Université Grenoble Alpes, Master 1 Physique.
TD de physique statistique. Frédéric Faure.

1 Entropie et suites standard. Solution.

Dans cet exercice on défini la notion d'entropie selon Shannon [1] (1949). Références : (Shannon, MacMillan 1949)(cf Khinchin p.17, Zinmeister p.22).

Considérons un ensemble de n évènements possibles $\{1, 2 \ldots n\}$ ayant chacun la probabilité $p_i \geq 0$, $i = 1, \ldots n$ vérifiant $\sum_{i=1}^{n} p_i = 1$, appelée "loi de probabilité finie". Soit $N \geq 1$. Considérons une suite de N réalisations aléatoires **indépendantes** de ces évènements, notée

$$\vec{i} := (i_1, i_2, \ldots, i_N) \tag{1.1}$$

Comme exemple simple, on pourra considérer un jeu de pile/face, c'est à dire $n = 2$ évènements possibles, mais de probabilités non égale (appelé loi de Bernouilli) :

$$p_1 = 2/3 \simeq 0.7 \quad > \quad p_2 = 1/3 \simeq 0.3.$$

1. Quel est le nombre total $\mathcal{N}_{\text{tot}}(N)$ de suites $\vec{i}$? Quelle est la probabilité $p\left(\vec{i}\right)$ pour que la suite $\vec{i} := (i_1, i_2, \ldots, i_N)$ apparaisse ? Dans l'exemple pile/face, quelle est la suite la plus probable ? la moins probable ?

   Réponse : Il y a $\mathcal{N}_{\text{tot}}(N) = n^N = e^{N \log n}$ suites possibles. Comme les évènements de la suite $\vec{i}$ sont indépendants, la probabilité pour que cette suite $\vec{i}$ apparaisse est le produit :
$$p\left(\vec{i}\right) = p_{i_1}.p_{i_2}.\ldots.p_{i_N} \tag{1.2}$$
   Dans l'exemple pile/face, la suite qui est individuellement la plus probable est
$$\vec{i} = (1, 1, 1 \ldots) \tag{1.3}$$
   de probabilité $p\left(\vec{i}\right) = p_1^N = e^{-N \log(1/p_1)} = e^{-N\,0.4}$. La suite la moins probable est
$$\vec{i} = (2, 2, 2, 2 \ldots) \tag{1.4}$$
   de probabilité $p\left(\vec{i}\right) = p_2^N = e^{-N \log(1/p_2)} = e^{-N\,1.1}$.

2. Par définition, on appelle "**suite standard**" une suite $\vec{i} := (i_1, i_2, \ldots, i_N)$ très longue, i.e. $N \gg 1$ et où le nombre d'apparition N_i de chaque chiffre i est $N_i \approx p_i N$ (précisément $N_i = p_i N + o(N)$). Dans l'exemple de pile/face, une suite standard ressemblerait à
$$\vec{i} = (1, 2, 1, 1, 2, 1, 2, 1, 1, 1, 2, 1 \ldots, 1) \tag{1.5}$$
   avec le nombre de 1 : $N_1 \approx p_1 N = \frac{2}{3}N$ et $N_2 \approx \frac{1}{3}N$. Est-ce que la suite la plus probable (respect. moins probable) de (1) sont standard ? Montrer que la probabilité d'apparition d'une suite standard $\vec{i}$ donnée est
$$p\left(\vec{i}\right) = e^{-NS} = \left(e^{-S}\right)^N$$

1. (ref : wikipedia "history of entropy") L'histoire raconte que en 1949 Claude Shannon qui travaillait en télécomunication au BELL labs a développé la théorie qui suit, en définissant une fonction H qu'il voulait appeler "incertitude". Il rencontre le physicien-mathématicien Von Neumann. Ce dernier lui fait remarquer qu'il devrait appeler sa fonction "entropie" pour deux raisons : 1) car cela correspond au concept de l'entropie développée depuis Boltzmann et Gibbs 1872 qui mesure le "désordre", et 2) car il deviendrait célèbre en donnant enfin une explication claire de ce qu'est l'entropie.

avec

$$S := -\sum_{i=1}^{n} p_i \log p_i = \sum_{i=1}^{n} p_i \log(1/p_i)$$

appelée **entropie**. Aide : utiliser que $p\left(\vec{i}\right) = p_{i_1}.p_{i_2}.\dots.p_{i_N} = p_1^{N_1} p_2^{N_2} \dots p_n^{N_n}$. Calculer l'entropie de l'exemple ?

Réponse : Les suites particulières $\vec{i} = (1,1,1\dots)$ et $\vec{i} = (2,2,2,2\dots)$ ne sont pas standard. Pour une suite standard, on écrit

$$\begin{aligned} p\left(\vec{i}\right) &= p_{i_1}.p_{i_2}.\dots.p_{i_N} = p_1^{N_1} p_2^{N_2} \dots p_n^{N_n} \\ &= e^{N_1 \log p_1} \dots e^{N_n \log p_n} \\ &= e^{N(p_1 \log p_1 + \dots p_n \log p_n)} = e^{-NS}. \end{aligned}$$

Pour l'exemple, $S = -\sum_{i=1}^{n} p_i \log p_i = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.63.. > 0$

3. Pour $N \gg 1$, on note $\mathcal{N}_N$ le **nombre de suites standard** de longueur N. Montrer que

$$\mathcal{N}_N \approx e^{NS} \Leftrightarrow S \approx \frac{1}{N}\log \mathcal{N}_N$$

c'est à dire que l'entropie est le taux exponentiel du nombre de suites standard. Aide : montrer que $\mathcal{N}_N = \frac{N!}{N_1!N_2!\dots N_n!}$ et utiliser la formule de Stirling $\ln N! \simeq N\ln N - N + o(N)$ pour $N \gg 1$.

Réponse : On choisit N_1 parmi N puis N_2 parmi $(N - N_1)$ etc. On a donc

$$\begin{aligned} \mathcal{N}_N &= C_N^{N_1} C_{N-N_1}^{N_2} \dots C_{N-(N_1+\dots+N_{n-1})}^{N_n} \\ &= \frac{N!}{(N-N_1)!N_1!}\frac{(N-N_1)!}{(N-(N_1+N_2))!N_2!}\cdots\frac{(N-(N_1+\dots N_{n-1}))!}{(N-(N_1+N_2+\dots N_n))!N_n!} \\ &= \frac{N!}{N_1!N_2!\dots N_n!} \\ &= e^{N\ln N - N + \sum_{i=1}^{n}(-N_i \log N_i + N_i) + o(N)} \\ &= e^{N\ln N - N + \sum_{i=1}^{n}(-p_i N \log(p_i N) + p_i N) + o(N)} \\ &= e^{N\ln N - N + \sum_{i=1}^{n}(-p_i N \log p_i - p_i N \log N + p_i N) + o(N)} \\ &= e^{N\ln N - N + NS - N\log N + N + o(N)} \\ &= e^{NS + o(N)} \\ &= e^{N(S + o(1))} \end{aligned}$$

4. D'après les résultats précédents les suites standard sont équiprobables et de probabilité $p\left(\vec{i}_s\right) = e^{-N(S+o(1))}$. Il y en a $\mathcal{N}_N = e^{N(S+o(1))}$. Cela semblent indiquer[2] que leur probabilité cumulée est $\mathcal{N}_N p\left(\vec{i}_s\right) \approx 1$ pour $N \gg 1$(la preuve sera dans la question 5). Quelle signification des suites standard et de l'entropie S cela suggère t-il ? Pour quelle(s) lois de probabilité $(p_i)_{i=1\to n}$ a t-on $S = \log n$? $S = 0$? Quelle est la valeur maximale de S ? Comparer $\mathcal{N}_N$ et $\mathcal{N}_{\text{tot}}(N)$? Dans quel cas a t-on $\mathcal{N}_N = \mathcal{N}_{\text{tot}}(N)$?

2. En fait, on ne peut pas déduire $\mathcal{N}_N p\left(\vec{i}_s\right) \approx 1$ car on a négligé des corrections $e^{No(1)} = e^{o(N)}$ qui peuvent être non négligeables, par exemple $\ln N = o(N)$ mais $e^{\ln N} = N \gg 1$.

Réponse : Signification des suites standard et de l'entropie : “presque surement” (cad avec probabilité 1) une suite sera une suite standard et l'entropie est le taux exponentiel de leur nombre. L'entropie mesure donc le taux exponentiel de réalisations possibles. L'entropie est une mesure de l'incertitude sur les évènements à venir.

- Pour une loi déterministe comme $p_1 = 1$, $p_{i>1} = 0$, on a $S = -p_1 \log p_1 = 0$. C'est la valeur minimale pour S. Il y a une seule suite standard $\vec{i} = (1, 1, 1 \dots 1)$. On n'a pas d'incertitude sur les évènements à venir qui sont forcément égaux à 1.
- Pour la loi équiprobable $p_i = \frac{1}{n}$, on a $S = -\sum_i p_i \log p_i = -\sum_i \frac{1}{n} \log \frac{1}{n} = \log n$. C'est la valeur maximale. Alors $\mathcal{N}_N = e^{SN} = e^{N \log n} = n^N = \mathcal{N}_{\mathrm{tot}}(N)$. L'incertitude est maximale.
- Dans un cas intermédiaire, $0 < S < \log n$ et donc $\mathcal{N}_N = e^{SN} \ll e^{N \log n} = \mathcal{N}_{\mathrm{tot}}(N)$. Le nombre de suite standard est très petit par rapport au nombre total de suites mais cependant il croit exponentiellement.

5. (Optionnel) Pour montrer que $\mathcal{N}_N p\left(\vec{i}_s\right) \approx 1$ pour $N \gg 1$, écrire $p\left(\vec{i}\right) = \exp\left(-\sum_{j=1}^{N} X_j\right)$ avec $X_j = -\log p_j$ et appliquer la loi des grands nombres.

Réponse : L'astuce est de faire apparaître une somme de variables aléatoires indépendantes. On écrit

$$\begin{aligned} p\left(\vec{i}\right) &= p_{i_1}.p_{i_2}.\dots.p_{i_N} = e^{\sum_{j=1}^{N} \log p_{i_j}} \\ &= e^{-\sum_{j=1}^{N} X_j} \end{aligned}$$

avec

$$X_j := -\log p_{i_j}.$$

Les variables aléatoires X_j sont indépendantes, de même loi, et de moyenne qui est précisément l'entropie :

$$\langle X_j \rangle = \sum_{i=1}^{n} p_i \left(-\log p_i\right) = S$$

Considérons la somme moins la moyenne :

$$Y_N := \frac{1}{N} \sum_{j=1}^{N} \left(X_j - S\right) \tag{1.6}$$

ainsi on peut exprimer :

$$p\left(\vec{i}\right) = e^{-NS(1+Y_N)}$$

La **loi (faible) des grands nombre** (cf Proposition 1.1) dit que Y_N “converge vers 0” au sens ou pour tout $\alpha > 0$ (arbitrairement petit) et pour $N \to \infty$,

$$\mathrm{proba}\left(|Y_N| < \alpha\right) = 1 - o(1) \tag{1.7}$$

Donc on définit les suites standard par

$$\begin{aligned} \mathcal{A}_\alpha : &= \left\{ \vec{i} \quad \mathrm{tq} \ -\alpha < Y_N < \alpha \right\} \\ &= \left\{ \vec{i} \quad \mathrm{tq} \ p\left(\vec{i}\right) \in \left[e^{-NS(1+\alpha)}, e^{-NS(1-\alpha)}\right] \right\} \end{aligned}$$

Cela donne

$$\forall \vec{i} \in \mathcal{A}, \quad e^{-NS(1+o(1))} \leq p\left(\vec{i}\right) \leq e^{-NS(1-o(1))} \tag{1.8}$$

On a d'après la loi des grands nombres

$$\sum_{\vec{i} \in \mathcal{A}_\alpha} p\left(\vec{i}\right) = 1 - o(1)$$

donnant

$$\sum_{\vec{i}\in\mathcal{A}} p\left(\vec{i}\right) = 1 - o(1), \qquad \sum_{\vec{i}\in\mathcal{B}} p\left(\vec{i}\right) = o(1). \tag{1.9}$$

Par conséquence

$$(\sharp\mathcal{A}).e^{-NS(1+\alpha)} \leq 1 - o(1) \quad \text{et} \quad (\sharp\mathcal{A}).e^{-NS(1-\alpha)} \geq 1 + o(1),$$

donnant

$$e^{NS(1-o(1))} \leq \sharp\mathcal{A} \leq e^{NS(1+o(1))}. \tag{1.10}$$

Il reste à justifier la loi des grands nombres utilisée dans (1.7).

Proposition 1.1. *"La loi faible des grands nombres". Supposons que $(X_j)_{j\in\mathbb{N}}$ sont des variables aléatoires indépendantes, de même moyenne $\langle X\rangle$ et même variance $\sigma^2 = \left\langle (X - \langle X\rangle)^2\right\rangle$. Pour $N \geq 1$, on pose*

$$Y_N := \frac{1}{N}\sum_{j=1}^{N} (X_j - \langle X\rangle)$$

qui est la "moyenne temporelle centrée". Alors pour tout $\alpha > 0$ (arbitrairement petit) et pour $N \to \infty$,

$$\text{proba}\,(|Y_N| > \alpha) = o(1)$$

Démonstration. Notons pour simplifier $Z = |Y_N|^2$ et $p(Z)\,dZ$ la densité de probabilité. On a

$$\begin{aligned}
\text{proba}\,(|Y_N| > \alpha) &= \text{proba}\left(|Y_N|^2 > \alpha^2\right) \\
&= \int_{\alpha^2}^{\infty} p(Z)\,dZ \leq \int_{\alpha^2}^{\infty}\left(\frac{Z}{\alpha^2}\right) p(Z)\,dZ \\
&= \frac{1}{\alpha^2}\int_{\alpha^2}^{\infty} Zp(Z)\,dZ \leq \frac{1}{\alpha^2}\int_{0}^{\infty} Zp(Z)\,dZ \\
&= \frac{1}{\alpha^2}\langle Z\rangle = \frac{1}{\alpha^2}\left\langle |Y_N|^2\right\rangle
\end{aligned}$$

où $\langle . \rangle$ désigne la moyenne. (Ce dernier calcul s'appelle "inégalité de Tchebychev"). On a $\langle (X_j - \langle X\rangle)\rangle = \langle X\rangle - \langle X\rangle = 0$ et donc pour $j \neq k$ $\langle (X_j - \langle X\rangle)(X_k - \langle X\rangle)\rangle = \langle (X_j - \langle X\rangle)\rangle\langle (X_k - \langle X\rangle)\rangle = 0$. Alors d'après (1.6)

$$\begin{aligned}
\left\langle |Y_N|^2\right\rangle &= \frac{1}{N^2}\sum_{j,k}\langle (X_j - \langle X\rangle)(X_k - \langle X\rangle)\rangle = \frac{1}{N^2}\sum_{j}\left\langle (X_j - \langle X\rangle)^2\right\rangle \\
&= \frac{1}{N}\sigma^2 \to 0
\end{aligned}$$

On déduit que $\text{proba}\,(|Y_N| > \alpha) = o(1)$. $\square$