Définitions statistiques de l'entropie. Applications à la Physique.

Armel Martin & Jean-Yves Tinevez
École Normale Supérieure de Lyon
Avril 2000

Résumé

Depuis plus d'un siècle, la notion d'entropie s"est développée par l'intermédiaire de définitions statistiques, pour s'étendre à un domaine plus vaste que celui de la seule Physique.

On présente dans cet article la première définition de l'entropie statistique, son statut et son utilisation en Physique, et son lien avec la théorie de l'information. On aborde le Principe du maximum d'entropie, sa signification et son application à la Physique. On montre aussi comment il apparaît plusieurs entropies statistiques dans l'étude des phénomènes dépendant du temps, ainsi que la méthode conduisant à la généralisation de la définition informationnelle et à des familles d'entropies généralisées (notamment celles de Rényi). Enfin on signale quelques problèmes posés par le MEP.

Introduction

L'entropie fait son apparition en Physique au début du XIX$^{\text{ème}}$ siècle, fruit de réflexions autour des notions telles que travail et chaleur, motivées par le développement des machines à vapeur. Elle bénéficie dans cette discipline d'un statut particulier, car elle est associée à l'unique principe permettant de définir la flèche du temps, c'est-à-dire de distinguer passé et futur : le second principe de la thermodynamique (l'entropie d'un système isolé est croissante au cours du temps). Jusqu'aux travaux de Boltzmann, l'entropie désignait une grandeur macroscopique, comme toutes les autres grandeurs de la thermodynamique. Ce dernier, en développant les base de la Physique statistique - dont le but est de décrire les propriétés physiques des corps en partant du comportement microscopique de leurs constituants - a été le premier à reconstruire l'entropie sur des bases statistiques. Depuis ses travaux, la croissance de l'entropie d'un système est souvent associés à l'idée de croissance du désordre interne, mais les avancées récentes de la Physique du non-équilibre (systèmes ouverts et structures dissipatives, voire à ce sujet les travaux de Prigogine) ont permis de mieux comprendre dans quelle mesure la complexification progressive du monde qui nous entoure n'est pas en désaccord avec le second principe et l'idée d'un désordre croissant (on parle de dégradation progressive des formes d'énergie).

Le développement de la théorie de l'information au milieu du XX$^{\text{ème}}$ siècle a donné à l'entropie un statut nettement plus général (qui devient un outil utilisable dans tout domaine utilisant la théorie des probabilités) et en apportant une méthode de traitement statistique très puissante : le principe du maximum d'entropie, qui lui aussi déborde nettement le champ d'application de la Physique.

Compte-tenu de l'aspect fondamental et général de ces notions, nous avons souhaité dans une première partie mettre au clair la définition de l'entropie statistique et le principe du maximum d'entropie ; ces notions constituent non seulement le fondement de le Physique statistique, mais intéressent comme on le verra un public plus large que les seuls physiciens. On montrera également comment l'ouverture de la Physique statistique aux problèmes dépendant du temps conduit à la multiplication des entropies statistiques.

La seconde partie - basée sur l'approche axiomatique de la définition de l'entropie en théorie de l'information - s'attache à présenter la démarche qui conduit à généraliser la première définition de l'entropie statistique, dite de Shannon-Gibbs, pour aboutir à des familles d'entropie paramétrées.

Enfin la dernière partie se propose de pousser plus loin l'analyse des problèmes liés au principe de maximisation d'entropie en présentant trois points de discussion qui nous ont paru importants.

I Entropie de Shannon, Principe du maximum d'entropie et Mécanique Statistique «standard»

1 L'information de Shannon

L'entropie dite de Shannon est liée à ses travaux sur le traitement du signal et la théorie de l'information [1, 2]. C'est pourquoi sa construction repose sur des notions de codage, utilisées dans le traitement numérique du signal (comme le codage binaire utilisé en informatique). L'origine de la notion d'information peut être comprise en considérant le problème du stockage des nombres en binaires sur une mémoire électronique.

Supposons que l'on ait un ensemble de A bits : $\{i_1,i_2,...,i_\mathrm{A}\}$ avec $i_j \in \{0,1\}$. Cet ensemble ordonné peut prendre $\mathrm{N} = 2^\mathrm{A}$ états différents, et on peut donc l'utiliser pour représenter et mémoriser par exemple N nombres. Dans le cas où ces nombres sont les N premiers entiers 0,1,...,N − 1, on peut les exprimer en base deux :

$$m = \sum_{k=0}^{\mathrm{A}-1} s_k \, 2^k$$

s_k est désigné $k^{\text{ème}}$ bit ; le bit apparaît comme unité élémentaire d'information. Donc pour coder l'ensemble des entiers inférieurs à N, il faut d'après les relations écrites plus haut $\mathrm{A} = \frac{\ln \mathrm{N}}{\ln 2} = \log_2(\mathrm{N})$ bits.

Intéressons nous maintenant à la probabilité d'un événement i appartenant à un ensemble d'événements. Dans le contexte du traitement numérique du signal, tous les nombres sont décimaux, donc représentables par une fraction, et l'on peut écrire p_i la probabilité de l'événement i comme le rapport de deux entiers N_i et N :

$$p_i = \mathrm{N}_i/\mathrm{N} \quad \text{avec} \quad \sum_i \mathrm{N}_i = \mathrm{N}$$

Cela correspond à l'interprétation fréquentielle de la probabilité, selon laquelle si l'on réalise N expériences, l'événement i aura lieu N_i fois.

Faisons maintenant le lien entre probabilité et information. Considérons un grand ensemble de N événements «élémentaires», tous équiprobables. On réalise une partition de cet ensemble, chaque sous-ensemble contenant N_i événements aléatoires. La probabilité qu'un événement appartienne au sous-ensemble i s'écrit comme précédemment $p_i = \mathrm{N}_i/\mathrm{N}$.

Supposons que l'on s'intéresse à un événement particulier parmi le grand ensemble. Comme on l'a vu précédemment, $\log_2 \mathrm{N}$ bits sont nécessaires pour pouvoir le sélectionner. De plus, ce nombre doit être indépendant de la façon dont on choisit de ranger les éléments ; notamment, le nombre de bits doit être le même si l'on choisit directement l'événement dans le grand ensemble où si l'on choisit d'abord un sous ensemble puis un événement du sous ensemble. Si l'on note b_i le nombre de bits nécessaire pour choisir l'événement dans le sous-ensemble i, il vient :

$$b_i + \ln \mathrm{N}_i = \ln \mathrm{N}$$

où $\ln \mathrm{N}_i$ est le nombre de bits nécessaires pour choisir un événement particulier parmi N_i.

d'où $$b_i = -\ln p_i$$

On définit l'information comme étant une mesure de la connaissance qu'un observateur acquiert lorsqu'un événement se produit, connaissant la probabilité de celui-ci et rien d'autre. On prendra l'information globale associée à une distribution de probabilité comme étant égale à la moyenne arithmétique de b_i :

$$\mathrm{I}(p) = -\sum p_i \, \ln p_i$$

b_i apparaît donc comme une information élémentaire, telle que celle-ci soit une fonction décroissante de p_i ; un événement qui se produit rarement (ou qui a peu de chance de se produire) apporte plus d'information à l'observateur qu'un événement plus fréquent. De plus l'information obtenue lorsque deux événements indépendants se produisent est la somme des deux informations des événements :

si $$p(ij) = p(i)\,p(j)$$

alors $$\mathrm{I}(ij) = \mathrm{I}(i) + \mathrm{I}(j)$$

Cette information moyenne est appelée *information* ou *entropie de Shannon*, notée $\mathrm{S_I}$. Notons qu'elle recouvre un concept différent de celle utilisée en thermodynamique classique.

2 Théorie de l'information et Mécanique statistique

2.1 Le principe du maximum d'entropie

Dans tout problème utilisant la théorie des probabilités, la première tâche est d'assigner une distribution de probabilité à l'ensemble d'événements considéré, et ceci malgré notre méconnaissance de ses caractéristiques. Supposons comme Jaynes [3] que l'on veuille attribuer à un dé pipé une probabilité pour chacune de ses faces, connaissant seulement la moyenne de ses jets ; ici après un grand nombre de tirages, on a obtenu $4,5$ (s'il n'était pas pipé, la moyenne serait $3,5$). On a donc deux conditions sur les six p_i que l'on va devoir choisir : $\sum p_i = 1$ et $\sum i \, p_i = 4,5$.

On pourrait prendre $p_4 = p_5 = 1/2$ et toutes les autres nulles, où bien même $(\forall i \in [1,5])$ $p_i = 1/10$ et $p_6 = 1/2$. Bien qu'elles satisfassent les exigences mathématiques données plus haut, notre bon-sens commun nous dit que ces deux distributions sont hautement non raisonnables. La première rend totalement improbable la plupart des faces et la deuxième fait jouer à la face 6 un rôle trop privilégié. Mais si l'on devait choisir l'une des deux, sans doute la seconde remporterait-elle notre adhésion car - moins biaisée que la première - elle fait jouer un rôle même minime à toutes les faces.

Ce qui a éveillé notre méfiance face à ces distributions est probablement le fait qu'elles donnent des valeurs si «discontinues» d'une face à l'autre, alors que l'on sait que pour obtenir la moyenne de $4,5$ on a fait un grand nombre de jets de dé, au cours desquels il est peu probable qu'une des faces soit aussi dépareillée ; on trouverait sans doute une distribution acceptable en cherchant celle qui donne les valeurs les plus continues et étalées possibles, afin d'obtenir la répartition la plus équitable. On a donc rajouté intuitivement un critère d'«étalement [1]» qui nous permet d'éliminer les distributions les moins raisonnables. On a donc besoin de construire un outils de mesure de cet étalement, afin de le maximiser par la suite. Dans son article fondateur, Jaynes a choisi pour cette mesure la fonctionnelle d'entropie de Shannon ($\mathrm{S_I} = -\sum p_i \, \ln p_i$) et admet que celle-ci s'impose. En effet elle présente la particularité d'être d'autant plus grande que le manque d'information sur l'ensemble des événement est grand ; c'est-à-dire que la distribution est étalée. On verra néanmoins que ce choix n'est pas le seul possible. Le «lissage»

1. «smoothness» et «spread» dans l'article de Shannon.

obtenu est une forme de critère d'objectivité : on a aucune raison *a priori* de préférer un état à un autre quand nos connaissances sur ceux ci sont nulles. C'est cette méthode, basée sur la théorie de l'information, que l'on appelle «principe du maximum d'entropie» (MEP).

2.2 Formalisation du MEP

La formalisation de ce principe se fait très simplement ; on suppose que l'on a une variable aléatoire x qui peut pendre les valeurs $(x_1,x_2,...,x_n)$ (avec n fini ou infini). Le problème est d'assigner à chaque état x_i une probabilité en connaissant seulement la valeur moyenne de m ($m < n$) fonctions f_k de x :

$$\sum_{i=1}^{n} p_i\, f_k(x_i) = < f_k > = \mathrm{F}_k \quad k = 1..m$$

en maximisant l'entropie donnée par la théorie de l'information : $\mathrm{S_I} = -\sum p_i \ln p_i$. Cela suppose d'une part que les contraintes peuvent s'écrire sous cette forme (valeur moyenne de $f_k = \mathrm{F}_k$), et plus généralement qu'elles puissent s'écrire à partir des probabilités, ce qui n'est pas toujours évident (voir à ce sujet [8]).

La méthode des multiplicateurs de Lagrange (maximisation sous contrainte) donne immédiatement la solution du problème ; on montre de plus qu'il s'agit bien d'un maximum absolu, donné par la distribution canonique généralisée, dite de Shannon-Gibbs :

$$p_i = \frac{1}{\mathrm{Z}} e^{-(\lambda_1 f_1(x_i) + \ldots + \lambda_m f_m(x_i))}$$

avec [2] $\mathrm{Z} = \sum \exp(-\lambda_1 f_1(x_i) + \ldots + \lambda_m f_m(x_i))$ et $\lambda_1,...,\lambda_m$ les multiplicateurs de Lagrange. Notamment s'il existe une seule contrainte sur l'énergie, définie par $< \mathrm{E} > = \sum p_i\, {x_i}^2 < +\infty$, on retrouve une distribution gaussienne (en l'occurrence celle qui exprime la statistique de Boltzmann), ce qui explique son omniprésence dans de nombreux domaines statistiques.

Notons que toutes les considérations développées jusqu'ici ne font pas intervenir intervenir d'argument physique ; elles pourront donc être appliquées sans aucune restriction à d'autres domaines (comme l'ingénierie, le traitement du signal, de l'image ou l'économie). Enfin, insistons sur deux points importants. D'abord, les résultats d'inférence que l'on peut obtenir à partir de cette distribution sont par construction optimales ; c'est-à-dire que si l'on voulait les rendre meilleurs, il faudrait ajouter plus d'information a priori, ce qui se traduirait par une contrainte supplémentaire. D'autre part, dans le cadre de cette approche, la distribution solution de ce problème ne représente en aucun cas la «vraie» distribution, mais la distribution la plus objective, celle qui a le plus de chance de bien décrire le système. En fait l'expression «vraie distribution» n'a aucun sens dans cette démarche, puisque c'est nous qui assignons une probabilité.

2.3 Lien avec l'approche ensembliste

Dans certaines situations, il est sensé de considérer la variable aléatoire x comme étant le résultat de la mesure d'une expérience pouvant être reprise de nombreuses fois. Une interprétation plus objective de ce formalisme est alors possible, et a été développée en grande partie par Boltzmann [4].

On suppose donc que N expériences imaginaires sont réalisées de manière identique. Chaque événement x_i, $i = 1...n$ se réalise N_i fois ; on peut lui assigner une fréquence relative $g_i = \frac{\mathrm{N}_i}{\mathrm{N}}$. On cherche à obtenir une estimation la plus précise possible de $p_i = \lim_{\mathrm{N}\to\infty} g_i$ que l'on peut interpréter comme étant la probabilité de l'événement x_i (interprétation fréquentielle de la probabilité). Comme précédemment, la seule information que l'on ait sur le système se réduit à la valeur moyenne sur les N expériences de certaines fonctions f_k :

$$\sum_i \frac{\mathrm{N}_i}{\mathrm{N}} f_k(x_i) = < f_k > = \mathrm{F}_k \quad k = 1,...,m$$

À une partition $\{\mathrm{N}_i\}$ donnée de N - c'est-à-dire à une distribution $\{g_i\}$ - correspond W résultats possibles de l'ensemble des N expériences :

$$\mathrm{W} = \frac{\mathrm{N}!}{\mathrm{N}_1!...\mathrm{N}_n!} = \frac{\mathrm{N}!}{(\mathrm{N}g_1)!...(\mathrm{N}g_n)!}$$

Le critère d'estimation de la distribution $\{p_i\}$ sera le suivant : quand N devient grand, le choix le plus raisonnable de la distribution $\{g_i\}$ sera celui qui a le plus de chances d'arriver, c'est-à-dire celui qui maximise W. Or on montre à l'aide de la formule de Stirling que :

$$\lim_{\mathrm{N}\to\infty} \frac{1}{\mathrm{N}} \log \mathrm{W} = -\sum_i p_i\, \ln p_i$$

On aboutit donc exactement au même problème mathématique que celui de la formalisation du MEP : la maximisation sous contrainte de la fonctionnelle $-\sum p_i \ln p_i$, qui se trouve être l'entropie de Shannon-Gibbs.

On dispose donc d'une nouvelle interprétation pour le formalisme du MEP, que l'on peut exprimer ainsi : *La probabilité p_i assignée par la théorie de l'information à l'événement x_i lors d'une expérience unique, est numériquement égale à une estimation de la fréquence relative g_i de cet événement lors d'un nombre infiniment grand d'expériences, estimation obtenue en énumérant tous les états compatibles avec les informations en notre possession et en choisissant la situation qui se produit le plus souvent.*

Même si cette interprétation peut paraître plus solide que l'interprétation informationnelle, elle n'est en fait pas exempte de critiques : en pratique cette répétition infinie, ou finie avec N grand, d'expériences identiques est très problématique voire impossible. Il faut alors postuler la propriété d'ergodicité pour pouvoir se ramener à l'étude d'une seule expérience. D'autre part le statut de la distribution obtenue est toujours le même : le critère de choix est la plausibilité maximale (W maximal) ; on obtient la distribution la plus raisonnable et non la «vraie».

2. Il s'agit de la fonction de partition, qu'on retrouve en thermodynamique statistique.

2.4 Application à la mécanique statistique

Le domaine d'application qui intéresse le plus Jaynes est la Physique. Le MEP est un outil efficace et puissant pour l'étude de l'équilibre thermodynamique d'un système physique soumis à certaines contraintes. D'autre part, son formalisme présente l'avantage d'être généralisable naturellement aux situations dépendant du temps en prenant notamment des multiplicateurs de Lagrange et des valeurs moyennes fonctions du temps et de l'espace.

Mais à ce stade, il nous faut préciser une notion importante. L'entropie informationnelle $\mathrm{S_I}$ construite jusqu'ici, même ramenée aux unités appropriées, n'a rien de physique : c'est un pur objet mathématique. Par contre, l'entropie thermodynamique S_e - introduite dans la thermodynamique classique comme une fonction d'état vérifiant le second principe, traduisant les échanges de chaleur et sur laquelle est basée la définition de la température - a un véritable sens physique dans la mesure où elle est accessible par l'expérience. Nous n'avons donc toujours pas fait le lien théorique entre $k_\mathrm{B}\mathrm{S_I}$ et S_e, c'est-à-dire entre la théorie de l'information et la thermodynamique. Pour faire ce lien, nous devons ajouter au MEP appliqué à la Physique le postulat suivant :

Hypothèse physique fondamentale :
Les propriétés d'un système à l'équilibre thermodynamique que l'on mesure expérimentalement sont en accord avec les résultats calculés à l'aide du MEP en théorie de l'information, en se plaçant dans un ensemble approprié (canonique, grand-canonique, T-P, ...).

ce qui revient à admettre qu'*à l'équilibre* les entropies thermodynamique et informationnelle coïncident :

$$\mathrm{S}_e = k_\mathrm{B}\,\mathrm{S_I}$$

Seule l'expérience est apte à justifier la pertinence de ce postulat, qui, jusqu'ici, s'est très souvent avéré juste au cours de ce siècle.

3 Cohérence du MEP dans son application à la Physique

Dans le cadre de son application à la Physique, le MEP pose un problème lors du calcul des distributions liées à un ensemble de Gibbs particulier (canonique, grand-canonique, T-P, ...). Tel qu'il est présenté jusqu'ici, c'est-à-dire selon Jaynes, le MEP constitue le postulat fondamental de la mécanique statistique, sous sa formulation dite *généralisée.* Bernard Diu [5, 6], dans son étude du problème en question, lui oppose la formulation dite *traditionnelle* basée sur l'ensemble microcanonique : *tous les états d'un système d'énergie* E *isolé sont équiprobables.* Dans cette seconde présentation, l'entropie statistique est définie comme celle de Shannon, mais on ne fait pas appel à la maximisation sous contrainte. On construit en fait les distributions standards des ensembles de Gibbs en introduisant la notion de «réservoir» (de chaleur, de particules, de volume, ...) qui permet à chaque fois de se ramener à l'ensemble microcanonique par l'intermédiaire d'un développement limité. Mais même cette approche reste tout aussi critiquable, comme on se propose de le montrer.

3.1 Difficultés dans la dérivation des ensembles canoniques

Comme on l'a vu, le MEP a ceci de particulier qu'il ne fait référence à aucun moment à un quelconque système extérieur - l'univers par exemple - qui englobe celui que l'on traite. Toutes les informations utilisées lors de l'étude proviennent exclusivement de l'ensemble auquel on s'intéresse alors qu'à l'opposé, lorsque l'on veut traiter l'ensemble canonique avec la formulation traditionnelle, on doit se référer à un ensemble dit «réservoir», suffisamment grand devant l'ensemble étudié, qui fixe la valeur moyenne de l'énergie. De plus, c'est ce réservoir qui permet la définition de la température.

Par exemple, dans le cas de l'ensemble canonique, on fixe la valeur moyenne d'une grandeur physique précise, ici l'énergie, choisie de manière assez arbitraire. Pourquoi ne pas considérer que l'on fixe la valeur moyenne de E^2 ? Mais alors on ne trouve plus la même distribution des probabilités, ce qui est inacceptable. Il faut donc considérer arbitrairement que l'on doit fixer la valeur moyenne de l'énergie si l'on veut déduire l'ensemble canonique de la formulation généralisée ; alors que la formulation traditionnelle, en introduisant la notion de réservoir, apporte une signification physique à ce choix.

D'autre part, si l'on applique la formulation généralisée à deux systèmes $\mathcal{S}$ et $\mathcal{S}'$ de même taille et en équilibre thermique, on trouve un résultat différent selon qu'on l'applique à $\mathcal{S}$ où $\mathcal{S}+\mathcal{S}'$ (c'est-à-dire que l'on fixe l'énergie de $\mathcal{S}$ ou $\mathcal{S}'$). Cette formulation a donc le défaut, comme le note l'auteur, d'être ambiguë et de manquer de cohérence. Mais il remarque également que la formulation traditionnelle a ses limites, comme on va le voir.

Celle-ci traite l'ensemble canonique en considérant que le système que l'on veut traiter (N particules) est en contact thermique avec un réservoir de taille très grande devant la sienne (N_r particules) qui fixe la valeur moyenne de l'énergie. La réunion des deux systèmes est isolée, et donc, selon le postulat de base de cette formulation traditionnelle, tous ses états sont équiprobables. De là on déduit la probabilité canonique pour un état l d'énergie E_l :

$$\mathrm{P}_l = \frac{1}{\mathrm{Z}}\, e^{-\frac{\mathrm{E}_l}{k\mathrm{T}_r}}$$

où Z est la fonction de partition de l'ensemble canonique $\mathrm{Z} = \sum_l \exp(-\frac{\mathrm{E}_l}{k\mathrm{T}_r})$ et T_r la température thermodynamique du réservoir.

Pour arriver à ce résultat, il a fallut faire deux approximations, justifiées par la taille du réservoir :

- $\mathrm{N} \ll \mathrm{N}_r$ pour permettre le tronquage du développement limité de l'entropie totale ;
- $\mathrm{N}^2 \ll \mathrm{N}_r$ pour permettre l'exponentiation de la relation qui en est déduite.

Si la première inégalité est tout à fait vérifiable puisque c'est ainsi que l'on a défini le réservoir, la seconde est irréalisable dès que N est de l'ordre du nombre d'Avogadro, tant le réservoir devrait être grand (au moins 10^{23} moles).

Les deux formulations butent donc toutes deux sur des difficultés lorsqu'il s'agit de traiter avec rigueur et honnêteté l'ensemble canonique.

3.2 Nouvelle formulation du postulat fondamental, proposée par Bernard Diu

L'auteur propose alors une nouvelle formulation du postulat fondamental, qui permet de s'affranchir de l'arbitraire de la formulation généralisée et des problèmes de conditions de la formulation traditionnelle :

Pour un système isolé, de taille macroscopique, en équilibre, la distribution de probabilité des états microscopiques est celle qui maximise l'entropie statistique, sous les contraintes imposées par l'extérieur.

Cette formulation se réduit aux systèmes macroscopiques, pour lesquels existaient des conditions d'applications irréalisables dans le cas de la formulation traditionnelle. Elle s'applique à des systèmes à l'équilibre et isolés seulement, pour éviter les contradictions de la formulation généralisée. La principale différence avec le postulat généralisé, formulé par Jaynes, consiste donc à faire référence explicitement au milieu extérieur, c'est-à-dire à un réservoir qui fixe la contrainte sous des conditions physiques de validité désormais connues, et qui permet notamment de définir la grandeur intensive associée au coefficient de Lagrange (température, potentiel chimique, pression, ...). On notera que le vide dénué de rayonnement ne fixe aucune contrainte ; cela correspond au cas microcanonique d'un système isolé, dont l'énergie est fixée comme un paramètre et n'est pas sensée fluctuer.

On peut montrer d'où ce postulat découle en traitant le cas canonique. Soit un système $\mathcal{S}$ qui échange de l'énergie avec un réservoir $\mathcal{R}$ beaucoup plus grand que lui, l'ensemble étant isolé. La probabilité que $\mathcal{S}+\mathcal{R}$ soit dans l'état $\ell+\mathrm{L}$ s'écrit $\mathrm{P}_{\ell,\mathrm{L}} = \mathrm{P}_\ell\, \mathrm{P}'^{(\ell)}_{\mathrm{L}}$, avec P_ℓ la probabilité que $\mathcal{S}$ soit dans l'état ℓ et $\mathrm{P}'^{(\ell)}_{\mathrm{L}}$ la probabilité conditionnelle que $\mathcal{R}$ soit dans l'état L sachant que $\mathcal{S}$ est dans l'état ℓ.

Nous voulons obtenir P_ℓ en maximisant l'entropie totale :

$$\mathrm{S}_t = -k \sum_{\ell,\mathrm{L}} \mathrm{P}_{\ell,\mathrm{L}} \,\ln \mathrm{P}_{\ell,\mathrm{L}}$$

sachant que $\mathcal{S}+\mathcal{R}$ est isolé, et que donc les seules contraintes a vérifier sont les suivantes :

$$\sum_\ell \mathrm{P}_\ell = 1 \quad \text{et} \quad \sum_{\mathrm{L}} \mathrm{P}'^{(\ell)}_{\mathrm{L}} = 1 \;\forall \ell$$

On déduit immédiatement de ces deux expressions une nouvelle écriture de l'entropie totale :

$$\mathrm{S}_t = \mathrm{S} + \sum_\ell \mathrm{P}_\ell\, \mathrm{S}_r^{(\ell)}$$

où $\mathrm{S} = -k\sum \mathrm{P}_\ell \ln \mathrm{P}_\ell$ est l'entropie statistique de $\mathcal{S}$ (celle que l'on maximisait dans la formulation généralisée) et

$$\mathrm{S}_r^{(\ell)} = -k \sum_{\mathrm{L}} \mathrm{P}'^{(\ell)}_{\mathrm{L}} \,\ln \mathrm{P}'^{(\ell)}_{\mathrm{L}}$$

est l'entropie du réservoir lorsque $\mathcal{S}$ est dans l'état ℓ.

On a besoin maintenant d'utiliser l'hypothèse de la taille du réservoir, et de considérer que E_ℓ l'énergie de $\mathcal{S}$ est beaucoup plus faible que l'énergie totale E_t, d'où :

$$\mathrm{S}_r^{(\ell)} = \mathrm{S}_r(\mathrm{E}_t - \mathrm{E}_\ell) \simeq \mathrm{S}_r(\mathrm{E}_t) - \mathrm{E}_\ell \frac{\partial \mathrm{S}_r}{\partial \mathrm{E}_r} = \mathrm{S}_r(\mathrm{E}_t) - \mathrm{E}_\ell \frac{1}{\mathrm{T}_r}$$

par définition de la température. Les conditions d'application de cette approximation sont raisonnables : il suffit que le réservoir soit tel que $\mathrm{N}_r \gg \mathrm{N}$.

L'entropie totale s'écrit maintenant :

$$\mathrm{S}_t = \mathrm{S} + \mathrm{S}_r(\mathrm{E}_t) - \frac{1}{\mathrm{T}_r} < \mathrm{E}_\ell >$$

où $< \mathrm{E}_\ell > = \sum_\ell \mathrm{P}_\ell \mathrm{E}_\ell$ est la valeur moyenne de l'énergie de $\mathcal{S}$ sur tous ses états. On doit donc maximiser la fonction :

$$\mathrm{S}_t = \mathrm{S}_r(\mathrm{E}_t) - k \sum_\ell \mathrm{P}_\ell \,\ln \mathrm{P}_\ell - \frac{1}{\mathrm{T}_r} \sum_\ell \mathrm{P}_\ell \mathrm{E}_\ell$$

Remarquons que la contrainte sur l'énergie moyenne fixée est apparue «naturellement», et que le «multiplicateur de Lagrange» qui lui est associé apparaît directement comme l'inverse de la température. On s'est affranchi ainsi du problème de l'exponentiation dont les conditions sont irréalisables. La maximisation se fait simplement en utilisant la méthode de Lagrange à l'aide de l'unique contrainte restante ($\sum \mathrm{P}_\ell = 1$) : on introduit la fonction auxiliaire

$$\tilde{\mathrm{S}} = \mathrm{S}_t + a(1 - \sum \mathrm{P}_\ell)$$

et on obtient immédiatement la distribution canonique.

Cette formulation donne donc le résultat attendu tout en étant rigoureuse. Précisons toutefois que selon Jaynes, la distribution obtenue ne doit pas être considérée comme la vraie, mais comme la plus raisonnable, subtilité qui n'apparaît pas dans cette dernière version du postulat fondamental.

4 Application aux régimes dépendant du temps : multiplicité des entropies

Dans le cas où l'on étudie les propriétés dynamiques d'un système physique - c'est-à-dire dans le cadre de la mécanique statistique du non-équilibre - on est confronté à la complexité bien plus grande du problème : on doit tenir compte des évolutions corrélées de toutes les particules. On est alors amené à faire des approximations différentes suivant le régime d'évolution dans lequel on se place, c'est-à-dire selon le temps caractéristique d'évolution locale, on est amené à considérer une entropie adéquate et particulière correspondant à chaque régime. Du fait de cette dépendance, la notion d'entropie statistique devient multiple. Roger Balian [7] présente une synthèse sur cette question.

i) Entropie de Von Neumann $\mathrm{S}(\widehat{\mathrm{D}})$

Lorsqu'on généralise l'entropie de Shannon aux phénomènes quantiques, on travaille avec l'opérateur densité - noté $\widehat{\mathrm{D}}$ - plutôt qu'avec la distribution de probabilité. L'entropie s'exprime alors, si l'on se ramène aux unités de la Physique :

$$\mathrm{S}(\widehat{\mathrm{D}}) = -k_{\mathrm{B}}\; \mathrm{Tr}\,(\widehat{\mathrm{D}} \ln \widehat{\mathrm{D}})$$

L'opérateur densité obéit à l'équation d'évolution suivante, dite de Liouville-Von Neumann et qui régit l'évolution microscopique du système :

$$i\hbar \frac{\mathrm{d}\widehat{\mathrm{D}}}{\mathrm{d}t} = [\widehat{\mathrm{H}},\widehat{\mathrm{D}}]$$

où $\widehat{\mathrm{H}}$ est l'hamiltonien du système.

Cette entropie est basée sur la description quantique exacte de l'évolution des particules, donc elle constitue la formulation statistique la plus complète que l'on puisse faire. Elle tient compte notamment de la corrélation qui existe entre l'instant présent t et l'instant initial t_0. On dit qu'elle conserve l'information initiale, et de ce fait, l'entropie $\mathrm{S}(\widehat{\mathrm{D}})$ reste constante au cours du temps, ce qui exprime non une violation du second principe mais la réversibilité des équations d'évolution microscopiques.

ii) Entropie de Boltzmann $\mathrm{S_B}$

Lorsqu'on s'intéresse à un gaz suffisamment dilué pour pouvoir l'étudier de manière classique, on travaille avec l'espace des phases classique position-impulsion $(\overrightarrow{r},\overrightarrow{p})$. La description la plus précise du gaz est alors donnée par la fonction densité de particules $f(\overrightarrow{r},\overrightarrow{p})$ dans l'espace des phases. L'entropie adéquate est alors l'entropie de Boltzmann :

$$\mathrm{S_B} = \int \mathrm{d}^3\overrightarrow{r}\, \mathrm{d}^3\overrightarrow{p}\; f(\overrightarrow{r},\overrightarrow{p})\,\{1 - \ln[h^3 f(\overrightarrow{r},\overrightarrow{p})]\}$$

Cette description est notamment plus précise que la description thermodynamique (cf iii) qui consisterait à faire une moyenne macroscopique des grandeurs physiques utiles. On peut en effet grâce à cette formulation décrire des phénomènes rapides (par exemple les *régimes balistiques*) tels que le temps caractéristique d'évolution de $f(\overrightarrow{r},\overrightarrow{p})$ est plus petit que le temps de thermalisation (durée entre deux collisions), ce qui correspond à des cas où les grandeurs thermodynamiques ne sont pas définies.

Cependant, cette description du système est incomplète comparée à la précédente car elle néglige les corrélations liées aux collisions précédentes. C'est d'ailleurs grâce à cette approximation que l'on peut écrire que $f(\overrightarrow{r},\overrightarrow{p})$ évolue selon *l'équation semi-phénoménologique de Boltzmann*. L'application du théorème [3] **H** montre alors que $\mathrm{S_B}$ est une fonction croissante du temps, ce qui exprime la perte d'information lors des collisions successives.

iii) Entropie thermodynamique $\mathrm{S_{th}}$

Malgré son appellation, cette entropie est toujours une entropie statistique et ne doit pas être confondue avec l'entropie thermodynamique, notée S_e et objet du 2.4. Cette entropie est utilisée lorsqu'on se place dans le *régime thermodynamique*, c'est-à-dire lorsque le système est assez proche de l'équilibre pour que l'on puisse le subdiviser en petit sous-systèmes qui, eux, sont à l'équilibre thermodynamique (on parle d'«équilibre thermodynamique local») et ainsi définir des grandeurs physiques locales. Cette description correspond au cas où ces dernières varient lentement par rapport au temps caractéristique des phénomènes microscopiques (par exemple la durée entre les collisions dans un gaz), qui est également le temps au bout duquel est perdue la mémoire d'une perturbation microscopique initiale.

Cette entropie est encore plus incomplète que la précédente (on fait des moyennes locales). On montre que c'est aussi une fonction croissante du temps, ce qui s'interprète cette fois par la dissipation due à la viscosité et à la conduction thermique.

iv) Entropie thermostatique $\mathrm{S_{eq}}$

Lorsque le système est à l'équilibre thermodynamique global, c'est-à-dire au bout d'un temps bien plus long que celui nécessaire pour atteindre l'équilibre local, on peut travailler avec $\mathrm{S_{eq}}$, qui n'est autre que $\mathrm{S_{th}}$ pour des contraintes (valeurs moyennes des grandeurs utiles) indépendantes du temps et de l'espace, et donc définies pour le système global. Comme elle ne dépend pas du temps, cette entropie est constante.

Les quatre descriptions précédentes étant de plus en plus incomplètes (c'est-à-dire décrivent le système de manière de plus en plus approximative), les entropies correspondantes sont de plus en plus grandes :

$$\mathrm{S}(\widehat{\mathrm{D}}) < \mathrm{S_B} < \mathrm{S_{th}} < \mathrm{S_{eq}}$$

En effet, on perd de l'information lorsque l'on passe d'une entropie à l'autre dans cet ordre. $\mathrm{S}(\widehat{\mathrm{D}})$ et $\mathrm{S_{eq}}$ sont les deux bornes entre lesquelles les fonctions du temps $\mathrm{S_B}(t)$ et $\mathrm{S_{th}}(t)$ évoluent pour tendre vers $\mathrm{S_{eq}}$. Si $\mathrm{S_B}$ ne décroît jamais, rien ne force cependant $\mathrm{S_{th}}$ à être croissante tant que l'équilibre local n'est pas atteint : on n'est pas encore dans son domaine d'application, la description thermodynamique est alors inadéquate. De même, l'utilisation de $\mathrm{S_B}$ est soumise à la condition de dilution du gaz ; les effets de mémoire sont alors absents et le théorème **H** est valable.

Dans des cas où les effets de mémoire ne sont pas négligeables (par exemple pour les systèmes de spins à écho cf [7]), il faut construire des entropies intermédiaires entre $\mathrm{S}(\widehat{\mathrm{D}})$ et $\mathrm{S_B}$ qui tiennent compte d'un certain nombre de corrélations.

On remarquera toutefois que la croissance de $\mathrm{S_B}(t)$ et $\mathrm{S_{th}}(t)$ ne signifie pas que l'on a démontré l'irréversibilité des phénomènes naturels, exprimé seulement par le second principe. L'irréversibilité ne saurait en effet être une conséquence d'approximations dans les équations du mouvement (pour plus de précisions, voire à ce sujet *Théorie du chaos* de Prigogine).

II Retour sur la définition informationnelle : vers d'autres entropies

Dans son article initial [3], Jaynes base la construction de son principe du maximum d'entropie sur l'hypothèse (qui lui semble tout à fait justifiée) selon laquelle la seule mesure possible de la quantité d'information ou d'incertitude (qu'il utilise comme mesure de l'«étalement» de la distribution) est celle présentée par Shannon : $-\sum p_i \ln p_i$. Depuis, beaucoup ont essayé de montrer que le choix de cette fonctionnelle d'entropie découlait de manière nécessaire de propriétés devant

3. Qu'on abordera pas ici.

être «raisonnablement» vérifiées, propriétés présentées comme s'imposant pour des raisons de cohérence. C'est là qu'intervient l'axiomatisation de la construction de la fonctionnelle d'entropie.

En explicitant les nombreux problèmes posés par le MEP, Jos Uffink [8] montre que les tentatives de justification de l'utilisation de la seule entropie de Shannon-Gibbs, sur des arguments de cohérences, ont échoué. Il évoque notamment la possibilité d'élargir le choix de la fonctionnelle d'entropie. En remettant en cause la démarche présentée au I.1, on peut poser deux questions :

- n'y-a-t'il pas d'autres manières de mesurer la quantité élémentaire d'information associée à une probabilité p_i que par la fonctionnelle $-\log_2 p_i$?
- dans l'hypothèse où l'on admet $\log_2 p_i$ comme quantité d'information élémentaire, n'est-il pas possible de calculer l'information moyenne autrement que par une moyenne arithmétique ?

La construction de l'entropie de Shannon repose donc sur une série d'axiomes, qu'on se propose maintenant d'étudier afin de voir ce qui peut y être modifié.

1 Déduction axiomatique de l'information - Entropie de Shannon

Il existe plusieurs ensembles d'axiomes qui permettent d'obtenir la fonctionnelle de Shannon. Le plus simple (c'est-à-dire le plus compact) fut donnée en 1957 par Fadeev :

Si $\mathrm{S}_n(p_1,...,p_n)$ *est la fonctionnelle d'entropie associée à la distribution discrète de probabilité* $\{p_1,...,p_n ; \sum p_i = 1\}$, *et si les axiomes suivants sont vérifiés :*

1 : Symétrie. $\mathrm{S}_n(p_1,...,p_n)$ *est invariante par permutation des* p_i.

2 : Continuité. $\mathrm{S}_2(p,1-p)$ *est une fonction continue de p pour* $0 \leqslant p \leqslant 1$.

3 : Récursivité. Pour $n \geqslant 2$,

$$\mathrm{S}_n(p_1,...,p_n) = \mathrm{S}_{n-1}(p_1+p_2,p_3,...,p_n) + (p_1+p_2)\,\mathrm{S}_2\left(\tfrac{p_1}{p_1+p_2},\tfrac{p_2}{p_1+p_2}\right)$$

ce qu'on peut écrire aussi avec $t \in [0,1]$:

$$\mathrm{S}_{n+1}(t\,p_1,(1-t)p_1,p_2,...,p_n) = \mathrm{S}_n(p_1,...,p_n) + p_1\,\mathrm{S}_2(t,1-t)$$

alors S_n *s'écrit nécessairement :*

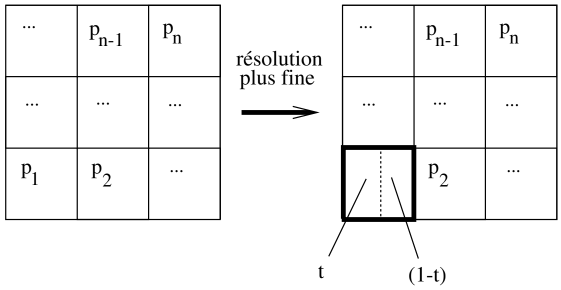
$$\mathrm{S}_n = -k \sum p_i \log_2 p_i$$

avec $k > 0$.

L'ajout d'un quatrième axiome permet de fixer la valeur de k à 1 :

4 : $\mathrm{S}_2(\frac{1}{2},\frac{1}{2}) = 1$

Ces axiomes paraissent effectivement raisonnables, compte-tenu de ce que l'on peut attendre d'une mesure de l'information. En particulier, l'axiome de récursivité (3) signifie que lorsque l'on change d'échelle - c'est-à-dire lorsque l'on prend un maillage de l'espace des événements plus fin -, l'information globale augmente de la moyenne arithmétique des informations contenues dans les mailles subdivisées (cf schéma).



Cependant, comme le montre Jos Uffink, on a souvent déclaré que ces axiomes s'imposaient par nécessité, par cohérence, alors qu'ils n'étaient pas absolument uniques. En effet, le troisième axiome, s'il est naturel et intuitif, n'en est pas moins quelque peu arbitraire et conventionnel, dans la mesure où il impose de travailler avec une moyenne arithmétique lors d'un changement de résolution de l'analyse du système.

Rényi [9] a montré comment on pouvait modifier cet ensemble d'axiomes, de manière à y introduire d'autres opérateurs de moyenne, aboutissant ainsi à une généralisation de l'entropie de Shannon.

2 Distributions incomplètes et famille des entropies généralisées de Rényi

2.1 Construction et dérivation axiomatique

Rényi proposa donc de remplacer l'axiome de récursivité (3) par un axiome d'additivité, beaucoup plus fondamental :

3_{R} *: Additivité. Soient* $\mathrm{P} = (p_1,...,p_n)$ *et* $\mathrm{Q} = (q_1,...,q_m)$ *deux distributions de probabilité associées à des systèmes ou des expériences indépendantes. On note* $\mathrm{P}\star\mathrm{Q} = \{p_i.q_j\}$ *la distribution associée à l'ensemble. Alors l'information totale est la somme des informations :*

$$\mathrm{S}(\mathrm{P}\star\mathrm{Q}) = \mathrm{S}(\mathrm{P}) + \mathrm{S}(\mathrm{Q})$$

Cet axiome est cependant moins fort que celui qu'il remplace, et il est nécessaire d'en introduire un cinquième pour caractériser la fonctionnelle. C'est cet axiome supplémentaire qui permet d'aboutir à d'autres entropies : il précise comment se construit l'information à partir de moyennes successives, celles-ci n'étant pas nécessairement arithmétiques.

Pour l'introduire, Rényi commence par généraliser la notion de distribution de probabilité en définissant ce qu'il appelle des distributions «incomplètes». Ces distributions généralisées sont en tout point comparables à celles définies dans la théorie des probabilités, à la différence près qu'elle ne sont pas normées : $0 < \sum p_i = \mathrm{W(P)} \leqslant 1$. Le «poids» $\mathrm{W(P)}$ de la distribution peut donc être strictement inférieur à 1, ce qui peut s'interpréter en disant que les événements possibles ne sont pas tous observables (argument qui peut être compris dans le cadre de la signification fréquentielle de la probabilité).

On peut alors associer deux distributions généralisées P et Q au même système par concaténation à condition que la distribution finale soit définie $(\mathrm{W(P)} + \mathrm{W(Q)} \leqslant 1)$:

$$\mathrm{P} \cup \mathrm{Q} = (p_1,...,p_n,q_1,...,q_m)$$

D'autre part la forme généralisée de la valeur moyenne d'un ensemble $\{x_1,...,x_n\}$, avec les poids associés $\{\mathrm{W}_1,...\mathrm{W}_2\}$, peut être écrite d'après Kolmogorov et Nagumo :

$$\bar{\mathrm{X}} = \phi^{-1}\left[\frac{\sum \mathrm{W}_k\, \phi(x_k)}{\sum \mathrm{W}_k}\right]$$

où ϕ est une fonction arbitraire, devant être strictement monotone et continue.

On étend donc la notion de fonctionnelle d'entropie pour les distributions incomplètes à l'aide des quatres axiomes vus précédemment et du cinquième que l'on peut énoncer comme suit :

5 : Soient P *et* Q *deux distributions généralisées, de poids* W(P) *et* W(Q). *Alors l'information obtenue lorsque l'on complète* P *par* Q *vaut la moyenne généralisée des informations respectives :*

$$\mathrm{S}(\mathrm{P} \cup \mathrm{Q}) = \phi^{-1}\left[\frac{\mathrm{W}(\mathrm{P})\, \phi[\mathrm{S}(\mathrm{P})] + \mathrm{W}(\mathrm{Q})\, \phi[\mathrm{S}(\mathrm{Q})]}{\mathrm{W}(\mathrm{P}) + \mathrm{W}(\mathrm{Q})}\right]$$

La fonctionnelle ainsi caractérisée dépend donc de choix de l'opérateur de moyenne ϕ. Rényi montre que si l'on reformule le problème du MEP en disant que l'on maximise une entropie conditionnelle (un gain d'information) et non seulement une entropie «absolue» (cf III.1), alors le choix de ϕ se restreint à deux possibilités :

1. ϕ est une fonction affine non constante : $\phi(x) = a\,x + b,\ a \neq 0$. Dans ce cas, on vérifie facilement que cela revient à prendre la moyenne arithmétique et mène à l'entropie de Shannon. En effet, en partant d'une information élémentaire $\mathrm{S}(p_i) = -\log p_i$, on peut directement écrire, avec $\mathrm{P}_i = \{p_i\}$:

$$\mathrm{S}(\cup \mathrm{P}_i) = \frac{-\sum p_i \log p_i}{\sum p_i}$$

ce qui est l'entropie de Shannon si $\sum p_i = 1$.

2. ϕ appartient à une famille de fonctions exponentielles du type $\phi_\alpha(x) = 2^{(\alpha-1)x}$ avec[4] $\alpha \geqslant 0$ et $\alpha \neq 1$. On obtient alors une famille continue d'entropies qui s'écrivent :

$$\mathrm{S}_\alpha(\mathrm{P}) = \frac{1}{1-\alpha} \log_2 \left[\frac{\sum p_i^\alpha}{\sum p_i}\right]$$

(*entropie d'ordre α de Rényi*, construite à partir d'une moyenne exponentielle).

2.2 Propriétés

Une propriété fondamentale de cette famille est de contenir l'entropie de Shannon ; en effet à la limite $\alpha \to 1$:

$$\mathrm{S}_1(\mathrm{P}) = \lim_{\alpha \to 1} \mathrm{S}_\alpha(\mathrm{P}) = -\sum p_i \log_2 p_i$$

L'entropie de Shannon n'apparaît donc que comme un cas particulier d'une famille continue de fonctions donnant une mesure de l'information.

On vérifie par exemple que l'on a bien annulation de $\mathrm{S}_\alpha(\mathrm{P})$ pour une distribution «certaine» (c'est-à-dire telle que $\{p_i = 1$ et $p_{j \neq i} = 0\}$), ainsi que la distribution uniforme comme maximum (ensemble microcanonique).

Signalons également que $\mathrm{S}_\alpha(\mathrm{P})$ est une fonction croissante de α :

$$\mathrm{S}_\alpha(\mathrm{P}) \leqslant \mathrm{S}_{\alpha'}(\mathrm{P}) \quad \text{si} \quad \alpha \leqslant \alpha'$$

ce qui permet notamment d'encadrer simplement l'entropie de Shannon. On a de la même manière une autre relation d'ordre, inversée par rapport à la première :

$$\frac{\alpha-1}{\alpha}\mathrm{S}_\alpha(\mathrm{P}) \geqslant \frac{\alpha'-1}{\alpha'}\mathrm{S}_{\alpha'}(\mathrm{P}) \quad \text{si} \quad \alpha \leqslant \alpha'$$

2.3 Applications

On peut a priori reconstruire toute la physique statistique à l'aide de cette famille - et ceci pour chaque valeur de α possible - en appliquant le MEP à la fonction $\mathrm{S}_\alpha(\mathrm{P})$ pour trouver les distributions à l'équilibre pour divers ensembles de Gibbs, qui seront différentes de celles utilisées habituellement. En pratique, les entropies de Rényi ont été effectivement appliquées en mécanique statistique des états d'équilibre (Schlögl [10], 1980), mais aussi en dynamique non-linéaire (Hentschel et Procaccia, 1983). Celles-ci s'avèrent être d'une grande utilité pour l'étude de systèmes dynamiques non-linéaires ayant un comportement chaotique (par exemple dans l'étude de l'attracteur étrange lorsqu'il existe, et dans la classification des systèmes chaotiques expérimentaux [2]).

Il apparaît d'ailleurs une séparation intéressante dans la famille des S_α : on prend $\alpha \leqslant 1$ ($\mathrm{S}_\alpha \geqslant 0$) pour l'étude des états d'équilibre, alors que se sont les fonctions $\mathrm{S}_\alpha(\mathrm{P})$ pour $\alpha > 1$ ($\mathrm{S}_\alpha \leqslant 0$) qui sont utilisées dans la science du chaos, ceci à cause du changement de concavité de ϕ de part et d'autre de la valeur $\alpha = 1$.

La famille des entropies de Rényi ne constitue pas, comme on le verra, la seule manière de généraliser l'entropie de Shannon, et par là même la statistique de Boltzmann-Gibbs. Mais c'est la seule à reposer sur une axiomatique en accord avec les principes usuels (tels que l'addiditivité).

3 Une autre famille : les entropies de Tsallis

Tsallis [11, 12] a postulé de manière ad-hoc une définition possible d'entropie généralisée :

$$\mathrm{S}_q = \frac{1-\sum p_i^q}{q-1} = \frac{\sum (p_i - p_i^q)}{q-1}, \quad q \in \mathbb{R}$$

On retrouve là encore dans cette famille l'entropie de Shannon. En effet :

$$\mathrm{S}_1 = \lim_{q \to 1} \mathrm{S}_q = -\sum p_i \ln p_i$$

4. Rényi a également montré que l'on pouvait définir S_α pour $\alpha < 0$, mais alors S_α ne peut plus être interprétée comme une mesure de l'information.

Même si cette forme d'entropie fut posée a priori sans justification, on peut trouver un fondement axiomatique à cette fonctionnelle, mais d'une manière assez surprenante. En effet l'axiome (3) sur l'additivité de l'information n'est plus vérifié, et pour deux distributions indépendantes P et Q Tsallis pose :

$$3_\mathrm{T} \;:\; \mathrm{S}_q(\mathrm{P} \star \mathrm{Q}) = \mathrm{S}_q(\mathrm{P}) + \mathrm{S}_q(\mathrm{Q}) + (1-q)\,\mathrm{S}_q(\mathrm{P})\,\mathrm{S}_q(\mathrm{Q})\,, \quad q \in \mathbb{R}$$

on abandonne donc l'idée intuitive et naturelle d'additivité de l'information (sauf bien sur pour $q = 1$).

Signalons que ce type d'entropies avait déjà été proposé dix ans auparavant par Havrda et Charvát [13] comme un outil de classification.

À partir de cette définition, on peut déduire les distributions d'équilibre des ensembles de Gibbs, la statistique de Boltzmann-Gibbs étant retrouvée à la limite $q \to 1$. Toute la structure mathématique de la thermodynamique statistique, basée sur la transformée de Legendre, reste invariante lorsque l'on modifie q, ce qui est un argument en faveur de la pertinence de cette fonctionnelle. On perd par contre la notion fondamentale d'additivité (et donc d'extensivité) des grandeurs thermodynamiques habituelles (entropie, énergie, ...). Pour être plus précis, celle-ci s'exprime à présent sous une nouvelle forme. Ce n'est plus la grandeur X en question qui a la propriété d'extensivité, mais la fonction :

$$f(\mathrm{X}) = \frac{\ln(1 + (1-q)\mathrm{X})}{q-1}$$

On ouvre ainsi le champ à une thermodynamique statistique «non-extensive».

Un autre comportement remarquable est que les distributions canoniques pour $q > 1$ s'annulent pour des énergies $\varepsilon_i \geqslant 1/(\beta(q-1))$. Cela correspond donc à des états rigoureusement «gelés», c'est-à-dire inaccessibles pour cause de température trop faible. Mais cela définit de plus un domaine de températures inaccessibles (telles que ε_0 = niveau fondamental $\geqslant 1/(\beta(q-1))$), domaine dont l'extension tend vers 0 quand q tend vers 1, ce qui est cohérent.

Ces entropies trouvent un terrain d'application dans les études des structures fractales (multifractales notamment), de par leur lien avec la notion de dimension généralisée.

III Autres problèmes posés par le MEP

On trouve dans l'article de Jos Uffink [8] un grand nombre de remises en causes, remarques, questionnements théoriques à l'égard du MEP. Trois points nous ont semblé particulièrement importants. On les mettra en évidence avec l'entropie de Shannon-Gibbs, mais cela est généralisable à toutes les autres entropies.

1 Le passage au continu

Nous avons travaillé jusqu'ici avec des distributions discrètes. Lorsqu'on en vient à travailler avec des distributions continues (p fonction de la variable continue x), la question se pose de savoir comment adapter le MEP. Le passage d'une somme discrète à une somme continue dans l'entropie de Shannon (par exemple) en faisant tendre vers 0 le pas de discrétisation pose deux problèmes :

- la limite est infinie ;
- si l'on retranche le terme divergent - ce qui est acceptable dans la mesure où les entropies informationnelles utilisées avec le MEP sont définies à une transformation affine près - on obtient l'entropie «différentielle» :

  $$\mathrm{S}_d(p) = -\int p(x)\,\ln p(x)\;\mathrm{d}x$$

  La densité de probabilité $p(x)$ prend alors pour dimension l'inverse d'une longueur, et S_d n'est plus invariante par changement d'unité ou de paramétrisation.

La solution consiste à reformuler le MEP d'une façon qui implique un changement conceptuel important. En effet, on remplace la notion d'entropie «absolue», utilisée jusqu'ici, par celle d'entropie «relative». On choisit une distribution dite «initiale» $m(x)$, et la distribution $p(x)$ cherchée est obtenue en maximisant sous contraintes l'entropie relative :

$$\mathrm{S}(p,m) = -\int p(x)\;\ln\frac{p(x)}{m(x)}\;\mathrm{d}x$$

qui représente en fait le *gain d'information* que constitue la connaissance de nouvelles contraintes. Par exemple, chercher la distribution canonique consiste à partir d'une distribution uniforme $m(x)$ correspondant à l'absence de contraintes, et à maximiser la gain d'information $\mathrm{S}(p,m)$ que constitue la connaissance de l'énergie moyenne $<\mathrm{E}>= \int p(x)\varepsilon(x)\,\mathrm{d}x$. On pourra appeler cette méthode Principe du maximum d'entropie relative (MREP [8]). Le changement conceptuel réside dans le fait que l'on maximise un gain d'information dû à l'apport de nouvelles contraintes, et non une information absolue. Cela peut être interprété en disant que l'on «met à jour» l'ancienne distribution au vu de l'information supplémentaire acquise. On remarquera que, si le pendant discret de S_d est l'entropie de Shannon, celui de $\mathrm{S}(p,m)$ est la divergence de Kullbak-Leibler :

$$\mathrm{KL}(\mathrm{P},\mathrm{Q}) = -\sum_i p_i\;\ln\frac{p_i}{q_i}$$

2 Problèmes concernant l'assimilation : information manquante - incertitude

L'interprétation informationnelle de la fonctionnelle d'entropie choisie (selon laquelle l'entropie représente l'information moyenne «cachée» dans le système, information que l'on gagne lorsqu'un événement se réalise) n'est pas complètement fiable. En effet, on a coutume de dire que si l'on acquiert de l'information supplémentaire sur le système, son entropie (au sens de son incertitude) décroît, ce que l'on exprime en utilisant l'entropie conditionnelle :

$$S(P/Q) \leqslant S(P)$$

Cela n'est pas toujours exact. Prenons deux exemples tirés de [8] :

- Dans certains cas, la connaissance supplémentaire sur une variable augmente l'incertitude sur une autre. Prenons une population de corbeaux telle que la probabilité qu'un corbeau soit noir vale $0,99$ et blanc $0,01$. On sait aussi que si une mère est blanche, alors les probabilités que son petit soit noir ou blanc valent toutes deux $0,5$. L'incertitude ou l'entropie, relative à la couleur d'un corbeau donné, augmente donc si l'on apprend que sa mère est blanche.
- Il existe aussi des cas où de l'information supplémentaire sur une variable augmente l'incertitude elle-même. L'entropie associé à la distribution de probabilité de la localisation d'un trousseau de clefs augmente à partir du moment où l'on découvre qu'elles ne sont pas à l'emplacement habituel et très probable.

De tels exemples ont été discutés également par Seidenfeld (1979). Ils montrent que entropie et information ne varient pas toujours dans le même sens.

3 Expression des contraintes

Nous avons signalé le fait que le MEP ne s'applique qu'à un problème où les contraintes peuvent s'exprimer à partir de la distribution de probabilité. Selon Uffink, il existe des systèmes pour lesquels cela est impossible : le système est influencé par certaines contraintes dont on doit tenir compte, mais différemment. Selon l'auteur, cette question doit faire l'objet d'un futur article.

Conclusion

Le principe du maximum d'entropie tel qu'il a été présenté par Jaynes (et basé sur l'entropie de Shannon), constitue en conclusion un moyen concret et efficace d'obtenir une unique distribution de probabilité dans de nombreux problèmes, et dont la validité est souvent vérifiée. Mais la principale lacune de cette méthode féconde réside dans sa justification, qui est incomplète et quelque peu subjective, précisément parce qu'elle ne fournit qu'une seule solution alors que que des approches plus rationnellement établies en fournissent plusieurs. On a vu en effet que l'on pouvait généraliser et construire une famille continue d'entropies (celles de Rényi) qui donne lieu a une classe continue de «Principe du maximum d'entropie de Rényi». Ces généralisations ouvrent le champ à l'étude de systèmes plus complexes, en particulier les systèmes non-linéaires et hors équilibre. Nous n'avons pas cherché l'exhaustivité dans ce développement ; les entropies présentées sont les plus couramment utilisées. Leurs applications, ainsi que les méthodes pour les estimer dans l'étude des systèmes expérimentaux, n'ont pas non plus été décrites ici. D'autre part, nous avons pu constater que le «second principe», s'il est univoque dans le cadre de la thermodynamique classique nécessite d'être précisé en nuancé en mécanique statistique. Et surtout, il ne doit pas être confondu avec les divers principes du maximum d'entropie, dont l'interprétation est purement statistique, et souvent informationnelle.

Bilan personnel

Cette recherche bibliographique laisse l'image d'un domaine complexe, inachevé sur le plan de ses fondements théoriques, de son interprétation, et encore en plein développement (la plupart des articles cités dans la seconde partie sont des contributions récentes). Mais curieusement, comme souvent dans la recherche scientifique en général, il se révèle en pratique d'une étonnante efficacité (et cela depuis plus d'un siècle), malgré ce manque de justification théorique. Les démarches actuelles reposent pour beaucoup sur la recherche de méthodes ad-hoc adaptées à un domaine d'application particulier.

Cela-dit, après ce petit triptyque sur les entropies statistiques, on peut rester un peu frustré de ne pas avoir vraiment trouvé une réponse satisfaisante à la question du lien profond entre information, description statistique, et monde réel. L'efficacité de cette théorie garde encore pour nous une part de mystère. Peut être un retour aux sources est-il nécessaire sous la forme d'une enquête un peu plus approfondie sur les travaux de Boltzmann et de Shannon, pour comprendre un peu mieux la genèse, et donc la signification, des éléments fondateurs. En effet, en Science on a parfois tendance à prendre pour acquis des idées que l'on utilise régulièrement, simplement parce qu'en fait elles commencent à dater. Un peu d'histoire peu alors parfois nous éclairer sur le sens de ce qu'on utilise quotidiennement, et éventuellement soulever des questions pertinentes, voire fécondes.

Références

[1] SHANNON — *Mathematical theory of communication*

[2] K. PAWELZIK, H.G. SHUSTER — *Generalized dimensions and entropies from a mesured time series* — Physical Review A, vol. 35, n° 1, Janvier 1987.

[3] E.T. JAYNES — *Information theory and statistical mechanics* — Brandier Lecture, 1962, vol. 3, pp 181-218.

[4] L. BOLTZMANN — *Leçons sur la théorie des gaz* — Gobay, 1957.

[5] B. DIU, C GUTHMANN, D. LEDERER, B. ROULET — *About the fundamental postulate of statistical mechanics* —

[6] B. DIU — *Fundamental postulate and canonical distribution in statistical mechanics* — Annales de Physique, vol. 15, pp 433-445, 1991.

[7] R. BALIAN — *Incomplete descriptions and relevant entropies* — American Journal of Physics, n° 67, pp1078-1090, December 1999.

[8] J. Uffink — *Can the maximum entropy principle be explained as a consistency requirement ?.* — non publié.

[9] A. Rényi — *On measures of entropy and information* — 4$^{\text{th}}$ Berkeley Symp. Math. - Statistics and Probabilities, vol. 1.

[10] C. Beck, F. Schlögl — *Thermodynamics of chaotic systems, an introduction* — Cambridge Non-linear Science, series 4.

[11] C. Tsallis — *Possible generalisation of Boltzmann-Gibbs statistics* — Journal of Statistical Physics, vol. 52, Nos, 1/2, 1988.

[12] F. Pennini, A.R. Plastino, A. Plastino — *Dynamical evolution and Tsallis generalized quantum thermostatics* — Physica A, 235 (1997), pp 388-406.

[13] J. Havrda, F. Charvát — *Quantification method of classification processes* — Kybernetica Ćislo, 1. Roćnik 3/1967.