

# Information et entropie

Autor(en): **Vittoz, Bernard**

Objektyp: **Article**

Zeitschrift: **Mémoires de la Société Vaudoise des Sciences Naturelles**

Band (Jahr): **17 (1978-1987)**

Heft 1

PDF erstellt am: **22.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-259563>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern. Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden. Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# Information et entropie

PAR

BERNARD VITTOZ \*

## INTRODUCTION

Un des problèmes fondamentaux des sciences naturelles est de comprendre et si possible de prédire l'évolution des systèmes matériels, que ceux-ci soient inanimés ou animés. Par système matériel, nous entendons tout ensemble d'un grand nombre de particules (atomes, molécules). Dans leur évolution, ces systèmes doivent obéir à des lois de conservation: masse, énergie, quantité de mouvement, moment cinétique. Remarquons qu'une loi de conservation ne prétend pas que la grandeur considérée soit strictement constante, mais que si elle varie c'est par un échange entre le système et son entourage (l'extérieur au système). Dans le cas simple où le système est isolé de l'extérieur, une loi de conservation implique que la grandeur correspondante est constante, le système évoluant ou non. Aussi, l'évolution ne peut pas être régie par des grandeurs qui doivent (dans certains cas) rester constantes. Il faut donc faire appel à une grandeur physique qui varie lorsque le système évolue. Cette grandeur est l'entropie, étroitement liée aux concepts de probabilité et d'information.

Utilisant le modèle atomique de la matière, dont la nécessité lui paraissait évidente, BOLTZMANN (1877) fut le premier à donner une interprétation probabiliste de l'entropie. Une excellente analyse de l'œuvre de Boltzmann a été faite par R. DUGAS (1959). La liaison entre l'entropie et l'information a été reconnue plus tard (SZILARD, 1929), (SHANNON et WEAVER, 1949), (BRILLOUIN, 1959).

## CONFIGURATIONS

Considérons le système matériel le plus simple possible. Il est formé de  $N$  particules identiques (atomes, molécules) contenues dans un récipient de

---

\* Laboratoire de Génie Atomique de l'Ecole Polytechnique Fédérale de Lausanne

volume  $V$ . Nous supposons de plus que ces particules n'exercent entre elles aucune interaction à distance. Seuls les chocs interviennent: entre particules elles-mêmes, entre particules et la paroi du récipient.

Intéressons-nous d'abord aux différentes positions que peuvent prendre les particules. Pour cela, divisons le volume  $V$  en un nombre de cases identiques, et de telle façon que dans une case il ne puisse y avoir que 0 ou 1 particule. Soit  $b$  le volume d'une case (il s'identifie au volume moléculaire). Le nombre de cases  $C$  est alors égal à:

$$C = \frac{V}{b} = \text{nombre de cases} \geq N \quad (1)$$

$$N = \text{nombre de particules}$$

Comme nous avons supposé qu'il n'y avait pas d'interaction à distance entre les particules, chaque particule peut se placer indifféremment dans l'une ou l'autre des cases vides. On appelle configuration (ou complexion) une réalisation possible de l'état du système, ici dans la position de ses particules. Une configuration est ici définie par la seule connaissance pour l'ensemble des cases de l'occupation (chiffre 1) ou de non-occupation (chiffre 0) par une particule. Une configuration est donc donnée par un ensemble de  $C$  chiffres (0 ou 1) écrits dans l'ordre de numérotation des cases:

$$\underbrace{0, 0, 1, 0, 0, 0, 1, 1, \dots, 1}_{C \text{ chiffres}} = \text{une configuration} \quad (2)$$

Avec, naturellement, la condition que la somme de ces chiffres, c'est-à-dire des 1, est égale au nombre  $N$  de particules.

Calculons le nombre  $W$  de configurations différentes que peut réaliser le système, une configuration (2) n'apparaissant qu'une seule fois. Par la suite, nous entendrons par configurations des configurations différentes. A partir d'une configuration donnée, on peut en créer d'autres en permutant les numéros des  $C$  cases, ce qui donne  $C!$  permutations. Mais pour chaque configuration ainsi obtenue, les permutations des  $N$  cases occupées ne donnent en fait qu'une configuration significative, les particules étant identiques, donc indiscernables. On a donc compté  $N!$  fois trop de permutations. De même, les permutations des  $(C-N)$  cases non-occupées donnent  $(C-N)!$  fois trop de permutations. D'où:

$$W = \frac{C!}{N!(C-N)!} = \text{nombre de configurations} \quad (3)$$

Comme  $C \geq N$ , il s'ensuit naturellement que  $W \geq 1$ .

*Exemple.*  $N = 7$  (particules),  $C = 100$  (cases):  $W = 168\,500\,640$

Comme chaque particule peut se placer dans n'importe quelle case non-occupée, qu'il y ait présence ou non d'une particule dans une case voisine (interaction nulle), toutes les configurations ont la même probabilité d'être réalisées par un système. On dit que les configurations sont équiprobables. Ce qui ne veut pas dire qu'une configuration prescrite ou de caractère particulier ait la même probabilité de réalisation qu'une configuration de caractère quelconque. Par exemple, les configurations où toutes les particules occupent des cases adjacentes au sens de la description (2) sont au nombre de  $(C-N)$ ; la probabilité de réalisation de telles configurations vaut donc  $(C-N)/W$ , ce qui donne dans l'exemple numérique ci-dessus, une probabilité de  $6 \cdot 10^{-7}$ . La loterie présente le même aspect: le numéro 111 111 a autant de chance de sortir que le numéro 237453, mais les numéros formés de chiffres identiques (comme le 111 111) ont beaucoup moins de chance de sortir que les numéros formés de chiffres quelconques.

#### MESSAGES ET INFORMATION

Toute configuration réalisée par le système physique constitue un message, si l'on suppose que l'on puisse mesurer cette configuration. Autant de configurations possibles, autant de messages que peut nous délivrer le système, autant l'information que nous pouvons recevoir du système est quantitativement riche. Le système est ainsi une source d'information. On définit la quantité d'information  $I$  comme étant une mesure du nombre de messages différents, ou de configurations différentes, que peut donner le système, sans juger de la qualité de tel ou tel message. C'est aussi la mesure de l'incertitude ou de l'ignorance dans laquelle on se trouve avant de recevoir le message. Ici, nous généralisons la notion de système à toute source d'information: gaz dans un récipient, morceau de cristal de cuivre, communication téléphonique, livre, etc.

*Configurations équiprobables.* Pour tout système à configurations équiprobables, nous soumettons la définition de la mesure de l'information à deux critères:

- A. La quantité d'information  $I$  est une fonction monotone croissante du nombre  $W$  de configurations équiprobables, c'est-à-dire que  $I$  croît si  $W$  augmente:

$$I = f(W) \quad \text{et} \quad \frac{dI}{dW} > 0 \quad (4)$$

La fonction  $f(W)$  étant indépendante du système considéré (à configurations équiprobables).

B. La quantité d'information contenue dans deux systèmes indépendants (1 et 2), est égale à la somme des quantités d'information de chaque système:

$$I = I_1 + I_2 \quad (5)$$

Au sens de la thermodynamique,  $I$  est ainsi une variable extensive.

Considérons alors deux systèmes indépendants, caractérisés chacun par  $W_1$  et  $W_2$  configurations équiprobables. Les quantités d'information  $I_1$  et  $I_2$  correspondantes valent :

$$I_1 = f(W_1) \quad I_2 = f(W_2) \quad (6)$$

Comme les deux systèmes sont indépendants, à toute configuration du 1<sup>er</sup> système peuvent correspondre  $W_2$  configurations équiprobables du 2<sup>e</sup> système, car la configuration qu'a prise le 1<sup>er</sup> système n'a pas d'influence sur le choix de configuration que peut prendre le 2<sup>e</sup> système. Le nombre total  $W$  de configurations que peut donner l'ensemble des deux systèmes vaut donc :

$$W = W_1 W_2 \quad (7)$$

On a donc :

$$I = f(W) = f(W_1 W_2) \quad (8)$$

ce qui donne avec (5) et (6) :

$$f(W_1 W_2) = f(W_1) + f(W_2) \quad (9)$$

En dérivant chaque membre de l'équation (9), partiellement par rapport à  $W_1$ , on trouve :

$$W_2 f' (W_1 W_2) = f' (W_1)$$

$f'$  signifiant la dérivée de  $f(x)$  par rapport à  $x$ .

La dérivation partielle par rapport à  $W_2$  donne :

$$W_1 f' (W_1 W_2) = f' (W_2)$$

d'où :

$$W_1 f' (W_1) = W_2 f' (W_2) \quad (10)$$

L'égalité (10) doit être satisfaite quelles que soient les valeurs de  $W_1$  et  $W_2$ , donc :

$$x f' (x) = \text{cste} = C \quad ; \quad f' (x) = \frac{C}{x}$$

Alors :

$$f(x) = C \log x + \text{cste}$$

Par le choix du zéro de la quantité d'information  $I = f(W)$ , on peut prendre la constante additive égale à zéro :

$$I = C \log W$$

ce qui donne une information nulle pour  $W = 1$ , c'est-à-dire pour le cas où le système ne peut donner qu'une seule configuration. Le choix du zéro de  $I$  est alors naturel, puisque dans ce dernier cas ( $W = 1$ ) on n'attend aucune information du système, celui-ci délivrant toujours le même message.

Avec la condition de croissance monotone (4), la constante multiplicative  $C$  doit être positive, car le logarithme est une fonction croissante:  $C > 0$ .

La quantité d'information  $I$  que peut donner un système à  $W$  configurations équiprobables est donc donnée par la formule :

$$I = C \log W \quad C > 0 \quad (11)$$

#### *Choix de la constante C et de la base du logarithme*

En théorie des communications (voir l'article de F. de Coulon), on prend :

$$C = 1 \quad \text{et} \quad \log = \log_2 \quad (\log_2 2 = 1)$$

I se mesure en bits (binary digit)

Un des systèmes simples en communication est le relais. Il peut être dans 2 états (configurations) différents (ouvert ou fermé) et délivrer ainsi 2 messages. S'ils sont équiprobables, l'information  $I_1$  contenue dans un relais vaut  $I_1 = 1$  bit. Et dans  $n$  relais indépendants:  $I_n = n$ , la quantité d'information s'additionnant, selon (5).

Nous verrons plus loin qu'en thermodynamique un autre choix est fait. Jusque-là, nous laissons ce choix ouvert.

#### *Probabilité*

Avec  $W$  configurations équiprobables, on peut exprimer la probabilité  $p$  de réalisation de l'une des configurations :

$$p = \frac{1}{W} \leq 1 \quad (12)$$

La formule (11) devient alors :

$$I = -C \log p \quad I \geq 0 \quad (13)$$

### Configurations non-équiprobables

Soit un système pouvant donner  $n$  configurations non-équiprobables (on remplace le symbole  $W$  par  $n$ ,  $W$  étant réservé pour des configurations équiprobables). Numérotons chaque configuration par l'indice  $a$  ( $a = 1, 2, \dots, n$ ) et soit  $p_a$  la probabilité que la configuration  $a$  apparaisse. On a la condition (ou contrainte) usuelle sur le champ de probabilités  $p_a$ :

$$\sum_a p_a = 1 \quad p_a \geq 0 \quad (14)$$

On peut considérer que la configuration  $a$  est un sous-système à  $1/p_a$  sous-configurations équiprobables. La quantité d'information  $I_a^o$  de ce sous-système est donnée par la formule (13):

$$I_a^o = -C \log p_a$$

Le système peut ainsi être considéré comme formé de  $n$  sous-systèmes  $a$  indépendants, à condition de leur attribuer les «poids» relatifs  $p_a$ . Le sous-système  $a$  apparaît avec la probabilité  $p_a$ , sa contribution  $I_a$  à l'information totale vaut donc:

$$I_a = p_a I_a^o = -C p_a \log p_a$$

Comme les sous-systèmes sont indépendants, l'information est additive:

$$I = -C \sum_a p_a \log p_a \quad I \geq 0 \quad (15)$$

Cette formule exprime la quantité d'information  $I$  que peut donner un système dont on connaît le champ de probabilité  $p_a$ . Dans le cas d'un champ continu de probabilités, les formules ci-dessus se généralisent, les sommes devenant des intégrales (ce que nous ne ferons pas ici).

### Conséquences

#### 1. Configurations équiprobables ( $n = W$ )

Dans ce cas, les probabilités  $p_a$  sont égales:  $p_a = p$ . La formule (14) donne:

$$\sum_a p_a = \sum_{a=1}^n p = np \Rightarrow p = \frac{1}{n} = \frac{1}{W}$$

Et la formule (15):

$$I = -C \sum_{a=1}^n p \log p = -C np \log p = -C \log p = C \log W$$

On retrouve bien les formules (11) et (13) du cas des configurations équiprobables.

## 2. Additivité de l'information

La quantité d'information totale  $I$  de deux systèmes  $S'$  et  $S''$  indépendants est la somme des quantités d'information  $I'$  et  $I''$ , même si les configurations ne sont pas équiprobables. En effet, soit :

$p'_a$  et  $p''_\beta$  = champs de probabilités des systèmes  $S'$  et  $S''$

$$\sum_a p'_a = 1 \quad \sum_\beta p''_\beta = 1$$

Les deux systèmes étant indépendants, la probabilité  $p_{a\beta}$  pour que les configurations  $a$  et  $\beta$  soient réalisées par  $S'$  et  $S''$  respectivement, vaut :

$$p_{a\beta} = p'_a p''_\beta \tag{16}$$

D'où l'information totale  $I$  :

$$\begin{aligned} I &= -C \sum_{a\beta} p_{a\beta} \log p_{a\beta} = -C \sum_{a\beta} p'_a p''_\beta [\log p'_a + \log p''_\beta] = \\ &= -C \underbrace{\sum_\beta p''_\beta}_1 \sum_a p'_a \log p'_a - C \underbrace{\sum_a p'_a}_1 \sum_\beta p''_\beta \log p''_\beta = \\ &= -C \sum_a p'_a \log p'_a - C \sum_\beta p''_\beta \log p''_\beta = I' + I''. \text{ On a bien :} \end{aligned}$$

$$I = I' + I''$$

## 3. Information maximale

Pour un nombre de configurations donné  $n$ , l'information sera maximale si les configurations sont équiprobables, c'est-à-dire si le champ de probabilités est uniforme :

$$n \text{ donné} \quad I_{\max} \Leftrightarrow p_a = p (= \frac{1}{n}) \tag{17}$$

La démonstration se fait par la méthode des multiplicateurs de Lagrange (annexe A), car le champ de probabilités  $p_a$  est soumis à la condition :

$$\sum_a p_a = 1$$

Si l'information est maximale, cela signifie que nous sommes dans la situation la plus incertaine vis-à-vis de la connaissance à priori du système. Rappelons en effet que la quantité d'information est la mesure du nombre de messages différents que le système pourra nous donner, ce qui est bien une mesure de l'incertitude dans laquelle nous sommes avant que le système



ne délivre un message. Or, si le champ de probabilités est uniforme, le système aura le maximum de choix entre les différents messages possibles, ce qui correspond bien à l'incertitude maximale, donc à l'information maximale.

En revanche, si le champ de probabilités n'est pas uniforme, le choix du système sera moins libre (à même nombre de messages possibles): tel message a plus de chance d'apparaître qu'un autre. Donc, plus le champ de probabilités est non-uniforme, plus nous avons une certaine connaissance à priori du système. A la limite, tous les  $p_a$  sont nuls sauf un qui doit être alors égal à l'unité: l'information est alors nulle, le système ne peut donner qu'un seul message, notre incertitude à priori est bien nulle et nous n'attendons aucune information du système.

### *L'information reçue $I_r$*

Dans le langage courant, l'information est la mesure du nombre de messages reçus. Cette information reçue  $I_r$  est directement reliée à l'information  $I$  définie plus haut. En effet, l'acquisition du message fait passer l'information, que peut donner le système, de la valeur initiale  $I_i$  à la valeur finale  $I_f = 0$ . Il est alors naturel de poser que l'information reçue est égale à la diminution de l'information  $I$  que peut donner le système :

$$I_r = I_{\text{initiale}} - I_{\text{finale}} \Leftrightarrow \Delta I_r = -\Delta I \quad (19)$$

Les deux significations que l'on peut donner au terme information prêtent souvent à confusion.

## L'ENTROPIE S EN PHYSIQUE

Considérons un système matériel dont l'état macroscopique est décrit par l'énergie interne  $U$ , le volume  $V$ , le nombre  $N$  de particules supposées identiques (on peut aussi recourir au nombre de moles). Au sens de la thermodynamique, ce système matériel dont l'état macroscopique est défini par trois variables ( $U$ ,  $V$ ,  $N$ ) est un système simple, tandis qu'un système composite est constitué de particules différentes. Un système simple peut réaliser l'état considéré selon un ensemble de configurations définies par exemple par la position et la vitesse des particules. Comme précédemment, pour simplifier, nous considérons un nombre fini de configurations possibles auxquelles est associé le champ de probabilités  $p_a$ . Boltzmann a donné une définition de l'entropie  $S$  du système tout à fait semblable à la définition donnée, 50 ans plus tard, à la quantité d'information  $I$ . Seules, la constante  $C$  et la base du logarithme sont différentes dans les formules (11) et (15):

$$\text{pour } W \text{ configurations équiprobables: } S = k \log W \quad (20)$$

$$\text{pour } n \text{ configurations, probabilités } p_a: S = -k \sum_{a=1}^n p_a \log p_a \quad (21)$$

avec:

$$k = \text{constante de Boltzmann } (1,38 \cdot 10^{-23} \text{ J K}^{-1})$$

$$\log = \text{logarithme naturel (base } e)$$

L'entropie  $S$  est ainsi proportionnelle à la quantité d'information  $I$ . Les propriétés vues pour  $I$  sont donc valables pour  $S$ .

La connaissance de l'état macroscopique ( $U, V, N$ ) constitue des contraintes (ou conditions) sur le champ de probabilités  $p_a$ , en plus de la contrainte  $\sum_a p_a = 1$ .

Grâce au 2<sup>e</sup> principe (de la thermodynamique), il est possible d'étudier les systèmes à l'équilibre (ils n'évoluent pas) et les systèmes dans leur évolution. Le 2<sup>e</sup> principe peut s'énoncer de la façon suivante:

1° à l'équilibre, l'entropie est maximale sous réserve que les contraintes imposées soient satisfaites.

$$U, V, N = \text{cstes et } \sum_a p_a = 1 \Rightarrow S = S_{\text{max}} \text{ à l'équilibre} \quad (22)$$

En termes de la théorie de l'information, on peut donner un autre énoncé (JAYNES, 1967):

la description la plus honnête de ce que l'on sait est de reconnaître qu'en dehors des contraintes imposées, notre ignorance est maximale vis-à-vis des configurations que peut prendre le système.

On dit aussi que l'état d'équilibre du système est celui du désordre maximal, compte tenu des contraintes imposées.

2° si le système est hors de l'équilibre, il tendra à évoluer dans le sens de l'augmentation de son entropie, sous réserve que les contraintes imposées soient satisfaites.

$$U, V, N = \text{cstes et } \sum_a p_a = 1 \Rightarrow \frac{dS}{dt} > 0 \text{ hors équilibre} \quad (23)$$

Ainsi, l'état le plus probable du système est celui d'information ou d'entropie maximale, de plus grand désordre, avec la condition que les contraintes imposées soient satisfaites.

Dans l'annexe B, l'augmentation de l'entropie est établie dans le cas où l'évolution du système se fait par processus de Markov.

Les énoncés ci-dessus se généralisent sans autres à des systèmes composites.

*Application: gaz parfait*

Le modèle du gaz parfait est un système constitué de particules identiques libres (sans interaction à distance). L'énergie interne  $U$  est alors simplement égale à l'énergie cinétique totale des particules. Intéressons-nous uniquement à la distribution d'énergie (cinétique) des particules du gaz à l'équilibre. Soit :

$p(E)$  = probabilité pour qu'une particule quelconque du gaz ait une énergie comprise entre  $E$  et  $E + dE$

$\bar{E}$  =  $U/N$  = énergie cinétique moyenne des particules

$\bar{S}$  =  $S/N$  = entropie moyenne par particule, due au champ de probabilités  $p(E)$ .

On a les contraintes :

$$\sum p(E) = 1 \quad \text{et} \quad \sum E p(E) = \bar{E} \quad (24)$$

et l'entropie moyenne  $\bar{S}$  par particule :

$$\bar{S} = -k \sum p(E) \log p(E) \quad (25)$$

qu'il faut rendre maximum compte tenu des contraintes (24). Avec la méthode des multiplicateurs de Lagrange, on trouve :

$$p(E) = A e^{-bE/k} \quad (26)$$

où  $A$  est proportionnel au nombre de particules qui à priori peuvent avoir une vitesse telle que leur énergie cinétique ( $mv^2/2$ ) soit comprise entre  $E$  et  $E + dE$ . Il est donné dans l'espace des vitesses  $\vec{v}$  par le volume d'une coquille sphérique de rayon  $v$ ,  $v + dv$  :

$$A = B 4\pi v^2 dv \quad (27)$$

Si l'on introduit les formules (26) et (27) dans les contraintes (24), on obtient deux équations aux inconnues  $b$  et  $B$ . Les sommations devenant des intégrales, on trouve :

$$B = \frac{1}{4} \left(\frac{m}{2\pi k}\right)^{\frac{3}{2}} b^{\frac{3}{2}} \quad (28)$$

$$b = \frac{3}{2} \frac{k}{\bar{E}} \quad (29)$$

Par définition, la température  $T$  d'un gaz de particules libres est proportionnelle à l'énergie cinétique moyenne  $\bar{E}$  des particules :

$$\text{D'où: } \quad \bar{E} = \frac{3}{2} k T \quad (30)$$

$$b = \frac{1}{T} \quad (31)$$

$$p(E) = A e^{-E/kT} \quad (32)$$

A est le nombre d'états d'énergie ( $E, E + dE$ ), tandis que  $e^{-E/kT}$  est la probabilité d'occupation de ces états. C'est la statistique de Boltzmann. Avec (27) et (28):

$$p(v) = \frac{1}{V \pi} \left( \frac{m}{2kT} \right)^{\frac{3}{2}} v^2 e^{-mv^2/2kT} dv \quad (33)$$

qui est la distribution de Maxwell pour les vitesses des particules.

#### ENTROPIE ET THERMODYNAMIQUE

L'entropie moyenne  $\bar{S}$  par particule devient d'après (25) et (32):

$$\bar{S} = \frac{\bar{E}}{T} + f(T)$$

Multipliant par le nombre  $N$  de particules, on trouve l'entropie  $S$  du système, avec  $N \bar{E} = U =$  énergie interne:

$$S = \frac{U}{T} + f(T, N, V)$$

où le terme additif  $f(T, N, V)$  provient de la contribution du volume  $V$  à l'entropie (configurations spatiales). Si l'on fait varier l'énergie interne  $U$  de  $dU$ , à volume et nombre de particules constants, on trouve:

$$dS = \frac{dU}{T} = \frac{\delta Q}{T} \quad (34)$$

où  $\delta Q$  est la quantité de chaleur fournie au système.

En effet, d'après le 1<sup>er</sup> principe,  $dU = \delta Q$  si  $V, N =$  cstes. Ainsi, pour le gaz parfait, on a retrouvé la définition usuelle de l'entropie, pour une transformation réversible, c'est-à-dire où le système est à l'équilibre. C'est bien ce que nous avons supposé dans les calculs ci-dessus.

L'entropie (21) au sens de Boltzmann, la même en théorie de l'information, s'identifie donc à l'entropie thermodynamique dans le cas de la trans-

formation réversible d'un gaz parfait. On considère alors que la définition (21) de Boltzmann, beaucoup plus générale que la définition usuelle ( $\delta Q/T$ ), donne l'expression de l'entropie des systèmes physiques.

### ANNEXE A – Information maximale

L'information due à un champ de probabilités  $p_a$  ( $a = 1, 2 \dots n$ ) est donnée par (15):

$$I = -C \sum_a p_a \log p_a \quad (\text{A1})$$

qu'il s'agit de rendre maximum, avec la condition (14):

$$\sum_a p_a - 1 = 0 \quad (\text{A2})$$

Avec la méthode des multiplicateurs de Lagrange, on construit la fonction F:

$$F = -C \sum_a p_a \log p_a + \lambda \left( \sum_a p_a - 1 \right) \quad (\text{A3})$$

Le problème revient à rendre F stationnaire selon les variables  $p_a$  et  $\lambda$ :

$$\frac{\partial F}{\partial p_a} = 0 = -C \log p_a - C + \lambda$$

D'où:  $\frac{\partial F}{\partial \lambda} = 0 = \sum_a p_a - 1$ : c'est la condition (A2)

$$\log p_a = \frac{\lambda - C}{C} = p: \text{ tous les } p_a \text{ sont égaux.}$$

Avec (A2):

$$p_a = p = \frac{1}{n} \quad (\text{A4})$$

et

$$I(p_a = p) = -C \log p = C \log n \quad (\text{A5})$$

Le champ de probabilités uniforme ( $p_a = p$ ) correspond à une information I optimale. Montrons que c'est un maximum. Pour cela, considérons un champ de probabilités voisin du champ uniforme:

$$p_a = p (1 + \varepsilon_a) \quad | \varepsilon_a | \ll 1 \quad (\text{A6})$$

La condition (A2) implique:

$$\sum_a \varepsilon_a = 0 \quad (\text{A7})$$

L'information  $I$  devient, avec l'approximation  $\log(1 + \varepsilon_a) = \varepsilon_a$ :

$$\begin{aligned} I &= -C \sum_a p(1 + \varepsilon_a) [\log p + \varepsilon_a] = \\ &= -C \log p - C p \underbrace{\sum_a \varepsilon_a}_0 - C p \log p \underbrace{\sum_a \varepsilon_a}_0 - C p \sum_a \varepsilon_a^2 \\ I &= -C \log p - C p \sum_a \varepsilon_a^2 \leq I(p = p_a) \end{aligned}$$

l'égalité n'ayant lieu que si  $\sum_a \varepsilon_a^2 = 0$ , c'est-à-dire si tous les  $\varepsilon_a$  sont nuls. L'information sera maximale si et seulement si le champ de probabilités est uniforme.

*ANNEXE B* – Inégalité de Jensen et entropie croissante

La définition de la quantité d'information et de l'entropie fait apparaître la fonction  $f(x)$  pour  $x$  compris dans l'intervalle  $[0, 1]$ :

$$f(x) = -x \log x \quad x \in [0, 1] \tag{B1}$$

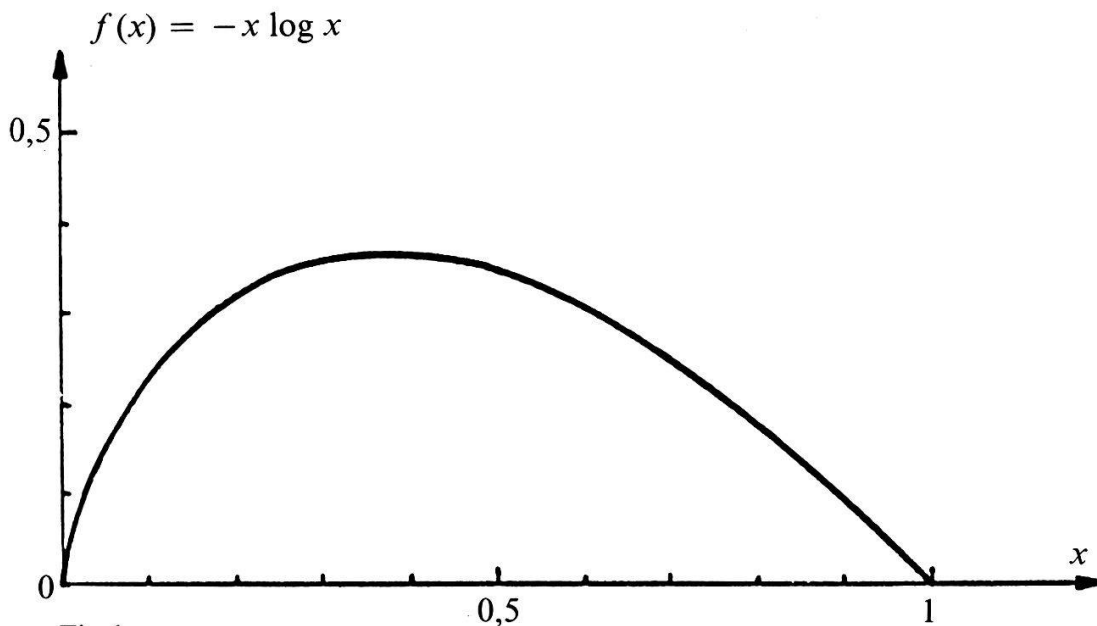


Fig. 1.

C'est une fonction concave (fig. 1). Or, toute fonction concave, dans un intervalle  $[a, b]$ , obéit à l'inégalité de Jensen :

$$\sum_i \lambda_i f(x_i) \leq f\left(\sum_i \lambda_i x_i\right) \quad x_i \in [a, b] \tag{B2}$$

avec :

$$i = 1, 2 \dots n; \quad 0 \leq \lambda_i \leq 1; \quad \sum_i \lambda_i = 1 \tag{B3}$$

L'égalité ne pouvant avoir lieu que si tous les coefficients  $\lambda_i$  sont nuls sauf un (disons  $\lambda_1$ ), égal à l'unité à cause de (B3), ou si tous les  $x_i$  sont égaux (ce qui revient au même):

$$\text{égalité} \Leftrightarrow \lambda_i = 0 \text{ sauf } \lambda_1 = 1 \text{ ou } x_i \text{ égaux} \quad (\text{B4})$$

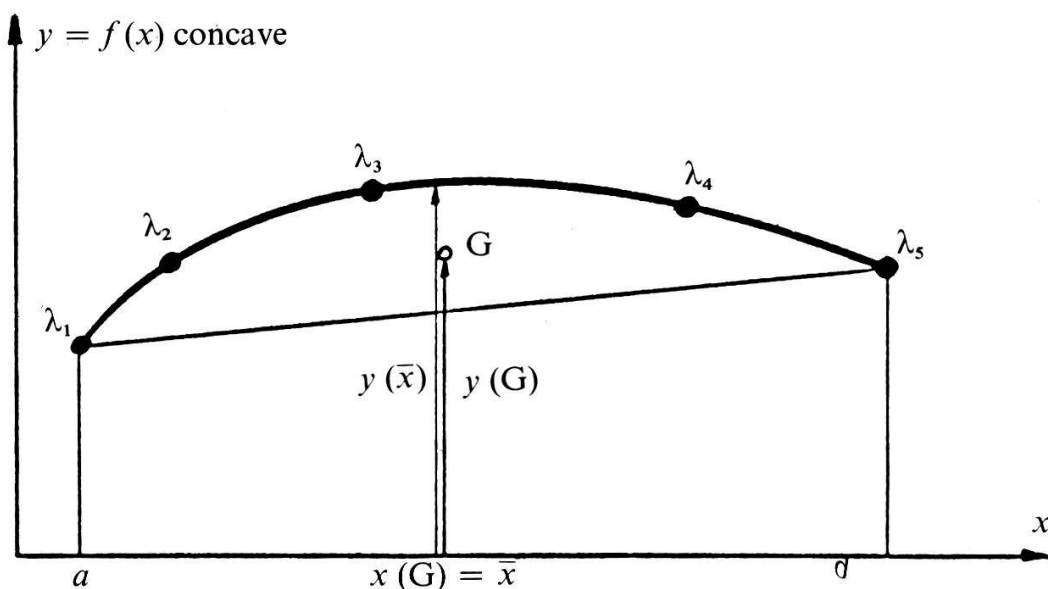


Fig. 2.

L'inégalité de Jensen peut se démontrer par l'analyse ou par la géométrie des masses (fig. 2). En effet, on peut considérer que les coefficients  $\lambda_i$  sont des masses ponctuelles disposées sur la courbe  $y = f(x)$  aux abscisses respectives  $x_i$ ; la masse totale valant l'unité d'après (B3). Alors:

$$\sum_i \lambda_i x_i = x(G) = \bar{x} = \text{abscisse du centre de masse (barycentre) ou abscisse moyenne}$$

$$\sum_i \lambda_i f(x_i) = \sum_i \lambda_i y_i = y(G) = \text{ordonnée du centre de masse}$$

Or, nous savons que le centre de masse d'un ensemble de masses ponctuelles est à l'intérieur de cette distribution, ici à l'intérieur de la ligne fermée définie par la courbe  $y = f(x)$  et la droite joignant les masses extrêmes. Donc:

$$y(G) \leq y(\bar{x})$$

l'égalité ne pouvant avoir lieu que si les masses sont concentrées en un seul point. Ce qui implique  $\lambda_i = 0$  sauf l'un d'eux, disons  $\lambda_1 = 1$ , ou (ce qui revient au même), tous les  $x_i$  sont égaux. L'inégalité de Jensen est ainsi démontrée.

*Applications*1. *Deux champs de probabilité*

Soit:

 $p_a, q_a =$  deux champs de probabilités;  $a = 1, 2 \dots n$ 

$$p_a, q_a \in [0, 1] \quad \sum_a p_a = \sum_a q_a = 1$$

Posons dans l'inégalité de Jensen ( $a$  remplaçant  $i$ ): $f(x) = \log x$ , qui est une fonction concave pour tout  $x > 0$ 

$$\lambda_a = q_a \quad \text{et} \quad x_a = \frac{p_a}{q_a}$$

$$\text{d'où:} \quad \sum_a q_a \log p_a \leq \sum_a q_a \log q_a \quad (\text{B5})$$

l'égalité ne pouvant avoir lieu que si  $q_a = p_a$ .2. *Entropie maximale sous la seule condition*  $\sum_a p_a = 1$ 

On a l'entropie:

$$S = -C \sum_a p_a \log p_a \quad ; \quad C > 0$$

Posons dans (B5):

$$p_a = \frac{1}{n} = p \quad \text{champ de probabilités uniforme}$$

Alors:

$$\log p \leq \sum_a q_a \log q_a, \quad \text{ou:} \quad -S(p_a = p) \leq -S(q_a), \quad \text{donc:}$$

$$S(p_a = p) \geq S(q_a) \quad (\text{B6})$$

l'égalité n'ayant lieu que si tous les  $q_a$  sont égaux à  $p$ . Ainsi, l'entropie est bien maximale pour des configurations équiprobables.3. *Processus de Markov et entropie croissante*

On considère un système qui évolue selon des processus de Markov, c'est-à-dire que nous supposons que la seule connaissance du champ de probabilités  $p_a$  à l'instant  $t$  permet de prévoir le champ de probabilités  $q_a$  à l'instant ultérieur  $t' > t$ :

$$q_a = \sum_{\beta} a_{a\beta} p_{\beta} \quad a, \beta = 1, 2 \dots n \quad (\text{B7})$$

La matrice  $a_{a\beta}$  est appelée la matrice de transfert. Elle obéit aux relations:

$$0 \leq a_{a\beta} \leq 1 \quad \sum_{\beta} a_{a\beta} = 1 \quad (\text{B8})$$



En prenant  $C = 1$ , les entropies  $S$  et  $S'$  aux instants  $t$  et  $t' > t$ , s'écrivent :

$$S = - \sum_a p_a \log p_a$$

$$S' = - \sum_a q_a \log q_a = - \sum_\beta p_\beta \sum_a a_{a\beta} \log q_a$$

D'après l'inégalité de Jensen (B2) :

$$\sum_a a_{a\beta} \log q_a \leq \log \left( \sum_a a_{a\beta} q_a \right) = \log b_\beta$$

avec :

$$b_\beta = \sum_a a_{a\beta} q_a; \quad 0 \leq b_\beta \leq 1; \quad \sum_\beta b_\beta = 1$$

D'où :

$$S' \geq - \sum_\beta p_\beta \log b_\beta \geq - \sum_\beta p_\beta \log p_\beta, \text{ d'après (B5).}$$

On a bien :

$$S' \geq S$$

l'égalité ne pouvant avoir lieu que si  $b_\beta = p_\beta$ , ce qui correspond à l'état d'équilibre, le champ de probabilités restant le même, à des permutations internes arbitraires près.

#### RÉFÉRENCES

BOLTZMANN, L., 1877. – *Wiener Berichte* 76, 373.

BRILLOUIN, L., 1959. – *La science et la théorie de l'information* (Masson, Paris).

DUGAS, R., 1959. – *La théorie physique au sens de Boltzmann et ses prolongements modernes* (éd. du Griffon, Neuchâtel).

JAYNES, E. T., 1967. – dans «*Delaware Seminar in the Foundations of Physics*». M. Bunge éd., Springer-Verlag, Berlin.

SHANNON, C. E. and WEAVER, W., 1949. – *The Mathematical Theory of Communication* (The University of Illinois Press, Urbana).

SZILLARD, L., 1929. – *Z. f. Physik* 53, 840.