprécises. Citons en particulier Dayantis [28] (détermination expérimentale d'une courbe), Hurwitz [29], Liboff [30] [31] et Hoyningen-Huene [44] (thermodynamique statistique), ainsi que Sempi [32], le point de vue exprimé par ce dernier étant proche du nôtre pour une classe particulière d'observations (cf. paragraphe 4.3.5). Rappelons enfin les importantes contributions de Kampé de Fériet, Forte [33 à 36], Picard [37] [38], Terrenoire et de leurs collaborateurs aux problèmes mathématiques liés à la mesure de l'information et à son extension à des événements de nature non statistique [39].

4 ENTROPIE D'UNE VARIABLE ALEATOIRE CONTINUE A RELATIVEMENT A UN PROCÉDE D'OBSERVATION

Malgré le bilan relativement négatif que nous venons d'établir concernant l'entropie des variables aléatoires continues, une voie semble se dégager, qui mérite d'être explorée plus avant, à savoir la prise en compte, d'une manière opérationnelle, des paramètres caractéristiques du procédé d'observation, et c'est elle que nous nous proposons de suivre maintenant. Pour ce faire, nous allons analyser et formaliser les propriétés habituellement reconnues à tout système expérimental, puis le processus d'observation lui-même. Tirant ensuite les conséquences de ces deux analyses, nous serons amenés à définir « l'information utile associée à la réalisation d'un événement α observé au moyen d'un procédé \mathcal{P} », et l'entropie $S(A, \mathcal{P})$ correspondante, pour laquelle d'intéressantes propriétés seront mises en évidence. Nous montrerons en particulier que la fonction $S(A, \mathcal{P})$ est exempte des défauts qui entachaient les formules de

Boltzmann, Gibbs, Shannon et Kolmogorov - Rényi, mais que ces dernières trouvent naturellement une place et une signification physique dans notre construction.

Il convient de préciser ici clairement le point de vue auquel nous allons nous placer maintenant. Nous admettons que l'ensemble \mathcal{A} qui constitue la « matière » physique observable est continu (le cas discret sera traité comme un cas limite) et que tout événement isolé α (au sens que nous avons donné à ce terme en 2.1.1) est a priori observable. L'observation d'un événement non isolé, c'est-à-dire d'un segment E de \mathcal{A} , sera alors éventuellement attribuée à une imperfection du procédé d'observation.

4.1 Caractéristiques d'un procédé d'observation

Il est bien connu que tout procédé d'observation (on peut dire aussi « appareil de mesure » en prenant ce terme dans un sens très large : à la limite, l'appareil peut être un observateur humain) 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.

L'observation fait donc intervenir d'une façon générale trois ensembles d'événements (ou de valeurs) : l'ensemble \mathcal{A} (continu ou discret) des valeurs α à mesurer ; l'ensemble \mathcal{A}_e (que nous supposons continu) des valeurs α_e enregistrées par l'appareil, et l'ensemble discret \mathcal{A}_m des valeurs mesurées α_i (affichées sur l'échelle de lecture) que nous représenterons sur les figures par des axes parallèles. La valeur enregistrée ne pourrait être identique à la valeur mesurée que si l'affichage était idéal (échelle de lecture continue, étalonnage parfait).

4.1.1 Sensibilité

La sensibilité se traduit par le fait que l'appareil ne permet pas de séparer les événements situés à l'intérieur d'un segment $\Delta_s(\alpha)$ de l'axe des α . En d'autres termes, pour tout événement α_o isolé, l'événement enregistré α_e est identique à α_o . Mais inversement, l'enregistrement d'une valeur $\alpha_{e,o}$ peut provenir, ou bien d'un événement isolé $\alpha_o \equiv \alpha_{e,o}$, ou d'un collectif $\{\alpha\}$ appartenant au segment $\Delta_s(\alpha_o)$ centré sur $\alpha_o \equiv \alpha_{e,o}$ (fig 1). Ainsi, l'indication par un thermomètre d'une température T_o cons-

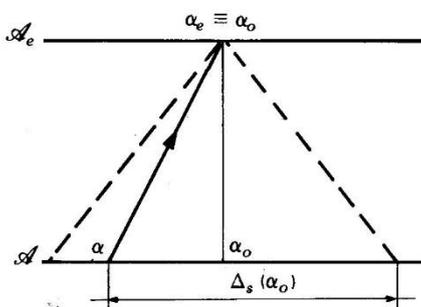


Fig. 1 Incidence de la limite de résolution Δ_s sur l'observation.

tante pendant un certain intervalle de temps ne permet pas de dire si la température T est restée rigoureusement égale à T_o , ou si elle a varié entre $T_o - \Delta_s(T_o)/2$ et $T_o + \Delta_s(T_o)/2$. De même, si une balance mesure un poids P_o cette indication n'est pas modifiée lorsqu'on fait varier P_o de ΔP_o pourvu que $\Delta P_o < \Delta_s(P_o)$. Mais si l'on avait appliqué directement $P_o + \Delta P_o$, la balance aurait affiché cette dernière valeur.

Le segment $\Delta_s(\alpha_o)$ est habituellement appelé « limite de résolution » de l'appareil pour la valeur α_o . Non seulement cette limite de résolution n'est jamais nulle, mais en pratique elle ne peut même pas être rendue arbitrairement petite (on se reportera par exemple aux remarques très pertinentes de Brillouin [26] sur la plus petite longueur mesurable).

4.1.2 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 de plus ou moins grande amplitude qui surviennent généralement pendant l'acte de mesure, comme dans la plupart des phénomènes physiques. La conséquence pratique en est que, à chaque valeur α à mesurer correspond tout un ensemble de valeurs enregistrées α_e possibles, formant un segment $\Delta_f(\alpha_{e,o})$ centré sur $\alpha_{e,o} \equiv \alpha$ (1), avec une densité de probabilité $f(\alpha_e)$ (fig. 2). Dans la suite, nous nous limiterons au cas simple où l'on ne dispose sur les fluctuations d'aucun autre renseignement que leur amplitude maximale $\Delta_f(\alpha)$. La valeur enregistrée α_e est donc a priori n'importe laquelle des valeurs constituant le segment $\Delta_f(\alpha_{e,o})$: ces dernières sont alors équiprobables et la densité conditionnelle $f(\alpha_e/\alpha)$ de α_e connaissant α est uniforme sur Δ_f et nulle en dehors.

$$\left. \begin{aligned} \alpha_e \in \Delta_f(\alpha_{e,o}) &\Rightarrow f(\alpha_e/\alpha) = \frac{1}{\Delta_f(\alpha)} \\ \alpha_e \notin \Delta_f(\alpha_{e,o}) &\Rightarrow f(\alpha_e/\alpha) = 0 \end{aligned} \right\} \quad (10)$$

Nous appellerons $\Delta_f(\alpha)$ « limite de fidélité » de l'appareil pour la valeur α . Désormais, nous considérerons ce terme comme une constante.

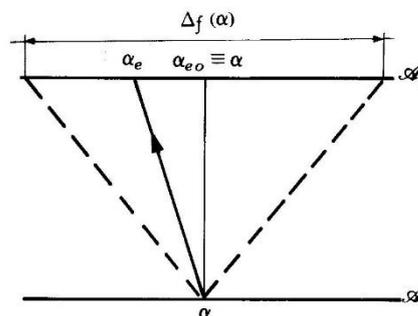


Fig. 2 Incidence de la limite de fidélité Δ_f sur l'observation.

(1) Nous le noterons aussi $\Delta_f(\alpha)$.

4.1.3 Précision

Dans un appareil de mesure, la précision proprement dite caractérise un écart systématique entre la valeur enregistrée α_e et la valeur mesurée α_i qui procède en fait de deux causes distinctes.

D'une part, l'ensemble \mathcal{A}_m des valeurs mesurées est toujours discret et dénombrable, soit que l'on utilise pour l'affichage un repère mobile se déplaçant par rapport à des graduations (on arrondit par exemple à la graduation la plus proche du repère), soit encore que l'on se serve d'un affichage numérique à lecture directe, limité à n chiffres. Toujours est-il que chaque valeur affichée α_i correspond en fait à tout un segment Δ_{pi} de l'ensemble \mathcal{A}_e des valeurs enregistrées : tous les événements α_e de Δ_{pi} auront la même valeur mesurée α_i . Quand à la position de Δ_{pi} par rapport à α_i , elle résulte d'une convention tout à fait arbitraire, et nous choisirons la plus commode qui consiste à prendre pour chaque α_e la valeur α_i immédiatement inférieure (ou égale) (fig. 3). Cela s'écrit, en notant $\alpha_{e,i}$ et $\alpha_{e,i+1}$ les bornes de Δ_{pi} et Δ_i l'intervalle $[\alpha_i, \alpha_{i+1}[$:

$$\Delta_{pi} = [\alpha_{e,i}, \alpha_{e,i+1}[\equiv [\alpha_i, \alpha_{i+1}[= \Delta_i$$

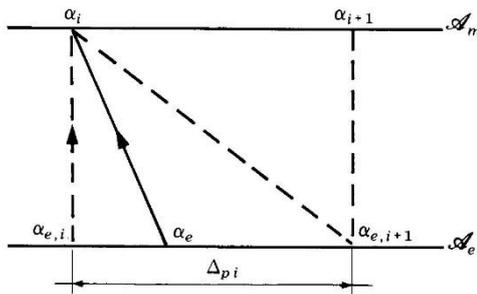


Fig. 3 Incidence du caractère discret de l'ensemble \mathcal{A}_m sur l'observation

Mais d'autre part, en raison d'un défaut d'étalonnage ou d'une détérioration de l'appareil, la valeur mesurée α_i peut correspondre à un ensemble Δ_{pi} (α_i) de valeurs enregistrées qui ne coïncide pas avec le segment $[\alpha_i, \alpha_{i+1}[$, les longueurs de ces deux segments pouvant même être différentes (fig. 4) (le décalage se réduit à une simple translation de \mathcal{A}_m par rapport à \mathcal{A}_e en cas d'égalité). En particulier, les erreurs de parallaxe entrent dans cette catégorie. On a donc :

$$\Delta_{pi} = [\alpha_{e,i}, \alpha_{e,i+1}[\neq [\alpha_i, \alpha_{i+1}[= \Delta_i$$

En dehors des expérimentalistes, qui connaissent bien ces problèmes, on aurait assez souvent tendance à identifier la sensibilité, la fidélité et la précision proprement dite sous le terme vague et global de précision, la confusion étant totale au niveau du langage courant. Les trois notions précédentes sont pourtant indispensables à une analyse en profondeur de la relation observation-information, parce qu'elles sont de natures bien différentes,

et parce que les paramètres Δ_s , Δ_f et Δ_{pi} ne peuvent généralement pas être considérés comme négligeables.

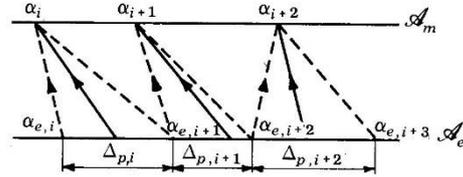


Fig. 4 Aspect général de la précision d'une observation.

Pour désigner de façon commode un procédé d'observation \mathcal{P} de caractéristiques Δ_s , Δ_f , Δ_{pi} , Δ_i , nous le noterons dorénavant $\mathcal{P}\{\Delta_s, \Delta_f, \Delta_{pi}, \Delta_i\}$.

4.2 Structure du processus de mesure

Poursuivons à présent notre analyse en considérant, non plus le détail du mécanisme d'observation, mais l'aspect global de l'acte de mesure. Il apparaît alors que l'observation n'est que le premier stade, pour ainsi dire passif, de la mesure car au-delà de l'observation, il y a l'interprétation, qui est active et conduit à l'expression ultime de l'information recueillie. Le processus de mesure est celui qui réalise la synthèse de ces deux démarches.

4.2.1 Processus d'observation

Considérons donc un procédé $\mathcal{P}\{\Delta_s, \Delta_f, \Delta_{pi}, \Delta_i\}$, et supposons d'abord que l'on observe un événement isolé α_o . Compte tenu de la limite de fidélité, l'événement enregistré correspondant α_e est l'un quelconque des événements appartenant au segment $\Delta_f(\alpha_o)$ (cf. paragraphe 4.1.2). Cette valeur α_e appartient à l'un des intervalles de précision Δ_{pk} , et c'est la valeur correspondante α_k de \mathcal{A}_m qui sera affichée avec une probabilité $p(\alpha_k)$ (fig. 5).

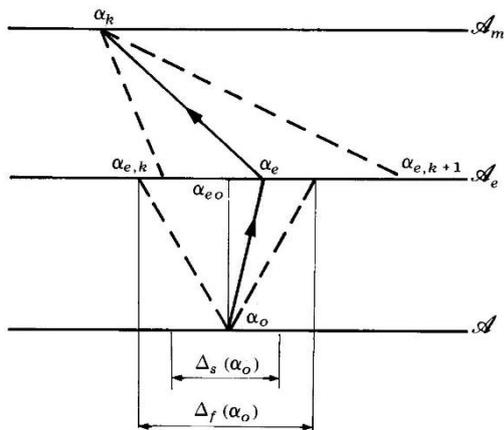


Fig. 5 Processus d'observation.

Si maintenant nous avons affaire à un ensemble $\{\alpha\}$ d'événements appartenant au segment $\Delta_s(\alpha_o)$ (limite de résolution de l'appareil en α_o), la probabilité d'affichage $p(\alpha_k)$ de la valeur α_k reste inchangée.

Le processus $(\alpha \in \Delta_s(\alpha_o), \alpha_k)$ qui conduit d'un événement α à la valeur mesurée α_k constitue l'observation proprement dite, et nous l'appellerons « processus d'observation ».

4.2.2 Processus interprétatif

La mesure serait cependant incomplète (et l'information obtenue très partielle) si on la limitait au processus d'observation, car il faut maintenant « interpréter » le résultat obtenu. Or, que nous apprend l'affichage de la valeur α_k ? d'abord que l'événement enregistré α_e appartenait au segment $\Delta_{p,k}$ associé, sans que l'on puisse préciser davantage. En particulier, α_e pouvait être la borne inférieure $\alpha_{e,k}$ (fig. 6). Dans ce cas extrême, connaissant la limite de fidélité Δ_f , nous en déduisons que α pouvait appartenir au segment $[\alpha_{e,k} - \Delta_f/2, \alpha_{e,k} + \Delta_f/2[$. En tenant le même raisonnement au voisinage de la borne supérieure $\alpha_{e,k+1}$ de $\Delta_{p,k}$, on constate que α était obligatoirement un événement de l'intervalle $[\alpha_{e,k} - \Delta_f/2, \alpha_{e,k+1} + \Delta_f/2[$.⁽¹⁾

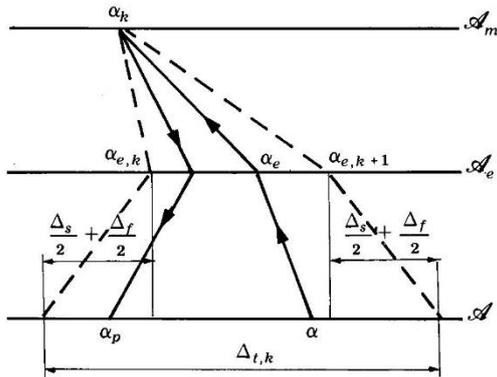


Fig. 6 Processus de mesure avec un procédé $\mathcal{P}(\Delta_s, \Delta_f, \Delta_{p,i}, \Delta_j)$.

Si l'on était certain qu'un événement isolé a été observé lors de la mesure, l'interprétation serait achevée. Mais comme cela n'est généralement pas le cas, il reste encore à tenir compte de la sensibilité. Pour ce faire, reprenons un raisonnement analogue au précédent : considérant le cas limite où α aurait eu la valeur $\alpha_{e,k} - \Delta_f/2$, et connaissant la limite de résolution $\Delta_s(\alpha)$, nous voyons que l'observation d'un ensemble d'événements situés à l'intérieur du segment $\Delta_s(\alpha_{e,k} - \Delta_f/2)$ aurait pu conduire à la même valeur mesurée α_k . Comme on peut dire la même chose au point $\alpha_{e,k+1} + \Delta_f/2$, nous concluons que la lecture de la valeur α_k implique l'appartenance du ou des événements α observés au segment :

$$\Delta_{t,k} = \left[\alpha_{e,k} - \frac{\Delta_f}{2} - \frac{\Delta_s}{2}, \alpha_{e,k+1} + \frac{\Delta_f}{2} + \frac{\Delta_s}{2} \right]$$

(1) Ceci n'est exact que si Δ_f est indépendant de α_e .

Toutes les valeurs situées à l'intérieur de cet intervalle sont des valeurs possibles pour α (fig. 6).

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

$$\alpha \in \Delta_{t,k} \quad (11)$$

Le processus par lequel, de la valeur mesurée α_k , on remonte à une valeur possible α_p de α sera appelé « processus interprétatif », et noté $(\alpha_k, \alpha_p \in \Delta_{t,k})$. Quant au segment $\Delta_{t,k}$, il représente la « marge totale d'incertitude du procédé \mathcal{P} pour la valeur mesurée α_k ».

On doit souligner que si les processus d'observation est un phénomène physique ordinaire, le processus interprétatif est un processus « intelligent » au sens propre du mot, qui nécessite l'appréhension des paramètres $\Delta_s, \Delta_f, \Delta_{p,i}, \Delta_j$, et ne peut donc être réalisé que par un système cybernétique. Cela illustre clairement l'écart qui sépare l'objet d'observation $\{\alpha\}$, le résultat d'observation α_k et l'objet de connaissance $\Delta_{t,k}$ qui est une sorte de fenêtre découpée dans \mathcal{A} autour de $\{\alpha\}$.

4.2.3 Processus de mesure

En fait, le véritable processus de mesure n'est pas le processus d'observation, dont l'événement final α_k (valeur mesurée) représente en quelque sorte une « mesure brute ». Il est certes courant dans la pratique de s'en tenir à cette mesure brute, et nous analyserons dans une autre publication le problème d'information correspondant. Mais la mesure, pour être complète, doit être interprétée, ce qui correspond au processus $(\alpha \in \Delta_s(\alpha_o), \alpha_p \in \Delta_{t,k})$ qui est la succession des processus d'observation et d'interprétation, et que nous appellerons « processus de mesure » :

$$(\alpha \in \Delta_s(\alpha_o), \alpha_k) \oplus (\alpha_k, \alpha_p \in \Delta_{t,k}) = (\alpha \in \Delta_s(\alpha_o), \alpha_p \in \Delta_{t,k}) \quad (12)$$

Sauf exception, ce processus n'est pas identique puisqu'on a en général $\alpha \neq \alpha_p$, ce qui revient encore à dire que le processus interprétatif n'est pas l'inverse du processus d'observation [40].

4.3 Entropie des distributions continues

4.3.1 Information acquise par mesure

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.

Signalons au préalable, sur un plan qualitatif et sans entrer dans le détail, que l'information apportée par le processus de mesure est maximale s'il s'agit d'un processus identique (α_o, α_o) , et qu'elle est d'autant plus faible que le processus de mesure est plus « éloigné » de ce processus identique. [40].

Par la lecture d'une valeur α_j de l'ensemble des valeurs mesurées \mathcal{A}_m , nous apprenons donc que $\alpha \in \Delta_{t,i}$. La

probabilité pour que α appartienne à cet intervalle Δ_{ti} est :

$$p(\alpha \in \Delta_{ti}) = \int_{\Delta_{ti}} \rho(\alpha) d\alpha = \overline{\rho_{ti}} \cdot \Delta_{ti} \quad (13)$$

où $\overline{\rho_{ti}}$ est la valeur moyenne de la densité $\rho(\alpha)$ sur le segment Δ_{ti} . Il semble donc logique d'appeler « information utile relative à l'événement α_i » la grandeur :

$$I_u(\alpha_i) = -\log_2 \left\{ p(\alpha \in \Delta_{ti}) \right\} = -\log_2 \left\{ \overline{\rho_{ti}} \cdot \Delta_{ti} \right\} \quad (14)$$

qui est une fonction de α_i .

4.3.2 Entropie : expression générale

La probabilité précédente $p(\alpha \in \Delta_{ti})$ ne doit pas être confondue avec la probabilité $p(\alpha_i)$ de lire la valeur α_i lors d'une mesure. Cette dernière est aussi la probabilité d'obtenir un événement enregistré α_e qui appartienne au segment Δ_{pi} associé à α_i :

$$p(\alpha_i) = p(\alpha_e \in \Delta_{pi}) \quad (15)$$

L'information moyenne que l'on peut recueillir en utilisant le procédé \mathcal{P} est évidemment la moyenne de la fonction $I_u(\alpha_i)$, qui a pour valeur : $\sum_i p(\alpha_i) I_u(\alpha_i)$; nous appellerons cette expression « entropie utile de la variable aléatoire A relativement à un procédé d'observation \mathcal{P} $\{\Delta_s, \Delta_f, \Delta_{pi}, \Delta_i\}$ » et nous la noterons $S(A, \Delta_s, \Delta_f, \Delta_{pi}, \Delta_i)$ ou plus brièvement $S(A, \mathcal{P})$:

$$S(A, \mathcal{P}) = \sum_i p(\alpha_i) I_u(\alpha_i) \quad (16)$$

Il reste encore à calculer $p(\alpha_i)$. Ayant déjà appelé $f(\alpha_e)$ la densité de probabilité de α_e , nous pouvons remplacer (15) par :

$$p(\alpha_i) = \int_{\Delta_{pi}} f(\alpha_e) d\alpha_e \quad (17)$$

Pour exprimer $f(\alpha_e)$, on aura recours à un calcul classique de probabilités. Soient $f(\alpha_e, \alpha)$ la densité de probabilité jointe de α_e et α , et $f(\alpha_e/\alpha_o)$ la densité de α_e sachant que $\alpha = \alpha_o$. La définition que nous avons donnée de la sensibilité nous permet de poser :

$$f(\alpha_e/\alpha_o) \equiv f(\alpha_e/\alpha \in \Delta_s(\alpha_o))$$

Il apparaît donc que la limite de résolution Δ_s ne figurera pas dans le calcul au niveau du processus d'observation. Pour une valeur quelconque α de α_o , nous pouvons alors écrire la relation classique :

$$f(\alpha_e, \alpha) = f(\alpha_e/\alpha) \rho(\alpha)$$

En utilisant la relation (10), il vient :

$$\begin{cases} f(\alpha_e, \alpha) = \frac{\rho(\alpha)}{\Delta_f} & \forall \alpha_e \in \Delta_f(\alpha_{eo} \equiv \alpha) \\ f(\alpha_e, \alpha) = 0 & \forall \alpha_e \notin \Delta_f(\alpha_{eo} \equiv \alpha) \end{cases}$$

Pour faire apparaître explicitement la variable α_e dans les relations précédentes, il est commode d'introduire la « fonction saut » de Heaviside $u(x)$, définie par :

$$\begin{cases} x \leq 0 \Rightarrow u(x) = 0 \\ x > 0 \Rightarrow u(x) = 1 \end{cases}$$

On vérifiera aisément que $f(\alpha_e, \alpha)$ s'écrit alors :

$$f(\alpha_e, \alpha) = \frac{\rho(\alpha)}{\Delta_f} \left\{ u(\alpha_e - \alpha + \frac{\Delta_f}{2}) - u(\alpha_e + \frac{\Delta_f}{2} - \alpha) \right\} \quad (18)$$

le crochet étant égal à 1 si $\alpha \in \Delta_f(\alpha_e)$ (1) et nul en dehors. En fixant α_e et en intégrant sur α , on obtient :

$$f(\alpha_e) = \int_{-\infty}^{+\infty} f(\alpha_e, \alpha) d\alpha = \int_{\Delta_f(\alpha_e)} \frac{\rho(\alpha)}{\Delta_f} d\alpha \quad (19)$$

D'après la formule (17), la probabilité $p(\alpha_i)$ s'écrit alors :

$$p(\alpha_i) = \int_{\Delta_{pi}} d\alpha_e \int_{\Delta_f(\alpha_e)} \frac{\rho(\alpha)}{\Delta_f} d\alpha \quad (20)$$

En regardant la relation (20), on notera une nouvelle fois le rôle tout particulier joué par la limite de résolution, puisque celle-ci intervient uniquement dans le cadre du processus interprétatif (formule (14)). En effet, Δ_s est inclus dans l'évaluation de Δ_{ti} , donc dans l'information utile $I(\alpha_i)$, alors qu'elle ne figure pas dans la probabilité $p(\alpha_i)$.

On obtient enfin l'expression générale de l'entropie $S(A, \mathcal{P})$ en remplaçant $p(\alpha_i)$ et $I_u(\alpha_i)$ dans l'expression (16) :

$$S(A, \mathcal{P}) = - \sum_i \int_{\Delta_{pi}} d\alpha_e \int_{\Delta_f(\alpha_e)} \frac{\rho(\alpha)}{\Delta_f} d\alpha \log_2 \int_{\Delta_{ti}} \rho(\alpha) d\alpha \quad (21)$$

La limite de fidélité Δ_f ayant été supposée indépendante de α , on peut encore sortir le terme Δ_f de l'intégrale dans l'expression (19) :

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

$f(\alpha_e)$ est donc la valeur moyenne de $\rho(\alpha)$ sur le segment Δ_f centré sur α_e , que nous avons notée $\overline{\rho_f(\alpha_e)}$. La probabilité $p(\alpha_i)$ s'écrit alors :

$$p(\alpha_i) = \int_{\Delta_{pi}} \overline{\rho_f(\alpha_e)} d\alpha_e$$

(1) Rappelons que $\Delta_f(\alpha_e) = [\alpha_e - \Delta_f/2, \alpha_e + \Delta_f/2] = \Delta_f(\alpha \equiv \alpha_e)$.

et l'expression générale (21) de l'entropie prend la forme équivalente :

$$S(A, \mathcal{P}) = - \sum_i \int_{\Delta_{pi}} \overline{\rho_f(\alpha_e)} d\alpha_e \log_2 \int_{\Delta_{ti}} \rho(\alpha) d\alpha \quad (22)$$

4.3.3 Propriétés essentielles de la fonction $S(A, \mathcal{P})$

En regroupant les expressions (14) et (16), on voit que l'entropie utile peut être mise sous la forme :

$$S(A, \mathcal{P}) = - \sum_i p(\alpha_i) \log_2 \left\{ p(\alpha \in \Delta_{ti}) \right\}$$

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 de même pour $S(A, \mathcal{P})$, qui possède en particulier l'invariance relativiste,
- le logarithme d'une probabilité étant ≤ 0 , on a : $S(A, \mathcal{P}) \geq 0$.

Les critiques adressées aux entropies de Boltzmann, ou Shannon ne s'appliquent donc plus à l'entropie utile, celle-ci possédant les mêmes caractères que lorsque la variable aléatoire A est discrète.

La formule (21) possède un caractère de grande généralité, mais dans de nombreuses situations usuelles, elle peut être notablement simplifiée, comme nous allons le voir maintenant en examinant quelques cas limites.

4.3.4 Incidence de la précision seule [41].

Considérons le cas d'un procédé $\mathcal{P}\{0, 0, \Delta_{pi}, \Delta_i\}$ où n'interviendrait que la précision (fig. 7). Alors il n'y a plus lieu de distinguer les événements observés α et enregistrés α_e , et le processus de mesure (12) s'écrit : $(\alpha, \alpha_p \in \Delta_{pk})$. On voit immédiatement à partir de la relation (22) que l'entropie se réduit à la forme simple :

$$S(A, \mathcal{P}) = - \sum_i \int_{\Delta_{pi}} \rho(\alpha) d\alpha \log_2 \int_{\Delta_{pi}} \rho(\alpha) d\alpha$$

soit encore, en reprenant la définition (15) :

$$S(A, \mathcal{P}) = - \sum_i p(\alpha_i) \log_2 p(\alpha_i) \quad (23)$$

En se reportant au paragraphe 3.2, on constate que cette grandeur généralise en la précisant l' ϵ - entropie de Kolmogorov-Rényi. Elle la généralise car nous n'avons pas postulé le caractère totalement borné de l'ensemble \mathcal{A} , la seule condition étant évidemment que la somme (23) existe (tout ou plus y a-t-il lieu de convenir que le segment Δ_{p1} associé à la plus petite valeur mesurée α_1 est la demi-droite $]-\infty, \alpha_{e,2}[$). Elle la précise enfin, car le critère de construction de l' ϵ - partition se trouve maintenant clairement défini par la donnée de l'ensemble \mathcal{A}_m des valeurs mesurées et par les caractéristiques de l'étalement (segments Δ_{pi}).

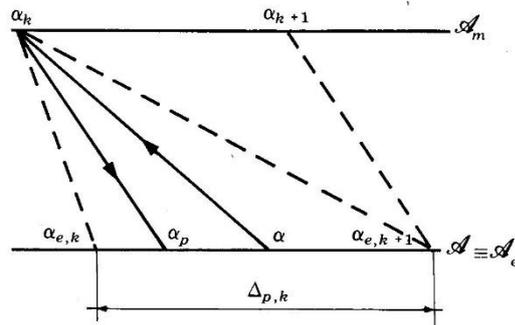


Fig. 7 Mesure avec un procédé $\mathcal{P}\{0, 0, \Delta_{pi}, \Delta_i\}$.

4.3.5 Incidence de la fidélité seule

S'agissant maintenant d'un procédé $\mathcal{P}\{0, \Delta_f, 0, 0\}$, la valeur mesurée α_m devient identique à la valeur enregistrée α_e , et le processus de mesure est défini par $(\alpha, \alpha_p \in \Delta_f(\alpha_e))$ (fig. 8). Il est alors aisé de montrer par un passage à la limite dans l'expression (22) que l'entropie devient :

$$S(A, \mathcal{P}) = - \int_{-\infty}^{+\infty} \overline{\rho_f(\alpha)} \log_2 \left\{ \overline{\rho_f(\alpha)} \cdot \Delta_f \right\} d\alpha \quad (24)$$

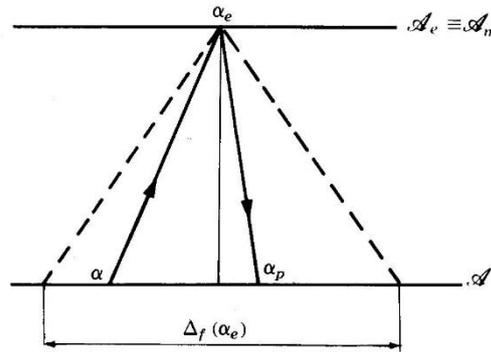


Fig. 8 Mesure avec un procédé $\mathcal{P}\{0, \Delta_f, 0, 0\}$.

On notera, en comparant avec la fonction $H(A)$ de Shannon, le remplacement de $\rho(\alpha)$ par la valeur moyenne $\overline{\rho_f(\alpha)}$ prise sur la limite de fidélité $\Delta_f(\alpha)$, et la présence de la mesure Δ_f dans le logarithme, dont l'argument est à présent une probabilité.

4.3.6 Incidence de la sensibilité seule

Avec un procédé d'observation $\mathcal{P}\{\Delta_s, 0, 0, 0\}$, il y a encore identité entre les valeurs enregistrées et les valeurs mesurées (fig. 9), le processus de mesure étant de la forme : $(\alpha \in \Delta_s(\alpha_o), \alpha_p \in \Delta_s(\alpha_o))$. Par un passage à la limite analogue au précédent, on obtient pour l'entropie :

$$S(A, \mathcal{P}) = - \int_{-\infty}^{+\infty} \rho(\alpha) \log_2 \left\{ \overline{\rho_s(\alpha)} \Delta_s(\alpha) \right\} d\alpha \quad (25)$$

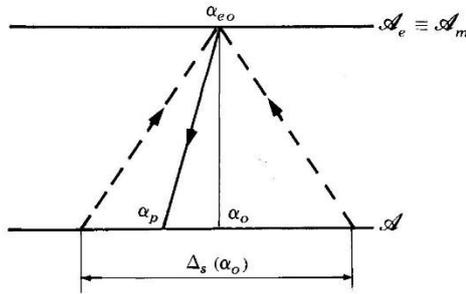


Fig. 9 Mesure avec un procédé $\mathcal{P}\{\Delta_s, 0, 0\}$.

Il est intéressant de remarquer, en comparant cette formule à la précédente, qu'il n'apparaît ici qu'une seule valeur moyenne (dans le logarithme), calculée cette fois sur la limite de résolution $\Delta_s(\alpha)$.

Nous avons déjà publié une démonstration directe pour ce cas particulier [42], et Sempi [32] [46] en a proposé presque simultanément une formulation équivalente, qui fait intervenir la fonction de répartition et la mesure de Lebesgue, Stieltjes (1). On relève néanmoins que dans ce dernier texte, Sempi considère « l'erreur de mesure » (jouant le rôle de notre terme Δ_s) comme une variable aléatoire, sans distinguer nettement les rôles respectifs du pouvoir de résolution et des fluctuations tels que nous les avons analysés. Notons cependant que si, en physique expérimentale, il existe généralement une dépendance fonctionnelle entre $\Delta_s(\alpha)$ et α qui ne présente aucun caractère fondamentalement aléatoire, rien n'interdit a priori d'envisager le cas où la limite de résolution $\Delta_s(\alpha)$ serait une variable aléatoire.

4.3.7 Cas mixtes

On peut bien entendu envisager également les combinaisons deux à deux des paramètres qui définissent le procédé \mathcal{P} . Nous ne nous attacherons ici qu'à la plus intéressante, où interviennent la sensibilité et la fidélité, le procédé étant alors du type $\mathcal{P}\{\Delta_s, \Delta_f, 0, 0\}$, et nous continuerons à considérer $\Delta_f(\alpha)$ comme une constante.

Posons

$$\Delta_{sf}(\alpha) = \left[\alpha - \frac{\Delta_s(\alpha)}{2} - \frac{\Delta_f}{2}, \alpha + \frac{\Delta_s(\alpha)}{2} + \frac{\Delta_f}{2} \right]$$

On voit que le processus de mesure est défini cette fois par : ($\alpha \in \Delta_s(\alpha_o)$, $\alpha_p \in \Delta_{sf}(\alpha_e)$) (fig. 10), et que l'entropie (22) a pour expression :

$$S(A, \mathcal{P}) = - \int_{-\infty}^{+\infty} \overline{\rho_f(\alpha)} \log_2 \left\{ \overline{\rho_{sf}(\alpha)} \cdot \Delta_{sf}(\alpha) \right\} d\alpha \quad (26)$$

La première valeur moyenne $\overline{\rho_f(\alpha)}$ est prise sur la limite de fidélité Δ_f alors que la seconde $\overline{\rho_{sf}(\alpha)}$ est calculée sur le segment Δ_{sf} .

(1) Si $F(\alpha)$ est la fonction de répartition associée à la densité $\rho(\alpha)$, on a évidemment dans (24) et (25) :

$$\overline{\rho_f(\alpha)} \cdot \Delta_f = F(\alpha + \Delta_f/2) - F(\alpha - \Delta_f/2)$$

$$\overline{\rho_{sf}(\alpha)} \cdot \Delta_s = F(\alpha + \Delta_s(\alpha)/2) - F(\alpha - \Delta_s(\alpha)/2).$$

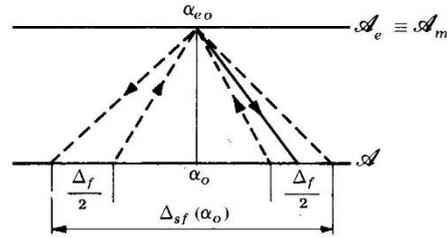


Fig. 10 Mesure avec un procédé $\mathcal{P}\{\Delta_s, \Delta_f, 0, 0\}$.

En conclusion de ce paragraphe, nous pouvons retenir que les investigations auxquelles nous nous sommes livrés sur le procédé d'observation et sur le processus de mesure, puis les conclusions que nous en avons tirées, nous ont mis en possession d'un outil - l'entropie utile - dont le domaine d'application est très vaste, et qui se révèle intéressant aussi bien sur le plan conceptuel que d'un point de vue pratique, car il intègre foncièrement les caractéristiques de l'observateur et de l'observé, tout en conduisant à des valeurs numériques parfaitement calculables. La notion d'entropie en physique s'en trouve donc sensiblement élargie.

5 LES TROIS NIVEAUX D'INFORMATION

L'analyse que nous venons de faire de la relation observation - observé nous amène maintenant à distinguer trois niveaux d'information, auxquels correspondent trois expressions différentes de l'entropie.

5.1 Niveaux d'information locale

5.1.1 Information absolue

Soit α_μ un événement élémentaire d'un ensemble \mathcal{A} discret, de probabilité $p(\alpha_\mu) = p_\mu \neq 0$.

On peut appeler « Information absolue relative à l'évènement α_μ » la grandeur :

$$I(\alpha_\mu) = - \log_2 p_\mu \quad (27)$$

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. C'est encore l'information associée à un processus identique. Evidemment, cette expression n'est mathématiquement définie que pour un ensemble discret d'événements. Si \mathcal{A} est continu, l'information absolue $I(\alpha)$ est alors infinie.

5.1.2 Information mesurable

Considérons maintenant un ensemble d'événements (continu ou discret) observé par un procédé $\mathcal{P}\{\Delta_s, \Delta_f, 0, 0\}$ de précision parfaite. Ayant obtenu pour valeur mesurée α_e , nous savons à l'issue du processus interprétatif que les valeurs possibles de α sont définies par : $\alpha_p \in \Delta_{sf}(\alpha_e)$ (paragraphe 4.3.5), leur probabilité globale étant $p(\alpha \in \Delta_{sf}(\alpha_e))$.

Par définition, nous dirons que « l'information mesurable apportée par l'enregistrement de la valeur α_e » est la grandeur :

$$I_m(\alpha_e) = -\log_2 p(\alpha \in \Delta_{sf}(\alpha_e)) \quad (28)$$

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, qui peut toujours en être dissocié et être traité de façon indépendante.

5.1.3 Information utile

Dans le cas le plus général, qui est aussi celui que l'on rencontre en pratique, tous les défauts du moyen d'observation doivent être inclus dans l'expression de l'information recueillie. Puisque le renseignement apporté par le processus interprétatif après lecture d'une valeur α_i est : $\alpha_p \in \Delta_{ti}$, nous avons appelé « information utile apportée par la valeur mesurée α_i » la grandeur :

$$I_u(\alpha_i) = -\log_2 p(\alpha \in \Delta_{ti}) \quad (29)$$

qui est la valeur efficace, utilisable, de l'information recueillie (paragraphe 4.3.1).

Il est intéressant de remarquer que si $\alpha_\mu \in \Delta_{ti}$, et puisque $\Delta_{ti} \supseteq \Delta_{sf}$, on a :

$$I_u(\alpha_i) \leq I_m(\alpha_e \equiv \alpha_i) \leq I(\alpha_\mu) \quad (30)$$

L'information absolue est bien l'information locale maximale que l'on puisse obtenir, y compris lorsque \mathcal{A} est discret.

A partir de là, trois conceptions de l'entropie peuvent être dégagées.

5.2 Entropie absolue

5.2.1 Distributions discrètes

Si l'objet d'observation est un ensemble discret d'événements α_μ , nous appellerons « entropie absolue $S(A)$ de la variable aléatoire A » la moyenne des informations absolues susceptibles d'être fournies par l'observation :

$$S(A) = -\sum_{\mu} p_{\mu} \log_2 p_{\mu} \quad (31)$$

Cette expression n'est autre évidemment que l'entropie habituelle de la thermodynamique et de la théorie de l'information, qui fait abstraction des propriétés du procédé d'information, ou ce qui revient au même, sous-entend un procédé idéal $\mathcal{P}\{0, 0, 0, 0\}$.

5.2.2 Distributions continues

L'extension de la définition précédente aux distributions continues conduit évidemment à une entropie absolue qui n'est pas bornée. On vérifierait d'ailleurs sans difficulté, à partir de la formule (26) par exemple, que $S(A) \rightarrow \infty$ lorsque Δ_s , Δ_f , Δ_{pi} et Δ_i tendent simultanément vers zéro. Mais il est remarquable de constater que la différence des entropies absolues de deux varia-

bles aléatoires A et A' est en général parfaitement définie.

Partons par exemple de la formule (26), dans laquelle on a déjà fait $\Delta_{pi} = \Delta_i = 0$. Ajoutons-y l'hypothèse $\Delta_s \rightarrow 0$ et $\Delta_f \rightarrow 0$: alors $\Delta_{sf} \rightarrow 0$, et les valeurs moyennes de la fonction $\rho(\alpha)$ sur Δ_f et Δ_{sf} tendent vers la valeur ponctuelle en α :

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

On a donc, en vertu du fait que $\int_{-\infty}^{+\infty} \rho(\alpha) d\alpha = 1$, et à condition que $\Delta_{sf} = \text{cte}$ (1)

$$S(A) = -\int_{-\infty}^{+\infty} \rho(\alpha) \log_2 \rho(\alpha) d\alpha = \lim_{\Delta_{sf} \rightarrow 0} \log_2 \Delta_{sf}$$

C'est à la présence du dernier terme qu'est due généralement la divergence de $S(A)$. Mais supposons qu'une autre variable aléatoire A' de densité $\rho'(\alpha)$ soit observée dans les mêmes conditions, c'est-à-dire avec le même procédé \mathcal{P} . Son entropie absolue est alors :

$$S(A') = -\int_{-\infty}^{+\infty} \rho'(\alpha) \log_2 \rho'(\alpha) d\alpha = \lim_{\Delta_{sf} \rightarrow 0} \log_2 \Delta_{sf}$$

et la différence des deux entropies absolues se révèle être égale à la différence des fonctions de Shannon de A et A' .

$$S(A) - S(A') = H(A) - H(A') \quad (32)$$

Cette différence a donc un sens, pourvu que les intégrales $H(A)$ et $H(A')$ convergent. La fonction empirique de Shannon se voit ainsi attribuer une signification physique précise.

En particulier, la fonction \mathcal{H} de Boltzmann doit être replacée dans ce contexte. Sa signification est la suivante :

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

D'autre part, puisque les fonctions $H(A)$ sont à valeurs réelles on peut définir une structure d'ordre dans l'ensemble $\{A\}$ de toutes les variables aléatoires à partir de leur entropie absolue. Ceci justifie a posteriori une pratique largement en usage en théorie de l'information, mais qui n'avait pas des bases parfaitement solides. Ainsi, prenons pour exemple quatre variables aléatoires A_1 à A_4 avec des distributions usuelles :

$$A_1 : \text{distribution de Cauchy} : \rho_1(\alpha) = \frac{a}{\pi(a^2 + \alpha^2)}$$

A_2 : distribution de Gauss :

$$\rho_2(\alpha) = \left(\frac{1}{2\pi}\right)^{1/2} \frac{1}{\sigma} \exp\left(-\frac{\alpha^2}{2\sigma^2}\right)$$

A_3 : distribution exponentielle :

$$\rho_3(\alpha) = \frac{1}{2\epsilon} \exp\left(-\frac{|\alpha|}{\epsilon}\right)$$

(1) Dans le cas présent, cette condition est très peu restrictive.

A_4 : distribution uniforme :

$$\rho_4(\alpha) = \begin{cases} \frac{1}{2b} & \text{pour } |\alpha| \leq b \\ 0 & \text{pour } |\alpha| > b \end{cases}$$

Les paramètres a, σ, ϵ, b ont même dimension que α , et un calcul simple montre que les écart-types des trois dernières distributions sont respectivement : $\sigma_2 = \sigma$, $\sigma_3 = \epsilon \sqrt{2}$, $\sigma_4 = b/\sqrt{3}$, la distribution de Cauchy étant sans écart-type.

En prenant pour référence la distribution de Cauchy et en attribuant aux trois autres distributions un même écart-type σ , on obtient :

$$\begin{aligned} S(A_1) - S(A_2) &= \log_2(12,87 a) - \log_2(\sigma \sqrt{2\pi e}) \\ S(A_1) - S(A_3) &= \log_2(12,87 a) - \log_2(\sigma e \sqrt{2}) \\ S(A_1) - S(A_4) &= \log_2(12,87 a) - \log_2(2 \sigma \sqrt{3}) \end{aligned}$$

De ces trois relations, on déduit les inégalités suivantes :

$$a > \frac{\sigma}{4,13} \Rightarrow S(A_1) > S(A_2) > S(A_3) > S(A_4)$$

qui traduisent un caractère statistique de moins en moins dispersif (1) lorsqu'on va de A_1 à A_4 .

5.3 Entropie mesurable

5.3.1 Définition

Corrélativement à la notion d'information mesurable, nous définirons une « entropie mesurable $S_m(A, \mathcal{P})$ de la variable aléatoire A relativement à un procédé d'observation \mathcal{P} » qui sera la moyenne de la fonction information mesurable $I_m(\alpha_e)$. L'entropie mesurable est donc la valeur que prend l'entropie utile $S(A, \mathcal{P})$ lorsque $\mathcal{P} = \mathcal{P}(\Delta_s, \Delta_f, 0, 0)$, dont nous avons déjà établi l'expression en (4.3.7).

$$\begin{aligned} S(A, \Delta_s, \Delta_f, 0, 0) &= S_m(A, \mathcal{P}) = \\ &= \int_{-\infty}^{+\infty} \overline{\rho_f(\alpha)} \log_2 \{ \overline{\rho_{sf}(\alpha)} \cdot \Delta_{sf}(\alpha) \} d\alpha \end{aligned} \quad (33)$$

Quant aux formules (24) et (25), elles définissent l'entropie mesurable pour deux procédés particuliers $\mathcal{P}(0, \Delta_f, 0, 0)$ et $\mathcal{P}(\Delta_s, 0, 0, 0)$.

5.3.2 Application à quelques distributions usuelles

Il serait intéressant d'étudier les variations de $S_m(A, \mathcal{P})$ en fonction de Δ_s et Δ_f pour des distributions variées, afin de pouvoir effectuer des comparaisons entre elles. Ce travail est en cours, et nous donnons seulement ici à titre d'exemples les entropies mesurables des variables aléatoires A_1, A_2, A_3, A_4 définies au paragraphe précédent, pour un procédé $\mathcal{P}(\Delta_s, 0, 0, 0)$ où n'intervient que le pouvoir de résolution. On a porté en

(1) On dit en mécanique statistique « de plus en plus ordonné ».

abscisses le paramètre sans dimension σ/Δ_s (σ étant l'écart-type de la distribution), sauf pour la distribution de Cauchy, qui n'a pas d'écart-type, et pour laquelle nous avons pris le paramètre a/Δ_s , puisque a est homogène à α (fig. 11 et 12) et caractérise l'étalement de A_1 .

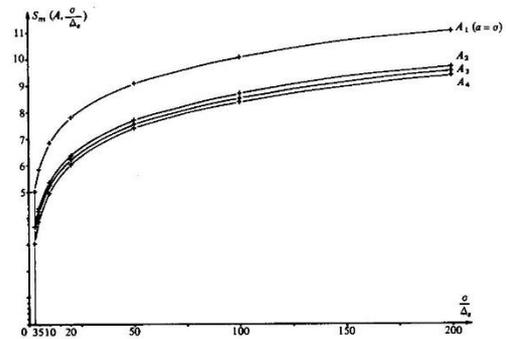


Fig. 11 Entropie mesurable $S_m(A, \sigma/\Delta_s)$ pour un procédé $\mathcal{P}(\Delta_s, 0, 0, 0)$. A_1 : distribution de Cauchy ; A_2 : distribution de Gauss ; A_3 : distribution exponentielle ; A_4 : distribution uniforme. Domaine $\sigma/\Delta_s \geq 3$. Les croix indiquent les valeurs calculées numériquement.

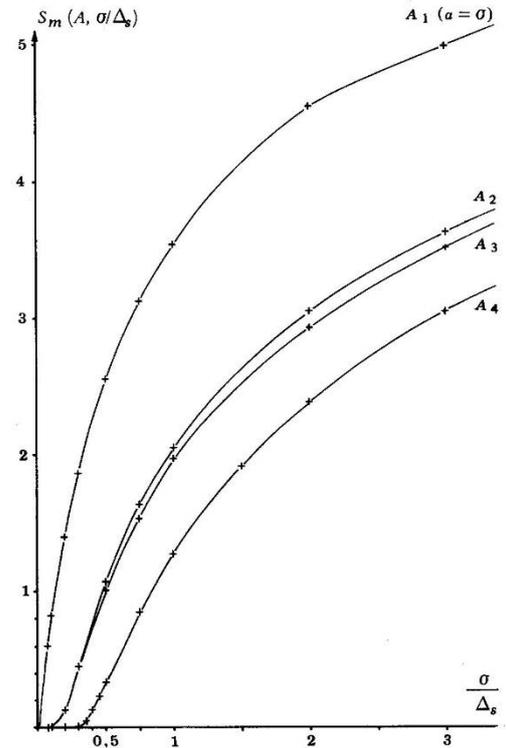


Fig. 12 Entropie mesurable $S_m(A, \sigma/\Delta_s)$. Domaine $0 \leq \sigma/\Delta_s < 3$.

Nous avons dû avoir recours au calcul sur ordinateur pour tracer les courbes $S_m(\sigma/\Delta_s)$ et $S_m(a/\Delta_s)$ car l'entropie $S(A, \Delta_s, 0, 0, 0)$ des distributions de Cauchy, de Gauss et exponentielle n'est pas calculable par des méthodes analytiques. A l'examen de ces courbes, on constate que la hiérarchie apparue entre les quatre variables aléatoires par référence à l'entropie absolue s'étend à l'entropie mesurable ; en effet, $\forall \sigma/\Delta_s$ donné, on a toujours : $S_m(A_2) > S_m(A_3) > S_m(A_4)$. La comparaison soulève quelques difficultés pour la distribution de Cauchy en raison de l'absence d'écart-type, mais il apparaît que pour $a = \sigma$ (courbe 1) on a également $S_m(A_1) > S_m(A_2)$.

Par ailleurs, on voit que pour les quatre distributions, $S_m \rightarrow 0$ quand $\sigma/\Delta_s \rightarrow 0$, ce qui correspond au cas de certitude presque totale, l'essentiel de la distribution se trouvant soit à l'intérieur, soit à l'extérieur de Δ_s [8]. A la limite, pour une distribution de Dirac (événement unique de probabilité 1) : $S_m = 0$.

5.3.3 Mise au point sur l'entropie de Gibbs

Revenons brièvement à l'entropie de Gibbs rappelée dans la formule (7). Nous voyons qu'elle fait implicitement référence à l'entropie mesurable $S(A, \Delta_s, 0, 0, 0)$ donnée par la relation (25), la cellule de volume C dans l'espace des phases jouant le rôle de la limite de résolution Δ_s . Mais elle est néanmoins incorrecte en raison de l'absence d'une densité moyenne sous le logarithme, et elle devrait s'écrire : $S = -k \int f(p, q) \log\{C \cdot \overline{f_c(p, q)}\} dp dq$, avec un terme $\overline{f_c(p, q)}$ qui serait la moyenne de la fonction de distribution f sur la cellule C . Sous sa forme (7), l'entropie de Gibbs n'a de sens que lorsque $C \rightarrow 0$; elle s'identifie alors à l'entropie absolue.

5.4 Entropie utile

L'utilisation d'un procédé \mathcal{P} tout à fait quelconque nous impose enfin de définir une troisième entropie : « l'entropie utile », qui est la moyenne des informations locales utiles $I_u(\alpha_i)$; c'est à elle que nous avons consacré une étude générale au paragraphe 4.3, elle est calculable à partir de la formule (21).

Une autre expression de l'entropie issue des travaux de Gibbs entre dans ce cadre. Il s'agit de l'entropie « grossière » ou « granulaire » (coarse-grained entropy), étudiée dans quelques publications très récentes [30] [31] [44]. L'espace des phases Γ étant divisé en cellules C_j mesurables (au sens physique) de manière analogue à une ϵ -partition, on considère que sur chaque cellule C_j , la mesure fournit la densité moyenne $\overline{f_j(p, q)} = f_j$ de $f(p, q)$. Si p_j est la probabilité qu'un point (p, q) de Γ appartienne à C_j , l'entropie grossière est définie comme la moyenne des $\log f_j$, soit :

$$S_g = -k \int_{\Gamma} \overline{f_j(p, q)} \log \overline{f_j(p, q)} dp dq = -k \sum_j p_j \log f_j$$

Cette entropie, pour intéressante qu'elle soit, prête à la même critique que la précédente, puisque l'argument

du logarithme est une densité et non une probabilité ; la situation n'est pas modifiée du fait qu'il s'agisse d'une densité moyenne au lieu d'une densité ponctuelle.

Des deux nouveaux concepts d'entropie que nous avons introduits, celui d'entropie mesurable est certainement le plus commode et le plus intéressant à utiliser, car il est en général relativement facile d'améliorer un appareil en rendant les paramètres de précision Δ_{p_i} et Δ_i petits devant Δ_s et Δ_f .

6 DISTRIBUTIONS CATASTROPHIQUES

Nous ne voudrions pas conclure cette étude sans évoquer au moins brièvement le problème de la frontière entre le cas discret et le cas continu. Nous aurons recours pour le traiter à une catégorie particulière de variables aléatoires continues, définies seulement sur des domaines disjoints, et dont les variables aléatoires discrètes apparaissent comme des cas limites. La notion de discernabilité de ces domaines va nous permettre de réinterpréter l'entropie de Shannon pour les variables aléatoires discrètes et d'élargir sa signification, tout en développant quelques propriétés nouvelles.

6.1 Position du problème

Nous appellerons « événements catastrophiques » ou « catastrophes », au sens de Thom [43] des événements caractérisés par un important gradient de densité $d\rho(\alpha)/d\alpha$, et nous nommerons « distributions catastrophiques » des distributions formant une collection dénombrable de domaines catastrophique C_μ (correspondant à des pics de la fonction $\rho(\alpha)$) séparés par des zones de densité nulle (fig. 13). A chacun de ces domaines C_μ , on peut associer une moyenne α_μ . Soit p_μ la probabilité de présence d'un événement α dans l'ensemble C_μ . On a évidemment : (1)

$$\int_{-\infty}^{+\infty} \rho(\alpha) d\alpha = \sum_{\mu} p_{\mu} = 1 \quad (34)$$

Si l'on suppose que $\rho(\alpha) \rightarrow 0$ presque partout (sauf aux points $\alpha_\mu, \alpha_{\nu}, \dots$) en conservant la propriété (34), on obtient à la limite une distribution discrète, que l'on peut exprimer à l'aide de la fonction de Dirac $\delta(\alpha)$:

$$\rho(\alpha) = p(\alpha) \delta(\alpha - \alpha_\mu) \quad (\mu = 1, 2, \dots) \quad (35)$$

(on aura $\rho(\alpha) = 0$ si $\alpha \neq \alpha_\mu$ et on posera $p(\alpha_\mu) = p_\mu$).

L'expression générale (21) de l'entropie aussi bien que les formules partielles qui suivent restent donc applicables aux distributions catastrophiques. Certaines particularités intéressantes apparaissent toutefois, que nous voudrions mettre en évidence.

(1) Dans la pratique, on rencontre des distributions telles que $\rho(\alpha) = \epsilon \neq 0$ entre les domaines C_μ . Nous limitons notre définition au cas où ϵ est assez petit pour être négligeable.

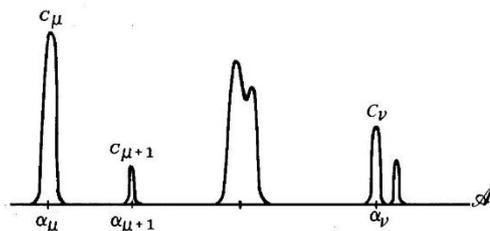


Fig. 13 Exemple de distribution catastrophique.

6.2 Evènements ou ensembles d'évènements discernables par un procédé d'observation

Soit E_μ l'ensemble d'indices $\{\dots, \mu, \dots, \nu, \dots\}$ caractérisant les moyennes $\alpha_\mu, \dots, \alpha_\nu, \dots$ des domaines catastrophiques C_μ, C_ν, \dots dans une distribution catastrophique, et E_i l'ensemble d'indices $\{\dots, i, \dots, j, \dots\}$ qui caractérise les événements mesurés $\alpha_i, \alpha_j, \dots$ dans l'ensemble \mathcal{A}_m . En particulier, pour une distribution discrète, $\alpha_\mu, \alpha_\nu, \dots$ sont les éléments de l'ensemble \mathcal{A} . Ces notations vont nous aider à introduire la notion de discernabilité des ensembles C_μ , par la définition suivante :

Définition :

Les domaines catastrophiques C_μ d'une distribution catastrophique sont discernables par un procédé d'observation $\mathcal{P} \{\Delta_s, \Delta_f, \Delta_{p_i}, \Delta_i\}$ s'il existe une application injective M de E_μ dans E_i : $E_\mu \rightarrow E_i$.

Cela revient à dire que :

- 1) chaque domaine catastrophique C_μ est mesuré par une valeur α_i et une seule : $M(\mu) = i$.
- 2) A chaque élément de l'ensemble des valeurs mesurées \mathcal{A}_m on peut attribuer au plus un domaine catastrophique C_μ : $M^{-1}(i) = \mu$. (1)

En termes de processus, cette définition signifie que pour chaque événement (dans le cas discret) ou chaque domaine catastrophique (dans le cas général), le processus interprétatif est l'inverse du processus d'observation. Le processus de mesure est donc un processus identique.

Nous allons maintenant établir quelques propriétés des distributions qui répondent à la définition précédente.

Proposition 1 :

Une condition nécessaire pour que les éléments d'une distribution catastrophique soient discernables par un procédé $\mathcal{P} \{\Delta_s, \Delta_f, \Delta_{p_i}, \Delta_i\}$ est que, pour tout couple μ, i d'indices tel que $M(\mu) = i$, on ait : $\Delta_{p_i} \supset \Delta_f(\alpha) \forall \alpha \in C_\mu$.

Démonstration :

Si la proposition n'est pas vérifiée, et compte tenu d'une limite de fidélité Δ_f non nulle, les événements α d'un domaine catastrophique C_μ peuvent conduire à des événements enregistrés α_e extérieurs à Δ_{p_i} , et qui sont donc susceptibles d'être mesurés par des valeurs autres

(1) L'application inverse M^{-1} est définie si $i \in M(E_\mu)$, image de E_μ par M . Si $i \notin M(E_\mu)$, alors $M^{-1}(i) = \emptyset$ (ensemble vide).

que α_i . Alors, l'ensemble C_μ n'est plus discernable des ensembles voisins puisque la propriété 1) n'est plus satisfaite.

Proposition 2 :

L'entropie $S(A, \mathcal{P})$ d'une distribution catastrophique dont les éléments C_μ sont discernables par un procédé d'observation $\mathcal{P} \{\Delta_s, \Delta_f, \Delta_{p_i}, \Delta_i\}$ est donnée par la formule :

$$S(A, \mathcal{P}) = - \sum_{\mu} p_{\mu} \log_2 p_{\mu} \quad (36)$$

où p_{μ} est la probabilité d'avoir α dans le domaine C_{μ} . Elle est invariante par toute transformation de \mathcal{P} qui conserve cette discernabilité.

Démonstration :

- 1) D'après la définition que nous avons posée, il est possible que certains éléments $\alpha_q, \dots, \alpha_r, \dots$ de l'ensemble des valeurs mesurées \mathcal{A}_m ne correspondent à aucun domaine catastrophique. Ils ne seront donc jamais obtenus au cours de l'observation et leurs probabilités $p(\alpha_q), \dots, p(\alpha_r), \dots$ sont nulles. Pour les autres éléments $\alpha_i, \dots, \alpha_k, \dots$, l'hypothèse de discernabilité implique :

$$i = M(\mu) \Rightarrow \begin{cases} p(\alpha_i) = p_{\mu} \text{ (d'après 1)} \\ p(\alpha \in \Delta_{p_i}) = p_{\mu} \text{ (d'après 2)} \end{cases} \quad (37)$$

L'information utile relative à α_i est donc $I_u(\alpha_i) = -\log_2 p_{\mu}$, et l'on obtient en revenant à la définition (16) de l'entropie :

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

On pourrait d'ailleurs retrouver ce résultat à partir de l'expression générale (21).

- 2) En ce qui concerne la seconde partie de la proposition, considérons un autre procédé \mathcal{P}' qui conserve le caractère discernable des domaines catastrophiques. L'ensemble \mathcal{A}_m des événements α_i pourra être remplacé par un autre ensemble \mathcal{A}'_m d'événements $\alpha_{i'}$, $\dots, \alpha_{j'}$, \dots , caractérisé par un nouvel ensemble $E_{i'}$ d'indices $\{\dots, i', \dots, j', \dots\}$, et il existe une application injective M' de E'_{μ} dans $E_{i'}$, telle que :

$$i' = M'(\mu) \Rightarrow \begin{cases} p(\alpha_{i'}) = p_{\mu} \\ p(\alpha \in \Delta_{p_{i'}}) = p_{\mu} \end{cases} \quad (38)$$

Les relations (37) et (38) sont indépendantes des paramètres qui caractérisent \mathcal{P} tant que la propriété de discernabilité reste vérifiée. On voit en les rapprochant que l'entropie est donc conservée :

$$S(A, \mathcal{P}) = - \sum_i p(\alpha_{i'}) I_u(\alpha_{i'}) = - \sum_{\mu} p_{\mu} \log_2 p_{\mu}$$

Cas particulier :

S'il s'agit d'une distribution discrète, alors p_{μ} est la probabilité de l'événement α_{μ} , et l'expression (36) s'identifie à l'entropie absolue définie au paragraphe 5.2. On peut alors énoncer la proposition 2 sous la forme suivante :

Corollaire :

L'entropie absolue $S(A)$ d'une distribution discrète est l'entropie utile $S(A, \mathcal{P})$ de cette distribution lorsque les événements α_μ sont discernables par le procédé d'observation \mathcal{P} .

En particulier, si \mathcal{A} est réduit à un seul événement dont la probabilité est égale à 1, (événement certain), la mesure ne peut apporter aucune information et l'on a (cf. paragraphe 5.3.2) :

$$S(A, \mathcal{P}) = S(A) = 0$$

Proposition 3 :

L'entropie absolue $S(A)$ d'une variable aléatoire A définie sur un champ discret de probabilités est la valeur maximale de l'entropie utile $S(A, \mathcal{P})$.

Démonstration :

En se reportant aux inégalités (30), on établit immédiatement la proposition. En effet :

$$I_u(\alpha_i) < I(\alpha_i) \quad \forall i \Rightarrow \sum_i p(\alpha_i) I_u(\alpha_i) \leq \sum_i p(\alpha_i) I(\alpha_i)$$

d'où : $S(A, \mathcal{P}) < S(A)$.

Cette proposition est conforme aux résultats généraux relatifs à l'information apportée par une partition (cf. Comyn et Losfeld [39]).

6.3 Ressemblance de deux ensembles d'évènements

Soient deux variables aléatoires A et A' associées à deux distributions catastrophiques de densités $\rho(\alpha)$ et $\rho'(\alpha)$ définies sur le même champ de probabilités, dont les domaines catastrophiques ont pour moyennes respectives $\{\dots \alpha_\mu \dots, \alpha_\nu \dots\}$ et $\{\dots \alpha_{\mu'} \dots, \alpha_{\nu'} \dots\}$. On appelle E_μ et $E_{\mu'}$ les ensembles d'indices correspondants, et l'on suppose que A et A' sont toutes deux discernables par un procédé d'observation \mathcal{P} .

Proposition 4 :

S'il existe une bijection m de E_μ dans $E_{\mu'}$, et si $\mu' = m(\mu) \Rightarrow p_\mu = p_{\mu'} \quad \forall \mu$, alors A et A' ont même entropie : $S(A, \mathcal{P}) = S(A', \mathcal{P})$.

Démonstration :

D'après la proposition 2, les entropies de A et A' sont respectivement :

$$S(A, \mathcal{P}) = - \sum_{\mu} p_{\mu} \log_2 p_{\mu}$$

$$S(A', \mathcal{P}) = - \sum_{\mu'} p_{\mu'} \log_2 p_{\mu'}$$

Si $\forall \mu, \mu'$ tel que $p_{\mu'} = p_{\mu}$, on a : $S(A, \mathcal{P}) = S(A', \mathcal{P})$

L'observation de ces deux distributions a donc le même contenu informationnel. En particulier, si $i = M(\mu) = M(\mu')$, c'est-à-dire schématiquement si \mathcal{A}' ne diffère de \mathcal{A} que par une « translation » qui laisse chaque domaine catastrophique C_μ en correspondance avec la même valeur mesurée α_i , les observations relatives à \mathcal{A} et \mathcal{A}' par \mathcal{P} sont identiques. Nous dirons alors que, pour le procédé d'observation \mathcal{P} , les ensembles \mathcal{A} et \mathcal{A}' se ressemblent. On reconnaît là une extension de la « ressemblance entre deux événements » [40].

Le résultat le plus intéressant dans ce dernier paragraphe est peut-être la signification élargie attribuée à l'entropie « classique » par le corollaire de la Proposi-

tion 2. Il apparaît en effet que, lorsque la variable aléatoire est discrète, il n'est nullement indispensable de disposer d'un procédé d'observation idéal pour recueillir l'information maximale. Cette remarque peut être étendue aux distributions catastrophiques dans la mesure où l'on ne s'intéresse que globalement à chaque domaine C_μ .

7 CONCLUSION ET PERSPECTIVES

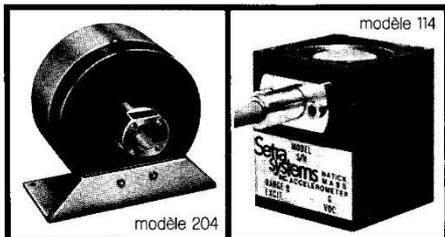
Le but que nous nous étions initialement fixé était d'aboutir à une expression pleinement satisfaisante de l'entropie pour les distributions continues rencontrées en physique. L'étude précédente montre qu'il a fallu pour l'atteindre intégrer les propriétés du procédé d'observation \mathcal{P} à la définition de l'entropie, et elle peut être étendue sans difficulté au cas des distributions discrètes, pour lesquelles toutefois les paramètres caractéristiques de \mathcal{P} n'interviennent plus explicitement dès lors que les événements sont discernables.

Les notions d'entropie utile et d'entropie mesurable que nous avons introduites intègrent comme cas limites ou particuliers les diverses évaluations de l'entropie proposées par Boltzmann, Gibbs, Shannon, Kolmogorov et Renyi. Grâce à leur domaine d'application assez large, elles devraient pouvoir être exploitées avec fruit dans plusieurs branches de la physique : thermodynamique statistique, théorie de l'information, mécanique quantique, description des états de surface, et plus généralement partout où intervient un ensemble statistique de données, pour conduire à une meilleure évaluation de l'information accessible et à une amélioration des procédés d'observation. Enfin, il serait intéressant, en se fondant sur l'axiomatique élaborée par Kampé de Fériet [33] [36], de les élargir à des ensembles de nature non statistique, dont les événements ne peuvent être répétés.

BIBLIOGRAPHIE

- [1] J.D. Fast, *Entropie*. Editions Techniques Philips (1962).
- [2] P. Chambadal, *Evolution et applications du concept d'entropie*. Dunod (1963).
- [3] C.E. Shannon, *The mathematical theory of communication I*, Bell Syst. Tech. J. 27, p 379 (1948).
- [4] C.E. Shannon, *The mathematical theory of communication II*. Bell Syst. Techn. J. 27, p 623 (1948).
- [5] N. Kolmogorov, *On the Shannon theory of information transmission in the case of continuous signals*, IRE. Trans. on Information Theory 12, p 102 (1956).
- [6] R. Giles, *Mathematical foundations of thermodynamics*. Pergamon Press (1964).
- [7] N. Boccara, *Les principes de la thermodynamique classique*. P.U.F. (1968).
- [8] E.T. Jaynes, *Information theory and statistical mechanics-I*. Phys. Rev. 106, 4, p. 620 (1957) II - Phys. Rev., 108, 2, p 171 (1957).
- [9] A. Blanc-Lapierre, *Mécanique statistique*. Masson (1967).
- [10] I. Prigogine, *Entropie et dynamique*. Entropie 57, p 5 (1974).
- [11] J.M. Souriau, *Structure des systèmes dynamiques*. Dunod (1970).
- [12] M. Harrison, *Entropy concepts in physics*. Dans Entropy and informations in time and philosophy. Kubat et Zeman Ed., Elsevier (1975).
- [13] A. Spataru, *Théorie de la transmission de l'information*. Masson (1970).
- [14] E. Roubine, *Introduction à la théorie de la communication*. Masson (1970).
- [15] S. Giasu, R. Theodorescu, *La théorie mathématique de l'information*. Dunod (1968).
- [16] B. Picinbono, *Eléments sur les théories du signal, de la détection et de l'information*. Cours DEA de Mathématiques statistiques. Faculté des Sciences de Paris.

capteurs Setra



leur principe capacitif original leur confère des propriétés inégalées :

- excellente linéarité et fidélité
- très faible hystérésis, quasi nulle pour les modèles en quartz.
- signal directement de haut niveau, permettant de se passer de tout ampli intermédiaire.

Pression, modèle 204

Non linéarité 0,1% P.E. Hystérésis < 0,05% P.E. Signal de sortie 5 V, impédance < 10 ohms, temps de réponse très court < 1 ms. Calibres 0..1,75 à 0..700 bars, relatifs ou absolus.

Pression, modèle 239, pour faibles Δp
Calibres $\pm 0,7$ mb à $\pm 3,5$ bars différentiels.

Pression, modèle 250, en quartz

Précision $\pm 0,05$ % P.E. Stabilité à court et à long terme inégalable, Modèle barométrique.

Pression, modèle 237, à membrane affleurante.
Va dans les Scanivalves.

Pression modèle 228, wet-wet, faible différentielle
(± 35 mb) sous forte pression totale (210 bars), réponse rapide (< 50 ms).

Manomètres électroniques complets série 300

Affichage directement dans les unités de votre choix. Finies les visées pénibles et imprécises, les dépouillements.

Accélération, modèle 114

Masse 50 g. Bande passante : 0 Hz à 4 Hz pour le calibre $\pm 0,2G$ (le plus faible); 0 Hz à 3 000 Hz pour le calibre $\pm 10 000 G$ (le plus élevé). Signal de sortie $\pm 1,5$ V.

Autres modèles : pour emplois embarqués, pour montages triaxiaux, pour essais de collision automobile.

Agents exclusifs pour la France : **Ets A. KOVACS**
177 R. de la Convention 75015 PARIS - Tél. 250-89-70

Veuillez me documenter en détail sur le capteur SETRA modèle _____
Nom _____
Firme _____
Adresse _____

- [17] S. Ikeda, *Continuity and characterization of Shannon. Wiener information measure for continous probability distributions.* Ann. Inst. Stat. Math. Tokyo, 11, p. 131 (1959).
- [18] R. Leipnik, *Entropy and the uncertainty principle.* Information and Control, 2, p. 64 (1959).
- [19] R. Bourret, *A note on an information theoretic form of the uncertainty principle.* Information and Control, 1, p. 398 (1958).
- [20] R. Vallée, *Expression asymptotique, pour les grandes valeurs du temps, de l'information associée à la fonction d'onde dans le cas d'un corpuscule libre.* C.R. Acad. Sc., 267, B, p. 529 (1968).
- [21] I. Hirschmann, *A note on entropy.* Am. J. of Math. p 152 (1957).
- [22] I. Bialynicki, J. Mycielski, *Uncertainty relations for information entropy in wave mechanics.* Comm. Math. Phys., 44, p. 129 (1975).
- [23] J. Gibbs, *Elementary principles in statistical Mechanics.* (Vol. II, Oeuvres Complètes). New Haven (1948).
- [24] R. Vallée, *Opérateurs d'observation et théorie de l'information.* C.R. Acad. Sc., 233, B, p. 1428 (1951).
- [25] J. Rothstein, *Information, Measurement and quantum mechanics.* Science, 114, p. 171, (1951).
- [26] L. Brillouin, *La science et la théorie de l'information.* Masson (1959).
- [27] A. Renyi, *On measures of entropy and information.* Proc. 4th Berkeley Symp. Math. Statis. and Prob., University of California Press, 1, p. 547 (1960).
- [28] J. Dayantis, *Essai de quantification de l'information obtenue lors de mesures à un appareil de mesure.* J. de Chimie Phys. (70), 6, p. 865 (1973).
- [29] H. Hurwitz, *Entropy reduction in Bayesian analysis of measurements.* Phys. Rev., A, (12), 2, p. 698 (1975).
- [30] R. Liboff, *Gibbs vs. Shannon entropies.* J. of Stat. Phys. (11), 4, p. 343 (1974).
- [31] H. Hirsch, *Note on the paper « Gibbs vs Shannon entropies » by R. Liboff.* J. of Stat. Phys., (13), 3, p. 273 (1975).
- [32] C. Sempì, *The entropies with error.* Rendiconti di Matematica. (1), 9, Série VI, p. 1 (1976).
- [33] J. Kampe de Fériet, *La théorie généralisée de l'information et la mesure subjective de l'information.* Lectures notes in mathematics, n° 398, Springer (Berlin) (1974).
- [34] J. Kampe de Fériet, B. Forte, *Information et probabilité.* C.R. Acad. Sc, 265, p. 110 (1967).
- [35] J. Kampe de Fériet, B. Forte, *Information et probabilité.* C.R. Acad. Sc., 265, p. 142 (1967).
- [36] J. Kampe de Fériet, *Mesure de l'information fournie par un événement.* Colloques internationaux du CNRS, n° 186, Clermont-Ferrand, p. 191 (1969).
- [37] C. Picard, *Rencontres en théorie de l'information.* Lectures Notes in Mathematics, n° 398, Springer (Berlin) (1974).
- [38] N. Aggarwal, C. Picard, *Des mesures de l'information et de leurs applications.* Journées Lyonnaises des questionnaires C.N.R.S., 12 et 13 juin 1975., p. 31.
- [39] Structures de l'information, *Journées Lyonnaises des questionnaires.* C.N.R.S., 12 et 13 juin 1975.
- [40] R. Vallée, *Sur la formalisation mathématique en théorie de l'observation.* Actes VIIe Congrès Int. de Cybernétique, Namur (1973).
- [41] M. Mugur-Schachter, C. Padet, J. Padet, *Le concept de quantité d'information accessible par mesure correspondant à une distribution de probabilités continue.* C.R. Acad. Sc. 282, A, p. 487 (1976).
- [42] C. Padet, M. Mugur-Schachter, J. Padet, *Une nouvelle expression de l'entropie pour les distributions statistiques continues.* C.R. Acad. Sc. 281, A, p. 993 (1975).
- [43] R. Thom, *Modèles mathématiques de la morphogénèse.* Coll. 10/18, Paris (1974).
- [44] P. Hoyningen-Huene, *On the time and cell-dependence of the coarse-grained entropy.* Physica A, 82, n° 3, p. 417 (1976).
- [45] U. Dinur, R. Levine, *On the entropy of a continuous distribution.* Chem. Physics, 9, p. 17-27 (1975).
- [46] C. Sempì, *On the limit of the entropies with error.* (communication personnelle, (1976).