Leçon 9 Exercices corrigés

(*Une étoile * désignera une question de difficulté supérieure.*)

Exercice 1. Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs réelles positives ou nulles. Démontrer que

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq t)dt.$$

(*Indication* : utiliser par exemple le théorème de Tonelli.)

Démontrer que $\mathbb{E}(X) < \infty$ si et seulement si pour un, ou tout, $\varepsilon > 0$, $\sum_{n \in \mathbb{N}} \mathbb{P}(X \geq \varepsilon n) < \infty$. En déduire que si $\mathbb{E}(X) < \infty$ alors $\lim_{t \to \infty} t\,\mathbb{P}(X \geq t) = 0$.

Corrigé. D'après la définition de l'intégrale d'une fonction indicatrice,

$$\mathbb{P}(X \geq t) = \int_\Omega \mathbb{1}_{\{X \geq t\}}\, d\mathbb{P}.$$

Donc, par le théorème de Fubini-Tonelli, toutes les variables étant positives ou nulles,

$$\int_0^\infty \mathbb{P}(X \geq t)dt = \int_0^\infty \left(\int_\Omega \mathbb{1}_{\{X \geq t\}} d\mathbb{P} \right) dt = \int_\Omega \left(\int_0^\infty \mathbb{1}_{\{X \geq t\}} dt \right) d\mathbb{P}.$$

Or $\mathbb{1}_{\{X \geq t\}} = \mathbb{1}_{[0,X]}(t)$ de sorte que

$$\int_0^\infty \mathbb{1}_{\{X \geq t\}} dt = \int_0^\infty \mathbb{1}_{[0,X]}(t)dt = X,$$

d'où le résultat. Il est à noter que $\int_0^\infty \mathbb{P}(X \geq t)dt$ est une intégrale impropre au sens de Riemann de la fonction décroissante bornée $t \to \mathbb{P}(X \geq t)$.

Quitte à remplacer X par $\frac{1}{\varepsilon}X$, il suffit de considérer $\varepsilon = 1$. Par une comparaison entre intégrale et série,

$$\begin{aligned}\int_0^\infty \mathbb{P}(X \geq t)dt &= \int_0^1 \mathbb{P}(X \geq t)dt + \sum_{n=0}^\infty \int_{2^n}^{2^{n+1}} \mathbb{P}(X \geq t)dt \\ &\leq 1 + \sum_{n=0}^\infty \mathbb{P}(X \geq 2^n) \int_{2^n}^{2^{n+1}} 1\, dt \\ &= 1 + \sum_{n=0}^\infty 2^n\, \mathbb{P}(X \geq 2^n),\end{aligned}$$

et de façon similaire pour la minoration. L'équivalence souhaitée en découle en vertu du premier point. Si $\mathbb{E}(X) < \infty$, la série précédente est donc convergente, et ainsi son terme général tend vers 0, c'est-à-dire $\lim_{n\to\infty} 2^n\, \mathbb{P}(X \geq 2^n) = 0$. Ainsi, pour tout $\eta > 0$, il existe n_0 tel que si $n \geq n_0$, $2^n\, \mathbb{P}(X \geq 2^n) \leq \eta$. Pour tout $t \geq 2$ réel, il existe un entier n tel que $2^n \leq t < 2^{n+1}$ de sorte que, par croissance,

$$t\, \mathbb{P}(X \geq t) \leq 2^{n+1}\, \mathbb{P}(X \geq 2^n).$$

Donc si $t \geq 2^{n_0}$, $t\, \mathbb{P}(X \geq t) \leq 2\eta$, ce qui permet de conclure que $\lim_{t\to\infty} t\, \mathbb{P}(X \geq t) = 0$.

Exercice 2. Soit X une variable aléatoire de loi uniforme sur $[-1, +1]$; poser $Y = |2X - 1|$. Calculer $\mathbb{E}(Y)$, $\mathbb{E}(X + Y)$, $\mathbb{E}((3Y + X^2)^2)$.

Indication. Par le théorème de transport,

$$\mathbb{E}(Y) = \mathbb{E}\big(|2X - 1|\big) = \frac{1}{2} \int_{[-1,+1]} |2x - 1|\, d\lambda$$

puisque la loi de X a pour densité $\frac{1}{2}\mathbb{1}_{[-1,+1]}$ par rapport à la mesure de Lebesgue λ sur $\mathbb{R}$.

Exercice 3. Soit X une variable aléatoire de loi uniforme sur l'intervalle $[-4, 2]$. Déterminer la loi de la variable aléatoire $Y = \frac{1}{2}(X + 3 + |X - 3|)$. Calculer l'espérance et la variance de Y de deux façons différentes.

Indications. Si a est un nombre réel, $a + |a| = 2a_+ = 2\max(a, 0)$. Ainsi $Y = \max(X + 3, 0)$ (et la loi de $X + 3$ est uniforme sur $[-1, 5]$). Le calcul de $\mathbb{E}(Y)$ et $\mathrm{Var}(Y)$ peut se faire à partir de la loi de Y ou, plus efficacement, directement par le théorème de transport pour X.

Exercice 4. Si x est un réel positif, $\lfloor x \rfloor$ désigne sa partie entière et $x - \lfloor x \rfloor$ sa partie décimale.

Soit X_n une variable aléatoire de loi uniforme $\mathcal{U}(0, n)$ sur l'intervalle $[0, n]$, où n est un entier ≥ 1.

a) Démontrer que $\mathbb{E}(X_n) = n\,\mathbb{E}(X_1)$ et calculer $\mathbb{E}(X_1)$.

b) Quelle est la loi de $\lfloor X_n \rfloor + 1$? de $X_n - \lfloor X_n \rfloor$? de $\lfloor X_n \rfloor + 1 - X_n$?

c) Soit Y_n une variable aléatoire de loi uniforme $\mathcal{U}(\{1, \ldots, n\})$ sur $\{1, \ldots, n\}$. Justifier que $\mathbb{E}(Y_n) = \mathbb{E}(X_n) + \mathbb{E}(X_1)$. En déduire la formule pour $\sum_{k=1}^{n} k$.

Corrigé. a) Comme X_n a pour densité $\frac{1}{n}\mathbb{1}_{[0,n]}$ par rapport à la mesure de Lebesgue λ sur $\mathbb{R}$, par le théorème de transport, pour tout $n \geq 1$,

$$\mathbb{E}(X_n) = \frac{1}{n}\int_{[0,n]} x\,d\lambda = n\int_{[0,1]} y\,d\lambda = n\,\mathbb{E}(X_1)$$

après le changement de variable $y = \frac{x}{n}$. Par ailleurs $\mathbb{E}(X_1) = \int_{[0,1]} x\,d\lambda = \frac{1}{2}$.
b) La variable aléatoire $\lfloor X_n \rfloor + 1$ prend (presque sûrement car $\mathbb{P}(X_n = n) = 0$) ses valeurs dans $\{1, \ldots, n\}$. Si $k \in \{1, \ldots, n\}$,

$$\mathbb{P}\big(\lfloor X_n \rfloor + 1 = k\big) = \mathbb{P}(k - 1 \leq X_n < k) = \frac{1}{n}$$

puisque X_n est uniforme sur $[0, n]$. Ainsi $\lfloor X_n \rfloor + 1$ suit la loi uniforme sur $\{1, \ldots, n\}$. La variable aléatoire $X_n - \lfloor X_n \rfloor$ prend (presque sûrement) ses valeurs dans $[0,1[$. Pour tout intervalle (a,b) de $[0,1[$,

$$\begin{aligned}
\mathbb{P}\big(X_n - \lfloor X_n \rfloor \in (a,b)\big) &= \sum_{k=1}^{n} \mathbb{P}\big(X_n \in [k-1,k[\,, X_n - \lfloor X_n \rfloor \in (a,b)\big) \\
&= \sum_{k=1}^{n} \mathbb{P}\big(X_n \in [k-1,k[\,, X_n \in (a+k-1, b+k-1)\big) \\
&= \sum_{k=1}^{n} \mathbb{P}\big(X_n \in (a+k-1, b+k-1).\big)
\end{aligned}$$

Comme la loi de X_n est uniforme sur $[0,n]$, il s'ensuit que

$$\mathbb{P}\big(X_n - \lfloor X_n \rfloor \in (a,b)\big) = \sum_{k=1}^{n} \frac{1}{n}(b-a) = (b-a).$$

Ceci ayant lieu pour tous les intervalles $(a,b) \subset [0,1[$, la loi de $X_n - \lfloor X_n \rfloor$ est donc la mesure uniforme (de Lebesgue) sur $[0,1[$. Il en va de même de $\lfloor X_n \rfloor + 1 - X_n$. c) Comme

$$\lfloor X_n \rfloor + 1 = X_n + \big(\lfloor X_n \rfloor + 1 - X_n\big),$$

en prenant l'espérance il s'ensuit que $\mathbb{E}(Y_n) = \mathbb{E}(X_n) + \mathbb{E}(X_1)$ puisque $\lfloor X_n \rfloor + 1$ suit la loi uniforme sur $\{1, \ldots, n\}$ et $\lfloor X_n \rfloor + 1 - X_n$ la loi uniforme sur $[0,1[$. Mais $\mathbb{E}(Y_n) = \frac{1}{n}\sum_{k=1}^{n} k$ et $\mathbb{E}(X_n) + \mathbb{E}(X_1) = (n+1)\,\mathbb{E}(X_1) = \frac{1}{2}(n+1)$. Donc $\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$.

Exercice 5 (*Loi symétrique.*) Une variable aléatoire réelle X est dite de loi symétrique, ou plus simplement symétrique, si $-X$ a même loi que X.

a) Vérifier que la loi normale $\mathcal{N}(0, \sigma^2)$ est symétrique ; donner d'autres exemples.

b) Démontrer que si X est une variable aléatoire de carré intégrable symétrique,

$$\mathbb{E}(X^2) \,=\, 4 \int_0^\infty t\, \mathbb{P}(X \geq t) dt.$$

(*Indication* : $\mathbb{P}(|X| \geq t) = 2\, \mathbb{P}(X \geq t)$, $t > 0$.)

Exercice 6. Si X est une variable aléatoire de carré intégrable, démontrer que

$$\mathrm{Var}(X) \,=\, \inf_{a \in \mathbb{R}} \mathbb{E}\big([X - a]^2\big).$$

Soient $a_1 \leq a_2 \leq \cdots \leq a_n$ des réels, et $m = \frac{1}{n}(a_1 + \cdots + a_n)$; démontrer que

$$\frac{1}{n} \sum_{k=1}^{n} (a_k - m)^2 \,\leq\, \frac{1}{4}\, (a_n - a_1)^2.$$

(*Indication* : trouver la variable aléatoire X et le réel a !)

Corrigé. En développant le carré, par linéarité de l'espérance, pour tout $a \in \mathbb{R}$,

$$\mathbb{E}\big([X - a]^2\big) \,=\, \mathbb{E}(X^2) - 2a\, \mathbb{E}(X) + \big[\mathbb{E}(X)\big]^2.$$

Le minimum du trinôme en $a \in \mathbb{R}$ est atteint pour $a = \mathbb{E}(X)$ ce qui démontre la première affirmation. La forme de m suggère une moyenne, et donc une variable aléatoire X de loi $\mathbb{P}(X = a_k) = \frac{1}{n}$, $k = 1, \ldots, n$. Comme $\mathbb{E}(X) = m$,

$$\mathrm{Var}(X) \,=\, \mathbb{E}\Big(\big[X - \mathbb{E}(X)\big]^2\Big) \,=\, \frac{1}{n} \sum_{k=1}^{n} (a_k - m)^2.$$

Donc, d'après la première question, pour tout $a \in \mathbb{R}$,

$$\frac{1}{n}\sum_{k=1}^{n}(a_k - m)^2 \;\leq\; \frac{1}{n}\sum_{k=1}^{n}(a_k - a)^2.$$

Si alors $a = \frac{1}{2}\,(a_1 + a_n)$, pour tout $k = 1, \ldots, n$,

$$\frac{1}{2}(a_1 - a_n) \;\leq\; a_k - a \;\leq\; \frac{1}{2}(a_n - a_1)$$

puisque $a_1 \leq \cdots \leq a_k \leq \cdots \leq a_n$. La conclusion annoncée s'ensuit.

Exercice 7*. Soit X une variable aléatoire intégrable et M une médiane de la loi de X ; l'objectif de l'exercice est d'établir que

$$\mathbb{E}\big(|X - M|\big) \;=\; \inf_{a \in \mathbb{R}} \mathbb{E}\big(|X - a|\big).$$

a) Si Z est une variable aléatoire intégrable, démontrer que

$$\mathbb{E}\big(|Z|\big) \;=\; \int_0^\infty \mathbb{P}(Z \geq t)dt + \int_{-\infty}^0 \mathbb{P}(Z \leq t)dt.$$

b) Déduire de la question précédente que si $a < b$,

$$\mathbb{E}(|X - b|) - \mathbb{E}(|X - a|) \;=\; \int_a^b \psi(t)dt$$

où $\psi(t) = \mathbb{P}(X \leq t) - \mathbb{P}(X \geq t)$, $t \in \mathbb{R}$.

c) Vérifier que si $t > M$, $\psi(t) \geq 0$, et si $t < M$, $\psi(t) \leq 0$.

d) Conclure au résultat annoncé.

Corrigé. a) D'après l'Exercice 1,

$$\begin{aligned}
\mathbb{E}\big(|Z|\big) &= \int_0^\infty \mathbb{P}\big(|Z| \geq t\big)dt \\
&= \int_0^\infty \mathbb{P}(Z \geq t)dt + \int_0^\infty \mathbb{P}(-Z \geq t)dt \\
&= \int_0^\infty \mathbb{P}(Z \geq t)dt + \int_{-\infty}^0 \mathbb{P}(Z \leq t)dt.
\end{aligned}$$

b) En appliquant cette formule à $Z = X - b$ et $Z = X - a$, après des changements de variables, la différence des deux expressions fournit

$$\mathbb{E}(|X - b|) - \mathbb{E}(|X - a|) = \int_a^b \psi(t)dt$$

comme annoncé. c) Par définition d'une médiane M de X, $\mathbb{P}(X \leq M) \geq \frac{1}{2}$ et $\mathbb{P}(X \geq M) \geq \frac{1}{2}$. Donc

$$\psi(t) = \mathbb{P}(X \leq t) + \mathbb{P}(X < t) - 1 \geq 0$$

si $t > M$ et

$$\psi(t) = 1 - \mathbb{P}(X > t) - \mathbb{P}(X \geq t) \leq 0$$

si $t < M$. d) En conséquence, si $a > M$,

$$\mathbb{E}\big(|X - a|\big) - \mathbb{E}\big(|X - M|\big) = \int_M^a \psi(t)dt \geq 0$$

et si $a < M$,

$$\mathbb{E}\big(|X - M|\big) - \mathbb{E}\big(|X - a|\big) = \int_a^M \psi(t)dt \leq 0.$$

Ainsi, dans tous les cas, $\mathbb{E}(|X - M|) \leq \mathbb{E}(|X - a|)$, ce qui établit le résultat.

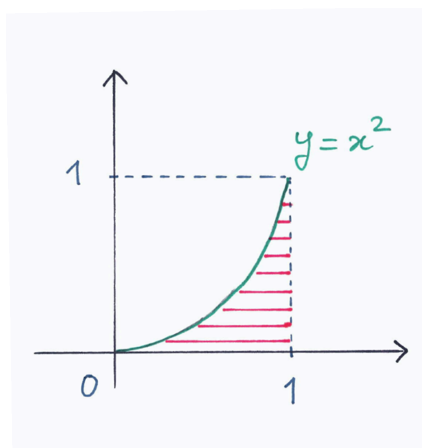
Exercice 8. Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs positives, intégrable ; démontrer que

$$\big[1 + \mathbb{E}(X)\big]\, \mathbb{E}\Big(\frac{1}{1+X}\Big) \;\geq\; 1.$$

Indication. Utiliser l'inégalité de Jensen pour la fonction convexe $\phi(x) = \frac{1}{1+x}$, $x \in \mathbb{R}_+$.

Exercice 9. Soit (X, Y) un vecteur aléatoire dans $\mathbb{R}^2$ dont la loi a pour densité par rapport à la mesure de Lebesgue $\lambda^2 = \lambda \otimes \lambda$ sur $\mathbb{R}^2$,

$$f(x,y) = \frac{4y}{x^3}\, \mathbb{1}_{\{0<x<1, 0<y<x^2\}}, \quad (x,y) \in \mathbb{R}^2.$$



Déterminer les lois de X et Y. Calculer $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, $\mathrm{Cov}(X, Y)$.

Corrigé. La loi de X a pour densité $x \mapsto \int_{\mathbb{R}} f(x,y) d\lambda(y)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$, soit

$$\int_{\mathbb{R}} f(x,y) d\lambda(y) \;=\; \bigg(\frac{1}{x^3} \int_{]0,x^2[} 4y\, d\lambda(y)\bigg)\, \mathbb{1}_{]0,1[}(x) \;=\; 2x\, \mathbb{1}_{]0,1[}(x).$$

De la même manière, la loi de Y a pour densité $y \mapsto \int_{\mathbb{R}} f(x,y)d\lambda(x)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$, soit

$$\int_{\mathbb{R}} f(x,y)d\lambda(x) \;=\; \left(4y \int_{]\sqrt{y},1[} \frac{1}{x^3}\, d\lambda(x)\right) \mathbb{1}_{]0,1[}(x) \;=\; 2(1-y)\, \mathbb{1}_{]0,1[}(y).$$

$\mathbb{E}(X) = \frac{2}{3}$, $\mathbb{E}(Y) = \frac{1}{3}$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{1}{18}$, $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y) = \frac{2}{45}$.

Exercice 10. Si X et Y sont deux variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ de carré intégrable, démontrer que

$$\mathrm{Cov}(X,Y) \;\leq\; \sigma(X)\,\sigma(Y),$$

où il est rappelé que $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ et $\sigma(Y) = \sqrt{\mathrm{Var}(Y)}$ sont les écarts-types de X et Y respectivement (et de même pour $|\mathrm{Cov}(X,Y)|$).

Déduire de cet exercice que si A et B sont des événements quelconques de $\mathcal{A}$,

$$\big|\mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B)\big| \;\leq\; \frac{1}{4}\,.$$

Donner un exemple pour lequel la valeur $\frac{1}{4}$ est atteinte.

Corrigé. Utiliser l'inégalité de Cauchy-Schwarz. Dans l'application, choisir $X = \mathbb{1}_A$ et $Y = \mathbb{1}_B$ et vérifier que $\mathrm{Var}(\mathbb{1}_A) = \mathbb{P}(A)(1-\mathbb{P}(A)) \leq \frac{1}{4}$.

Il peut être intéressant de s'interroger, de la même façon, sur les valeurs minimale et maximale de

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) - \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$$

lorsque $A_1, \ldots, A_n$ parcourent une tribu $\mathcal{A}$ pour un entier $n \geq 2$ quelconque. La démonstration est un peu plus approfondie. Quelques exemples donnent à penser que la valeur maximale pourrait être atteinte lorsque tous les A_k, $k = 1, \ldots, n$, sont égaux (pour rendre $\mathbb{P}(A_1 \cap \cdots \cap A_n)$ le plus grand possible).

Si $A_k \in \mathcal{A}$, $k = 1, \ldots, n$, sont quelconques, et $A = A_1 \cap \cdots \cap A_n$, il est naturel alors de minorer chaque $\mathbb{P}(A_k)$ par $\mathbb{P}(A)$ (puisque $A_k \supset A$), de sorte que

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) - \mathbb{P}(A_1) \cdots \mathbb{P}(A_n) \,\leq\, \mathbb{P}(A) - \mathbb{P}(A)^n.$$

Une optimisation simple montre que le maximum de la fonction $p - p^n$ lorsque $p \in [0,1]$ est atteint en $\frac{1}{n^{1/(n-1)}}$, valeur pour laquelle il vaut $\frac{n-1}{n^{n/(n-1)}}$ ($= \frac{1}{4}$ si $n = 2$). Le maximum est précisément atteint si tous les A_k, $k = 1, \ldots, n$, sont égaux de probabilité $\frac{1}{n^{1/(n-1)}}$. (Noter que l'espace de probabilité devra être suffisamment riche pour contenir des événements atteignant la valeur souhaitée, comme l'espace $[0,1]$ muni de la mesure de Lebesgue, un espace discret pourrait ne pas convenir.) La quantité $\frac{n-1}{n^{n/(n-1)}}$ tend vers 1 quand $n \to \infty$.

D'autres exemples indiquent que le minimum pourrait être obtenu lorsque $\mathbb{P}(A) = 0$, et même de façon ensembliste, $A = A_1 \cap \cdots \cap A_n = \emptyset$. Fort de cette intuition, pour tous $A_1, \ldots, A_n \in \mathcal{A}$, par passage au complémentaire d'une intersection,

$$1 - \mathbb{P}(A) \,=\, \mathbb{P}(A^c) \,=\, \mathbb{P}\big((A_1 \cap \cdots \cap A_n)^c\big) \,=\, \mathbb{P}(A_1^c \cup \cdots \cup A_n^c).$$

Puis, par sous-additivité d'une mesure de probabilité,

$$1 - \mathbb{P}(A) \,\leq\, \mathbb{P}(A_1^c) + \cdots + \mathbb{P}(A_n^c).$$

Comme $\mathbb{P}(A_k^c) = 1 - \mathbb{P}(A_k)$, $k = 1, \ldots, n$, il résulte de cette inégalité que

$$\mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) \,\leq\, n - (1 - p)$$

où $p = \mathbb{P}(A)$. D'après l'inégalité entre moyenne arithmétique et moyenne géométrique (conséquence de la concavité de la fonction logarithme),

$$\frac{1}{n}\big[\mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n)\big] \,\geq\, \big[\mathbb{P}(A_1) \cdots \mathbb{P}(A_n)\big]^{\frac{1}{n}}.$$

Avec l'inégalité précédente, en réorganisant les termes, il vient que

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) - \mathbb{P}(A_1) \cdots \mathbb{P}(A_n) \,\geq\, p - \Big(\frac{n - (1 - p)}{n}\Big)^n.$$

La fonction de $p \in [0,1]$ à droite de l'inégalité est croissante en p, donc plus grande que sa valeur en 0 égale à $-\big(1-\frac{1}{n}\big)^n$ $(= -\frac{1}{4}$ si $n=2)$. Cette valeur est précisément atteinte si les événements A_k^c, $k=1,\ldots,n$, forment une partition de Ω et vérifient $\mathbb{P}(A_k^c) = \frac{1}{n}$, $k=1,\ldots,n$, situation dans laquelle $p=0$ et toutes les inégalités précédentes sont des égalités. Lorsque $n \to \infty$, le minimum tend vers $-\frac{1}{e}$.

Exercice 11 (*Entropie*). Si P est une mesure de probabilité sur $\{1,\ldots,n\}$,

$$\mathrm{H}(P) \;=\; -\sum_{k=1}^{n} p_k \ln(p_k),$$

où $p_k = P(\{k\})$ avec la convention $0\ln(0)=0$, désigne l'entropie de P.

a) Montrer que H est à valeurs dans $\mathbb{R}_+$. Trouver P telle que $\mathrm{H}(P)=0$. Montrer que la mesure uniforme $\mathcal{U}(\{1,\ldots,n\})$ sur $\{1,\ldots,n\}$ réalise le maximum de H.

b) Si P est une mesure de probabilité sur $\mathbb{N}$, $\mathrm{H}(P) = -\sum_{n\in\mathbb{N}} p_n \ln(p_n)$ désigne de même l'entropie de P, éventuellement infinie. Montrer que la loi géométrique $\mathcal{G}(p)$ de paramètre $p \in]0,1[$ réalise le maximum de l'entropie sur l'ensemble des mesures de probabilité sur $\mathbb{N}$ de moyenne inférieure ou égale à $\frac{1}{p}$. (*Indication* : écrire $p_n \ln(p_n) = p_n \ln(\frac{p_n}{(1-p)^{n-1}}) + p_n \ln((1-p)^{n-1})$.)

c) Si P est une mesure de probabilité sur les boréliens de $\mathbb{R}$, de densité f par rapport à la mesure de Lebesgue, soit l'entropie

$$\mathrm{H}(P) \;=\; -\int_{\mathbb{R}} f \ln(f) d\lambda$$

lorsque cette intégrale a un sens, $\mathrm{H}(P) = +\infty$ sinon. Calculer l'entropie de la mesure gaussienne standard $\mathcal{N}(0,1)$ et démontrer qu'elle maximise l'entropie de toute mesure de densité f vérifiant $\int_{\mathbb{R}} x^2 f(x) d\lambda \le 1$. (*Indication* : vérifier que pour toute densité g, $\int_{\mathbb{R}} f \ln(\frac{f}{g}) d\lambda \ge 0$, et appliquer le résultat à la densité gaussienne.)

Corrigé. a) Comme $p_k \in [0,1]$, $k = 1, \ldots, n$, $\mathrm{H}(P) \geq 0$. Si P est concentrée en un point, par exemple $P = \delta_1$, alors $p_1 = 1$ et $p_2 = \cdots = p_n = 0$. Pour cette mesure, $\mathrm{H}(P) = 0$. Si $P = \mathcal{U}(\{1, \ldots, n\})$ est la mesure de probabilité uniforme sur $\{1, \ldots, n\}$, $p_k = \frac{1}{n}$ pour tout k et donc $\mathrm{H}(\mathcal{U}(\{1, \ldots, n\})) = \ln(n)$. Si P est une autre probabilité de poids p_k, $k = 1, \ldots, n$, puisque $\sum_{k=1}^n p_k = 1$, la concavité de la fonction logarithme entraîne que

$$\mathrm{H}(P) \;=\; \sum_{k=1}^n p_k \ln\left(\frac{1}{p_k}\right) \;\leq\; \ln\left(\sum_{k=1}^n p_k \cdot \frac{1}{k}\right) \;=\; \ln(n) \;=\; \mathrm{H}\big(\mathcal{U}(\{1, \ldots, n\})\big).$$

b) L'entropie de la loi géométrique $\mathcal{G}(p)$ est donnée par

$$\begin{aligned}
\mathrm{H}\big(\mathcal{G}(p)\big) \;&=\; -\sum_{n\in\mathbb{N}} (1-p)^{n-1} p \ln\big((1-p)^{n-1}p\big) \\
&=\; -\ln(1-p) \sum_{n\in\mathbb{N}} (n-1)(1-p)^{n-1}p - \ln(p) \\
&=\; -\big(\tfrac{1}{p} - 1\big)\ln(1-p) - \ln(p)
\end{aligned}$$

puisque $\sum_{n\in\mathbb{N}} n(1-p)^{n-1}p = \frac{1}{p}$ est la moyenne de $\mathcal{G}(p)$. Si P est une autre probabilité de poids p_n, $n \in \mathbb{N}$, et de moyenne $\sum_{n\in\mathbb{N}} np_n \leq \frac{1}{p}$, écrire pour tout $n \in \mathbb{N}$ comme indiqué

$$-p_n \ln(p_n) \;=\; p_n \ln\left(\frac{(1-p)^{n-1}}{p_n}\right) - p_n \ln\big((1-p)^{n-1}\big)$$

de sorte que

$$\mathrm{H}(P) \;=\; \sum_{n\in\mathbb{N}} p_n \ln\left(\frac{(1-p)^{n-1}}{p_n}\right) - \sum_{n\in\mathbb{N}} p_n \ln\big((1-p)^{n-1}\big).$$

À nouveau en vertu de la concavité de la fonction logarithme,

$$\mathrm{H}(P) \;\leq\; \ln\left(\sum_{n\in\mathbb{N}} (1-p)^{n-1}\right) - \sum_{n\in\mathbb{N}} p_n \ln\big((1-p)^{n-1}\big).$$

Comme $\sum_{n\in\mathbb{N}} p_n = 1$ et $\sum_{n\in\mathbb{N}} np_n \leq \frac{1}{p}$, il vient après quelques détails

$$\mathrm{H}(P) \;\leq\; -\ln(p) - \big(\tfrac{1}{p} - 1\big)\ln(1-p) \;=\; \mathrm{H}\big(\mathcal{G}(p)\big)$$

comme souhaité. c) Si P est $\mathcal{N}(0,1)$, alors $f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$, $x \in \mathbb{R}$, de sorte que

$$\mathrm{H}\big(\mathcal{N}(0,1)\big) \;=\; \frac{1}{2}\int_{\mathbb{R}} x^2 dP + \ln\big(\sqrt{2\pi}\,\big) \;=\; \frac{1}{2}\big(1 + \ln(2\pi)\big).$$

Si f et g sont des densités de probabilité, d'après l'inégalité de Jensen pour la fonction convexe $-\ln$ et la mesure de probabilité $f d\lambda$,

$$\int_{\mathbb{R}} f \ln\Big(\frac{f}{g}\Big) d\lambda \;=\; \int_{\mathbb{R}} (-\ln)\Big(\frac{g}{f}\Big) f\, d\lambda \geq -\ln\bigg(\int_{\mathbb{R}} g\, d\lambda\bigg) \;=\; 0$$

(car $\int_{\mathbb{R}} g\, d\lambda = 1$). Autrement dit $\int_{\mathbb{R}} f\ln(f) d\lambda \geq \int_{\mathbb{R}} f \ln(g) d\lambda$. Si alors P est de densité f quelconque par rapport à λ, vérifiant $\int_{\mathbb{R}} x^2 f d\lambda \leq 1$, et si $g(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$, $x \in \mathbb{R}$, est la densité de la loi $\mathcal{N}(0,1)$,

$$\begin{aligned}
\mathrm{H}(P) \;&=\; -\int_{\mathbb{R}} f \ln(f) d\lambda \\
&\leq\; -\int_{\mathbb{R}} f \ln(g) d\lambda \\
&=\; \frac{1}{2}\int_{\mathbb{R}} \big(x^2 + \ln(2\pi)\big) f\, d\lambda \\
&\leq\; \frac{1}{2}\big(1 + \ln(2\pi)\big) \;=\; \mathrm{H}\big(\mathcal{N}(0,1)\big)
\end{aligned}$$

puisque $\int_{\mathbb{R}} x^2 f d\lambda \leq 1$.