Aspects rigoureux de la mécanique statistique à l'équilibre
TD n$^{\text{o}}$ 3 : Entropie et principe variationnel

Jérémie Bouttier et Guilhem Semerjian

Avril 2015

1 Entropie de Shannon

On considère un ensemble dénombrable d'évènements élémentaires, et l'ensemble des lois de probabilités p qui donnent une probabilité non-nulle à un nombre fini d'évènements. Pour simplifier on notera $p = (p_1, \dots, p_M)$, avec les $p_i \in [0,1]$ et $\sum_{i=1}^{M} p_i = 1$.

En 1948 Shannon a introduit une fonction entropie $S(p)$, qui quantifie le "manque d'information a priori" sur un tirage d'un évènement aléatoire avec la loi p. Il a imposé les conditions suivantes sur S :

(P1) la fonction S est une fonction continue des probabilités $p_1, \dots, p_M$

(P2) $S(p_1, \dots, p_M) \leq S(1/M, \dots, 1/M)$

(P3) $S(p_1, \dots, p_M, 0) = S(p_1, \dots, p_M)$

(P4) propriété d'additivité : soient A l'évènement global correspondant aux évènements de 1 à m de probabilité $q_A = \sum_{i=1}^{m} p_i$ et B l'évènement global correspondant aux évènements de $m+1$ à M de probabilité $q_B = \sum_{i=m+1}^{M} p_i$. Alors on a :

$$S(p_1, \dots, p_M) = S(q_A, q_B) + q_A\, S\left(\frac{p_1}{q_A}, \dots, \frac{p_m}{q_A}\right) + q_B\, S\left(\frac{p_{m+1}}{q_B}, \dots, \frac{p_M}{q_B}\right)$$

On va montrer que l'expression $S(p_1, \dots, p_M) = -\sum_{i=1}^{M} p_i \ln p_i$ est, à une constante multiplicative près, la seule fonction vérifiant ces propriétés. On utilise la convention $0 \ln 0 = 0$. On notera $\sigma(M) = S(1/M, \dots, 1/M)$ l'entropie du cas équiprobable.

1. Montrer que $\sigma(M)$ est une fonction croissante de M.
2. Montrer que $S(1) = 0$. Commenter.
3. Généraliser la quatrième propriété pour montrer que

$$S(p_1, \dots, p_{m_1}, \dots, p_{m_2}, \dots, p_{m_\alpha}) = S(q_1, \dots, q_\alpha) + \sum_{i=1}^{\alpha} q_i\, S\left(\frac{p_{m_{i-1}+1}}{q_i}, \dots, \frac{p_{m_i}}{q_i}\right)$$

   avec $q_i = p_{m_{i-1}+1} + \ldots + p_{m_i}$ $(m_0 = 0)$.
4. En déduire que $\sigma(MN) = \sigma(M) + \sigma(N)$.
5. En déduire que $\sigma(M) = k \ln M$, où k est une constante. Quel doit être son signe ?
6. En déduire que, pour $(p_1, \dots, p_M) \in \mathbb{Q}^M$, $S(p_1, \dots, p_M) = -k \sum_{i=1}^{M} p_i \ln p_i$.
7. En déduire l'expression de l'entropie de Shannon pour des probabilités quelconques.

L'entropie de Shannon joue un rôle fondamental en physique statistique, en probabilité (en particulier pour la théorie des grandes déviations) et en théorie de l'information. Dans ce dernier contexte elle apparaît notamment dans les théorèmes de Shannon, qui montrent (i) que le taux de compression possible sans perte d'information d'une suite de symboles générés par une source aléatoire en un fichier binaire est précisément l'entropie de la source, (ii) et que la vitesse de communication au travers d'un canal bruité est bornée par l'information mutuelle (une notion reliée à l'entropie que l'on va définir ci-dessous) entre l'entrée et la sortie du canal Pour ces aspects de théorie de l'information on pourra consulter : *Elements of Information Theory* de T. Cover et J. Thomas, ainsi que *Information Theory, Inference, and Learning Algorithms* de D. MacKay, disponible gratuitement sur le site `http://www.inference.phy.cam.ac.uk/mackay/itila/book.html`

2 Propriétés de l'entropie

Pour faire le lien avec le cours on note à partir de maintenant $\underline{\sigma} \in \Sigma = \chi^\Delta$ les configurations dans une partie Δ *finie* de $\mathbb{Z}^d$, et μ une loi de probabilité sur ces configurations, donnée de manière élémentaire par les $\mu(\underline{\sigma}) \in [0,1]$ avec $\sum\limits_{\underline{\sigma}} \mu(\underline{\sigma}) = 1$. On note alors l'entropie de cette loi

$$S(\mu) = -\sum_{\underline{\sigma}} \mu(\underline{\sigma}) \ln \mu(\underline{\sigma}) \ . \tag{1}$$

1. Montrer que S est concave, i.e. $S(\alpha\mu + (1-\alpha)\nu) \geq \alpha S(\mu) + (1-\alpha)S(\nu)$ si μ, ν sont deux lois de probabilité sur le même espace de configurations, et $\alpha \in [0,1]$. *Indication :* montrer d'abord que $g(x) = -x \ln x$ est une fonction concave sur $[0,1]$.
2. Montrer que S est "presque convexe", dans le sens que
$$S(\alpha\mu + (1-\alpha)\nu) \leq \alpha S(\mu) + (1-\alpha)S(\nu) - \alpha \ln \alpha - (1-\alpha)\ln(1-\alpha) \ . \tag{2}$$
Indication : montrer que g est sous-additive, i.e. $g(a+b) \leq g(a) + g(b)$, et que $g(uv) = ug(v) + vg(u)$.
3. On introduit la "distance" de Küllback-Leibler entre deux mesures de probabilité par
$$D(\mu||\nu) = \sum_{\underline{\sigma}} \mu(\underline{\sigma}) \ln \left(\frac{\mu(\underline{\sigma})}{\nu(\underline{\sigma})} \right) \ , \tag{3}$$
qui n'est pas une distance puisqu'elle n'est pas symétrique. Montrer que $D(\mu||\nu) \geq 0$. Dans quel cas a-t-on $D(\mu||\nu) = 0$?
4. Montrer que $S(\mu) \in [0, |\Delta| \ln |\chi|]$. Dans quels cas ces bornes sont-elles atteintes ?
5. On considère un sous-ensemble de variables $\Lambda \subset \Delta$ et son complémentaire $\Lambda^c = \Delta \setminus \Lambda$, et les lois marginales de μ
$$\mu_\Lambda(\underline{\sigma}_\Lambda) = \sum_{\underline{\sigma}_{\Lambda^c}} \mu((\underline{\sigma}_\Lambda, \underline{\sigma}_{\Lambda^c})) \ , \qquad \mu_{\Lambda^c}(\underline{\sigma}_{\Lambda^c}) = \sum_{\underline{\sigma}_\Lambda} \mu((\underline{\sigma}_\Lambda, \underline{\sigma}_{\Lambda^c})) \ . \tag{4}$$
On note $S_\Lambda(\mu)$ et $S_{\Lambda^c}(\mu)$ les entropies de ces deux lois. Montrer que
$$S(\mu) \leq S_\Lambda(\mu) + S_{\Lambda^c}(\mu) \ . \tag{5}$$
Dans quel cas la borne est-elle atteinte ? On définit l'information mutuelle entre Λ et Λ^c comme $I_{\Lambda,\Lambda^c}(\mu) = S_\Lambda(\mu) + S_{\Lambda^c}(\mu) - S(\mu) = I_{\Lambda^c,\Lambda}(\mu)$, qui est donc positive d'après l'inégalité ci-dessus.
6. La loi conditionnelle dans Λ sachant Λ^c est définie par $\mu(\underline{\sigma}_\Lambda | \underline{\sigma}_{\Lambda^c}) = \frac{\mu((\underline{\sigma}_\Lambda, \underline{\sigma}_{\Lambda^c}))}{\mu_{\Lambda^c}(\underline{\sigma}_{\Lambda^c})}$, et l'entropie conditionnelle selon
$$S_{\Lambda|\Lambda^c}(\mu) = \sum_{\underline{\sigma}_{\Lambda^c}} \mu_{\Lambda^c}(\underline{\sigma}_{\Lambda^c}) \left[-\sum_{\underline{\sigma}_\Lambda} \mu(\underline{\sigma}_\Lambda | \underline{\sigma}_{\Lambda^c}) \ln \mu(\underline{\sigma}_\Lambda | \underline{\sigma}_{\Lambda^c}) \right] \ . \tag{6}$$
Interpréter cette définition, exprimer l'entropie conditionnelle à l'aide des entropies marginales et de l'information mutuelle, et justifier le moyen mnémotechnique "le conditionnement réduit l'entropie".

3 Principe variationnel de Gibbs

Toujours dans le cadre d'un espace de configurations fini on considère maintenant un Hamiltonien $H(\underline{\sigma})$ et on définit l'énergie libre $F_{\rm c}(\beta, H)$ et la loi de probabilité de l'ensemble canonique $\mu_{\rm c}(\beta, H)$ selon

$$F_{\rm c}(\beta, H) = -\frac{1}{\beta} \ln Z(\beta, H) \ , \qquad Z(\beta, H) = \sum_{\underline{\sigma}} e^{-\beta H(\underline{\sigma})} \ , \qquad \mu_{\rm c}(\beta, H, \underline{\sigma}) = \frac{e^{-\beta H(\underline{\sigma})}}{Z(\beta, H)} \ . \tag{7}$$

On définit par ailleurs l'énergie libre de Gibbs $F_{\rm Gibbs}(\beta, H, \mu)$ qui associe un réel à la donnée d'une température inverse β, d'un Hamiltonien H et d'une mesure de probabilité μ :

$$F_{\rm Gibbs}(\beta, H, \mu) = \sum_{\underline{\sigma}} \mu(\underline{\sigma}) H(\underline{\sigma}) - \frac{1}{\beta} S(\mu) \ . \tag{8}$$

On notera $\mathcal{M}$ l'ensemble des mesures de probabilités sur cet espace de configurations.

1. Montrer que $F_{\rm Gibbs}(\beta, H, \mu) \geq F_{\rm c}(\beta, H)$ pour tout $\mu \in \mathcal{M}$, et que $F_{\rm Gibbs}(\beta, H, \mu_{\rm c}(\beta, H)) = F_{\rm c}(\beta, H)$. On a donc la caractérisation variationnelle de l'énergie libre et de la loi canonique,
$$F_{\rm c}(\beta, H) = \min_{\mu \in \mathcal{M}} F_{\rm Gibbs}(\beta, H, \mu) \ , \qquad \mu_{\rm c}(\beta, H) = \underset{\mu \in \mathcal{M}}{\rm argmin} \ F_{\rm Gibbs}(\beta, H, \mu) \ . \tag{9}$$
2. Montrer que la loi canonique est celle qui maximise l'entropie $S(\mu)$ lorsque l'on contraint l'énergie moyenne $\sum_{\underline{\sigma}} \mu(\underline{\sigma}) H(\underline{\sigma})$ à prendre une valeur donnée. Quelle est l'interprétation de la température dans ce calcul ?
3. Montrer que pour une loi de probabilité arbitraire $\mu \in \mathcal{M}$ on a l'expression variationnelle suivante de l'entropie :
$$S(\mu) = \min_H \left[\beta \sum_{\underline{\sigma}} \mu(\underline{\sigma}) H(\underline{\sigma}) - \beta F_{\rm c}(\beta, H) \right] \ . \tag{10}$$
On pourra commencer par supposer pour simplifier que $\mu(\underline{\sigma}) > 0$ pour toutes les configurations.
4. On considère un modèle d'Ising sur une portion rectangulaire Δ de $\mathbb{Z}^d$, avec des conditions aux bords périodiques, interagissant selon l'Hamiltonien
$$H(\underline{\sigma}) = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - h \sum_{i \in \Delta} \sigma_i \ , \tag{11}$$
où la première somme porte sur les liens de Δ. Donner une borne supérieure à l'énergie libre par spin $F_{\rm c}(\beta, H)/|\Delta|$, en considérant pour μ une mesure produit de spins indépendants.