L'ENTROPIE DE BOLTZMANN ET L'ENTROPIE DE SHANNON, MEME CONCEPT ?

José-Philippe PEREZ
Laboratoire d'astrophysique
Observatoire Midi-Pyrénées
31400 Toulouse

Résumé :

On analyse les liens entre l'entropie de Boltzmann telle qu'elle est définie en physique statistique et l'entropie de Shannon telle qu'elle est introduite dans la théorie de l'information, pour aboutir à l'identité des concepts. Dans ce contexte, l'entropie apparaît naturellement comme une mesure de l'information manquante sur le système et la méthode du maximum de l'entropie permet de retrouver les distributions statistiques de Maxwell-Boltzmann, de Fermi-Dirac et de Bose-Einstein.

1 Introduction

Le lien entre l'entropie statistique et l'information n'est pas immédiat et peut même surprendre [1]. Comment, en effet, un concept tel que l'entropie thermodynamique, relié depuis Boltzmann au désordre d'un système, peut-il être associé à l'information au sens de la théorie de la communication de Shannon ?

Nous proposons ici de rappeler la relation entre l'entropie statistique et l'entropie telle qu'elle est définie par Shannon, ceci afin d'interpréter l'entropie comme une mesure du manque d'information sur le système. Cet aspect important de l'interprétation de l'entropie thermodynamique est malheureusement quasi inexistant dans l'enseignement de la physique. Seuls quelques rares ouvrages signalent l'existence d'une relation étroite entre le point de vue de la physique statistique et celui de la théorie de la communication [2], [3], [4].

Ce lien apparaît clairement lorsque l'on restitue les distributions bien connues de la thermodynamique statistique, à l'aide de la méthode du maximum d'entropie appliquée en théorie de la communication.

2 Entropie en théorie de l'information

Dans la théorie de la communication, développée par l'ingénieur américain Shannon en 1949 [5], le concept d'entropie est introduit à partir de celui d'information transmise par un canal depuis la source d'information, à l'entrée d'un système, jusqu'au récepteur à sa sortie.

On considère un ensemble de e messages x_e (e pour entrée ou états) émis par une source à l'entrée avec des probabilités respectives P_e, en général différentes. À la sortie, on s'intéresse à la probabilité conditionnelle $P(s|e)$ de détecter le message y_s (s pour sortie), sachant que la source à l'entrée a émis le message x_e. La théorie de Shannon, comme la physique statistique, est donc essentiellement probabiliste.

Par définition, l'information contenue dans le message x_e à l'entrée, de probabilité P_e, est :

$$I_e = \mathrm{lb}\left(\frac{1}{P_e}\right) = -\mathrm{lb}P_e$$

lb étant le logarithme binaire (base 2). Le mot information n'a pas ici le sens vague habituel, mais au contraire il est très précis et exclut tout contenu sémantique associé à un groupe de messages provenant de la source.

Cette définition de l'information satisfait aux deux conditions suivantes :

(1) l'information contenue dans le message x_e, de probabilité P_e, est d'autant plus grande que sa probabilité est faible. Lorsque le message est totalement prévisible ($P_e = 1$), l'information qu'il contient est nulle ; en revanche, lorsqu'il est original

($P_e \simeq 0$), l'information est très grande. En bref, selon Shannon, être informé, c'est être surpris.

(2) Si les probabilités d'un événement composite sont les produits des probabilités d'événements élémentaires, comme c'est le cas s'ils sont indépendants, les informations s'ajoutent, ce qui justifie le choix d'une fonction logarithmique.

L'entropie de la source est la valeur moyenne de l'information (espérance mathématique au sens des probabilités) :

$$H = \sum_s P_e\, I_e = \sum_e P_e \mathrm{lb}\left(\frac{1}{P_e}\right) = -\sum_e P_e\, \mathrm{lb} P_e$$

L'unité de H est le bit, qui est l'entropie d'un système élémentaire à deux états, de même probabilité 1/2, puisque :

$$H = -\frac{1}{2} \times \mathrm{lb}\left(\frac{1}{2}\right) - \frac{1}{2} \times \mathrm{lb}\left(\frac{1}{2}\right) = \mathrm{lb}\, 2 = 1 \mathrm{bit}$$

3 Entropie statistique

3.1 Définition

Si on considère un système thermodynamique, qui peut se trouver dans plusieurs états microscopiques e, de probabilités respectives P_e, l'entropie statistique est la quantité :

$$S = k_B \sum_e P_e \ln\left(\frac{1}{P_e}\right) = -k_B \sum_e P_e \ln P_e$$

Ainsi, comme l'entropie en théorie de la communication, l'entropie statistique est la valeur moyenne de l'information $\ln(1/P_e)$. Entre S et H la relation est simple et directe ; puisque $\ln P_e = \ln 2 \times \mathrm{lb} P_e$, on a :

$$S = k_B \ln 2\, H = 0,956 \times 10^{-23}\, H$$

Le logarithme népérien et le coefficient k_B (constante de Boltzmann) n'introduisent qu'un banal changement d'unité : alors que H s'exprime en bit, l'unité de S est le joule par kelvin.

3.2 Entropie statistique dans l'hypothèse microcanonique

Dans l'hypothèse microcanonique, valable pour des systèmes isolés, tous les états microscopiques sont également probables. Par conséquent :

$$P_e = \frac{1}{\Omega} \quad \text{d'où} \quad S = -k_B \sum_e \frac{1}{\Omega} \ln\left(\frac{1}{\Omega}\right) = -k_B \ln\left(\frac{1}{\Omega}\right) = k_B \ln \Omega$$

Ainsi, dans ce cas, l'entropie statistique S est reliée au nombre d'états accessibles Ω par la relation $S = k_B \ln\Omega$ introduite par Boltzmann dès 1867. On montre que l'hypothèse microcanonique est celle qui réalise, pour un système isolé, la valeur maximale de l'entropie statistique définie plus haut [6].

4 Entropie statistique et information manquante

Considérons l'exemple bien connu du système à deux états, constitué par un gaz de molécules contenues dans un récipient à deux compartiments dont on escamote la cloison.

On sait que l'entropie est maximale lorsque les nombres de molécules dans chaque compartiment sont égaux. Selon l'interprétation statistique, le nombre d'états microscopiques correspondant à l'état macroscopique, compatible avec les contraintes sur le nombre de particules et sur l'énergie, est lui aussi maximal. Dans l'analyse de Boltzmann, on dit que le désordre est maximal. En revanche, l'entropie est minimale si toutes les particules sont dans un seul état, c'est-à-dire dans un seul compartiment ; elle vaut précisément 0 dans ce cas, puisque le nombre d'états microscopiques se réduit alors à 1. Là, le désordre est minimal.

Du point de vue de la théorie de l'information, ce système thermodynamique se comporte comme une source qui n'envoie aucun message vers un récepteur ; aussi l'entropie de la source mesure-t-elle *l'information qui manque* au récepteur, du fait de l'absence de message.

Dans le premier cas, où les les nombres de molécules dans chaque compartiment sont égaux, l'information manquante est maximale puisqu'un grand nombre d'états microscopiques réalisent l'état macroscopique le plus probable. Dans le second, où toutes les molécules sont dans le même compartiment, il n'y a qu'un seul état microscopique ; l'information manquante est nulle.

Ainsi, on peut dire, comme le biophysicien H. Atlan, que l'entropie statistique est proportionnelle à l'information moyenne que nous recevrions du système si nous pouvions savoir, par la mesure, dans quel état microscopique il se trouve, grâce à un message [7]. C'est dans ce contexte que Brillouin [8] introduisit la néguentropie $N = -S$. Après analyse, il nous semble préférable d'utiliser l'expression d'information manquante pour souligner la signification informationnelle de l'entropie et de cantonner la néguentropie dans son rôle de potentiel thermodynamique associé à un système isolé [6].

5 Les statistiques classiques et quantiques déduites de la théorie de l'information

5.1 Méthode du maximum d'entropie

Comme son nom l'indique, la méthode du maximum d'entropie consiste à chercher le maximum de l'entropie d'un système, laquelle est exprimée en assignant une loi de probabilité compatible avec certaines contraintes, c'est-à-dire avec certaines connaissances partielles que l'on a sur le système. Cette méthode revient donc à rechercher la loi de probabilité qui rend maximale l'information manquante associée au système, c'est-à-dire la loi qui, en n'introduisant aucune information supplémentaire, en dehors de celles exprimées sous forme de contraintes, est la plus objective.

La mise en œuvre de cette méthode s'appuie sur la technique des multiplicateurs de Lagrange [2]. On cherche à déterminer l'ensemble des n variables d'une fonction $H(x_1, \ldots, x_i, \ldots, x_n)$, qui rend cette fonction extrémale (minimale ou maximale), sachant que ces variables sont soumises à K contraintes écrites sous la forme :

$$c_k(x_1, \ldots x_i, \ldots x_n) = 0 \quad \text{avec} \quad 1 \leq k \leq K$$

On montre que cet ensemble est celui qui réalise l'extrémum de la fonction de Lagrange :

$$L = H + \sum_k \lambda_k c_k$$

les coefficients λ_k étant des des quantités arbitraires, appelés multiplicateurs de Lagrange. En différentiant par rapport à x_i, on obtient :

$$\frac{\partial H}{\partial x_i} + \sum_k \lambda_k \frac{\partial c_k}{\partial x_i} = 0$$

5.2 Application à la thermodynamique statistique

Les démonstrations qui suivent s'inspirent de l'excellent ouvrage de Kapur [9] sur les applications de la méthode du maximum d'entropie.

5.2.1 Distribution de Maxwell-Boltzmann

Considérons un système de N particules dont chacune peut occuper un état d'énergie ε_i.

Si P_i représente la probabilité d'occupation par une particule du niveau d'énergie ε_i, on a, en sommant sur les n niveaux :

$$\sum_i P_i = 1$$

Il s'agit dans ce problème de déterminer la probabilité P_i en fonction de ε_i. Pour cela, on applique la méthode du maximum d'entropie en tenant compte de la contrainte précédente sur les probabilités et celle sur l'énergie qui s'écrit, si ε est l'énergie moyenne :

$$\sum_i P_i \varepsilon_i = \varepsilon$$

Pour trouver le maximum de $H = -\sum_i P_i \mathrm{lb} P_i$, ou de $H \ln 2 = -\sum_i P_i \ln P_i$, sous les deux contraintes précédentes, on cherche l'extrémum de la fonction de Lagrange suivante :

$$L = -\sum_i P_i \ln P_i + \lambda_1 \left(\sum_i P_i - 1 \right) + \lambda_2 \left(\sum_i P_i \varepsilon_i - \varepsilon \right)$$

En annulant la dérivée de L par rapport à P_i, on trouve :

$$\frac{\partial L}{\partial P_i} = -\ln P_i - 1 + \lambda_1 + \lambda_2 \varepsilon_i = 0 \quad \text{d'où} \quad \ln P_i = -1 + \lambda_1 + \lambda_2 \varepsilon_i \quad \text{et} \quad P_i = \exp(-1 + \lambda_1 + \lambda_2 \varepsilon_i)$$

Comme $\sum_i P_i = 1$, il vient :

$$\sum_i \exp(-1 + \lambda_1 + \lambda_2 \varepsilon_i) = 1 \quad \text{soit} \quad Z = \sum_i \exp(\lambda_2 \varepsilon_i) = \exp(1 - \lambda_1)$$

Ainsi, en introduisant la fonction de partition Z, les probabilités P_i sont données par l'expression :

$$P_i = \frac{\exp(\lambda_2 \varepsilon_i)}{Z}$$

Or, en physique, on définit la température thermodynamique d'un système de N particules, d'énergie totale $U = N\varepsilon$, par l'expression :

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U} \right)_{N,\ldots} \quad \text{avec} \quad S = N k_B \ln 2 \, H \quad \text{et} \quad U = N\varepsilon$$

Par conséquent :

$$\frac{1}{T} = k_B \left(\frac{\partial H \ln 2}{\partial \varepsilon} \right)_{N,\ldots}$$

Il vient, en exprimant $H \ln 2$ et en effectuant cette dérivation :

$$H \ln 2 = -\sum_i P_i \ln P_i = -\frac{1}{Z} \sum_i \exp(\lambda_2 \varepsilon_i)(\lambda_2 \varepsilon_i - \ln Z) = -\lambda_2 \sum_i P_i \varepsilon_i + \ln Z = -\lambda_2 \varepsilon + \ln Z$$

Par conséquent :

$$\frac{1}{T} = k_B \left(\frac{\partial H \ln 2}{\partial \varepsilon} \right)_{N,\ldots} = -k_B \lambda_2 \quad \text{d'où} \quad \lambda_2 = -\frac{1}{k_B T} = -\beta$$

Finalement :

$$P_i = \frac{\exp(-\beta \varepsilon_i)}{Z} \quad \text{avec} \quad Z = \sum_i \exp(-\beta \varepsilon_i) \quad \text{et} \quad \beta = \frac{1}{k_B T}$$

5.2.2 Statistiques quantiques

Dans l'analyse précédente, les particules étaient supposées distinctes puisque, même identiques, elles pouvaient occuper des niveaux d'énergie distincts, ce qui les rendait discernables.

Ici, le caractère probabiliste est d'une double nature : d'abord, les différents niveaux d'énergie ε_i sont affectés de probabilités G_i associées aux facteurs de dégénérescence g_i puisque la sommation porte sur les états et non sur les énergies. On a naturellement :

$$G_i = \frac{g_i}{\sum_{i=1}^{n} g_i}$$

Ensuite chaque niveau d'énergie ε_i peut être occupé par un nombre entier aléatoire j de particules identiques *indiscernables* avec la probabilité jointe $P_{i,j}$ pour que j particules, compris entre 0 et m, aient l'énergie ε_i. La probabilité jointe satisfait à la relation suivante :

$$\sum_{j=0}^{m} P_{i,j} = 1$$

puisqu'un niveau d'énergie ε_i est occupé avec certitude par un nombre quelconque j de particules, compris entre 0 et m.

Dans ce problème, on doit prendre en compte plusieurs contraintes ; d'abord celles sur les probabilités qui s'écrivent simplement :

$$\sum_{i=1}^{n} G_i = 1 \quad \text{et} \quad \sum_{j=0}^{m} P_{i,j} = 1$$

Les autres contraintes concernent l'énergie totale U et le nombre N de particules :

$$\sum_{i=1}^{n} G_i \varepsilon_i \sum_{j=0}^{m} jP_{ij} = U \quad \text{et} \quad \sum_{i=1}^{n} G_i \sum_{j=0}^{m} jP_{ij} = N$$

L'entropie a pour expression :

$$H = -\sum_{i=1}^{n} G_i H_i = -\sum_{i=1}^{n} G_i \sum_{j=0}^{m} P_{i,j} \mathrm{lb} P_{i,j} \quad \text{puisque} \quad H_i = \sum_{j=0}^{m} P_{i,j} \mathrm{lb} P_{i,j}$$

La recherche du maximum de H ou de $H\ln 2$, avec contraintes, revient à déterminer les conditions dans lesquelles la fonction de Lagrange L est extrémale :

$$L = -\sum_{i=1}^{n} G_i \sum_{j=0}^{m} P_{i,j} \ln P_{i,j} + \sum_{k} \lambda_k c_k .$$

avec :

$$\sum_k \lambda_k c_k = \lambda_1 \left(\sum_j P_{i,j} - 1\right) + \lambda_2 \left(\sum_i G_i - 1\right) + \lambda_3 \left(\sum_i G_i \sum_j jP_{i,j} - N\right) + \lambda_4 \left(\sum_i G_i \varepsilon_i \sum_j jP_{i,j} - U\right)$$

Annulons la dérivée de L par rapport à $P_{i,j}$. On trouve :

$$\frac{\partial L}{\partial P_{i,j}} = -G_i(1+\ln P_{i,j}) + \lambda_1 + \lambda_2 G_i + \lambda_3 jG_i + \lambda_4 j\varepsilon_i G_i = 0 \quad \text{d'où} \quad \ln P_{i,j} = \frac{\lambda_1}{G_i} + \lambda_2 + j(\lambda_3 + \lambda_4 \varepsilon_i)$$

Par conséquent :

$$P_{i,j} = a_i \exp(-jb_i) \quad \text{avec} \quad b_i = -\lambda_3 - \lambda_4 \varepsilon_i \quad \text{et} \quad a_i^{-1} = \sum_{j=0}^{m} \exp(-jb_i)$$

On en déduit le nombre moyen de particules d'énergie ε_i :

$$\overline{N_i} = \sum_{j=0}^{m} jP_{i,j} = a_i \sum_{j=0}^{m} j \exp(-jb_i)$$

soit :

$$\overline{N_i} = \frac{\exp(-b_i) + 2\exp(-2b_i) + \cdots + m\exp(-mb_i)}{1 + \exp(-b_i) + \exp(-2b_i) + \cdots + \exp(-mb_i)}$$

Le cas particulier où $m = 1$ correspond à la statistique quantique de Fermi-Dirac, celui où $m = \infty$ à celle de Bose-Einstein.

5.2.3 Distribution de Fermi-Dirac

Si l'on fait $m = 1$ dans l'expression précédente, on obtient :

$$\overline{N_i} = \frac{\exp(-b_i)}{1 + \exp(-b_i)} = \frac{1}{\exp b_i + 1}$$

On exprime b_i en fonction de la température absolue T et du potentiel chimique μ en rappelant les définitions thermodynamique de ces quantités :

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_{N,\dots} = k_B \left(\frac{\partial H \ln 2}{\partial U}\right)_{N,\dots} \quad \text{et} \quad -\frac{\mu}{T} = \left(\frac{\partial S}{\partial N}\right)_{U,\dots} = k_B \left(\frac{\partial H \ln 2}{\partial N}\right)_{U,\dots}$$

Or :

$$H \ln 2 = -\sum_{i=1}^{n} G_i \sum_{j=0}^{m} P_{i,j} \ln P_{i,j} = -\sum_{i=1}^{n} G_i \sum_{j=0}^{m} P_{i,j}(-jb_i + \ln a_i) = -\sum_{i=1}^{n} G_i(\ln a_i - jb_i P_{i,j})$$

soit, en remplaçant b_i par son expression :

$$H\ln 2 = -\sum_{i=1}^{n} G_i j P_{i,j}(\lambda_3 + \lambda_4 \varepsilon_i) - \sum_{i=1}^{n} G_i \ln a_i = -(\lambda_3 N + \lambda_4 U) - \sum_{i=1}^{n} G_i \ln a_i$$

On en déduit :

$$\lambda_4 = -\frac{1}{k_B T} = -\beta \qquad \lambda_3 = \frac{\mu}{k_B T} = \beta\mu \quad \text{et} \quad b_i = \beta(\varepsilon_i - \mu)$$

Finalement :

$$\overline{N_i} = \frac{1}{\exp \beta(\varepsilon_i - \mu) + 1}$$

5.2.4 Distribution de Bose-Einstein

Dans ce cas, $m = \infty$. Les séries au numérateur et au dénominateur convergent dans le cas où $b_i > 0$, ce qui donne :

$$\overline{N_i} = \frac{\exp(-b_i) + 2\exp(-2b_i) + \cdots + m\exp(-mb_i)}{1 + \exp(-b_i) + \exp(-2b_i) + \cdots + \exp(-mb_i)}$$

Comme :

$$1 + x + x^2 + \cdots + x^m = \frac{1}{(1-x)} \quad \text{et} \quad x + 2x^2 + \cdots + mx^m = \frac{x}{(1-x)^2}$$

il vient :

$$\overline{N_i} = \frac{\exp(-b_i)}{1 - \exp(-b_i)} = \frac{1}{\exp b_i - 1}$$

En reliant comme précédemment b_i à la température T et au potentiel chimique μ, on trouve :

$$\overline{N_i} = \frac{1}{\exp[\beta(\varepsilon_i - \mu)] - 1}$$

6 Conclusion

Rappelons les résultats qui nous semblent essentiels.

– L'entropie de Boltzmann et l'entropie de Shannon représentent finalement le même concept exprimé dans des unités physiques différentes.

– La théorie de l'information de Shannon permet d'interpréter l'entropie d'un système thermodynamique comme l'information objectivement manquante, du fait de l'absence de communication.

– Il est possible de déduire de la théorie de l'information les résultats importants de la physique statistique : on cherche pour cela les conditions dans lesquelles l'entropie est maximale, compte tenu des contraintes [10],[11]. C'est dire que la théorie de Shannon offre un cadre très général qui englobe celui de la physique statistique. Finalement, ce cadre est riche et bien moins lourd que celui des ensembles canonique et grand-canonique de Gibbs [6].

La relation entre entropie et information manquante joue un rôle essentiel dans la résolution des problèmes inverses, tels qu'ils se posent en théorie du signal et des images [9]. Il s'agit là de restituer objectivement une information-objet à partir d'une donnée-image affectée par du bruit, connaissant au moins partiellement le processus de formation de cette image. On utilise des méthodes de recherche du maximum de l'entropie définie en théorie de l'information. Aussi, ce concept apparaît-il à certains auteurs comme plus fondamental encore que celui de l'énergie [12].

Bibliographie

[1] Rothstein J, 1951, *Information, measurement and quantum mechanics*, Science, Vol 114, p 171

[2] Diu B, Guthmann C, Lederer D, Roulet, 1989, *Physique statistique*, Hermann, p 8 et p 98

[3] Réfrégier Ph, 1993, *Théorie du signal*, Masson, p 227

[4] Frieden B R, 1983, *Probability, statistical optics and data processing*, Springer-Verlag, p 279

[5] Shannon C, 1949, *The mathematical theory of communication*, University of Illinois press, Urbana, p 50

[6] Pérez J-Ph, 1997, *Thermodynamique, fondements et applications*, 2^e édition, Masson, p 279

[7] Atlan H, 1992, *L'organisation biologique et la théorie de l'information*, Hermann, p 180

[8] Brillouin L, 1959, *La science et la théorie de l'information*, Masson, Jacques Gabbay, p 147

[9] Kapur J N, 1989, *Maximum-entropy models in science and engineering*, John Wiley, p 146

[10] Landau L et Lifchitz E, 1967, *Physique statistique*, Édition Mir, p 141

[11] Maréchal P et Lannes A, 1997, *Unification of some deterministic and probabilistic methods for the solution of linear inverse problems via the principle of maximum entropy on the mean*, Inverse problems, 13, p 135

[12] Jaynes E T, 1957, *Information theory and statistical mechanics*, Physical Review, 106, 4, p 620.