

PREMIER PRÉCEPTORAT DE PHYSIQUE STATISTIQUE:  
Entropie et probabilités

Devant l'impossibilité de connaître toutes les variables microscopiques d'un système macroscopique, la thermodynamique classique introduit une description probabiliste de systèmes parfaitement déterministes (voir par exemple la théorie des gaz parfaits développée par Maxwell). Dans ce cadre, on définit parfois l'entropie comme la mesure du manque d'information liée à cette description probabiliste. On se propose dans ce préceptorat d'étudier quelques propriétés de l'entropie telle que l'a définie Shannon.

## 1 Propriétés de l'entropie

Soit  $\{\Omega, \mathcal{A}, P\}$  un espace probabilisé. On se limite au cas où  $\text{Card}(\mathcal{A}) = n$  est fini. On note alors

$$\mathcal{A} = \{A_1, A_2, \dots, A_n\}$$
$$P(A_i) = p_i, 1 \leq i \leq n$$

Suivant Shannon, on définit l'entropie de  $\{\Omega, \mathcal{A}, P\}$  comme:

$$(1) \quad H(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \ln p_i$$

1. Montrer que  $H(p_1, p_2, \dots, p_n) = 0$  si et seulement si l'un des  $p_i$  vaut 1 et tous les autres valent 0.
2. Montrer que pour toute distribution, on a

$$H(p_1, p_2, \dots, p_n) \leq H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right).$$

3. On considère deux espaces probabilisés  $\{\Omega, \mathcal{A}, P\}$ , et  $\{\Omega', \mathcal{B}, Q\}$ . On suppose à nouveau que  $\text{Card}(\mathcal{A}) = n$  et  $\text{Card}(\mathcal{B}) = m$  sont finis. On introduit enfin l'espace probabilisé  $\{\Omega \times \Omega', \mathcal{A} \times \mathcal{B}, \Pi\}$ . On note:

$$\mathcal{A} = \{A_1, A_2, \dots, A_n\} \quad , \quad \mathcal{B} = \{B_1, B_2, \dots, B_m\}$$
$$P(A_i) = p_i, 1 \leq i \leq n \quad , \quad Q(A_j) = q_j, 1 \leq j \leq m.$$

Enfin, on note

$$\pi_{ij} = \Pi(A_i, B_j),$$

et on introduit la probabilité conditionnelle  $q_{ij}$  définie par:

$$\pi_{ij} = p_i q_{ij}.$$

Les processus  $A$  et  $B$  sont indépendants si, par définition,  $q_{ij} = q_j$ . Montrer qu'alors

$$H(A, B) = H(A) + H(B).$$

4. D'une manière générale, montrer que pour deux processus  $A$  et  $B$  quelconques

$$H(A, B) \leq H(A) + H(B).$$

Donner un exemple où l'inégalité est stricte.

5. On souhaite nuancer la définition de l'entropie afin de pouvoir mesurer le gain d'information apportée par la réalisation de l'expérience  $B$  après l'expérience  $A$ . Pour cela, introduisons l'entropie  $H_i(B)$ , qui est l'entropie qui mesure le manque d'information lié à  $B$ , sachant que le processus  $A$  a donné pour résultat  $A_i$ . Que vaut  $H_i(B)$ ?
6. On introduit alors la variable aléatoire qui prend la valeur  $H_i(B)$  si le avec la probabilité  $p_i$ . On note  $H_A(B)$  la valeur de l'espérance de cette variable.

Montrer que

$$H(A, B) = H(A) + H_A(B).$$

On voit apparaître la signification de  $H_A(B)$ : c'est le *gain d'information moyen apporté par la réalisation de  $B$  après la réalisation de  $A$* .

## 2 Application à un système simple

1. On considère un spin  $\frac{1}{2}$ . En champ magnétique nul, la probabilité  $p_+$  d'avoir  $S_z = +\frac{1}{2}$  et la probabilité  $p_-$  d'avoir  $S_z = -\frac{1}{2}$  sont égales. Calculer l'entropie.
2. On place le spin dans un champ magnétique  $\vec{B} = B\vec{e}_z$ , de sorte que  $p_+ = 2p_-$ . Que vaut désormais l'entropie? Commenter.
3. On se place à nouveau en champ magnétique nul. On considère désormais deux spins  $A$  et  $B$  couplés par un couplage ferromagnétique, de sorte que la probabilité que les deux spins aient la même valeur est deux fois plus importante que celle qu'ils aient des valeurs opposées. Que vaut l'entropie? Calculer  $H_A(B)$ . Comparer à la question précédente.

## 3 Théorème d'unicité

On va voir que certaines des propriétés vues dans la première partie, et qui apparaissent comme des propriétés "naturelles" d'une fonction  $H(p_1, \dots, p_n)$  mesurant le manque d'information lié à la distribution  $(p_1, \dots, p_n)$  permettent en fait de définir  $H$  à une constante multiplicative près.

On impose donc à  $H$  de vérifier les propriétés suivantes:

1. Pour tout  $n$ , pour toute distribution  $(p_1, \dots, p_n)$ , la fonction  $H(p_1, \dots, p_n)$  est maximum si  $p_k = \frac{1}{n}$ , pour  $1 \leq k \leq n$ .
2. Pour deux processus  $A$  et  $B$ ,  $H$  vérifie la propriété 6 de la première partie:

$$H(A, B) = H(A) + H_A(B).$$

3. Si on fabrique à partir d'une distribution de probabilité comportant  $n$  événements une distribution de probabilité comportant les mêmes événements, avec la même probabilité, plus un événement impossible, on obtient la même valeur de l'entropie. C'est à dire:

$$H(p_1, \dots, p_n, 0) = H(p_1, \dots, p_n).$$

**Théorème d'unicité:**

Soit  $H(p_1, \dots, p_n)$  une fonction définie pour tout  $n \geq 1$  et pour tout  $(p_1, \dots, p_n) \in [0, 1]^n$  tels que  $\sum_{i=1}^n p_i = 1$ . Si  $H$  vérifie les propriétés 1, 2, et 3 vue ci-dessus, alors  $\exists \lambda \in \mathbb{R}_+^*$  tel que

$$H(p_1, p_2, \dots, p_n) = -\lambda \sum_{i=1}^n p_i \ln p_i.$$

**Démonstration:**

On pose

$$L(n) = H\left(\frac{1}{n}, \dots, \frac{1}{n}\right).$$

1. Montrer que  $L$  est une fonction croissante.

2. Montrer que

$$\forall (m, r) \in \mathbb{N}^2, L(r^m) = mL(r)$$

3. On se donne désormais deux entiers  $r$  et  $s$  strictement supérieur à 1. Soit alors  $n \in \mathbb{N}$ , et  $m$  l'entier défini par:

$$r^m \leq s^n \leq r^{m+1}.$$

Montrer que:

$$\frac{m}{n} \leq \frac{L(s)}{L(r)} \leq \frac{m}{n} + \frac{1}{n}.$$

4. En déduire:

$$\left\| \frac{L(s)}{L(r)} - \frac{\ln s}{\ln r} \right\| \leq \frac{1}{n}.$$

5. En déduire que  $\exists \lambda \in \mathbb{R}_+^*$  tel que  $\forall n \in \mathbb{N}, L(n) = \lambda \ln n$ . On a donc prouvé le théorème d'unicité dans le cas d'une distribution de probabilité *uniforme*.

On se propose de généraliser la démonstration au cas d'une distribution de probabilité  $(p_1, \dots, p_n)$  pour laquelle tous les  $p_i$  sont rationnels. On pose alors

$$p_i = \frac{g_i}{g}, \quad 1 \leq i \leq n.$$

La condition de normalisation devient alors:  $\sum_{i=1}^n g_i = g$ . On fabrique alors le processus

$B$  défini de la façon suivante:  $B$  est composé de  $g$  événements  $B_1, B_2, \dots, B_g$ , que l'on divise en  $n$  groupes comportant  $g_1, g_2, \dots, g_n$  événements. Si le processus  $A$  a pour

résultat  $A_k$ , alors la probabilité d'obtenir chacun des événements du groupe  $k$  vaut  $\frac{1}{g_k}$ , et la probabilité des autres  $B_i$  est nulle:

$$\begin{array}{ccc}
 \underbrace{B_1, \dots, B_{g_1},}_{\text{Si } A_1,} & & \dots, B_g \\
 \downarrow & & \dots \\
 P(B_1) = \frac{1}{g_1}, \dots, P(B_{g_1}) = \frac{1}{g_1}, & & \\
 P(B_{g_1+1}) = 0, \dots, P(B_g) = 0 & & 
 \end{array}$$

6. Calculer  $H_k(B)$  (voir question 5 de la première partie).
7. En déduire  $H_A(B)$ .
8. Calculer la probabilité d'un événement  $A_k B_l$ .
9. En déduire  $H(AB)$ .
10. Conclure que:

$$H(A) = -\lambda \sum_{i=1}^n p_i \ln p_i.$$

11. Achever la démonstration en faisant appel à un argument de continuité.