Aspects rigoureux de la mécanique statistique à l'équilibre
Corrigé du TD n° 3 : Entropie et principe variationnel

Jérémie Bouttier et Guilhem Semerjian

Avril 2015

1 Entropie de Shannon

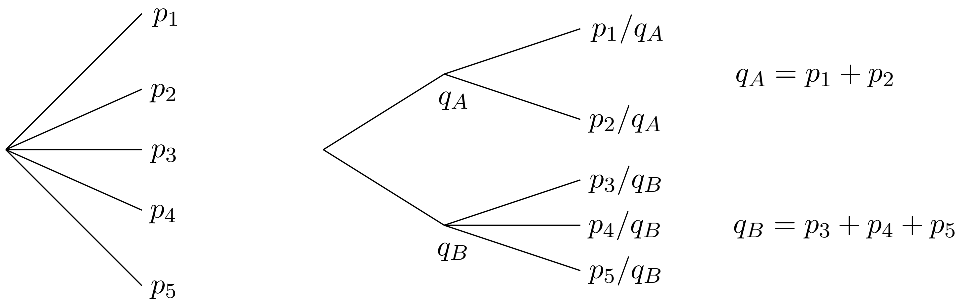
Commentons les hypothèses faites sur S :

(P1) La continuité est “naturelle”, changer infinitésimalement une probabilité ne peut provoquer de grand changement dans le manque d'information.

(P2) Le manque d'information est maximal si les évènements susceptibles de se produire sont équiprobables, il n'y a pas de stratégie possible pour deviner l'évènement qui va se produire.

(P3) Rajouter un évènement qui ne se produira jamais ne change pas le manque d'information sur le tirage.

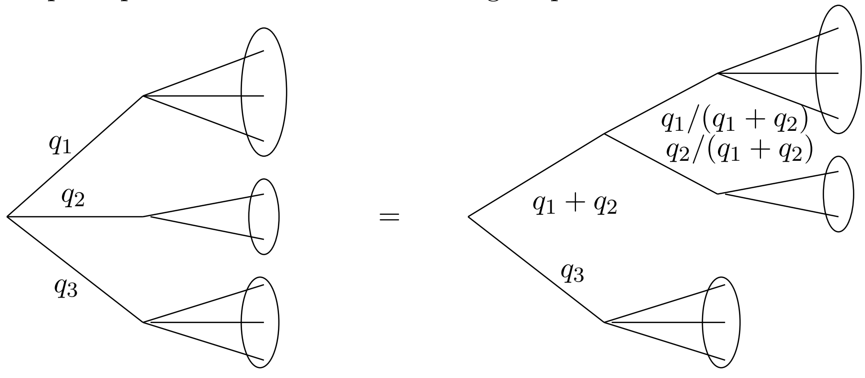
(P4) Tirer un évènement dans $[1, M]$ est équivalent à tirer A ou B avec les probabilités q_A et q_B, puis, conditionné à ce premier tirage, choisir l'évènement élémentaire à l'intérieur du groupe choisi avec la probabilité conditionnelle. Le manque d'information est donc additif sous cette décomposition.



1.
$$\sigma(M) = S\left(\frac{1}{M}, \dots, \frac{1}{M}\right) = S\left(\frac{1}{M}, \dots, \frac{1}{M}, 0\right) \leq S\left(\frac{1}{M+1}, \dots, \frac{1}{M+1}\right) = \sigma(M+1) \quad (1)$$

2. En appliquant la règle de décomposition avec $M = 2$, $q_A = p_1$, $q_B = p_2$, il vient
$$S(p_1, p_2) = S(p_1, p_2) + (p_1 + p_2) S(1) = S(p_1, p_2) + S(1) \ . \quad (2)$$
Il n'y a en effet aucun manque d'information sur un évènement dont le tirage est déterministe.

3. Comme l'hypothèse (P4) correspond à cette formule pour $\alpha = 2$ on peut faire une récurrence sur α, dont le principe est schématisé sur cette figure pour $\alpha = 3$:



Explicitement,

$$\begin{aligned}
S(p_1,\ldots,p_{m_1},\ldots,p_{m_2},\ldots,p_{m_\alpha}) &= S(q_1+\cdots+q_{\alpha-1},q_\alpha) \\
&\quad +(q_1+\cdots+q_{\alpha-1})S\left(\frac{p_1}{q_1+\cdots+q_{\alpha-1}},\ldots,\frac{p_{m_{\alpha-1}}}{q_1+\cdots+q_{\alpha-1}}\right) \\
&\quad +q_\alpha S\left(\frac{p_{m_{\alpha-1}+1}}{q_\alpha},\ldots,\frac{p_{m_\alpha}}{q_\alpha}\right) \qquad (3)
\end{aligned}$$

en utilisant (P4). On utilise ensuite l'hypothèse de récurrence au rang $\alpha-1$ pour transformer le S de la deuxième ligne, ce qui donne

$$\begin{aligned}
& S(q_1+\cdots+q_{\alpha-1},q_\alpha) \\
& +(q_1+\cdots+q_{\alpha-1})\left[S\left(\frac{q_1}{q_1+\cdots+q_{\alpha-1}},\ldots,\frac{q_{\alpha-1}}{q_1+\cdots+q_{\alpha-1}}\right)\right. \\
& \qquad\qquad \left.+\sum_{i=1}^{\alpha-1}\frac{q_i}{q_1+\cdots+q_{\alpha-1}}S\left(\frac{p_{m_{i-1}+1}}{q_i},\ldots,\frac{p_{m_i}}{q_i}\right)\right] \\
& +q_\alpha S\left(\frac{p_{m_{\alpha-1}+1}}{q_\alpha},\ldots,\frac{p_{m_\alpha}}{q_\alpha}\right) \\
= \;& S(q_1+\cdots+q_{\alpha-1},q_\alpha)+(q_1+\cdots+q_{\alpha-1})S\left(\frac{q_1}{q_1+\cdots+q_{\alpha-1}},\ldots,\frac{q_{\alpha-1}}{q_1+\cdots+q_{\alpha-1}}\right)+q_\alpha S(\frac{q_\alpha}{q_\alpha}) \\
& +\sum_{i=1}^{\alpha} q_i S\left(\frac{p_{m_{i-1}+1}}{q_i},\ldots,\frac{p_{m_i}}{q_i}\right)
\end{aligned}$$

En utilisant (P4) on voit finalement que la première ligne est $S(q_1,\ldots,q_\alpha)$, ce qui conclut la preuve.

4. Il suffit de considérer l'entropie de MN évènements équiprobables, regroupés en $\alpha = M$ blocs de N évènements.
5. En itérant cette relation il vient $\sigma(l^n) = n\sigma(l)$. On veut exprimer $\sigma(l)$ en terme de $\sigma(2)$, on définit donc $m(n) = \lfloor n\frac{\ln l}{\ln 2}\rfloor$, de sorte que $2^m \le l^n < 2^{m+1}$. En utilisant aussi la croissance de σ on obtient

$$\begin{aligned}
\sigma(2^m) \le & \quad \sigma(l^n) & \le \sigma(2^{m+1}) \qquad (4) \\
m\sigma(2) \le & \quad n\sigma(l) & \le (m+1)\sigma(2) \qquad (5) \\
\frac{1}{n}\left\lfloor n\frac{\ln l}{\ln 2}\right\rfloor\sigma(2) \le & \quad \sigma(l) & \le \left(\frac{1}{n}\left\lfloor n\frac{\ln l}{\ln 2}\right\rfloor+\frac{1}{n}\right)\sigma(2)\ . \qquad (6)
\end{aligned}$$

En prenant la limite $n\to\infty$ à l fixé il vient $\sigma(l) = \frac{\sigma(2)}{\ln 2}\ln l = k\ln l$. Il faut que k soit positive pour que l'entropie du cas équiprobable soit croissante.

6. En réduisant au même dénominateur on peut écrire $p_i = N_i/D$, où D et les N_i sont des entiers. En utilisant la décomposition de l'entropie pour D évènements équiprobables groupés en M blocs contenant chacun N_i évènements élémentaires, il vient

$$S\left(\frac{1}{D},\ldots,\frac{1}{D}\right) = S\left(\frac{N_1}{D},\ldots,\frac{N_M}{D}\right)+\sum_{i=1}^{M}\frac{N_i}{D}S\left(\frac{1}{N_i},\ldots,\frac{1}{N_i}\right)\ , \qquad (7)$$

soit

$$S\left(\frac{N_1}{D},\ldots,\frac{N_M}{D}\right) = \sigma(D)-\sum_{i=1}^{M}\frac{N_i}{D}\sigma(N_i) = -k\sum_{i=1}^{M}\frac{N_i}{D}\ln\left(\frac{N_i}{D}\right) = -k\sum_{i=1}^{M}p_i\ln p_i\ , \qquad (8)$$

en utilisant l'expression précédemment établie pour σ.

7. Comme $\mathbb{Q}^M$ est dense dans $\mathbb{R}^M$ et que S est supposée continue cette expression reste valable pour des probabilités arbitraires.

2 Propriétés de l'entropie

1. On peut remarquer que $g(x) = -x\ln x$ est une fonction concave sur $[0,1]$, sa dérivée seconde $-1/x$ étant négative. Pour une valeur de $\underline{\sigma}$ on a donc
$$-(\alpha\mu(\underline{\sigma}) + (1-\alpha)\nu(\underline{\sigma}))\ln(\alpha\mu(\underline{\sigma}) + (1-\alpha)\nu(\underline{\sigma})) \geq -\alpha\mu(\underline{\sigma})\ln\mu(\underline{\sigma}) - (1-\alpha)\nu(\underline{\sigma})\ln\nu(\underline{\sigma}) \ , \quad (9)$$
qu'il suffit ensuite de sommer sur $\underline{\sigma}$ pour obtenir $S(\alpha\mu + (1-\alpha)\nu) \geq \alpha S(\mu) + (1-\alpha)S(\nu)$.
2. Notons que pour $a, b \geq 0$
$$g(a+b) = -(a+b)\ln(a+b) = -a\ln(a+b) - b\ln(a+b) \leq -a\ln a - b\ln b = g(a) + g(b) \ , \quad (10)$$
car le logarithme est croissant. De plus
$$g(uv) = -uv\ln(uv) = -uv\ln u - uv\ln v = ug(v) + vg(u) \ , \quad (11)$$
donc
$$g(\alpha x + (1-\alpha)y) \leq g(\alpha x) + g((1-\alpha)y) = \alpha g(x) + (1-\alpha)g(y) + xg(\alpha) + yg(1-\alpha) \ . \quad (12)$$
pour $\alpha, x, y \in [0,1]$.
En appliquant cette inégalité avec $x = \mu(\underline{\sigma})$, $y = \nu(\underline{\sigma})$ et en sommant sur $\underline{\sigma}$ il vient bien le résultat demandé.
3. Le logarithme étant une fonction concave, pour toute fonction positive $f(\underline{\sigma})$ et toute mesure de probabilité $\mu(\underline{\sigma})$ l'inégalité de Jensen conduit à
$$\sum_{\underline{\sigma}} \mu(\underline{\sigma})\ln(f(\underline{\sigma})) \leq \ln\left(\sum_{\underline{\sigma}} \mu(\underline{\sigma})f(\underline{\sigma})\right) \ . \quad (13)$$
En appliquant cette inégalité à $f(\underline{\sigma}) = \frac{\nu(\underline{\sigma})}{\mu(\underline{\sigma})}$,
$$-D(\mu||\nu) \leq \ln\left(\sum_{\underline{\sigma}} \mu(\underline{\sigma})\frac{\nu(\underline{\sigma})}{\mu(\underline{\sigma})}\right) = 0 \ . \quad (14)$$
Le logarithme étant strictement concave il n'y a égalité que si f est constante, autrement dit $D(\mu||\nu) = 0 \Rightarrow \mu = \nu$.
Notons qu'on peut aussi prouver l'inégalité à partir de $\ln x \leq x - 1$:
$$-D(\mu||\nu) = \sum_{\underline{\sigma}} \mu(\underline{\sigma})\ln\left(\frac{\nu(\underline{\sigma})}{\mu(\underline{\sigma})}\right) \leq \sum_{\underline{\sigma}} (\nu(\underline{\sigma}) - \mu(\underline{\sigma})) = 0 \quad (15)$$
4. Comme $\mu(\underline{\sigma}) \in [0,1]$ les termes de la somme sont non-négatifs, d'où $S(\mu) \geq 0$. Pour que $S(\mu)$ soit égale à 0 il faut que $\mu(\underline{\sigma})\ln(\mu(\underline{\sigma})) = 0$ pour tout $\underline{\sigma}$, i.e. que $\mu(\underline{\sigma}) \in \{0,1\}$, comme μ est normalisée cela implique que $\mu(\underline{\sigma}) = \delta_{\underline{\sigma},\underline{\tau}}$, la mesure est supportée par une seule configuration $\underline{\tau}$.
Notons ν la mesure uniforme sur les $|\chi|^{|\Delta|}$ configurations, avec $\nu(\underline{\sigma}) = \frac{1}{|\chi|^{|\Delta|}}$. Alors
$$D(\mu||\nu) = -S(\mu) + |\Delta|\ln|\chi| \geq 0 \quad \Rightarrow S(\mu) \leq |\Delta|\ln|\chi| \ . \quad (16)$$
On peut aussi le voir comme une conséquence de l'hypothèse (P2) de l'exercice précédent. La borne n'est atteinte que pour $\mu = \nu$, la mesure uniforme.
5. Comme $\mu_\Lambda\mu_{\Lambda^c}$ est une mesure de probabilité sur le même espace de configurations que μ on peut calculer
$$\begin{aligned} D(\mu||\mu_\Lambda\mu_{\Lambda^c}) &= -S(\mu) - \sum_{\underline{\sigma}_\Lambda,\underline{\sigma}_{\Lambda^c}} \mu((\underline{\sigma}_\Lambda,\underline{\sigma}_{\Lambda^c}))\ln\mu_\Lambda(\underline{\sigma}_\Lambda) - \sum_{\underline{\sigma}_\Lambda,\underline{\sigma}_{\Lambda^c}} \mu((\underline{\sigma}_\Lambda,\underline{\sigma}_{\Lambda^c}))\ln\mu_{\Lambda^c}(\underline{\sigma}_{\Lambda^c}) \\ &= -S(\mu) - \sum_{\underline{\sigma}_\Lambda} \mu_\Lambda(\underline{\sigma}_\Lambda)\ln\mu_\Lambda(\underline{\sigma}_\Lambda) - \sum_{\underline{\sigma}_{\Lambda^c}} \mu(\underline{\sigma}_{\Lambda^c})\ln\mu_{\Lambda^c}(\underline{\sigma}_{\Lambda^c}) \\ &= -S(\mu) + S_\Lambda(\mu) + S_{\Lambda^c}(\mu) \geq 0 \end{aligned} \quad (17)$$
La borne est atteinte si $\mu = \mu_\Lambda\mu_{\Lambda^c}$, c'est-à-dire si les variables dans Λ sont indépendantes de celles dans Λ^c, l'information mutuelle est alors nulle.

6. L'entropie conditionnelle est la moyenne, sur la variable conditionnante, de l'entropie de la loi conditionnelle. D'après la définition,

$$S_{\Lambda|\Lambda^c}(\mu) = -\sum_{\underline{\sigma}} \mu(\underline{\sigma}) \ln\left(\frac{\mu(\underline{\sigma})}{\mu_{\Lambda^c}(\underline{\sigma}_{\Lambda^c})}\right) = S(\mu) - S_{\Lambda^c}(\mu) = S_\Lambda(\mu) - I_{\Lambda,\Lambda^c}(\mu) \leq S_\Lambda(\mu) \qquad (18)$$

puisque l'information mutuelle est positive. On peut ainsi interpréter

$$I_{\Lambda,\Lambda^c}(\mu) = S_\Lambda(\mu) - S_{\Lambda|\Lambda^c}(\mu) = S_{\Lambda^c}(\mu) - S_{\Lambda^c|\Lambda}(\mu) \qquad (19)$$

comme la réduction (moyenne) du manque d'information sur une des deux variables aléatoires quand l'autre est connue.

3 Principe variationnel de Gibbs

1. Notons que l'énergie libre de Gibbs $F_{\text{Gibbs}}(\beta, H, \mu)$ a bien la forme "énergie (moyenne) - température $\times$ entropie". Calculons la distance de Kullback-Leibler entre μ et $\mu_c(\beta, H)$:

$$D(\mu||\mu_c(\beta,H)) = \sum_{\underline{\sigma}} \mu(\underline{\sigma}) \ln\left(\mu(\underline{\sigma}) \frac{Z(\beta,H)}{e^{-\beta H(\underline{\sigma})}}\right) = -S(\mu) + \ln Z(\beta,H) + \beta \sum_{\underline{\sigma}} \mu(\underline{\sigma}) H(\underline{\sigma}) \ , \qquad (20)$$

soit

$$F_{\text{Gibbs}}(\beta, H, \mu) = F_c(\beta, H) + \frac{1}{\beta} D(\mu||\mu_c(\beta,H)) \ . \qquad (21)$$

Comme $D \geq 0$ on a $F_{\text{Gibbs}}(\beta, H, \mu) \geq F_c(\beta, H)$, et comme $D(\mu_c(\beta,H)||\mu_c(\beta,H)) = 0$ la borne est atteinte pour $\mu = \mu_c(\beta, H)$. Comme $D(\mu||\nu) = 0$ seulement si $\mu = \nu$, $\mu_c(\beta, H)$ est l'unique point où le minimum est atteint.

Une preuve alternative de l'inégalité consiste à écrire

$$Z(\beta,H) = \sum_{\underline{\sigma}} e^{-\beta H(\underline{\sigma})} = \sum_{\underline{\sigma}} \mu(\underline{\sigma}) e^{-\beta H(\underline{\sigma}) - \ln \mu(\underline{\sigma})} \geq e^{-\beta \sum_{\underline{\sigma}} \mu(\underline{\sigma}) H(\underline{\sigma}) - \sum_{\underline{\sigma}} \mu(\underline{\sigma}) \ln \mu(\underline{\sigma})} \ , \qquad (22)$$

en appliquant l'inégalité de Jensen à la fonction exponentialle qui est convexe. Il suffit ensuite de prendre le logarithme et de diviser par $-\beta$.

Pour vérifier que $F_{\text{Gibbs}}(\beta, H, \mu_c(\beta, H)) = F_c(\beta, H)$ on aurait aussi pu calculer l'entropie de la loi canonique,

$$S(\mu_c(\beta,H)) = -\sum_{\underline{\sigma}} \mu_c(\beta,H,\underline{\sigma}) \ln\left(\frac{e^{-\beta H(\underline{\sigma})}}{Z(\beta,H)}\right) = \ln Z(\beta,H) + \beta \sum_{\underline{\sigma}} \mu_c(\beta,H,\underline{\sigma}) H(\underline{\sigma}) \ . \qquad (23)$$

2. Pour maximiser une fonction en imposant des contraintes on utilise la méthode des multiplicateurs de Lagrange. Ici on doit maximiser $S(\mu)$ par rapport à μ, en imposant la normalisation $\sum_{\underline{\sigma}} \mu(\underline{\sigma}) = 1$ et $\sum_{\underline{\sigma}} \mu(\underline{\sigma}) H(\underline{\sigma}) = E$, une énergie donnée. La méthode des multiplicateurs de Lagrange consiste donc à chercher un point stationnaire de $S(\mu) + A \sum_{\underline{\sigma}} \mu(\underline{\sigma}) + B \sum_{\underline{\sigma}} \mu(\underline{\sigma}) H(\underline{\sigma})$, où A et B sont les multiplicateurs de Lagrange. En annulant la dérivée par rapport à $\mu(\underline{\sigma})$ de cette fonction on obtient

$$\mu(\underline{\sigma}) = e^{A + B\, H(\underline{\sigma})} \ . \qquad (24)$$

Les multiplicateurs de Lagrange A et B sont ensuite fixés pour satisfaire les contraintes. On retrouve alors la loi canonique, avec $B = -\beta$, et $e^A = 1/Z$. On peut voir aussi l'énergie libre de Gibbs, à un facteur $-1/\beta$ près, comme la fonction de Lagrange à maximiser. Le fait que l'entropie soit maximale (et non minimale) à cet extrémum vient de sa concavité.

3. La fonction qui à H associe
$$\beta \sum_{\underline{\sigma}} \mu(\underline{\sigma}) H(\underline{\sigma}) - \beta F_{\rm c}(\beta, H) \tag{25}$$
est convexe et dérivable (c'est évident pour le premier terme, et cela a été démontrée en cours pour le deuxième). Le minimum est donc atteint au point $\widehat{H}$ où ses dérivées par rapport aux $H(\underline{\sigma})$ s'annulent, autrement dit $\widehat{H}(\underline{\sigma}) = -\frac{1}{\beta} \ln \mu(\underline{\sigma}) + C$, où C est une constante indépendante de $\underline{\sigma}$. On a donc un Hamiltonien $\widehat{H}$ tel que $\mu_{\rm c}(\beta, \widehat{H}) = \mu$. Comme $F_{\rm Gibbs}(\beta, \widehat{H}, \mu_{\rm c}(\beta, \widehat{H}) = \mu) = F_{\rm c}(\beta, \widehat{H})$, il vient
$$\sum_{\underline{\sigma}} \mu(\underline{\sigma}) \widehat{H}(\underline{\sigma}) - \frac{1}{\beta} S(\mu) = F_{\rm c}(\beta, \widehat{H}) \ , \qquad \text{soit} \ \ S(\mu) = \beta \sum_{\underline{\sigma}} \mu(\underline{\sigma}) \widehat{H}(\underline{\sigma}) - \beta F_{\rm c}(\beta, \widehat{H}) \ . \tag{26}$$

4. Notons μ la mesure produit dans Δ, où chaque spin a indépendamment l'aimantation m, $\mu(\underline{\sigma}) = \prod\limits_{i \in \Delta} \frac{1+m\sigma_i}{2}$. On a alors
$$\frac{1}{|\Delta|} F_{\rm Gibbs}(\beta, H, \mu) = -Jdm^2 - hm - \frac{1}{\beta} s(m) \ , \qquad s(m) = -\frac{1+m}{2} \ln\left(\frac{1+m}{2}\right) - \frac{1-m}{2} \ln\left(\frac{1-m}{2}\right) \ ,$$
car il y a $2d$ liens autour de chaque site mais il ne faut les compter qu'une fois pour chacune des deux extrémités. La borne étant vraie pour tout m, on peut donc conclure
$$\frac{1}{|\Delta|} F_{\rm c}(\beta, H) \leq \inf_{m \in [-1,1]} f_{\rm cm}(m) \ , \qquad f_{\rm cm}(m) = -Jdm^2 - hm - \frac{1}{\beta} s(m) \ , \tag{27}$$
où l'on a reconnu l'expression de l'énergie libre de champ moyen établi au TD 1.

Plus généralement, le principe variationnel de Gibbs est souvent exploité en mécanique statistique, sous le nom de "méthodes variationnelles". En effet en général on est incapable de calculer exactement $F_{\rm c}(\beta, H)$ pour un Hamiltonien $H(\underline{\sigma})$ un peu compliqué. On peut néanmoins calculer $F_{\rm Gibbs}[\beta, H, \mu]$ pour des mesures μ suffisament simples. En cherchant le minimum de cette fonction dans un sous-espace de $\mathcal{M}$ on obtient donc une borne supérieure de la quantité inaccessible $F_{\rm c}$. Quand μ se cantonne aux mesures produits on obtient ainsi l'approximation de champ moyen.

Dans ce contexte le principe variationnel est souvent exprimé sous la forme
$$F_{\rm c}(\beta, H) \leq F_0 + \langle H - H_0 \rangle_0 \ , \tag{28}$$
où la mesure variationnelle est prise de la forme $\mu = e^{-\beta H_0}/Z_0$, les moyennes avec cette mesure étant notées $\langle \bullet \rangle_0$, et F_0 étant l'énergie libre associée au Hamiltonien variationnel H_0.