DS7 correction (Maths II) ESSEC II 2019

Un modèle probabiliste d'une expérience aléatoire représente dans un certain sens le désordre qui intervient dans l'expérience et il est donc naturel que des outils soient introduits qui permettent de mesurer l'intensité de ce désordre. C'est le cas de la notion d'entropie qui fait l'objet du présent problème. On considèrera différentes situations et notamment la façon dont on mesure l'information que deux variables aléatoires s'apportent mutuellement.
Dans la première partie on étudie le cas plus simple techniquement de variables dont la loi admet une densité. Les deuxièmes et troisièmes parties sont consacrées au cas discret. Dans la deuxième partie, on introduit les différentes notions d'entropie pour le cas de variables discrètes et dans la troisième partie, on examine comment on peut mesurer l'information apportée mutuellement par deux variables aléatoires.
Toutes les variables aléatoires intervenant dans le problème sont définies sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$. Pour toute variable aléatoire Y, on notera $\mathbb{E}(Y)$ son espérance lorsqu'elle existe.

Première partie : Entropie différentielle d'une variable à densité

1. La fonction logarithme de base 2, notée $\log_2$, est définie sur $\mathbb{R}_+^*$ par : $\log_2(x) = \dfrac{\ln(x)}{\ln(2)}$.

a) Montrer que pour tout (x, y) élément de $\mathbb{R}_+^* \times \mathbb{R}_+^*$, on a : $\log_2(x\,y) = \log_2(x) + \log_2(y)$.

Démonstration.
Soit $(x, y) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$.

$$\begin{aligned}
\log_2(x\,y) &= \frac{\ln(x\,y)}{\ln(2)} \\
&= \frac{\ln(x) + \ln(y)}{\ln(2)} && \textit{(par propriété de la fonction } \ln\textit{)} \\
&= \frac{\ln(x)}{\ln(2)} + \frac{\ln(y)}{\ln(2)} \\
&= \log_2(x) + \log_2(y)
\end{aligned}$$

$$\boxed{\forall (x, y) \in \mathbb{R}_+^* \times \mathbb{R}_+^*,\ \log_2(x\,y) = \log_2(x) + \log_2(y)}$$

□

b) Vérifier que pour tout réel α : $\log_2\left(2^\alpha\right) = \alpha$.

Démonstration.
Soit $\alpha \in \mathbb{R}$.

$$\log_2\left(2^\alpha\right) = \frac{\ln\left(2^\alpha\right)}{\ln(2)} = \frac{\alpha \cancel{\ln(2)}}{\cancel{\ln(2)}} = \alpha$$

$$\boxed{\forall \alpha \in \mathbb{R},\ \log_2\left(2^\alpha\right) = \alpha}$$

□

c) Montrer que la fonction $\log_2$ est concave sur $\mathbb{R}_+^*$.

Démonstration.
La fonction $\log_2$ est de classe $\mathcal{C}^2$ sur $\mathbb{R}_+^*$ car la fonction ln l'est.
Soit $x \in \mathbb{R}_+^*$.

$$(\log_2)'(x) = \frac{1}{\ln(2)} \frac{1}{x}$$

$$(\log_2)''(x) = \frac{1}{\ln(2)} \frac{-1}{x^2} < 0$$

On en déduit que la fonction $\log_2$ est concave sur $\mathbb{R}_+^*$.

Commentaire

- La fonction $\log_2$ est, à constante (positive) multiplicative près, égale à la fonction ln. De ce fait, la fonction $\log_2$ hérite de certaines propriétés de la fonction ln. C'est évidemment le cas du caractère régulier (ces deux fonctions sont de classes $\mathcal{C}^\infty$ sur $]0, +\infty[$), de la monotonie (stricte croissance sur $]0, +\infty[$) et de la concavité.
- On utilise ici la caractérisation de la convexité pour les fonctions deux fois dérivables. C'est la caractérisation la plus simple et donc celle qu'il faut utiliser dès que la fonction considérée est suffisamment régulière. Rappelons toutefois que la notion de convexité est une notion géométrique qui n'exige pas d'hypothèse de régularité. Plus précisément, une focntion est convexe si pour tout couple de points (A, B) situés sur la courbe représentative de f, l'arc de courbe joignant A à B est situé en dessous de la corde de f joignant A à B. Cette propriété s'exprime naturellement à l'aide d'une inégalité, à savoir :

$$\text{La fonction } \log_2 \text{ est concave sur } \mathbb{R}_+^*$$

$$\Updownarrow$$

$$\forall \lambda \in [0,1],\ \forall (x,y) \in \mathbb{R}_+^* \times \mathbb{R}_+^*,\ \log_2\big(\lambda\, x + (1-\lambda) y\big) \geqslant \lambda \log_2(x) + (1-\lambda)\log_2(y)$$

- À l'aide de cette définition, il est d'ailleurs simple de démontrer que la fonction $\log_2$ hérite du caractère concave de ln.
  Pour ce faire, considérons $\lambda \in [0,1]$ et $(x,y) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$. On a alors :

$$\begin{aligned}
&\log_2\big(\lambda\, x + (1-\lambda) y\big) \geqslant \lambda \log_2(x) + (1-\lambda)\log_2(y) \\
\Leftrightarrow\ &\frac{\ln\big(\lambda\, x + (1-\lambda) y\big)}{\ln(2)} \geqslant \lambda \frac{\ln(x)}{\ln(2)} + (1-\lambda)\frac{\ln(y)}{\ln(2)} \\
\Leftrightarrow\ &\ln\big(\lambda\, x + (1-\lambda) y\big) \geqslant \lambda \ln(x) + (1-\lambda)\ln(y) \qquad (\textit{car } \ln(2) > 0)
\end{aligned}$$

  On démontre ainsi :

$$\text{La fonction } \log_2 \text{ est concave sur } \mathbb{R}_+^* \Leftrightarrow \text{La fonction ln est concave sur } \mathbb{R}_+^*$$

□

2. Soit X une variable aléatoire réelle à densité, et soit f une densité de X. On appelle **support** de f l'ensemble $I = \{x \in \mathbb{R} \mid f(x) > 0\}$, et on suppose que I est un intervalle de $\mathbb{R}$ d'extrémités a et b ($a < b$, a et b finis ou infinis). L'**entropie différentielle** de X est, sous réserve d'existence, le réel :

$$h(X) \;=\; -\int_a^b f(x)\,\log_2\big(f(x)\big)\,dx$$

Montrer : $h(X) = -\mathbb{E}\Big(\log_2\big(f(X)\big)\Big)$.

Commentaire

- Dans cet énoncé, on définit la notion de support d'une fonction f. Par définition, le support de f est l'ensemble des réels x pour lesquels $f(x)$ est strictement positive. L'énoncé fait l'hypothèse que le support de la fonction f étudiée est un intervalle. Rappelons qu'un intervalle est un ensemble qui s'écrit sous la forme $]a,b[$, $]a,b]$, $[a,b[$ ou $[a,b]$, avec a et b des réels finis ou infinis. Ces notations englobent donc les cas $]-\infty,+\infty[$, $]-\infty,b]$, $[a,+\infty[$.
- La fonction f introduite dans cette question est une densité d'une v.a.r. X. Par définition, une densité est une fonction positive sur $\mathbb{R}$. On en conclut donc que la fonction f est nulle en dehors de son support. En effet, si $x \notin I$ alors $f(x) \not> 0$. Autrement dit : $f(x) \leqslant 0$.
  Et comme $f(x) \geqslant 0$, on en déduit : $f(x) = 0$.
  Ainsi, si I est un intervalle d'extrémités a et b, on a en particulier :

  $$\int_{-\infty}^{+\infty} f(t)\,dt \;=\; \int_a^b f(t)\,dt$$

  car f est nulle en dehors de I. On met ici en avant l'un des intérêts de la notion de support d'une densité f : cela permet de restreindre l'intervalle d'intégration. C'est pour ces considérations d'intervalle d'intégration que l'énoncé introduit la notion de support d'une fonction (cette notion n'est pas développée dans le programme officiel ECE).
- Il est à noter que si f est une densité d'une v.a.r. X alors tout fonction $\tilde{f}$ obtenue en modifiant un nombre fini de valeurs de f par des valeurs positives est toujours une densité de X. Par exemple, les fonctions :

  $$f : x \mapsto \begin{cases} 0 & \text{si } x < 0 \\ \lambda\,\mathrm{e}^{-\lambda x} & \text{si } x \geqslant 0 \end{cases} \qquad \text{et} \qquad \tilde{f} : x \mapsto \begin{cases} 5 & \text{si } x = -1 \\ 0 & \text{si } x \in\,]-\infty,-1[\,\cup\,]-1,0[\\ \lambda\,\mathrm{e}^{-\lambda x} & \text{si } x \in [0,3[\,\cup\,]3,+\infty[\\ 0 & \text{si } x = 3 \end{cases}$$

  sont deux densités différentes de toute v.a.r. X suivant la loi $\mathcal{E}(\lambda)$ (où $\lambda > 0$). On constate que ces deux fonctions ont des supports différents :

  $$f \text{ a pour support } [0,+\infty[\quad \text{et} \quad \tilde{f} \text{ a pour support } \{-1\} \cup [0,3[\,\cup\,]3,+\infty[$$

  C'est bien dommage : si toutes les densités de X admettaient le même support, cet ensemble pourrait définir la notion de support de la v.a.r. X.
- La notion de support d'une v.a.r. X admettant une densité f_X est en fait un peu plus technique à définir. Il s'agit de l'ensemble des points au voisinage desquels la fonction f_X n'est pas identiquement nulle. Autrement dit, un point x est dans $\mathrm{Supp}(X)$ (le support de la v.a.r. X) s'il n'existe pas de réel $\alpha > 0$ qui permettrait de vérifier la propriété : $\forall t \in\,]x-\alpha, x+\alpha[$, $f(t) = 0$. Cette définition ne dépend pas de la densité f_X étudiée. Si on reprend l'exemple précédent ($X \hookrightarrow \mathcal{E}(\lambda)$), on obtient : $\mathrm{Supp}(X) = [0,+\infty[$.

Démonstration.

- Considérons la fonction $g : x \mapsto \log_2\big(f(x)\big)$.
  La fonction g est continue sur I sauf (éventuellement) en un nombre fini de points car elle est la composée $g = \log_2 \circ f$ où :
  - f est :
    - continue sur I sauf (éventuellement) en un nombre fini de points,
    - telle que : $f(I) \subset]0, +\infty[$.
  - $\log_2$ est continue sur $]0, +\infty[$.

  La fonction g est continue sur I sauf (éventuellement) en un nombre fini de points.

- Par théorème de transfert, la v.a.r. $g(X) = \log_2\big(f(X)\big)$ admet une espérance si et seulement si l'intégrale $\displaystyle\int_a^b g(x)\,f(x)\,dx$ est absolument convergente.
  Ainsi, sous réserve de convergence absolue :

$$\mathbb{E}\big(g(X)\big) \;=\; \int_a^b g(x)\,f(x)\,dx \;=\; \int_a^b \log_2\big(f(x)\big)\,f(x)\,dx \;=\; \int_a^b f(x)\,\log_2\big(f(x)\big)\,dx$$

Finalement, sous réserve de convergence absolue :

$$h(X) \;=\; -\int_a^b f(x)\,\log_2\big(f(x)\big)\,dx \;=\; -\mathbb{E}\Big(\log_2\big(f(X)\big)\Big).$$

Commentaire

- On peut noter que la fonction $g : x \mapsto \log_2\big(f(x)\big)$ est définie sur l'intervalle I, support de la fonction f. On ne peut donc pas, en toute rigueur, considérer l'intégrale : $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\log_2\big(f(x)\big)\,dx$. L'introduction de la notion de support permet d'éviter l'écriture de cette intégrale.
- Ce n'est pas parce que la fonction $x \mapsto f(x)\,\log_2\big(f(x)\big)$ est définie sur l'intervalle I que l'on peut en conclure que l'intégrale $\displaystyle\int_a^b f(x)\,\log_2\big(f(x)\big)\,dx$ est bien définie ! On dit que l'intégrale $\displaystyle\int_a^b f(x)\,\log_2\big(f(x)\big)\,dx$ est bien définie (ou existe) si la fonction $x \mapsto f(x)\,\log_2\big(f(x)\big)$ est continue (par morceaux) sur le **segment** $[a, b]$. Si ce n'est pas le cas, on a affaire à une intégrale généralisée (ou impropre) et il convient alors de déterminer la nature (convergence / divergence) de cette intégrale.

□

3. Soit X une variable aléatoire de densité f de support I, intervalle de $\mathbb{R}$ d'extrémités a et b. On suppose que X admet une entropie différentielle.

a) Soit c un réel, et soit Y la variable aléatoire définie par $Y = c + X$.

(i) Déterminer une densité de Y.

Démonstration.

- La v.a.r. $Y = c + X$ est une transformée affine de la v.a.r. à densité X.
  On en conclut que la v.a.r. Y est une v.a.r. à densité.

- De plus, d'après le cours, la fonction :

$$f_Y : x \mapsto \frac{1}{|1|}\, f_X\left(\frac{x-c}{1}\right)$$

est une densité de Y.

La fonction $f_Y : x \mapsto f_X(x-c)$ est une densité de Y.

Commentaire

- On utilise dans cette question le résultat suivant.

  Soit X une v.a.r. à densité. On note f_X une densité de X.

  Soient $\alpha \neq 0$ et β deux réels. On note $Y = \alpha\, X + \beta$ (c'est une transformée affine de X).

  Alors $Y = \alpha X + \beta$ est une v.a.r. à densité.
  De plus, la fonction :

$$f_Y : x \mapsto \frac{1}{|\alpha|}\, f_X\left(\frac{x-\beta}{\alpha}\right)$$

  On utilise ici ce résultat avec $\alpha = 1$ et $\beta = c$.

- Il n'est pas nécessaire de connaître ce résultat par cœur. Ce qui importe c'est de savoir le retrouver rapidement. On peut d'ailleurs mener toute l'étude au brouillon et conserver la rédaction détaillée ci-dessus. Rappelons les différentes étapes permettant d'obtenir une densité f_Y de la v.a.r. Y.
- Déterminons tout d'abord la fonction de répartition de la v.a.r. Y.
  Soit $x \in \mathbb{R}$.

$$F_Y(x) \;=\; \mathbb{P}([Y \leqslant x]) \;=\; \mathbb{P}([c + X \leqslant x]) \;=\; \mathbb{P}([X \leqslant x - c]) \;=\; F_X(x-c)$$

Finalement : $F_Y : x \mapsto F_X(x-c)$.

- La fonction F_Y est continue sur $\mathbb{R}$ car elle est la composée $F_Y = F_X \circ g_1$ où :
  - × $g_1 : x \mapsto x - c$ est :
    - continue sur $\mathbb{R}$,
    - telle que : $g_1(\mathbb{R}) \subset \mathbb{R}$.
  - × F_X est continue sur $\mathbb{R}$, car X est une v.a.r. à densité.
- La fonction F_Y est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ sauf (éventuellement) en un nombre fini de points avec des arguments similaires à ceux de la continuité sur $\mathbb{R}$.

On en déduit que Y est une v.a.r. à densité.

- Pour déterminer une densité f_Y de Y, on dérive la fonction F_Y en les points où elle est de classe $\mathcal{C}^1$. Soit $x \in \mathbb{R}$.
  - × Si F_X est de classe $\mathcal{C}^1$ en x :

$$\begin{aligned} f_Y(x) \;&=\; F_Y'(x) \;=\; (F_X \circ g_1)'(x) \;=\; (F_X' \circ g_1)(x) \times g_1'(x) \\ &=\; F_X'\big(g_1(x)\big) \times 1 \;=\; F_X'(x-c) \;=\; f(x-c) \end{aligned}$$

  - × Si F_X n'est pas de classe $\mathcal{C}^1$ en x, on choisit : $f_Y(x) = f(x-c)$.

Finalement : $f_Y : x \mapsto f(x-c)$.

Commentaire

On peut détailler le dernier point précédent concernant les points où F_X est de classe $\mathcal{C}^1$.

- Si F_X est de classe $\mathcal{C}^1$ sur $\mathbb{R}$, alors la fonction F_Y est également de classe $\mathcal{C}^1$ sur $\mathbb{R}$ en tant composée $F_Y = F_X \circ g_1$ (même démonstration que la continuité de F_Y sur $\mathbb{R}$).
- Si F_X n'est pas de classe $\mathcal{C}^1$ sur $\mathbb{R}$ tout en entier, on peut déterminer plus précisément les points où la fonction F_Y n'est pas de classe $\mathcal{C}^1$.
  - × Comme X est une v.a.r. à densité, sa fonction de répartition F_X est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ sauf en un nombre fini de points. Notons $\{x_1, \ldots, x_n\}$ l'ensemble des points où la fonction F_X n'est pas de classe $\mathcal{C}^1$.
  - × Comme $F_Y : x \mapsto F_X(x-c)$, on en déduit que la fonction F_Y est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ sauf en $x_1 + c$, …, $x_n + c$.

□

(ii) Justifier l'existence de l'entropie différentielle $h(Y)$, et la déterminer en fonction de $h(X)$.

Démonstration.

- Commençons tout d'abord par déterminer le support de f_Y.
  Soit $x \in \mathbb{R}$.

$$\begin{aligned}
& x \text{ est un élément du support de } f_Y \\
\Leftrightarrow\ & f_Y(x) > 0 \\
\Leftrightarrow\ & f_X(x-c) > 0 \\
\Leftrightarrow\ & x - c \text{ est un élément de } I \text{ (intervalle d'extrémités } a \text{ et } b\text{), support de } f_X \\
\Leftrightarrow\ & x \text{ est un élément de } J\text{, intervalle d'extrémités } a+c \text{ et } b+c
\end{aligned}$$

Commentaire

Pour bien comprendre la dernière étape, on peut remarquer :

$$a < x - c < b \quad \Leftrightarrow \quad a + c < x < b + c$$

Ainsi, f_Y a pour support un intervalle J d'extrémités $a+c$ et $b+c$.

- Étudions maintenant l'entropie différentielle de Y.

$$\begin{aligned}
& \text{La v.a.r. } Y \text{ admet une entropie différentielle} \\
\Leftrightarrow\ & \text{L'intégrale } \int_{a+c}^{b+c} f_Y(x)\ \log_2\big(f_Y(x)\big)\ dx \text{ est convergente} && \textit{(d'après l'énoncé)} \\
\Leftrightarrow\ & \text{L'intégrale } \int_{a+c}^{b+c} f_X(x-c)\ \log_2\big(f_X(x-c)\big)\ dx \text{ est convergente} && \textit{(d'après la question précédente)}
\end{aligned}$$

Or, sous réserve de convergence, en posant le changement de variable $\boxed{u = x - c}$:

$$\int_{a+c}^{b+c} f_X(x-c)\ \log_2\big(f_X(x-c)\big)\ dx \ = \ \int_a^b f_X(u)\ \log_2\big(f_X(u)\big)\ du$$

Par hypothèse, X admet une entropie différentielle.
On en déduit que l'intégrale $\displaystyle\int_a^b f_X(u)\ \log_2\big(f_X(u)\big)\ du$ converge. D'après l'égalité précédente, il en est de même de l'intégrale $\displaystyle\int_{a+c}^{b+c} f_X(x-c)\ \log_2\big(f_X(x-c)\big)\ dx$.

- On en conclut enfin :

$$h(Y) \;=\; -\int_{a+c}^{b+c} f_Y(x)\;\log_2\big(f_Y(x)\big)\;dx \;=\; -\int_a^b f_X(u)\;\log_2\big(f_X(u)\big)\;du \;=\; h(X)$$

La v.a.r. $Y = X + c$ admet une entropie différentielle et $h(Y) = h(X)$.

Commentaire

On n'a pas détaillé ici le changement de variable qui consiste ici simplement à opérer une translation. Il faut évidemment connaître la manière de procéder que l'on rappelle ci-dessous.

$u = x - c \quad$ (et donc $x = u + c$)
$\hookrightarrow du = dx$
- $x = a + c \;\Rightarrow\; u = (a + c) - c = a$
- $x = b + c \;\Rightarrow\; u = (b + c) - c = b$

Ce changement de variable est valide car $\psi : u \mapsto u + c$ est de classe $\mathcal{C}^1$ sur I.

□

b) Soit α un réel strictement positif, et soit Z la variable aléatoire définie par $Z = \alpha\, X$.

(i) Déterminer une densité de Z.

Démonstration.

- La v.a.r. $Z = \alpha\, X$ est une transformée affine de la v.a.r. à densité X.
  On en conclut que la v.a.r. Z est une v.a.r. à densité.
- De plus, d'après le cours, la fonction :

$$f_Z : x \mapsto \frac{1}{|\alpha|}\; f_X\left(\frac{x}{\alpha}\right)$$

est une densité de Z.

La fonction $f_Z : x \mapsto \dfrac{1}{\alpha}\; f_X\left(\dfrac{x}{\alpha}\right)$ est une densité de Y.

Commentaire

- On utilise de nouveau dans cette question le résultat du cours donnant l'expression d'une densité d'une transformée affine de la v.a.r. X. Rappelons comment retrouver ce résultat dans le cas précis de l'énoncé.
- Déterminons la fonction de répartition de la v.a.r. Z.
  Soit $x \in \mathbb{R}$.

$$\begin{aligned} F_Z(x) &= \mathbb{P}([Z \leqslant x]) \\ &= \mathbb{P}([\alpha\, X \leqslant x]) \\ &= \mathbb{P}\left(\left[X \leqslant \frac{x}{\alpha}\right]\right) \qquad \textit{(car } \alpha > 0\textit{)} \\ &= F_X\left(\frac{x}{\alpha}\right) \end{aligned}$$

Finalement : $F_Z : x \mapsto F_X\left(\dfrac{x}{\alpha}\right)$.

Commentaire

- La fonction F_Z est continue sur $\mathbb{R}$ car elle est la composée $F_Z = F_X \circ g_2$ de fonctions continues sur $\mathbb{R}$ (avec $g_2 : x \mapsto \frac{x}{\alpha}$).
- La fonction F_Z est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ sauf (éventuellement) en un nombre fini de points avec des arguments similaires à ceux de la continuité sur $\mathbb{R}$.
- Pour déterminer une densité f_Z de Z, on dérive la fonction F_Z en les points où elle est de classe $\mathcal{C}^1$. Soit $x \in \mathbb{R}$.
  - × Si F_X est de classe $\mathcal{C}^1$ en x :

$$\begin{aligned} f_Z(x) &= F_Z'(x) = (F_X \circ g_2)'(x) = (F_X' \circ g_2)(x) \times g_2'(x) \\ &= F_X'\big(g_2(x)\big) \times \frac{1}{\alpha} = \frac{1}{\alpha}\, F_X'\left(\frac{x}{\alpha}\right) = \frac{1}{\alpha}\, f\left(\frac{x}{\alpha}\right) \end{aligned}$$

  - × Si F_X n'est pas de classe $\mathcal{C}^1$ en x, on choisit : $f_Z(x) = \frac{1}{\alpha}\, f\left(\frac{x}{\alpha}\right)$.

Finalement : $f_Z : x \mapsto \frac{1}{\alpha}\, f\left(\frac{x}{\alpha}\right)$.

□

(ii) Justifier l'existence de l'entropie différentielle $h(Z)$, et la déterminer en fonction de $h(X)$.

Démonstration.

- Commençons tout d'abord par déterminer le support de f_Z.
  Soit $x \in \mathbb{R}$.

x est un élément du support de f_Z

$\Leftrightarrow$ $f_Z(x) > 0$

$\Leftrightarrow$ $\frac{1}{\alpha}\, f_X\left(\frac{x}{\alpha}\right) > 0$

$\Leftrightarrow$ $f_X\left(\frac{x}{\alpha}\right) > 0$ *(car $\alpha > 0$)*

$\Leftrightarrow$ $\frac{x}{\alpha}$ est un élément de I (intervalle d'extrémités a et b), support de f_X

$\Leftrightarrow$ x est un élément de J, intervalle d'extrémités $\alpha\, a$ et $\alpha\, b$

Commentaire

Pour bien comprendre la dernière étape, on peut remarquer, comme $\alpha > 0$:

$$a < \frac{x}{\alpha} < b \quad \Leftrightarrow \quad \alpha\, a < x < \alpha\, b$$

Ainsi, f_Z a pour support un intervalle J d'extrémités $\alpha\, a$ et $\alpha\, b$.

- Étudions maintenant l'entropie différentielle de Y.

La v.a.r. Z admet une entropie différentielle

$\Leftrightarrow$ L'intégrale $\displaystyle\int_{\alpha\, a}^{\alpha\, b} f_Z(x)\, \log_2\big(f_Z(x)\big)\, dx$ est convergente *(d'après l'énoncé)*

$\Leftrightarrow$ L'intégrale $\displaystyle\int_{\alpha\, a}^{\alpha\, b} \frac{1}{\alpha}\, f_X\left(\frac{x}{\alpha}\right)\, \log_2\big(\frac{1}{\alpha}\, f_X\left(\frac{x}{\alpha}\right)\big)\, dx$ est convergente *(d'après la question précédente)*

Sous réserve de convergence, on pose le changement de variable $\boxed{u = \dfrac{x}{\alpha}}$:

$$\left|\begin{array}{l} u = \frac{x}{\alpha} \quad (\text{et donc } x = \alpha\, u) \\ \hookrightarrow du = \dfrac{1}{\alpha}\, dx \quad \text{et donc} \quad dx = \alpha\, du \\ \bullet\ x = \alpha\, a \ \Rightarrow\ u = \frac{\alpha\, a}{\alpha} = a \\ \bullet\ x = \alpha\, b \ \Rightarrow\ u = \frac{\alpha\, b}{\alpha} = b \end{array}\right.$$

Ce changement de variable est valide car $\psi : u \mapsto \alpha\, u$ est de classe $\mathcal{C}^1$ sur I.
On obtient :

$$\int_{\alpha\, a}^{\alpha\, b} \frac{1}{\alpha}\, f_X\left(\frac{x}{\alpha}\right)\, \log_2\left(\frac{1}{\alpha}\, f_X\left(\frac{x}{\alpha}\right)\right)\, dx \;=\; \int_a^b \frac{1}{\cancel{\alpha}}\, f_X(u)\, \log_2\left(\frac{1}{\alpha}\, f_X(u)\right)\, \cancel{\alpha}\, du$$

- Démontrons la convergence de cette dernière intégrale.
  Tout d'abord, pour tout $u \in I$:

$$\begin{aligned} \log_2\left(\frac{1}{\alpha}\, f_X(u)\right) &= \log_2\left(\frac{1}{\alpha}\right) + \log_2\big(f_X(u)\big) \\ &= \frac{\ln\left(\frac{1}{\alpha}\right)}{\ln(2)} + \log_2\big(f_X(u)\big) \\ &= -\frac{\ln(\alpha)}{\ln(2)} + \log_2\big(f_X(u)\big) \;=\; \log_2\big(f_X(u)\big) - \log_2(\alpha) \end{aligned}$$

$$\boxed{\text{On en déduit, pour tout } u \in I : f_X(u)\, \log_2\left(\frac{1}{\alpha}\, f_X(u)\right) \;=\; f_X(u)\, \log_2\big(f_X(u)\big) - f_X(u)\, \log_2(\alpha).}$$

- Or :
  × l'intégrale $\displaystyle\int_a^b f_X(u)\, \log_2\big(f_X(u)\big)\, du$ est convergente car X admet une entropie différentielle,
  × l'intégrale $\displaystyle\int_a^b f_X(u)\, du$ est convergente car f_X est une densité de X, nulle en dehors de I.
  Ainsi :

$$\int_a^b f_X(u)\, du \;=\; \int_{-\infty}^{+\infty} f_X(u)\, du \;=\; 1$$

  On en déduit que l'intégrale $\displaystyle\int_a^b f_X(u)\, \log_2\left(\frac{1}{\alpha}\, f_X(u)\right)\, du$ est convergente.

$$\boxed{\text{La v.a.r. } Z \text{ admet donc une entropie différentielle.}}$$

- De plus :

$$\begin{aligned} h(Z) &= -\int_a^b f_X(u)\, \log_2\left(\frac{1}{\alpha}\, f_X(u)\right)\, du \\ &= -\int_a^b f_X(u)\, \log_2\big(f_X(u)\big)\, du + \log_2(\alpha) \int_a^b f_X(u)\, du \;=\; h(X) + \log_2(\alpha) \end{aligned}$$

$$\boxed{\text{Finalement : } h(Z) = h(X) + \log_2(\alpha).}$$

Commentaire

- Le but de la question ***3.*** est d'étudier comment se comporte la notion d'entropie différentielle vis à vis de la transformation affine. On opére en deux temps :
  - × on étudie d'abord la v.a.r. $Y = X + c$ (translatée de la v.a.r. X),
  - × on étudie ensuite la v.a.r. $Z = \alpha\, X$ (obtenue par homothétie à partir de la v.a.r. X).
- On peut déduire de ces deux étapes que si X est une v.a.r. à densité admettant une entropie différentielle, alors, toute v.a.r. $T = \alpha\, X + \beta$ (où $\alpha > 0$ et $\beta \in \mathbb{R}$) admet une entropie différentielle. En effet :
  - × $\alpha\, X$ admet une entropie différentielle car X en admet une (d'après la question ***3.b)***),
  - × $\alpha\, X + \beta = (\alpha\, X) + \beta$ admet une entropie différentielle car $\alpha\, X$ est une v.a.r. à densité qui en admet une (d'après la question ***3.a)***).

  De plus :

$$\begin{aligned} h(T) &= h(\alpha\, X + \beta) \\ &= h\big((\alpha\, X) + \beta\big) \\ &= h\big(\alpha\, X\big) \qquad \textit{(d'après la question } \textbf{3.a)} \\ &= h(X) + \log_2(\alpha) \end{aligned}$$

$$\boxed{h(\alpha\, X + \beta) \;=\; h(X) + \log_2(\alpha)}$$

□

4. On détermine dans cette question l'entropie différentielle de quelques variables aléatoires suivant des lois classiques.

a) Soit $a > 0$. On considère X une variable aléatoire de loi uniforme sur $[0, a]$.

(i) Donner une densité de X.

Démonstration.

$$\boxed{\text{Une densité } f \text{ de } X \text{ est : } f : x \mapsto \begin{cases} 0 & \text{si } x \in \,]-\infty, 0[\\ \dfrac{1}{a} & \text{si } x \in [0, a] \\ 0 & \text{si } x \in \,]a, +\infty[\end{cases}}$$

□

Commentaire

- Une bonne connaissance du cours est une condition *sine qua non* de réussite au concours. En effet, on trouve dans toutes les épreuves de mathématiques (même pour les écoles les plus prestigieuses), des questions d'application directe du cours. C'est le cas en particulier de cette question qui nécessite simplement de connaître la loi uniforme.
- On trouve parfois la présentation plus compacte :

$$f : x \mapsto \begin{cases} \dfrac{1}{a} & \text{si } x \in [0, a] \\ 0 & \text{sinon} \end{cases}$$

  Il faut faire attention avec ce type de présentation qui introduit une dissymétrie qui n'a pas lieu d'être entre fonction de répartition définie à l'aide de trois cas (nulle sur $]-\infty, 0[$ et constante égale à 1 sur $]a, +\infty[$) et la densité f choisie.

(ii) Justifier l'existence de l'entropie différentielle $h(X)$, et la déterminer.

Démonstration.

- On commence par remarquer que, par définition de f, son support I est l'intervalle $[0, a]$. Alors, d'après la définition de l'entropie, X admet une entropie différentielle si et seulement si l'intégrale $\displaystyle\int_0^a f(x)\,\log_2\big(f(x)\big)\,dx$ est convergente.
- Or la fonction $x \mapsto f(x)\,\log_2\big(f(x)\big)$ est continue par morceaux sur le **segment** $[0, a]$. On en déduit que l'intégrale $\displaystyle\int_0^a f(x)\,\log_2\big(f(x)\big)\,dx$ converge.

Ainsi, la v.a.r. X admet une entropie différentielle.

- De plus :

$$\begin{aligned}
h(X) &= -\int_0^a f(x)\log_2\big(f(x)\big)\,dx \\
&= -\int_0^a \frac{1}{a}\,\log_2\left(\frac{1}{a}\right)\,dx && \textit{(par définition de } f \textit{ sur } [0,a]\textit{)} \\
&= \frac{-1}{a}\,\log_2\left(\frac{1}{a}\right)\,\int_0^a 1\,dx \\
&= \frac{1}{a}\,\log_2(a)\,\big[\,x\,\big]_0^a \\
&= \frac{1}{a}\,\log_2(a)\,(a-0) \\
&= \log_2(a)
\end{aligned}$$

$h(X) = \log_2(a)$

□

(iii) Déterminer une condition nécessaire et suffisante sur a pour que $h(X) > 0$.

Démonstration.
D'après la question précédente :

$$\begin{aligned}
h(X) > 0 &\Leftrightarrow \log_2(a) > 0 \\
&\Leftrightarrow \frac{\ln(a)}{\ln(2)} > 0 \\
&\Leftrightarrow \ln(a) > 0 && \textit{(car } \ln(2) > 0\textit{)} \\
&\Leftrightarrow a > 1 && \textit{(par stricte croissance de la fonction } \exp \textit{ sur } \mathbb{R}\textit{)}
\end{aligned}$$

$h(X) > 0 \;\Leftrightarrow\; a > 1$

□

b) On considère Y une variable aléatoire de loi normale centrée réduite. Montrer que Y admet une entropie différentielle et : $h(Y) = \dfrac{1}{2}\ \log_2(2\pi\mathrm{e})$.

Démonstration.

- Une densité de Y est : $f_Y : x \mapsto \dfrac{1}{\sqrt{2\pi}}\ \mathrm{e}^{-\frac{1}{2}x^2}$.

  Le support de la fonction f_Y est l'intervalle $I =]-\infty, +\infty[$.
- Soit $x \in \mathbb{R}$.

$$\begin{aligned}
\log_2\big(f_Y(x)\big) &= \log_2\left(\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}x^2}\right) \\
&= \log_2\left(\frac{1}{\sqrt{2\pi}}\right) + \log_2\left(\mathrm{e}^{-\frac{1}{2}x^2}\right) \qquad \textit{(d'après la question } \textbf{\textit{1.a)}}\textit{)} \\
&= \frac{\ln\left(\frac{1}{\sqrt{2\pi}}\right)}{\ln(2)} + \frac{\ln\left(\mathrm{e}^{-\frac{1}{2}x^2}\right)}{\ln(2)} \\
&= \frac{-\frac{1}{2}\ln(2\pi)}{\ln(2)} + \frac{-\frac{1}{2}x^2}{\ln(2)} \\
&= -\frac{1}{2}\,\frac{\ln(2\pi)}{\ln(2)} - \frac{1}{\ln(2)}\,\frac{x^2}{2} \;=\; -\frac{1}{2}\ \log_2(2\pi) - \frac{1}{2\ln(2)}\ x^2
\end{aligned}$$

En particulier :

$$\log_2\big(f_Y(Y)\big) \;=\; -\frac{1}{2}\ \log_2(2\pi) - \frac{1}{2\ln(2)}\ Y^2 \quad (*)$$

Commentaire

- Formellement, cette égalité $(*)$ est obtenue en remplaçant x par Y dans l'égalité précédente. Mais c'est plus subtil qu'il n'y paraît. Pour le comprendre, il faut revenir à la définition d'une transformée $g(Y)$ d'une v.a.r. Y. Rappelons que si $g : Y(\Omega) \to \mathbb{R}$ alors on utilise la notation $g(Y)$ pour désigner $g \circ Y$. Autrement dit, $g(Y)$ est la v.a.r. définie par :

$$\begin{array}{rcl} g(Y) \;:\; \Omega & \to & \mathbb{R} \\ \omega & \mapsto & g\big(Y(\omega)\big) \end{array}$$

  Ainsi, la v.a.r. $\log_2\big(f_Y(Y)\big)$ n'est autre que l'application $\omega \mapsto \log_2\big(f_Y(Y(\omega))\big)$. Ainsi, pour démontrer $(*)$, il faut pouvoir remplacer x par $Y(\omega)$ pour tout $\omega \in \Omega$. Cela ne revêt aucune difficulté ici car x est supposé être un réel quelconque.
- Mais cela peut être plus délicat. Par exemple, en question ***4.a)***, on a introduit une densité f de la loi uniforme sur $[0, a]$ et on a démontré, pour tout $x \in [0, a]$:

$$\log_2\big(f(x)\big) \;=\; -\log_2(a)$$

  On ne peut pour autant en conclure : $\log_2\big(f(X)\big) = -\log_2(a)$. En effet, on réalise ici une hypothèse forte : l'égalité initiale n'est démontrée que pour $x \in [0, a]$. De ce fait, l'égalité $\log_2\Big(f\big(X(\omega)\big)\Big) = -\log_2(a)$ n'est vérifiée que si ω est tel que : $X(\omega) \in [0, a]$. C'est le cas si l'on suppose que X est à valeurs dans $[0, a]$ (on a alors : $\forall \omega \in \Omega$, $X(\omega) \in [0, a]$). Or, dans l'énoncé, on ne fait jamais une telle supposition. On touche ici une ambiguïté classique en Probabilités. On ne sait pas (si l'énoncé ne le précise pas) que X est à valeurs dans $[0, a]$ mais on sait que cette propriété est vérifiée avec probabilité 1 :

$$\mathbb{P}\big([0 \leqslant X \leqslant a]\big) = 1$$

Commentaire

- Dans le point précédent, il est à noter que sans hypothèse $X(\Omega) = [0, a]$ la v.a.r. $\log_2\big(f(X)\big)$ n'est pas bien définie mais est seulement presque sûrement bien définie. Dans certains sujets, l'ensemble image des v.a.r. étudiées sera précisé (« On considère une v.a.r. à valeurs strictement positives » par exemple).
  Si ce n'est pas le cas, si X est une v.a.r. de densité f_X, on pourra agir comme suit :
  - × si X suit une loi usuelle, on peut se référer à l'ensemble image donné en cours. Par exemple, si $X \hookrightarrow \mathcal{U}([0,1])$, on se permet d'écrire :

    « Comme $X \hookrightarrow \mathcal{U}([0,1])$, on **considère** : $X(\Omega) = [0,1]$. »

  - × si X ne suit pas une loi usuelle, on étudie l'ensemble : $I = \{x \in \mathbb{R} \mid f_X(x) > 0\}$. On se permet alors d'écrire :

    « Dans la suite, on **considère** : $X(\Omega) = I$. »

    En réalité, il serait plus judicieux de considérer le support de X plutôt que le support de f_X. Il est fréquent que ces deux supports coïncident (l'exemple de la densité $\tilde{f}$ en remarque de la question ***3.*** est donné à titre pédagogique mais il n'y a pas de raison qu'un énoncé considère une construction aussi tordue).

  En **décrétant** la valeur de $X(\Omega)$, on ne commet pas une erreur mais on décide d'ajouter une hypothèse qui ne fait pas partie de l'énoncé. Cette audace permet de travailler avec un ensemble image connu, ce qui permet de structurer certaines démonstrations (l'ensemble image étant connu, on se rappelle que la fonction de répartition, par exemple, s'obtient par une disjonction de cas) ou de résoudre certains problèmes de définitions.

- Comme $Y \hookrightarrow \mathcal{N}(0,1)$, Y admet un moment d'ordre 2.
  Autrement dit, la v.a.r. Y^2 admet une espérance et il en est de même de la v.a.r. $\log_2\big(f_Y(Y)\big)$, transformée affine de la v.a.r. Y^2 (d'après l'égalité $(*)$).

  On en conclut, d'après la question ***2.***, que la v.a.r. Y admet une entropie différentielle.

- De plus :

$$\begin{aligned}
h(Y) &= -\mathbb{E}\Big(\log_2\big(f_Y(Y)\big)\Big) \\
&= -\mathbb{E}\left(-\frac{1}{2}\,\log_2(2\pi) - \frac{1}{2\ln(2)}\,Y^2\right) \\
&= -\left(-\mathbb{E}\left(\frac{1}{2}\,\log_2(2\pi)\right) - \frac{1}{2\ln(2)}\,\mathbb{E}\left(Y^2\right)\right) && \textit{(par linéarité de l'espérance)} \\
&= \frac{1}{2}\,\log_2(2\pi) + \frac{1}{2\ln(2)}\,\mathbb{E}\left(Y^2\right)
\end{aligned}$$

  Or, d'après la formule de Kœnig-Huygens : $\mathbb{V}(Y) = \mathbb{E}\left(Y^2\right) - \big(\mathbb{E}(Y)\big)^2$.

$$\begin{aligned}
\mathbb{E}\left(Y^2\right) &= \mathbb{V}(Y) + \big(\mathbb{E}(Y)\big)^2 \\
&= 1 + 0 \;=\; 1
\end{aligned}$$

- Finalement :

$$\begin{aligned} h(Y) &= \frac{1}{2}\,\log_2(2\pi) + \frac{1}{2\,\ln(2)} \\ &= \frac{1}{2}\,\left(\log_2(2\pi) + \frac{1}{\ln(2)}\right) \\ &= \frac{1}{2}\left(\log_2(2\pi) + \frac{\ln(\mathrm{e})}{\ln(2)}\right) \\ &= \frac{1}{2}\big(\log_2(2\pi) + \log_2(\mathrm{e})\big) \;=\; \frac{1}{2}\,\log_2(2\pi\mathrm{e}) \end{aligned}$$

On a bien : $h(Y) = \dfrac{1}{2}\,\log_2(2\pi\mathrm{e})$.

Commentaire

- On a utilisé dans cette question la caractérisation de $h(Y)$ à l'aide de l'espérance. Il était aussi possible de procéder par étude d'intégrales.
- Plus précisément, la v.a.r. Y admet une entropie différentielle si et seulement si l'intégrale $\displaystyle\int_{-\infty}^{+\infty} f_Y(x)\,\log_2\big(f_Y(x)\big)\,dx$ est convergente. Or, pour tout $x \in \mathbb{R}$:

$$f_Y(x)\,\log_2\big(f_Y(x)\big) \;=\; -\frac{1}{2}\,\log_2(2\pi)\,f_Y(x) - \frac{1}{2\ln(2)}\,x^2\,f_Y(x)$$

- On sait que :
  - × l'intégrale $\displaystyle\int_{-\infty}^{+\infty} f_Y(x)\,dx$ est convergente car f_Y est une densité.
    De plus : $\displaystyle\int_{-\infty}^{+\infty} f_Y(x)\,dx \;=\; 1.$
  - × l'intégrale $\displaystyle\int_{-\infty}^{+\infty} x^2\,f_Y(x)\,dx$ est convergente car la v.a.r. Y admet un moment d'ordre 2.
    De plus : $\displaystyle\int_{-\infty}^{+\infty} x^2\,f_Y(x)\,dx \;=\; \mathbb{E}(Y^2) = 1.$

  Ainsi, l'intégrale $\displaystyle\int_{-\infty}^{+\infty} f_Y(x)\,\log_2\big(f_Y(x)\big)\,dx$ est convergente.
- Enfin :

$$\begin{aligned} h(Y) &= -\int_{-\infty}^{+\infty}\left(-\frac{1}{2}\,\log_2(2\pi)\,f_Y(x) - \frac{1}{2\ln(2)}x^2\,f_Y(x)\right)\,dx \\ &= \frac{1}{2}\,\log_2(2\pi)\,\int_{-\infty}^{+\infty} f_Y(x)\,dx + \frac{1}{2\ln(2)}\,\int_{-\infty}^{+\infty} x^2\,f_Y(x)\,dx \\ &= \frac{1}{2}\,\log_2(2\pi)\times 1 + \frac{1}{2\ln(2)}\times 1 \;=\; \frac{1}{2}\,\log_2(2\pi\mathrm{e}) \end{aligned}$$

- Profitons enfin de cette remarque pour mettre en avant un point de calcul effectué dans cette question, à savoir : $\dfrac{1}{\ln(2)} = \dfrac{\ln(\mathrm{e})}{\ln(2)} = \log_2(\mathrm{e})$. □

c) On considère Z une variable aléatoire de loi exponentielle de paramètre λ ($\lambda > 0$). Justifier l'existence de l'entropie différentielle $h(Z)$ et la déterminer.

Démonstration.

- Une densité de Z est :
$$f_Z : x \mapsto \begin{cases} 0 & \text{si } x \in]-\infty, 0[\\ \lambda\,\mathrm{e}^{-\lambda x} & \text{si } x \in [0, +\infty[\end{cases}$$
Le support de la fonction f_Z est l'intervalle $I = [0, +\infty[$.
Ainsi, la v.a.r. Z admet une entropie différentielle si et seulement si l'intégrale $\displaystyle\int_0^{+\infty} f_Z(x)\,\log_2\big(f_Z(x)\big)\,dx$ est convergente.
- Soit $x \in [0, +\infty[$.
$$\begin{aligned} \log_2\big(f_Z(x)\big) &= \log_2\big(\lambda\,\mathrm{e}^{-\lambda x}\big) \\ &= \log_2(\lambda) + \log_2\big(\mathrm{e}^{-\lambda x}\big) && \textit{(d'après la question } \textbf{1.a)}\textit{)} \\ &= \log_2(\lambda) + \frac{\ln\big(\mathrm{e}^{-\lambda x}\big)}{\ln(2)} \\ &= \log_2(\lambda) + \frac{-\lambda x}{\ln(2)} \\ &= \log_2(\lambda) - \frac{\lambda}{\ln(2)}\,x \end{aligned}$$
Ainsi :
$$\begin{aligned} f_Z(x)\,\log_2\big(f_Z(x)\big) &= f_Z(x)\left(\log_2(\lambda) - \frac{\lambda}{\ln(2)}\,x\right) \\ &= \log_2(\lambda)\,f_Z(x) - \frac{\lambda}{\ln(2)}\,x\,f_Z(x) \end{aligned}$$
- Or,
  - × l'intégrale $\displaystyle\int_0^{+\infty} f_Z(x)\,dx$ est convergente car f_Z est une densité. De plus, comme f_Z est nulle en dehors de $[0, +\infty[$:
$$\int_0^{+\infty} f_Z(x)\,dx = \int_{-\infty}^{+\infty} f_Z(x)\,dx = 1$$
  - × l'intégrale $\displaystyle\int_0^{+\infty} x\,f_Z(x)\,dx$ est convergente car la v.a.r. Z admet une espérance.
  De plus, toujours comme f_Z est nulle en dehors de $[0, +\infty[$:
$$\int_0^{+\infty} x\,f_Z(x)\,dx = \int_{-\infty}^{+\infty} x\,f_Z(x)\,dx = \mathbb{E}(Z) = \frac{1}{\lambda}$$
- On en déduit que l'intégrale $\displaystyle\int_0^{+\infty} f_Z(x)\,\log_2\big(f_Z(x)\big)\,dx$ est convergente en tant que combinaison linéaire d'intégrales convergentes.

Ainsi, la v.a.r. Z admet une entropie différentielle.

- De plus :

$$\begin{aligned}
& h(Z) \\
= & -\int_0^{+\infty} f_Z(x)\,\log_2\big(f_Z(x)\big)\,dx \\
= & -\int_0^{+\infty} \log_2(\lambda)\,f_Z(x) - \frac{\lambda}{\ln(2)}\,x\,f_Z(x)\,dx && \textit{(d'après les calculs précédents)} \\
= & -\log_2(\lambda)\int_0^{+\infty} f_Z(x)\,dx + \frac{\lambda}{\ln(2)}\int_0^{+\infty} x\,f_Z(x)\,dx && \textit{(par linéarité de l'intégrale, car les intégrales en présence sont convergentes)} \\
= & -\log_2(\lambda)\times 1 + \frac{\cancel{\lambda}}{\ln(2)}\times\frac{1}{\cancel{\lambda}} \\
= & -\log_2(\lambda) + \log_2(\mathrm{e})
\end{aligned}$$

D'après la question ***1.a)***, on obtient : $h(Z) = \log_2\left(\frac{\mathrm{e}}{\lambda}\right)$.

Commentaire

- On a démontré que pour tout $x \in [0,+\infty[$: $\log_2\big(f_Z(x)\big) = \log_2(\lambda) - \frac{\lambda}{\ln(2)}\,x$.
  Comme on l'a dit en remarque précédente, cela ne permet pas, de manière rigoureuse, de conclure :
  $$\log_2\big(f_Z(Z)\big) = \log_2(\lambda) - \frac{\lambda}{\ln(2)}\,Z$$
  Pour pouvoir le faire, il faudrait savoir : $Z(\Omega) \subseteq [0,+\infty[$.
  Cela signifie : $\forall \omega \in \Omega,\ Z(\omega) \in [0,+\infty[$. Si c'est le cas, on peut écrire, pour tout $\omega \in \Omega$:
  $$\log_2\Big(f_Z\big(Z(\omega)\big)\Big) = \log_2(\lambda) - \frac{\lambda}{\ln(2)}\,Z(\omega)$$

- Il est possible de rédiger comme suit.
  Dans la suite, on **considère** $Z(\Omega) \subseteq [0,+\infty[$. Ainsi :
  $$\log_2\big(f_Z(Z)\big) = \log_2(\lambda) - \frac{\lambda}{\ln(2)}\,Z$$
  La v.a.r. $\log_2\big(f_Z(Z)\big)$ admet une espérance en tant que transformée affine de la v.a.r. Z qui admet une espérance. On en déduit que Z admet une entropie différentielle. De plus :

$$\begin{aligned}
h(Z) &= -\mathbb{E}\big(\log_2(\lambda) - \frac{\lambda}{\ln(2)}\,Z\big) \\
&= -\mathbb{E}\big(\log_2(\lambda)\big) + \frac{\lambda}{\ln(2)}\,\mathbb{E}(Z) && \textit{(par linéarité de l'espérance)} \\
&= -\log_2(\lambda) + \frac{\cancel{\lambda}}{\ln(2)}\,\frac{1}{\cancel{\lambda}} \\
&= \log_2\left(\frac{1}{\lambda}\right) + \log_2(\mathrm{e}) = \log_2\left(\frac{\mathrm{e}}{\lambda}\right)
\end{aligned}$$

□

d) Soit f la fonction définie sur $\mathbb{R}$ par $f(x) = \dfrac{1}{2}\,\lambda\,\mathrm{e}^{-\lambda\,|x|}$ $(\lambda > 0)$.

(i) Montrer que f est une densité de probabilité sur $\mathbb{R}$.

Démonstration.

- La fonction $x \mapsto \mathrm{e}^{-\lambda|x|}$ est continue sur $\mathbb{R}$ car elle est la composée $\exp \circ g_3$ où :
  - × $g_3 : x \mapsto -\lambda|x|$ est :
    - continue sur $\mathbb{R}$ en tant que produit de fonctions continues sur $\mathbb{R}$,
    - telle que : $g_3(\mathbb{R}) \subset \mathbb{R}$ (en fait on a même : $g_3(\mathbb{R}) \subset]-\infty, 0]$).
  - × exp est continue sur $\mathbb{R}$.

On en déduit que la fonction f est continue sur $\mathbb{R}$.

Commentaire

- La fonction g_3 est définie par cas (si $x \geqslant 0$, alors $|x| = x$ et si $x < 0$ alors $|x| = -x$). Le réflexe pour ce type de fonctions est d'étudier la continuité sur les intervalles ouverts $]-\infty, 0[$ et $]0, +\infty[$). Ce n'est pas nécessaire ici puisque la fonction valeur absolue est une fonction usuelle dont on sait qu'elle est continue sur $\mathbb{R}$ en entier.
- La fonction exp est continue sur $\mathbb{R}$ tout en entier. Ainsi, la détermination précise de l'ensemble $g_3(\mathbb{R})$ a peu d'importance dans la rédaction ci-dessus. Le domaine de continuité de g_3 donne le domaine de continuité de la fonction exp.
- Afin de permettre une bonne compréhension de ce point, on a rédigé en détails la continuité de la composée. Mais lorsque les deux fonctions en présence (exp et g_3 ici) sont continues sur $\mathbb{R}$ tout en entier, on peut se contenter de la rédaction plus succincte suivante :

  « La fonction f est continue sur $\mathbb{R}$ comme composée de deux fonctions continues sur $\mathbb{R}$ »

- De manière générale, il n'est pas nécessaire de rédiger aussi précisément les questions portant sur la régularité de fonctions. Il est conseillé :
  - × de rédiger très proprement la régularité d'une fonction pour les questions que l'on traite en premier. On démontre ainsi au correcteur sa capacité à rédiger ce type de questions.
  - × de rédiger très proprement la régularité lorsqu'il s'agit du cœur de la question (« Démontrer que la fonction est continue / de classe $\mathcal{C}^1$ sur ... »

  Dans les autres cas, on pourra se contenter d'écrire que la fonction f est continue sur J (intervalle à déterminer) car elle est la composée de fonctions continues sur les intervalles adéquats. Évidemment, cela n'apportera pas de point si l'intervalle J n'est pas le bon.

- Soit $x \in \mathbb{R}$.
  Comme $\lambda > 0$ et $\mathrm{e}^{-\lambda|x|} > 0$, on obtient : $f(x) > 0$.

$$\forall x \in \mathbb{R},\ f(x) \geqslant 0$$

- Montrons que l'intégrale $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx$ converge et vaut 1.
  - × Tout d'abord, l'intégrale $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx$ converge si et seulement si $\displaystyle\int_{-\infty}^{0} f(x)\,dx$ et $\displaystyle\int_{0}^{+\infty} f(x)\,dx$ convergent.

× Soit $B \in [0, +\infty[$.

$$\begin{aligned}
\int_0^B f(x)\,dx &= \int_0^B \frac{1}{2}\,\lambda\,\mathrm{e}^{-\lambda|x|}\,dx \\
&= \frac{1}{2}\int_0^B \lambda\,\mathrm{e}^{-\lambda\,x}\,dx \\
&= \frac{1}{2}\int_0^B f_Z(x)\,dx \\
&\underset{B\to+\infty}{\longrightarrow} \frac{1}{2}\int_0^{+\infty} f_Z(x)\,dx = \frac{1}{2}\int_{-\infty}^{+\infty} f_Z(x)\,dx = \frac{1}{2}
\end{aligned}$$

Le passage à la limite est justifié par le fait que l'intégrale $\displaystyle\int_0^{+\infty} f_Z(x)\,dx$ est convergente car f_Z est une densité de probabilité.

On en déduit que l'intégrale $\displaystyle\int_0^{+\infty} f(x)\,dx$ est convergente et vaut $\frac{1}{2}$.

Commentaire

- Il est conseillé de savoir repérer les intégrales usuelles issues du chapitre Probabilités. Cela permet souvent de simplifier les calculs et ainsi de gagner un temps précieux. Pour cela, il est nécessaire de maîtriser les définitions des lois usuelles à densités. Par exemple :
  × les moments d'ordre 0, 1 et 2 des lois exponentielles permettent d'obtenir :
  $$\int_0^{+\infty} \lambda\,\mathrm{e}^{-\lambda x} = 1 \quad , \quad \int_0^{+\infty} \lambda\,x\,\mathrm{e}^{-\lambda x} = \frac{1}{\lambda} \quad , \quad \int_0^{+\infty} \lambda\,x^2\,\mathrm{e}^{-\lambda x} = \frac{1}{\lambda^2} + \left(\frac{1}{\lambda}\right)^2$$
  × les moments d'ordre 0, 1 et 2 de la loi normale centrée réduite permet d'obtenir :
  $$\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}x^2}\,dx = \sqrt{2\pi}\,, \int_{-\infty}^{+\infty} x\,\mathrm{e}^{-\frac{1}{2}x^2}\,dx = 0\,, \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\,x^2\,\mathrm{e}^{-\frac{1}{2}x^2}\,dx = 1 + 0 = 1$$
  De plus, à l'aide de la parité des intégrandes, on peut aussi démontrer :
  $$\int_0^{+\infty} \mathrm{e}^{-\frac{1}{2}x^2}\,dx = \frac{\sqrt{2\pi}}{2} \quad , \quad \int_0^{+\infty} x^2\,\mathrm{e}^{-\frac{1}{2}x^2}\,dx = \frac{1}{2}$$
- Pour cette question, il était aussi possible de réaliser le calcul (fort simple) suivant :
  $$\int_0^B f(x)\,dx = \frac{1}{2}\int_0^B \lambda\,\mathrm{e}^{-\lambda\,x}\,dx = \frac{1}{2}\left[-\mathrm{e}^{-\lambda x}\right]_0^B = \frac{1}{2}\left(-\mathrm{e}^{-\lambda B} + 1\right) \underset{B\to+\infty}{\longrightarrow} \frac{1}{2}$$

- Remarquons que la fonction f est paire.
  Soit $x \in \mathbb{R}$, alors $-x \in \mathbb{R}$.
  $$f(-x) = \frac{1}{2}\,\lambda\,\mathrm{e}^{-\lambda\,|-x|} = \frac{1}{2}\,\lambda\,\mathrm{e}^{-\lambda\,|x|} = f(x)$$

- La fonction f est paire. Or l'intégrale $\displaystyle\int_0^{\infty} f(x)\,dx$ est convergente.

  On en déduit que l'intégrale $\displaystyle\int_{-\infty}^{0} f(x)\,dx$ est convergente et :

$$\int_{-\infty}^{0} f(x)\,dx \;=\; \int_0^{+\infty} f(x)\,dx$$

Commentaire

Rappelons que ce point se démontre à l'aide du changement de variable $\boxed{u = -x}$.

$$\left|\begin{array}{l} u = -x \text{ (donc } x = -u) \\ \hookrightarrow du = -\,dx \quad \text{et} \quad dx = -du \\ \bullet\ x = -\infty \;\Rightarrow\; u = +\infty \\ \bullet\ x = 0 \;\Rightarrow\; u = 0 \end{array}\right.$$

Ce changement de variable est valide car $\psi : u \mapsto -u$ est $\mathcal{C}^1$ sur $[0, +\infty[$. On obtient :

$$\int_{-\infty}^{0} f(x)\,dx \;=\; \int_{+\infty}^{0} f(-u)(-\,du) \;=\; -\int_{+\infty}^{0} f(u)\,du \;=\; \int_0^{+\infty} f(u)\,du$$

- On en conclut que l'intégrale $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx$ est convergente. De plus :

$$\int_{-\infty}^{+\infty} f(x)\,dx \;=\; \int_{-\infty}^{0} f(x)\,dx + \int_0^{+\infty} f(x)\,dx \;=\; 2\,\frac{1}{2} \;=\; 1$$

$$\boxed{\text{L'intégrale } \int_{-\infty}^{+\infty} f(x)\,dx \text{ converge. De plus : } \int_{-\infty}^{+\infty} f(x)\,dx \;=\; 1.}$$

$$\boxed{\text{On en déduit que la fonction } f \text{ est une densité de probabilité.}}$$

□

(ii) Soit W une variable aléatoire de densité f. Justifier l'existence de l'entropie différentielle $h(W)$ et la déterminer.

Démonstration.

- Par définition de la fonction f, son support est $]-\infty, +\infty[$.
- Soit $x \in \mathbb{R}$.

$$\begin{aligned}
\log_2\big(f(x)\big) &= \log_2\left(\frac{1}{2}\,\lambda\,\mathrm{e}^{-\lambda\,|x|}\right) \\
&= -\log_2(2) + \log_2(\lambda) + \log_2\left(\mathrm{e}^{-\lambda\,|x|}\right) && \textit{(d'après la question } \textbf{1.a)}\textit{)} \\
&= -1 + \log_2(\lambda) + \frac{\ln\left(\mathrm{e}^{-\lambda\,|x|}\right)}{\ln(2)} && \textit{(car } \log_2(2^1) = 1 \textit{ d'après } \textbf{1.b)}\textit{)} \\
&= -1 + \log_2(\lambda) - \frac{\lambda}{\ln(2)}\,|x|
\end{aligned}$$

En particulier :

$$\log_2\big(f(W)\big) \;=\; -1 + \log_2(\lambda) - \frac{\lambda}{\ln(2)}\,|W|$$

- La v.a.r. $\log_2\big(f(W)\big)$ est une transformée affine de la v.a.r. $|W|$.

Ainsi, la v.a.r. $\log_2\big(f(W)\big)$ admet une espérance si et seulement si la v.a.r. $|W|$ en admet une.

- La v.a.r. $|W|$ admet une espérance si et seulement si l'intégrale $\displaystyle\int_{-\infty}^{+\infty} |x|\, f(x)\, dx$ est convergente.

  × Démontrons tout d'abord que l'intégrale $\displaystyle\int_{0}^{+\infty} |x|\, f(x)\, dx$ est convergente.

  × La fonction $x \mapsto |x|\, f(x)$ est continue sur $[0, +\infty[$ en tant que produit de fonctions continues sur $[0, +\infty[$. Soit $B \in [0, +\infty[$.

$$\begin{aligned}\int_0^B |x|\, f(x)\, dx &= \frac{1}{2}\int_0^B x\, \lambda\, \mathrm{e}^{-\lambda x}\, dx \\ &= \frac{1}{2}\int_0^B x\, f_Z(x)\, dx \qquad (car\ Z \hookrightarrow \mathcal{E}(\lambda))\end{aligned}$$

  Or la v.a.r. Z admet une espérance. Ainsi, avec le même raisonnement qu'en question ***4.c)***, on en déduit que l'intégrale $\displaystyle\int_{0}^{+\infty} x\, f_Z(x)\, dx$ est convergente et :

$$\int_0^{+\infty} x\, f_Z(x)\, dx = \mathbb{E}(Z) = \frac{1}{\lambda}$$

Finalement, l'intégrale $\displaystyle\int_{0}^{+\infty} |x| f(x)\, dx$ est convergente et : $\displaystyle\int_{0}^{+\infty} |x|\, f(x)\, dx = \frac{1}{2} \times \frac{1}{\lambda}$.

  × Remarquons que la fonction $h : x \mapsto |x|\, f(x)$ est paire.
  Soit $x \in \mathbb{R}$, alors $-x \in \mathbb{R}$.

$$h(-x) = |-x| \times \frac{1}{2}\, \lambda\, \mathrm{e}^{-\lambda|-x|} = |x| \times \frac{1}{2}\, \lambda\, \mathrm{e}^{-\lambda|x|} = h(x)$$

  La fonction h est paire. Or l'intégrale $\displaystyle\int_{0}^{\infty} h(x)\, dx$ est convergente.
  On en déduit que l'intégrale $\displaystyle\int_{-\infty}^{0} h(x)\, dx$ est convergente et :

$$\int_{-\infty}^{0} h(x)\, dx = \int_{0}^{+\infty} h(x)\, dx$$

  On en conclut que l'intégrale $\displaystyle\int_{-\infty}^{+\infty} h(x)\, dx$ est convergente. De plus :

$$\int_{-\infty}^{+\infty} h(x)\, dx = \int_{-\infty}^{0} h(x)\, dx + \int_{0}^{+\infty} h(x)\, dx = 2\,\frac{1}{2\,\lambda} = \frac{1}{\lambda}$$

Finalement, la v.a.r. $|W|$ admet une espérance et ainsi la v.a.r. $\log_2\big(f(W)\big)$ en admet aussi une.

- On en conclut que W admet une entropie différentielle. De plus :

$$\begin{aligned} h(W) &= -\mathbb{E}\Big(\log_2\big(f(W)\big)\Big) \\ &= -\mathbb{E}\left(-1+\log_2(\lambda)-\frac{\lambda}{\ln(2)}\,|W|\right) \\ &= \mathbb{E}(1)-\mathbb{E}\big(\log_2(\lambda)\big)+\frac{\lambda}{\ln(2)}\,\mathbb{E}\big(|W|\big) && \textit{(par linéarité de l'espérance)} \\ &= 1-\log_2(\lambda)+\frac{\cancel{\lambda}}{\ln(2)}\,\frac{1}{\cancel{\lambda}} \\ &= 1-\log_2(\lambda)+\log_2(\mathrm{e}) \;=\; 1+\log_2\left(\frac{1}{\lambda}\right)+\log_2(\mathrm{e}) \end{aligned}$$

Finalement, on obtient : $h(W)=1+\log_2\left(\frac{\mathrm{e}}{\lambda}\right)$. □

Commentaire

Dans cette question, on étudie une loi classique mais hors programme. Plus précisément, la v.a.r. W suit la *loi de Laplace de paramètre* $(0,1)$. L'étude de cette loi est fréquente aux concours. On pourra par exemple regarder l'épreuve `ESSEC-I 2017` pour une étude plus détaillée de cette loi.

5. On dit qu'un couple (X,Y) de variables aléatoires est un couple gaussien centré si, pour tout $(\alpha,\beta)\in\mathbb{R}^2$, $\alpha X+\beta Y$ est une variable de loi normale centrée, c'est-à-dire qu'il existe $\gamma\in\mathbb{R}$ et une variable aléatoire Z de loi normale centrée réduite tels que $\alpha X+\beta Y$ a même loi que γZ. On considère un tel couple (X,Y) et on noté σ^2 la variance de X. On suppose : $\sigma^2>0$.

a) Montrer que X suit une loi normale centrée.

Démonstration.
Comme (X,Y) est un couple gaussien centré, alors **pour tout** $(\alpha,\beta)\in\mathbb{R}^2$, $\alpha X+\beta Y$ suit une loi normale centrée. **En particulier**, la v.a.r. $1\cdot X+0\cdot Y=X$ suit une loi normale centrée. De plus, d'après l'énoncé : $\mathbb{V}(X)=\sigma^2$.

Finalement : $X\hookrightarrow\mathcal{N}\left(0,\sigma^2\right)$. □

b) Calculer $h(X)$.

Démonstration.

- D'après la question précédente, X est une v.a.r. de loi normale centrée $\left(X\hookrightarrow\mathcal{N}\left(0,\sigma^2\right)\right)$. Il existe donc $\gamma\in\mathbb{R}$ et $Z\in\mathcal{N}(0,1)$ tels que X a même loi que γZ. On a ainsi :

$$\begin{array}{ccccccc} \mathbb{V}(X) & = & \mathbb{V}(\gamma Z) & = & \gamma^2\,\mathbb{V}(Z) & = & \gamma^2 \\ \| & & & & & & \\ \sigma^2 & & & & & & \end{array}$$

Finalement, X suit la même loi que la v.a.r. σZ, où : $Z\hookrightarrow\mathcal{N}(0,1)$.

- De plus, d'après la question ***4.b)***, la v.a.r. Z admet une entropie différentielle et :

$$h(Z) \;=\; \frac{1}{2}\log_2(2\pi\mathrm{e})$$

- Enfin, d'après la question ***3.b)(ii)*** :
  - × la v.a.r. X admet une entropie différentielle,
  - × on obtient de plus :

$$h(X) \;=\; h(Z) + \log_2(\sigma) \;=\; \frac{1}{2}\,\log(2\pi\mathrm{e}) + \frac{1}{2}\,\log_2(\sigma^2) \;=\; \frac{1}{2}\,\log_2(2\pi\,\mathrm{e}\,\sigma^2)$$

$$\boxed{h(X) \;=\; \frac{1}{2}\,\log_2(2\pi\,\mathrm{e}\,\sigma^2)}$$

Commentaire

- Rappelons que les lois normales sont stables par transformation affine. Plus précisément :

$$X \hookrightarrow \mathcal{N}\left(m, \sigma^2\right) \;\Leftrightarrow\; a\,X + b \hookrightarrow \mathcal{N}\left(am + b, a^2\,\sigma^2\right)$$

- En particulier : $Z \hookrightarrow \mathcal{N}(0,1) \;\Leftrightarrow\; \sigma\,Z \hookrightarrow \mathcal{N}\left(0, \sigma^2\right)$.
  Ainsi, si $X \hookrightarrow \mathcal{N}\left(0, \sigma^2\right)$ alors X a même loi que $\sigma\,Z$ où $Z \hookrightarrow \mathcal{N}(0,1)$.

□

c) On suppose désormais que X et Y suivent la même loi normale centrée de variance σ^2 et on admet que les propriétés de l'espérance des variables discrètes se généralisent aux variables aléatoires quelconques.

(i) Montrer que $\mathbb{E}(XY)$ existe.

Démonstration.
Les v.a.r. X et Y suivent une loi normale.
Elles admettent donc chacune un moment d'ordre 2.

On en déduit que la v.a.r. XY admet une espérance.

Commentaire

- On rappelle que, dans le cas général, la v.a.r. produit XY admet une espérance si les v.a.r. X et Y admettent **un moment d'ordre** 2.
- On peut se demander d'où provient cette hypothèse liée aux moments d'ordre 2.
  Elle est issue d'un théorème de domination. Détaillons ce point.
  Remarquons tout d'abord : $(X-Y)^2 \geqslant 0$.
  On en déduit : $X^2 - 2\,XY + Y^2 \geqslant 0$.
  Et, en réordonnant : $XY \leqslant \dfrac{1}{2}\,X^2 + \dfrac{1}{2}\,Y^2$. Ou encore :

$$0 \;\leqslant\; |XY| \;\leqslant\; \frac{1}{2}\,X^2 + \frac{1}{2}\,Y^2$$

  Comme X et Y admettent un moment d'ordre 2, la v.a.r. $\frac{1}{2}\,X^2 + \frac{1}{2}\,Y^2$ admet une espérance comme combinaison linéaire de v.a.r. qui en admettent une.
  Ainsi, par théorème de domination (présenté seulement dans le programme ECS), la v.a.r. $|XY|$ admet une espérance. Il en est de même de la v.a.r. XY.

□

(ii) Montrer de plus, pour tout réel λ : $\lambda^2\,\mathbb{E}\left(Y^2\right)+2\lambda\,\mathbb{E}(XY)+\mathbb{E}\left(X^2\right)\ \geqslant\ 0$.

Démonstration.
Soit $\lambda\in\mathbb{R}$.

$$\begin{aligned}\lambda^2\,\mathbb{E}(Y^2)+2\lambda\,\mathbb{E}(XY)+\mathbb{E}(X^2) &= \mathbb{E}\left(\lambda^2\,Y^2+2\lambda\,XY+X^2\right) && \textit{(par linéarité de l'espérance)}\\ &= \mathbb{E}\Big((\lambda\,Y+X)^2\Big)\end{aligned}$$

Or : $(\lambda\,Y+X)^2\geqslant 0$.
Ainsi, par croissance de l'espérance : $\mathbb{E}\Big((\lambda Y+X)^2\Big)\geqslant 0$.

On en déduit : $\lambda^2\,\mathbb{E}(Y^2)+2\lambda\,\mathbb{E}(XY)+\mathbb{E}(X^2)\geqslant 0$.

Commentaire

On prêtera attention à la différence de nature des deux inégalités citées plus haut :

× l'inégalité $(\lambda\,Y+X)^2\geqslant 0$ est une inégalité entre **variables aléatoires**, c'est-à-dire entre applications de Ω dans $\mathbb{R}$. En particulier, le membre droit « 0 » est ici la variable aléatoire constante égale à 0. Cette inégalité signifie :

$$\forall\omega\in\Omega,\ (\lambda\,Y+X)^2(\omega)=(\lambda\,Y(\omega)+X(\omega))^2\geqslant 0$$

Cette dernière inégalité est par contre une comparaison entre réels.

× l'inégalité $\mathbb{E}\Big((\lambda\,Y+X)^2\Big)\geqslant 0$ est une inégalité entre **réels**.
Ainsi, le membre droit « 0 » est ici tout simplement le réel 0.

□

(iii) En déduire : $\big(\mathbb{E}(XY)\big)^2\ \leqslant\ \mathbb{E}\left(X^2\right)\,\mathbb{E}\left(Y^2\right)$.

Démonstration.

- La fonction g définie par :

$$\forall\lambda\in\mathbb{R},\quad g(\lambda)=\lambda^2\,\mathbb{E}(Y^2)+2\lambda\,\mathbb{E}(XY)+\mathbb{E}(X^2)$$

est une fonction polynomiale de degré 2 en λ.

- De plus, d'après la question précédente : $\forall\lambda\in\mathbb{R}$, $g(\lambda)\geqslant 0$.
Cette fonction polynomiale étant de signe constant, on en déduit que le discriminant du polynôme associé est négatif. Or :

$$\Delta\ =\ \big(2\,\mathbb{E}(XY)\big)^2-4\,\mathbb{E}(Y^2)\,\mathbb{E}(X^2)\ =\ 4\left(\big(\mathbb{E}(XY)\big)^2-\mathbb{E}(Y^2)\,\mathbb{E}(X^2)\right)$$

On obtient donc :

$$\Delta\leqslant 0\quad\Leftrightarrow\quad\big(\mathbb{E}(XY)\big)^2\ \leqslant\ \mathbb{E}(Y^2)\,\mathbb{E}(X^2)$$

Ainsi : $\big(\mathbb{E}(XY)\big)^2\ \leqslant\ \mathbb{E}(Y^2)\,\mathbb{E}(X^2)$.

□

Commentaire

On reconnaît ici le principe de démonstration de l'inégalité de Cauchy-Schwarz.
D'après le cours, si X et Y sont des v.a.r. **discrètes** admettant un moment d'ordre 2 alors :

$$\big(\operatorname{Cov}(X,Y)\big)^2\ \leqslant\ \mathbb{V}(X)\,\mathbb{V}(Y)$$

Commentaire

- Rappelons que dans le programme ECE, l'existence de l'espérance d'un produit XY de v.a.r. est étudiée seulement dans deux cas :

× si les v.a.r. X et Y sont **discrètes**, alors, sous hypothèse de convergence absolue, on sait définir $\mathbb{E}(XY)$ (*cf* cours sur les couples de v.a.r.).

× si les v.a.r. X et Y sont indépendantes et admettent toutes deux une espérances alors la v.a.r. XY admet une espérance donnée par : $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$.

Dans cette question, on ne fait pas l'hypothèse d'indépendance de X et Y. C'est d'ailleurs pour cela que l'énoncé précise que « les propriétés de l'espérance des v.a.r. discrètes se généralisent aux v.a.r. quelconques ». Cela permet d'étudier l'existence de l'espérance du produit XY même si X et Y sont à densité.

- En particulier, dans le programme ECE, la notion de covariance de deux v.a.r. X et Y n'est elle aussi définie que dans le cas de deux v.a.r. **discrètes**. C'est ce qui explique que la notation $\mathrm{Cov}(X,Y)$ n'apparaisse pas. Finalement, de manière assez habile, l'énoncé permet d'établir le théorème de Cauchy-Schwarz pour des v.a.r. **centrées** à densité. Ce caractère centré ($\mathbb{E}(X) = 0$ et $\mathbb{E}(Y) = 0$) est l'élément qui permet d'établir de considérer : $\mathbb{E}(XY)$ en lieu et place de : $\mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y) = \mathrm{Cov}(X,Y)$.

(iv) On pose $\rho = \dfrac{\mathbb{E}(XY)}{\sigma^2}$. Montrer : $\rho \in [-1,1]$.

Démonstration.

- D'après la question précédente :

$$\big(\mathbb{E}(XY)\big)^2 \leqslant \mathbb{E}(Y^2)\,\mathbb{E}(X^2) \Leftrightarrow \sqrt{\big(\mathbb{E}(XY)\big)^2} \leqslant \sqrt{\mathbb{E}(Y^2)\,\mathbb{E}(X^2)} \quad \textit{(par stricte croissance de } \sqrt{\cdot} \textit{ sur } [0,+\infty[\textit{)}$$
$$\Leftrightarrow |\mathbb{E}(XY)| \leqslant \sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}$$

- De plus, par formule de Koenig-Huygens :

$$\begin{array}{ccccc} \mathbb{V}(X) & = & \mathbb{E}(X^2) & - & \big(\mathbb{E}(X)\big)^2 \\ \| & & & & \| \\ \sigma^2 & & & & 0 \end{array} \quad \textit{(car } X \textit{ est une v.a.r. centrée)}$$

Ainsi : $\mathbb{E}(X^2) = \sigma^2$. De même : $\mathbb{E}(Y^2) = \sigma^2$.

- On en déduit :

$$\begin{aligned} |\mathbb{E}(XY)| \leqslant \sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)} &\Leftrightarrow |\mathbb{E}(XY)| \leqslant \sqrt{\sigma^2}\sqrt{\sigma^2} \\ &\Leftrightarrow |\mathbb{E}(XY)| \leqslant \sigma^2 \\ &\Leftrightarrow \frac{|\mathbb{E}(XY)|}{\sigma^2} \leqslant 1 \qquad \textit{(car } \sigma^2 > 0\textit{)} \\ &\Leftrightarrow \left|\frac{\mathbb{E}(XY)}{\sigma^2}\right| \leqslant 1 \qquad \textit{(car } \sigma^2 \geqslant 0\textit{)} \\ &\Leftrightarrow |\rho| \leqslant 1 \end{aligned}$$

Ainsi : $-1 \leqslant \rho \leqslant 1$.

Commentaire

- On reconnaît la notation ρ utilisée pour définir le *coefficient de corrélation linéaire* de deux v.a.r. X et Y.
- Rappelons que si X et Y sont deux v.a.r. **discrètes** qui admettent chacune un moment d'ordre 2, alors X et Y admettent un *coefficient de corrélation linéaire* défini par :

$$\rho(X,Y) \;=\; \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)}\,\sqrt{\mathbb{V}(Y)}} \;=\; \frac{\mathbb{E}(XY)-\mathbb{E}(X)\,\mathbb{E}(Y)}{\sqrt{\mathbb{V}(X)}\,\sqrt{\mathbb{V}(Y)}}$$

- Plus précisément, on calcule ici $\rho(X,Y)$ dans le cas où X et Y sont deux v.a.r. :
  - × centrées (dans ce cas : $\mathbb{E}(XY)-\mathbb{E}(X)\,\mathbb{E}(Y)=\mathbb{E}(XY)$),
  - × de variance σ^2.
- Comme dit précédemment, le programme d'ECE ne définit ce coefficient que pour des v.a.r. discrètes. On peut cependant généraliser cette définition au cas de v.a.r. quelconques. Tout l'objectif de la question ***5.c)*** est de démontrer que l'on conserve les grandes propriétés du coefficient de corrélation linéaire dans le cas où les v.a.r. étudiées sont des v.a.r. à densité **centrées**.

□

(v) Que vaut ρ si X et Y sont indépendantes ?

Démonstration.
Si les v.a.r. X et Y sont indépendantes, alors :

$$\begin{aligned}\mathbb{E}(XY) &= \mathbb{E}(X)\,\mathbb{E}(Y)\\ &= 0 \qquad \textit{(car } X \textit{ et } Y \textit{ sont des v.a.r. centrées)}\end{aligned}$$

Ainsi, si les v.a.r. X et Y sont indépendantes, alors : $\rho \;=\; \dfrac{\mathbb{E}(XY)}{\sigma^2} \;=\; 0.$

□

d) On suppose $|\rho|<1$. On appelle **entropie jointe** du couple (X,Y) le réel :

$$h(X,Y) \;=\; \log_2\left(2\pi\,\mathrm{e}\,\sigma^2\,\sqrt{1-\rho^2}\right)$$

(i) À quelle condition $h(X,Y)$ est-elle nulle ?

Démonstration.
On a les équivalences suivantes :

$$\begin{aligned}h(X,Y)=0 \;&\Leftrightarrow\; \log_2\left(2\pi\,\mathrm{e}\,\sigma^2\,\sqrt{1-\rho^2}\right) \;=\; 0\\ &\Leftrightarrow\; \frac{\ln\left(2\pi\,\mathrm{e}\,\sigma^2\,\sqrt{1-\rho^2}\right)}{\ln(2)} \;=\; 0\\ &\Leftrightarrow\; \ln\left(2\pi\,\mathrm{e}\,\sigma^2\,\sqrt{1-\rho^2}\right) \;=\; 0\\ &\Leftrightarrow\; 2\pi\,\mathrm{e}\,\sigma^2\,\sqrt{1-\rho^2} \;=\; 1 \qquad \textit{(par bijectivité de la fonction } \exp \textit{ sur } \mathbb{R}\textit{)}\end{aligned}$$

De plus :

$$\begin{aligned}
2\pi\,\mathrm{e}\,\sigma^2\sqrt{1-\rho^2} \;=\; 1 \;&\Leftrightarrow\; \sqrt{1-\rho^2} \;=\; \frac{1}{2\pi\mathrm{e}\sigma^2} \\
&\Leftrightarrow\; 1-\rho^2 \;=\; \left(\frac{1}{2\pi\mathrm{e}\sigma^2}\right)^2 && \textit{(par bijectivité de la fonction } x\mapsto x^2 \textit{ sur } [0,+\infty[\textit{)} \\
&\Leftrightarrow\; \rho^2 \;=\; 1-\frac{1}{\left(2\pi\mathrm{e}\sigma^2\right)^2} \\
&\Leftrightarrow\; \sqrt{\rho^2} \;=\; \sqrt{1-\frac{1}{\left(2\pi\mathrm{e}\sigma^2\right)^2}} \\
&\Leftrightarrow\; |\rho| \;=\; \sqrt{1-\frac{1}{\left(2\pi\mathrm{e}\sigma^2\right)^2}}
\end{aligned}$$

$$\text{Finalement : } h(X,Y)=0 \;\Leftrightarrow\; \rho\in\left\{-\sqrt{1-\frac{1}{\left(2\pi\mathrm{e}\sigma^2\right)^2}}\;,\;\sqrt{1-\frac{1}{\left(2\pi\mathrm{e}\sigma^2\right)^2}}\right\}.$$

Commentaire

Cette question est un peu vague :
- × demande-t-on une condition nécessaire ? suffisante ? nécessaire et suffisante ?
- × demande-t-on une condition sur ρ ou sur un autre paramètre ?

Sans plus de précisions, on considèrera que les termes « nécessaire et suffisante » sont sous-entendus. On choisit par ailleurs dans ce corrigé de considérer qu'on demande bien une condition sur le paramètre ρ.

□

(ii) L'**information mutuelle** de X et Y est définie par :

$$I(X,Y) \;=\; h(X)+h(Y)-h(X,Y)$$

Calculer $I(X,Y)$.

Démonstration.
On calcule :

$$\begin{aligned}
I(X,Y) \;&=\; h(X)+h(Y)-h(X,Y) \\
&=\; \frac{1}{2}\log_2(2\pi\,\mathrm{e}\,\sigma^2)+\frac{1}{2}\log_2(2\pi\,\mathrm{e}\,\sigma^2)-\log_2\left(2\pi\,\mathrm{e}\,\sigma^2\sqrt{1-\rho^2}\right) && \textit{(d'après la question } \textbf{5.b)}\textit{)} \\
&=\; \log_2(2\pi\,\mathrm{e}\,\sigma^2)-\log_2\left(2\pi\,\mathrm{e}\,\sigma^2\sqrt{1-\rho^2}\right) \\
&=\; -\log_2\left(\frac{\cancel{2\pi\,\mathrm{e}\,\sigma^2}\sqrt{1-\rho^2}}{\cancel{2\pi\,\mathrm{e}\,\sigma^2}}\right) && \textit{(d'après la question } \textbf{1.a)}\textit{)} \\
&=\; -\log_2\left(\sqrt{1-\rho^2}\right)
\end{aligned}$$

$$\text{Finalement : } I(X,Y) \;=\; -\frac{1}{2}\log_2(1-\rho^2)$$

□

(iii) Montrer : $I(X,Y) \geqslant 0$.

Démonstration.
On a la succession d'équivalences suivante :

$$\begin{aligned}
I(X,Y) \geqslant 0 &\Leftrightarrow -\frac{1}{2}\log_2(1-\rho^2) \geqslant 0 && \textit{(d'après la question précédente)}\\
&\Leftrightarrow \log_2(1-\rho^2) \leqslant 0 &&\\
&\Leftrightarrow \frac{\ln(1-\rho^2)}{\ln(2)} \leqslant 0 &&\\
&\Leftrightarrow \ln(1-\rho^2) \leqslant 0 && \textit{(car } \ln(2) > 0\textit{)}\\
&\Leftrightarrow 1-\rho^2 \leqslant 1 && \textit{(par stricte croissance de la fonction } \exp \textit{ sur } \mathbb{R}\textit{)}\\
&\Leftrightarrow 0 \leqslant \rho^2 &&
\end{aligned}$$

La dernière inégalité est vraie. Ainsi, par équivalence, la première inégalité aussi.

Finalement : $I(X,Y) \geqslant 0$.

□

(iv) Quelle est la limite de $I(X,Y)$ quand ρ tend vers 1 ?

Démonstration.

- D'après la question ***5.d)(ii)*** :

$$I(X,Y) \;=\; -\frac{1}{2}\log_2(1-\rho^2) \;=\; -\frac{1}{2}\,\frac{\ln(1-\rho^2)}{\ln(2)}$$

- Tout d'abord, comme $\lim\limits_{\rho\to 1} 1-\rho^2 = 0$, alors : $\lim\limits_{\rho\to 1} \ln(1-\rho^2) \;=\; -\infty$.

Comme de plus $-\dfrac{1}{2\ln(2)} < 0$, on obtient : $\lim\limits_{\rho\to 1} I(X,Y) = +\infty$.

□

Deuxième partie : Généralités sur l'entropie des variables discrètes

Soit A un ensemble fini non vide. On dit que X est une variable aléatoire dont la loi est à support A, si X est à valeurs dans A et si pour tout $x \in A$: $\mathbb{P}([X = x]) > 0$.

Commentaire

- Profitons-en de cette définition pour faire un point sur la notation $X(\Omega)$.
  Rappelons qu'une v.a.r. X est une application $X : \Omega \to \mathbb{R}$.
  Comme la notation le suggère, $X(\Omega)$ est l'image de Ω par l'application X.
  Ainsi, $X(\Omega)$ n'est rien d'autre que l'ensemble des valeurs prises par la v.a.r. X :

$$\begin{aligned} X(\Omega) &= \{X(\omega) \mid \omega \in \Omega\} \\ &= \{x \in \mathbb{R} \mid \exists \omega \in \Omega,\ X(\omega) = x\} \end{aligned}$$

  Il faut bien noter que dans cette définition aucune application probabilité $\mathbb{P}$ n'apparaît.
- Il est toujours correct d'écrire : $X(\Omega) \subseteq]-\infty, +\infty[$.
  En effet, cette propriété signifie que toute v.a.r. X est à valeurs dans $\mathbb{R}$, ce qui est toujours le cas par définition de la notion de variable aléatoire.
- Dans le cas des v.a.r. **discrètes**, il est d'usage relativement courant de confondre :
  - × l'ensemble de valeurs possibles de la v.a.r. X (*i.e.* l'ensemble $X(\Omega)$),
  - × l'ensemble $\{x \in \mathbb{R} \mid \mathbb{P}\big([X = x]\big) \neq 0\}$, ensemble des valeurs que X prend avec probabilité non nulle. Dans le cas qui nous intéresse ici, à savoir X est une v.a.r. discrète, cet ensemble est appelé support de X et est noté $\mathrm{Supp}(X)$.
- L'énoncé introduit ici la notion de support d'une v.a.r. **discrète** et le note A de manière assez malhabile. Le support d'une v.a.r. X dépend évidemment de la v.a.r. X considéré et il est préférable de faire apparaître la dépendance dans la notation choisie.
- Si X est une v.a.r. **discrète**, il est à noter que toute valeur prise par X avec probabilité non nulle est une valeur prise par X. Autrement dit, on a toujours :

$$\boxed{\mathrm{Supp}(X) \subseteq X(\Omega)}$$

  En effet, si $x \in \mathrm{Supp}(X)$ alors $\mathbb{P}\big([X = x]\big) \neq 0$. On en déduit : $[X = x] \neq \varnothing$. Il existe donc (au moins) un élément $\omega \in \Omega$ tel que $X(\omega) = x$. La v.a.r. X prend donc la valeur x.
- La réciproque n'est pas forcément vérifiée : $X(\Omega) \not\subseteq \mathrm{Supp}(X)$.
  Autrement dit, une v.a.r. X peut prendre une valeur avec probabilité nulle. On peut par exemple penser à l'expérience consistant au lancer d'un dé à 6 faces. La v.a.r. X qui donne le résultat du dé a pour ensemble image $X(\Omega) = [\![1, 6]\!]$. Si on considère que le dé est truqué et ne renvoie que 6, alors le support de X est $\mathrm{Supp}(X) = \{6\}$.
- Dans l'énoncé, il est précisé qu'on considère que X est un v.a.r. dont la loi est à support A si X est à valeurs dans A. On fait donc l'hypothèse : $X(\Omega) \subseteq \mathrm{Supp}(X)$. Finalement, ce préambule ne sert qu'à affirmer qu'on fera dans la suite la confusion entre $X(\Omega)$ (l'ensemble des valeurs prises par la v.a.r. discrète X) et $\mathrm{Supp}(X)$ (l'ensemble des valeurs prises par X avec probabilité non nulle).

6. Soit X une variable aléatoire de loi à support $\{0,1,2,\dots,n\}$ où n est un entier naturel. On appelle **entropie** de X le réel :

$$H(X) \;=\; -\sum_{k=0}^{n} \mathbb{P}([X=k])\,\log_2\big(\mathbb{P}([X=k])\big)$$

a) On définit la fonction $g : \{0,\dots,n\} \to \mathbb{R}$ en posant $g(k) = \log_2\big(\mathbb{P}([X=k])\big)$ pour k élément de $\{0,1,\dots,n\}$. Montrer : $H(X) = -\mathbb{E}\big(g(X)\big)$.

Commentaire

- La deuxième partie de l'énoncé se concentre sur l'entropie des v.a.r. discrètes. On introduit alors la v.a.r. $g(X)$, transformée de la v.a.r. X. Comme dans le cas des v.a.r. à densité, il faut revenir à la définition pour bien comprendre la notation $g(X)$. Rappelons que si $g : X(\Omega) \to \mathbb{R}$ alors on utilise la notation $g(X)$ pour désigner $g \circ X$.
  Autrement dit, $g(X)$ est la v.a.r. définie par :

$$\begin{array}{rccc} g(X) & : & \Omega & \to & \mathbb{R} \\ & & \omega & \mapsto & g\big(X(\omega)\big) \end{array}$$

- D'après la définition de g, la v.a.r. $g(X)$ n'est autre que l'application :

$$\omega \mapsto g\big(X(\omega)\big) = \log_2\Big(\mathbb{P}\big(\,[X=X(\omega)]\,\big)\Big)$$

  Et comme $X(\Omega) = [\![0,n]\!]$ alors, lorsque ω décrit Ω, $X(\omega)$ décrit $[\![0,n]\!]$ (*i.e.* prend toutes les valeurs k de $[\![0,n]\!]$). Ceci est en accord avec la définition de $g(X)$: l'application g est définie sur $X(\Omega)$.

- Ceci étant établi, on comprend pourquoi on ne peut écrire :

$$g(X) \;\not=\; \log_2\big(\mathbb{P}([X=X])\big) = \log_2(1) = 0$$

Démonstration.

- La v.a.r. $g(X)$ admet une espérance car c'est une v.a.r. finie (car X est une v.a.r. finie).

  Commentaire

  Rappelons que par définition de la v.a.r. $g(X)$, on a :

$$\begin{aligned} \big(g(X)\big)(\Omega) &= g\big(X(\Omega)\big) \\ &= \{g(x) \mid x \in X(\omega)\} \;=\; \{g(k) \mid k \in [\![0,n]\!]\} \end{aligned}$$

  Ainsi, si $k \in [\![0,n]\!]$, $g(k)$ peut prendre **au plus** $n+1$ valeurs distinctes. L'ensmble $\big(g(X)\big)(\Omega)$ est donc bien un ensemble fini.

- Par théorème de transfert :

$$\begin{aligned} \mathbb{E}\big(g(X)\big) &= \sum_{x\in X(\Omega)} g(x)\,\mathbb{P}([X=x]) \\ &= \sum_{k=0}^{n} g(k)\,\mathbb{P}([X=k]) \\ &= \sum_{k=0}^{n} \log_2\big(\mathbb{P}([X=k])\big)\,\mathbb{P}([X=k]) \qquad \textit{(par définition de } g\textit{)} \end{aligned}$$

  Par définition de H : $H(X) = -\mathbb{E}\big(g(X)\big)$.

□

b) Montrer : $H(X) \geqslant 0$.

Démonstration.

- Soit $k \in [\![0, n]\!]$.
  Tout d'abord, comme le support de X est $[\![0, n]\!]$:

$$\begin{array}{lll} & 0 < \mathbb{P}([X=k]) \leqslant 1 & \\ \text{donc} & \ln\big(\mathbb{P}([X=k])\big) \leqslant 0 & \textit{(par croissance de ln sur }]0,+\infty[\textit{)} \\ \text{d'où} & \dfrac{\ln\big(\mathbb{P}([X=k])\big)}{\ln(2)} \leqslant 0 & \textit{(si } \ln(2) > 0\textit{)} \\ \text{ainsi} & \log_2\big(\mathbb{P}([X=k])\big) \leqslant 0 & \\ \text{puis} & \mathbb{P}([X=k])\,\log_2\big(\mathbb{P}([X=k])\big) \leqslant 0 & \textit{(car } \mathbb{P}([X=k]) \geqslant 0\textit{)} \end{array}$$

- On en déduit :

$$\sum_{k=0}^{n} \mathbb{P}([X=k])\,\log_2\big(\mathbb{P}([X=k])\big) \;\leqslant\; 0$$

$$\boxed{\text{Et ainsi : } H(X) = -\sum_{k=0}^{n} \mathbb{P}([X=k])\,\log_2\big(\mathbb{P}([X=k])\big) \geqslant 0.}$$ □

c) Soit p un réel tel que $0 < p < 1$.
On suppose dans cette question que X suit la loi de Bernoulli $\mathcal{B}(p)$.

(i) Calculer $H(X)$ en fonction de p. On note ψ la fonction qui, à p, associe $H(X)$.

Démonstration.

- La v.a.r. X est une v.a.r. finie. L'entropie $H(X)$ est donc bien définie.
- On obtient :

$$\begin{array}{rcll} H(X) & = & -\Big(\mathbb{P}([X=0])\,\log_2\big(\mathbb{P}([X=0])\big) + \mathbb{P}([X=1])\,\log_2\big(\mathbb{P}([X=1])\big)\Big) & \\ & = & -(1-p)\,\log_2(1-p) - p\,\log_2(p) & \textit{(car } X \hookrightarrow \mathcal{B}(p)\textit{)} \end{array}$$

$$\boxed{\text{Finalement : } \psi : p \mapsto -(1-p)\,\log_2(1-p) - p\,\log_2(p).}$$ □

(ii) Montrer que ψ est concave sur $]0, 1[$.

Démonstration.

- La fonction ψ est de classe $\mathcal{C}^2$ sur $]0, 1[$ en tant que somme et produit de fonctions de classe $\mathcal{C}^2$ sur $]0, 1[$.

Commentaire

Détaillons l'obtention de la régularité de la fonction $g : p \mapsto \log_2(1-p)$.
La fonction g est de classe $\mathcal{C}^2$ sur $]0, 1[$ car elle est la composée $g = h_2 \circ h_1$ où :

× $h_1 : p \mapsto 1 - p$ est :
  - de classe $\mathcal{C}^2$ sur $]0, 1[$ en tant que fonction affine,
  - telle que : $h_1(]0, 1[) \subset]0, +\infty[$.

× $h_2 : x \mapsto \log_2(x) = \dfrac{\ln(x)}{\ln(2)}$ est de classe $\mathcal{C}^2$ sur $]0, +\infty[$.

- Soit $p \in]0,1[$.

$$\begin{aligned}\psi'(p) &= -\left((-1)\times\log_2(1-p) + \cancel{(1-p)}\times\left(-\frac{1}{\cancel{(1-p)}\ln(2)}\right) + \log_2(p) + \cancel{p}\times\frac{1}{\cancel{p}\ln(2)}\right)\\ &= \log_2(1-p) - \log_2(p) + \cancel{\frac{1}{\ln(2)}} - \cancel{\frac{1}{\ln(2)}}\\ &= \log_2(1-p) - \log_2(p)\end{aligned}$$

On en déduit :

$$\psi''(p) = -\frac{1}{(1-p)\ln(2)} - \frac{1}{p\ln(2)} = -\frac{1}{\ln(2)}\left(\frac{1}{1-p}+\frac{1}{p}\right)$$

Or, comme $0<p<1$: $\dfrac{1}{1-p} > 0$ et $\dfrac{1}{p} > 0$. D'où : $\psi''(p) \leqslant 0$.

On en déduit que la fonction ψ est concave sur $]0,1[$.

□

(iii) Déterminer la valeur p_0 où ψ est maximale.

Démonstration.

- Soit $p \in]0,1[$. D'après la question précédente :

$$\begin{aligned}\psi'(p) \geqslant 0 &\Leftrightarrow \log_2(1-p) - \log_2(p) \geqslant 0\\ &\Leftrightarrow \frac{\ln(1-p)}{\ln(2)} \geqslant \frac{\ln(p)}{\ln(2)}\\ &\Leftrightarrow \ln(1-p) \geqslant \ln(p) && \textit{(car } \ln(2)>0\textit{)}\\ &\Leftrightarrow 1-p \geqslant p && \textit{(par stricte croissance de } \exp \textit{ sur } \mathbb{R}\textit{)}\\ &\Leftrightarrow 1 \geqslant 2p\\ &\Leftrightarrow \frac{1}{2} \geqslant p\end{aligned}$$

- On note $p_0 = \dfrac{1}{2}$. On obtient le tableau de variations suivant :

p	0		p_0		1
Signe de $\psi'(p)$		$+$	0	$-$	
Variations de ψ	0	↗	1	↘	0

- Détaillons les éléments de ce tableau.
  - × Tout d'abord :

$$\psi\left(\frac{1}{2}\right) \;=\; -\frac{1}{2}\,\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\,\log_2\left(\frac{1}{2}\right) \;=\; -\log_2\left(\frac{1}{2}\right) \;=\; \log_2(2) \;=\; 1 \qquad \textit{(d'après \textbf{1.b)})}$$

  - × De plus, pour tout $p \in]0,1[$:

$$\psi(p) \;=\; -\frac{1}{\ln(2)}\big((1-p)\,\ln(1-p) + p\,\ln(p)\big)$$

  Par croissances comparées : $\lim\limits_{p\to 0} p\,\ln(p) = 0$. Ainsi : $\lim\limits_{p\to 0} \psi(p) = 0$.

  - × Enfin, avec le changement de variable $u = 1-p$:

$$\lim_{p\to 1} (1-p)\,\ln(1-p) \;=\; \lim_{u\to 0} u\,\ln(u) \;=\; 0 \qquad \textit{(par croissances comparées)}$$

  On en déduit : $\lim\limits_{p\to 1} \psi(p) = 0$.

Finalement, sur $]0,1[$, la fonction ψ admet un maximum en $p_0 = \dfrac{1}{2}$.

□

d) On suppose dans cette question que la loi de X est à support $\{0,1,2,3\}$ avec les probabilités :

$$\mathbb{P}([X=0]) = \frac{1}{2} \quad ; \quad \mathbb{P}([X=1]) = \frac{1}{4} \quad ; \quad \mathbb{P}([X=2]) = \mathbb{P}([X=3]) = \frac{1}{8}$$

Calculer $H(X)$.

Démonstration.

- La v.a.r. X est une v.a.r. finie. L'entropie $H(X)$ est donc bien définie.
- On obtient :

$$\begin{aligned}
H(X) &= -\sum_{k=0}^{3} \mathbb{P}([X=k])\,\log_2\big(\mathbb{P}([X=k])\big) \\
&= -\left(\frac{1}{2}\,\log_2\left(\frac{1}{2}\right) + \frac{1}{4}\,\log_2\left(\frac{1}{4}\right) + \frac{1}{8}\,\log_2\left(\frac{1}{8}\right) + \frac{1}{8}\,\log_2\left(\frac{1}{8}\right)\right) \\
&= \frac{1}{2}\,\log_2(2) + \frac{1}{4}\,\log_2(4) + \frac{1}{4}\,\log_2(8) \\
&= \frac{1}{2}\,\log_2(2^1) + \frac{1}{4}\,\log_2(2^2) + \frac{1}{4}\,\log_2(2^3) \\
&= \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{4}\times 3 \qquad \textit{(d'après \textbf{1.b)})} \\
&= \frac{7}{4}
\end{aligned}$$

$H(X) = \dfrac{7}{4}$

□

7. On souhaite écrire une fonction en **Python** pour calculer l'entropie d'une variable aléatoire X dont le support de la loi est de la forme $A = \{0,1,\ldots,n\}$ où n est un entier naturel. On suppose que le vecteur `P` de **Python** est tel que pour tout k de A, `P[k]` $= \mathbb{P}([X=k])$. Compléter la fonction ci-dessous d'argument `P` qui renvoie l'entropie de X, c'est-à-dire $-\sum\limits_{k=0}^{n} \mathbb{P}([X=k])\,\log_2\big(\mathbb{P}([X=k])\big)$.

```
1 import numpy as np
2 def entropie(P):
3 ...
```

Si nécessaire, on pourra utiliser l'instruction `len(P)` qui donne le nombre d'éléments de `P`.

Commentaire

- Rappelons qu'en **Python** les vecteurs sont indexés par des entiers naturels. L'indice du premier élément d'un vecteur est donc 0.
- Ici, le vecteur `P` doit contenir les valeurs :

$$\Big[\ \mathbb{P}([X=0]), \mathbb{P}([X=1]), \dots, \mathbb{P}([X=n])\ \Big]$$

Ainsi :

× `P[0]` (premier élément du vecteur `P`) contient $\mathbb{P}([X=0])$.

× `P[1]` (deuxième élément du vecteur `P`) contient $\mathbb{P}([X=1])$.

× ...

× `P[n]` ($(n+1)^{\text{ème}}$ élément du vecteur `P`) contient $\mathbb{P}([X=n])$.

Autrement dit, pour tout `k` de $[\![0,n]\!]$, `P[k]` contient $\mathbb{P}([X=k])$.

Démonstration.
On propose la fonction suivante :

```
1 import numpy as np
2 def entropie(P):
3     n = len(P)-1
4     S = 0
5     for k in range(n+1):
6         S = S + P[k] * (np.log(P[k])/np.log(2))
7     return -S
```

Détaillons les éléments de ce script.

- **Début de la fonction**
  L'énoncé commence par préciser la structure de la fonction :
  × cette fonction se nomme `entropie`,
  × elle prend en paramètre d'entrée le vecteur `P`,
  × elle doit renvoyer l'entropie de la v.a.r. X dont la loi est encodée dans le vecteur `P`.

```
2 def entropie(P):
```

- **Initialisation**
  La variable `n + 1` contient le nombre d'éléments du vecteur `P`.
  Autrement dit, la variable `n + 1` contient la taille du vecteur `P`.
  La variable `S`, qui contiendra les valeurs successives de $\sum_{i=0}^{k} \mathbb{P}([X=i]) \log_2\big(\mathbb{P}([X=i])\big)$, est initialisée à 0 (choix naturel d'initialisation lorsqu'on souhaite coder une somme puisque 0 est l'élément neutre de l'opérateur de sommation).

```
3        n = len(P)-1
4        S = 0
```

- **Structure itérative**
  Les lignes 5 à 6 consistent à calculer les valeurs successives de $\sum_{i=0}^{k} \mathbb{P}([X=i]) \log_2\big(\mathbb{P}([X=i])\big)$.
  Pour cela, on utilise une structure itérative (boucle for) :

```
5        for k in range(n+1):
6            S = S + P[k] * (np.log(P[k])/np.log(2))
```

- **Fin du programme**
  À l'issue de cette boucle, la variable `S` contient la somme $\sum_{k=0}^{n} \mathbb{P}([X=k]) \log_2\big(\mathbb{P}([X=k])\big)$. On renvoie alors $-S = -\sum_{k=0}^{n} \mathbb{P}([X=k]) \log_2\big(\mathbb{P}([X=k])\big)$.

```
7        return -S
```

Commentaire

Afin de permettre une bonne compréhension des mécanismes en jeu, on a détaillé la réponse à cette question. Cependant, proposer un programme **Python** correct démontre la bonne compréhension de ces mécanismes et permet certainement d'obtenir la majorité des points alloués à cette question.

□

On souhaite maintenant démontrer quelques inégalités concernant l'entropie.

8. On commence par une inégalité générale, appelée **Inégalité de Jensen**.

a) Soit $N \geqslant 2$. Soit X une variable aléatoire de loi à support $\{x_1, x_2, \dots, x_N\}$ où les x_i sont des éléments distincts de $\mathbb{R}_+$. On pose $\mathbb{P}([X=x_i]) = p_i$.
Montrer que, pour tout $1 \leqslant i \leqslant N$, on a : $p_i < 1$.

Démonstration.
Soit $i \in [\![1, N]\!]$.

- Comme le support de X est $\{x_1, \dots, x_N\}$, alors $(\ [X=x_k]\ )_{k\in[\![1,N]\!]}$ est un système complet d'événement. On obtient alors :

$$\sum_{k=1}^{N} \mathbb{P}([X=x_k]) \;=\; 1$$

On en déduit : $p_i \;=\; 1 - \sum\limits_{\substack{k=1 \\ k\neq i}}^{N} p_k$.

- Or, toujours comme le support de X est $\{x_1, \dots, x_N\}$, on en déduit : $\forall k \in [\![1, N]\!]$, $p_k > 0$.
  D'où : $\sum_{\substack{k=1 \\ k \neq i}}^{N} p_k > 0$.

Et ainsi : $p_i = 1 - \sum_{\substack{k=1 \\ k \neq i}}^{N} p_k < 1$.

$$\boxed{\forall i \in [\![1, N]\!],\ p_i < 1}$$

□

On désire démontrer par récurrence la propriété suivante :

$\mathcal{P}(N)$: **Pour toute fonction φ convexe sur $\mathbb{R}_+$, si X est une variable aléatoire de loi à support $A \subset \mathbb{R}_+$ avec $\mathrm{Card}(A) = N$, on a :** $\mathbb{E}\big(\varphi(X)\big) \geqslant \varphi\big(\mathbb{E}(X)\big)$

b) Montrer que $\mathcal{P}(2)$ est vraie.

Démonstration.
Soient φ une fonction convexe sur $\mathbb{R}_+$ et X une v.a.r. de loi à support $A \subset \mathbb{R}_+$ avec $\mathrm{Card}(A) = 2$.
On note alors : $A = \{x_1, x_2\}$.

- Les v.a.r. X et $\varphi(X)$ admettent des espérances car ce sont des v.a.r. finies.
- D'une part, par théorème de transfert :

$$\mathbb{E}\big(\varphi(X)\big) = \varphi(x_1)\,\mathbb{P}([X = x_1]) + \varphi(x_2)\,\mathbb{P}([X = x_2])$$

Or, comme le support de X est $\{x_1, x_2\}$: $\mathbb{P}([X = x_1]) + \mathbb{P}([X = x_2]) = 1$. Ainsi :

$$\mathbb{E}\big(\varphi(X)\big) = \varphi(x_1)\,\mathbb{P}([X = x_1]) + \varphi(x_2)\,\big(1 - \mathbb{P}([X = x_1])\big)$$

ou encore, en notant $\lambda = \mathbb{P}([X = x_1])$:

$$\mathbb{E}\big(\varphi(X)\big) = \lambda\,\varphi(x_1) + (1 - \lambda)\,\varphi(x_2)$$

- D'autre part :

$$\begin{aligned} \varphi\big(\mathbb{E}(X)\big) &= \varphi\big(x_1\,\mathbb{P}([X = x_1]) + x_2\,\mathbb{P}([X = x_2])\big) \\ &= \varphi\big(\lambda\,x_1 + (1 - \lambda)\,x_2\big) \end{aligned}$$

- Par définition de λ : $0 \leqslant \lambda \leqslant 1$.
  Enfin, par définition de la convexité de φ :

$$\varphi\big(\lambda\,x_1 + (1 - \lambda)\,x_2\big) \leqslant \lambda\,\varphi(x_1) + (1 - \lambda)\,\varphi(x_2)$$

On en déduit : $\varphi\big(\mathbb{E}(X)\big) \leqslant \mathbb{E}\big(\varphi(X)\big)$.

$$\boxed{\text{D'où } \mathcal{P}(2).}$$

□

c) Soit $N \geqslant 3$. On suppose que $\mathcal{P}(N-1)$ est vérifiée. Soit X une variable aléatoire de loi à support $A = \{x_1, x_2, \ldots, x_N\}$ où les x_i sont des éléments distincts de $\mathbb{R}_+$. On pose : $\mathbb{P}([X = x_i]) = p_i$. Pour i tel que $1 \leqslant i \leqslant N-1$, on pose : $p'_i = \dfrac{p_i}{1-p_N}$.

Commentaire

- Comme annoncé en début de question ***8.***, on souhaite démontrer un propriété par récurrence. Plus précisément, il s'agit de démontrer : $\forall N \geqslant 2$, $\mathcal{P}(N)$. En question ***8.b)***, on a procédé à l'étape d'initialisation en démontrant $\mathcal{P}(2)$. Il s'agit maintenant de procéder à l'étape d'hérédité. On considère donc $N \geqslant 3$, on suppose $\mathcal{P}(N-1)$ et on souhaite démontrer $\mathcal{P}(N)$ (de manière équivalent, on aurait pu considérer $N \geqslant 2$, supposer $\mathcal{P}(N)$ et démontrer $\mathcal{P}(N+1)$).
- La propriété $\mathcal{P}(N)$ est doublement quantifiée universellement. Pour la démontrer il s'agit tout d'abord d'introduire une fonction φ convexe sur $\mathbb{R}_+$ (ce que l'énoncé oublie de faire) et une v.a.r. de loi à support A où A est un sous-ensemble de $\mathbb{R}_+$ de cardinal N.
- La grande question qui se pose lors de l'étape d'hérédité est de savoir comment passer de la propriété au rang $N-1$ à celle au rang N. Ici, le support de la v.a.r. est de cardinal N. Or on ne peut utiliser la propriété $\mathcal{P}(N-1)$ que pour une v.a.r. de support de cardinal $N-1$. C'est tout le but de cette question ***8.c)*** : on introduit une v.a.r. Y, construite à partir de la v.a.r. X, dont le support est de cardinal $N-1$. La construction est habile et permet, malgré la suppression d'un élément du support, de transporter suffisamment d'informations sur X pour que l'inégalité obtenue sur la v.a.r. Y fournisse l'inégalité attendue sur la v.a.r. X.

(i) Montrer : $\displaystyle\sum_{i=1}^{N-1} p'_i = 1$ et $\forall i \in [\![1, N-1]\!]$, $0 < p'_i < 1$.

Démonstration.

- Tout d'abord :
$$\sum_{i=1}^{N-1} p'_i \;=\; \sum_{i=1}^{N-1} \frac{p_i}{1-p_N} \;=\; \frac{1}{1-p_N} \sum_{i=1}^{N-1} p_i$$
Or $(\ [X = x_i]\ )_{i \in [\![1,N]\!]}$ est un système complet d'événements. D'où :
$$\sum_{i=1}^{N} \mathbb{P}([X = x_i]) \;=\; 1$$
Ainsi : $\displaystyle\sum_{i=1}^{N-1} p_i = 1 - p_N$. On en déduit :
$$\sum_{i=1}^{N-1} p'_i \;=\; \frac{1}{\cancel{1-p_N}} \cancel{(1-p_N)} \;=\; 1$$
$$\boxed{\sum_{i=1}^{N-1} p'_i = 1}$$
- Soit $i \in [\![1, N-1]\!]$.
  - × Comme le support de X est $\{x_1, \ldots, x_N\}$, on en déduit : $\forall k \in [\![1, N]\!]$, $p_k > 0$.
    Ainsi : $\forall k \in [\![1, N]\!]$, $p'_k = \dfrac{p_k}{1-p_N} > 0$.
  - × On vient également de démontrer : $p'_i = 1 - \displaystyle\sum_{\substack{k=1 \\ k \neq i}}^{N-1} p'_k$.

× De plus : $\forall k \in [\![1, N]\!]$, $p'_k > 0$. D'où : $\displaystyle\sum_{\substack{k=1\\k\neq i}}^{N-1} p'_k > 0$.

Ainsi : $p'_i = 1 - \displaystyle\sum_{\substack{k=1\\k\neq i}}^{N-1} p'_k \ < \ 1$.

Finalement : $\forall i \in [\![1, N]\!]$, $0 < p'_i < 1$.

□

(ii) Soit Y une variable aléatoire de loi à support $\{x_1, \ldots, x_{N-1}\}$ telle que $\mathbb{P}([Y = x_i]) = p'_i$ pour $1 \leqslant i \leqslant N-1$. Montrer : $\displaystyle\sum_{i=1}^{N-1} p'_i\,\varphi(x_i) \ \geqslant \ \varphi\left(\sum_{i=1}^{N-1} p'_i\,x_i\right)$.

Démonstration.

- Tout d'abord, remarquons que, d'après la question précédente, la v.a.r. Y est bien définie.
- On obtient alors :
  × la v.a.r. Y est une v.a.r. de loi à support $A = \{x_1, \ldots, x_{N-1}\} \subset \mathbb{R}_+$ avec $\mathrm{Card}(A) = N-1$,
  × la fonction φ est convexe sur $\mathbb{R}_+$.
  D'après $\mathcal{P}(N-1)$: $\mathbb{E}\big(\varphi(Y)\big) \geqslant \varphi\big(\mathbb{E}(Y)\big)$.
- Par théorème de transfert :

$$\mathbb{E}\big(\varphi(Y)\big) \ = \ \sum_{i=1}^{N-1} \varphi(x_i)\,\mathbb{P}([Y = x_i]) \ = \ \sum_{i=1}^{N-1} \varphi(x_i)\,p'_i$$

De plus :

$$\varphi\big(\mathbb{E}(Y)\big) \ = \ \varphi\left(\sum_{i=1}^{N-1} x_i\,\mathbb{P}([Y = x_i])\right) \ = \ \varphi\left(\sum_{i=1}^{N-1} x_i\,p'_i\right)$$

Finalement : $\displaystyle\sum_{i=1}^{N-1} p'_i\,\varphi(x_i) \geqslant \varphi\left(\sum_{i=1}^{N-1} p'_i\,x_i\right)$.

□

(iii) Montrer : $\mathbb{E}\big(\varphi(X)\big) \ \geqslant \ \varphi\big(\mathbb{E}(X)\big)$.

Démonstration.

- Les v.a.r. X et $\varphi(X)$ admettent une espérance car ce sont des v.a.r. finies.
- Par théorème de transfert :

$$\mathbb{E}\big(\varphi(X)\big) \ = \ \sum_{i=1}^{N} \varphi(x_i)\,\mathbb{P}([X = x_i]) \ = \ \sum_{i=1}^{N} \varphi(x_i)\,p_i$$

De plus :

$$\varphi\big(\mathbb{E}(X)\big) \ = \ \varphi\left(\sum_{i=1}^{N} x_i\,\mathbb{P}([X = x_i])\right) \ = \ \varphi\left(\sum_{i=1}^{N} x_i\,p_i\right)$$

- Par ailleurs :

$$\begin{aligned}
\mathbb{E}\big(\varphi(X)\big) &= \sum_{i=1}^{N} \varphi(x_i)\, p_i \\
&= \sum_{i=1}^{N-1} \varphi(x_i)\, p_i + \varphi(x_N)\, p_N \\
&= (1-p_N) \sum_{i=1}^{N-1} \varphi(x_i)\, \frac{p_i}{1-p_N} + \varphi(x_N)\, p_N \\
&= (1-p_N) \sum_{i=1}^{N-1} \varphi(x_i)\, p'_i + \varphi(x_N)\, p_N \\
&\geqslant (1-p_N)\, \varphi\left(\sum_{i=1}^{N-1} x_i\, p'_i\right) + p_N\, \varphi(x_N) \qquad \textit{(d'après la question précédente, car : } 1-p_N \geqslant 0\textit{)}
\end{aligned}$$

- Comme la fonction φ est convexe sur $\mathbb{R}_+$:

$$\forall (x,y) \in (\mathbb{R}_+)^2,\ \forall \lambda \in [0,1],\quad \lambda\,\varphi(x) + (1-\lambda)\varphi(y) \geqslant \varphi\big(\lambda\, x + (1-\lambda)\, y\big)$$

On applique cette propriété à $\lambda = 1-p_N$, $x = \sum_{i=1}^{N-1} x_i\, p'_i$ et $y = x_N$, on obtient :

$$(1-p_N)\, \varphi\left(\sum_{i=1}^{N-1} x_i\, p'_i\right) + p_N\, \varphi(x_n) \ \geqslant\ \varphi\left((1-p_N) \sum_{i=1}^{N-1} x_i\, p'_i + p_N\, x_N\right)$$

Or :

$$(1-p_N) \sum_{i=1}^{N-1} x_i\, p'_i + p_N\, x_N \ =\ \cancel{(1-p_N)} \sum_{i=1}^{N-1} x_i\, \frac{p_i}{\cancel{1-p_N}} + p_N\, x_N \ =\ \sum_{i=1}^{N} x_i\, p_i$$

- On a ainsi démontré, par transitivité :

$$\begin{array}{ccc}
\mathbb{E}\big(\varphi(X)\big) & \geqslant & \varphi\left(\displaystyle\sum_{i=1}^{N} x_i\, p_i\right) \\
 & & \shortparallel \\
 & & \varphi\big(\mathbb{E}(X)\big)
\end{array}$$

$$\boxed{\mathbb{E}\big(\varphi(X)\big) \geqslant \varphi\big(\mathbb{E}(X)\big)} \qquad \square$$

d) Montrer que, si φ est *concave* sur $\mathbb{R}_+$, on a : $\mathbb{E}\big(\varphi(X)\big) \ \leqslant\ \varphi\big(\mathbb{E}(X)\big)$.

Démonstration.
Si la fonction φ est concave sur $\mathbb{R}_+$, alors la fonction $\psi = -\varphi$ est convexe sur $\mathbb{R}_+$. Ainsi, d'après la récurrence précédente :

$$\begin{array}{ccc}
\mathbb{E}\big(\psi(X)\big) & \geqslant & \psi\big(\mathbb{E}(X)\big) \\
\shortparallel & & \shortparallel \\
\mathbb{E}\big(-\varphi(X)\big) & & -\varphi\big(\mathbb{E}(X)\big)
\end{array}$$

Par linéarité de l'espérance : $-\mathbb{E}\big(\varphi(X)\big) \geqslant -\varphi\big(\mathbb{E}(X)\big)$.

$$\boxed{\text{Finalement : } \mathbb{E}\big(\varphi(X)\big) \ \leqslant\ \varphi\big(\mathbb{E}(X)\big).}$$

Commentaire

Démontrons que, si φ est concave sur $\mathbb{R}_+$, alors $\psi = -\varphi$ est convexe sur $\mathbb{R}_+$.

- Soit $(x, y) \in (\mathbb{R}_+)^2$. Soit $\lambda \in [0, 1]$.
  Comme φ est concave sur $\mathbb{R}_+$:

$$\varphi\big(\lambda\, x + (1-\lambda)\, y\big) \;\geqslant\; \lambda\, \varphi(x) + (1-\lambda)\, \varphi(y)$$

  On en déduit :

$$-\varphi\big(\lambda\, x + (1-\lambda)\, y\big) \;\geqslant\; -\big(\lambda\, \varphi(x) + (1-\lambda)\, \varphi(y)\big)$$

  Ainsi :

$$\psi\big(\lambda\, x + (1-\lambda)\, y\big) \;\leqslant\; \lambda\, \psi(x) + (1-\lambda)\, \psi(y)$$

  On en déduit que $\psi = -\varphi$ est convexe sur $\mathbb{R}_+$.
- Si on sait que la fonction φ est de classe $\mathcal{C}^1$ (resp. $\mathcal{C}^2$) sur $\mathbb{R}_+$, on aurait également pu utiliser la caractérisation de la convexité portant sur φ' (resp. φ''). □

9. Soit X une variable aléatoire de loi à support $\{0, 1, \ldots, n\}$. On pose, pour k tel que $0 \leqslant k \leqslant n$, $p_k = \mathbb{P}([X = k])$.

a) Montrer : $\displaystyle\sum_{k=0}^{n} p_k \log_2\left(\frac{1}{(n+1)\, p_k}\right) \;\leqslant\; \log_2\left(\sum_{k=0}^{n} \frac{p_k}{(n+1)\, p_k}\right) \;=\; 0.$

Démonstration.

- L'inégalité présentée dans l'énoncé compare une somme de $\log_2$ au $\log_2$ d'une somme. Il semble donc que l'on souhaite appliquer l'inégalité de Jensen avec la fonction $\varphi = \log_2$ qui est concave sur $\mathbb{R}_+^*$ (d'après ***1.b)***).
- Pour cela, on doit introduire une v.a.r. Y à support de cardinal $n+1$ (car les sommes possèdent $n+1$ éléments). Une telle v.a.r. Y possède un ensemble image qui se note, de manière générique :

$$Y(\Omega) \;=\; \{y_0, y_1, \ldots, y_n\}$$

  Pour pouvoir appliquer l'inégalité de Jensen, il faudra vérifier que tous ces éléments y_i sont des réels positifs distincts.
- De manière générale, l'inégalité de Jensen appliquée à $\log_2$ et Y permet d'obtenir :

$$\begin{array}{ccc} \mathbb{E}\big(\log_2(Y)\big) & \leqslant & \log_2\big(\mathbb{E}(Y)\big) \\ \shortparallel & & \shortparallel \\ \displaystyle\sum_{k=0}^{n} \log_2(y_k)\; \mathbb{P}\big([Y = y_k]\big) & & \log_2\Big(\displaystyle\sum_{k=0}^{n} y_k\; \mathbb{P}\big([Y = y_k]\big)\Big) \end{array}$$

  Cette inégalité doit permettre d'obtenir celle de l'énoncé à savoir :

$$\sum_{k=0}^{n} p_k \log_2\left(\frac{1}{(n+1)\, p_k}\right) \quad\leqslant\quad \log_2\left(\sum_{k=0}^{n} \frac{p_k}{(n+1)\, p_k}\right)$$

  Plus précisément, on veut exprimer :

  × la quantité $\displaystyle\sum_{k=0}^{n} p_k \log_2\left(\frac{1}{(n+1)\, p_k}\right)$ sous la forme $\displaystyle\sum_{k=0}^{n} \log_2(y_k)\; \mathbb{P}\big([Y = y_k]\big)$. Pour que ces deux quantités coïncident, on n'a guère le choix que de trouver une v.a.r. Y telle que :

$$\forall k \in [\![0, n]\!],\; y_k = \frac{1}{(n+1)\, p_k} \quad \text{et} \quad p_k = \mathbb{P}\big([Y = y_k]\big)$$

× la quantité $\log_2\left(\sum_{k=0}^{n} \frac{p_k}{(n+1)\,p_k}\right)$ sous la forme $\log_2\left(\sum_{k=0}^{n} y_k\ \mathbb{P}\big([Y=y_k]\big)\right)$.

Ces deux quantités coïncident bien si on trouve une v.a.r. Y vérifiant les contraintes évoquées dans le point précédent.

Commentaire

On détaille ici longuement la manière de trouver la v.a.r. Y qui va permettre de résoudre la question. Ce travail doit plutôt se réaliser au brouillon. On l'a présenté ici afin de permettre une bonne compréhension de cette résolution et que l'introduction qui va suivre de la v.a.r. Y ne paraisse pas sortie du chapeau.

- Pour cela, on introduit la v.a.r. Y définie par :

$$Y : \omega \mapsto \begin{cases} \dfrac{1}{(n+1)\,p_0} & \text{si } X(\omega)=0 \\ \vdots & \\ \dfrac{1}{(n+1)\,p_n} & \text{si } X(\omega)=n \end{cases}$$

On a en particulier, pour tout $k \in [\![0,n]\!]$:

$$\left[Y = \frac{1}{(n+1)\,p_k}\right] \ = \ [X=k]$$

- Ainsi :
  × la fonction $\log_2$ est concave sur $\mathbb{R}_+^*$ (d'après ***1.c)***),
  × la loi de Y est à support $A = \{\frac{1}{(n+1)\,p_0}, \dots, \frac{1}{(n+1)\,p_n}\} \subset \mathbb{R}_+$ avec $\mathrm{Card}(A) = n+1$.

  D'après l'inégalité de Jensen (question ***8.d)***) :

$$\mathbb{E}\big(\log_2(Y)\big) \ \leqslant \ \log_2\big(\mathbb{E}(Y)\big)$$

- De plus :
  × d'une part :

$$\mathbb{E}(Y) \ = \ \sum_{k=0}^{n} \frac{1}{(n+1)\,p_k}\,\mathbb{P}\left(\left[Y=\frac{1}{(n+1)\,p_k}\right]\right) \ = \ \sum_{k=0}^{n} \frac{1}{(n+1)\,p_k}\,\mathbb{P}([X=k]) \ = \ \sum_{k=0}^{n} \frac{p_k}{(n+1)\,p_k}$$

  × d'autre part, par théorème de transfert :

$$\begin{aligned} \mathbb{E}\big(\log_2(Y)\big) \ &= \ \sum_{k=0}^{n} \log_2\left(\frac{1}{(n+1)\,p_k}\right)\ \mathbb{P}\left(\left[Y=\frac{1}{(n+1)\,p_k}\right]\right) \\ &= \ \sum_{k=0}^{n} \log_2\left(\frac{1}{(n+1)\,p_k}\right)\ \mathbb{P}([X=k]) \\ &= \ \sum_{k=0}^{n} \log_2\left(\frac{1}{(n+1)\,p_k}\right)\ p_k \end{aligned}$$

Finalement : $\displaystyle\sum_{k=0}^{n} p_k \log_2\left(\frac{1}{(n+1)\,p_k}\right) \leqslant \log_2\left(\sum_{k=0}^{n} \frac{p_k}{(n+1)\,p_k}\right)$.

Commentaire

- La v.a.r. Y est définie par cas. Elle peut donc s'exprimer naturellement à l'aide de variables aléatoires indicatrices :

$$Y = \sum_{k=0}^{n} \frac{1}{(n+1)\,p_k} \cdot \mathbb{1}_{[X=k]}$$

(l'égalité précédente est une égalité entre variables aléatoires)

- Les variables aléatoires indicatrices ne font pas partie du programme d'ECG maths appliquées. Donnons néanmoins certaines de leurs propriétés.
Soit A un événement. On note $\mathbb{1}_A$ la v.a.r. telle que :

$$\mathbb{1}_A : \omega \mapsto \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

  × Loi de $\mathbb{1}_A$.
  - Par définition de $\mathbb{1}_A$, cette v.a.r. ne prend comme valeur que 0 et 1.
    Donc $\mathbb{1}_A(\Omega) = \{0, 1\}$.
  - Soit $\omega \in \Omega$.

$$\omega \in [\mathbb{1}_A = 1] \;\Leftrightarrow\; \mathbb{1}_A(\omega) = 1 \;\Leftrightarrow\; \omega \in A$$

    D'où : $[\mathbb{1}_A = 1] = A$. Ainsi : $\mathbb{P}([\mathbb{1}_A = 1]) = \mathbb{P}(A)$.

  On en déduit : $\mathbb{1}_A \hookrightarrow \mathcal{B}(\mathbb{P}(A))$.

  × En particulier :

$$\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A) \qquad \text{et} \qquad \mathbb{V}(\mathbb{1}_A) = \mathbb{P}(A)\big(1 - \mathbb{P}(A)\big)$$

- Il peut aussi être utilise de savoir démontrer les propriétés suivantes.
Soient B et C deux événements.

  × $\mathbb{1}_{B\cap C} = \mathbb{1}_B \times \mathbb{1}_C$

  × $\mathbb{1}_B + \mathbb{1}_{\overline{B}} = 1$

  Pour la démonstration de ces deux propriétés, on pourra, par exemple, se référer au sujet `ESSEC-II 2018`.

- Enfin :

$$\log_2\left(\sum_{k=0}^{n} \frac{\cancel{p_k}}{(n+1)\,\cancel{p_k}}\right) = \log_2\left(\frac{1}{n+1}\sum_{k=0}^{n} 1\right) = \log_2\left(\frac{1}{\cancel{n+1}} \times \cancel{(n+1)}\right) = \frac{\ln(1)}{\ln(2)} = 0$$

$$\boxed{\log_2\left(\sum_{k=0}^{n} \frac{p_k}{(n+1)\,p_k}\right) = 0}$$

□

b) Montrer : $\sum_{k=0}^{n} p_k \log_2\big((n+1)\,p_k\big) \;=\; \log_2(n+1) - H(X)$.

Démonstration.
On remarque :

$$\begin{aligned}
&\sum_{k=0}^{n} p_k \log_2\big((n+1)\,p_k\big) \\
=\;& \sum_{k=0}^{n} p_k \left(\log_2(n+1) + \log_2(p_k)\right) && \textit{(d'après \textbf{1.a})} \\
=\;& \log_2(n+1) \sum_{k=0}^{n} p_k + \sum_{k=0}^{n} p_k \log_2(p_k) \\
=\;& \log_2(n+1) \times 1 - \left(-\sum_{k=0}^{n} p_k \log_2(p_k)\right) && \textit{(car } ([X=k])_{k\in[\![0,n]\!]} \textit{ forme un système complet d'événements)} \\
=\;& \log_2(n+1) - H(X) && \textit{(par définition de } H\textit{)}
\end{aligned}$$

$$\boxed{\sum_{k=0}^{n} p_k \log_2\big((n+1)\,p_k\big) \;=\; \log_2(n+1) - H(X)}$$ □

c) Montrer : $H(X) \leqslant \log_2(n+1)$.

Démonstration.

- D'après la question précédente :

$$\begin{aligned}
H(X) &= \log_2(n+1) - \sum_{k=0}^{n} p_k \log_2\big((n+1)\,p_k\big) \\
&= \log_2(n+1) + \sum_{k=0}^{n} p_k \log_2\left(\frac{1}{(n+1)\,p_k}\right)
\end{aligned}$$

- Or, d'après ***9.a)*** :

$$\sum_{k=0}^{n} p_k \log_2\left(\frac{1}{(n+1)\,p_k}\right) \;\leqslant\; 0$$

$$\boxed{\text{D'où : } H(X) \leqslant \log_2(n+1).}$$ □

d) On suppose que X suit la loi uniforme sur $\{0, 1, \ldots, N\}$. Calculer $H(X)$.

Démonstration.

- La v.a.r. X est une v.a.r. finie. L'entropie $H(X)$ est donc bien définie.
- De plus :

$$\begin{aligned}
H(X) &= -\sum_{k=0}^{N} \mathbb{P}([X=k]) \log_2\big(\mathbb{P}([X=k])\big) \\
&= -\sum_{k=0}^{N} \frac{1}{N+1} \log_2\left(\frac{1}{N+1}\right) \\
&= \frac{1}{N+1} \log_2(N+1) \sum_{k=0}^{N} 1 \\
&= \frac{1}{\cancel{N+1}} \log_2(N+1) \cancel{(N+1)}
\end{aligned}$$

$$\boxed{H(X) = \log_2(N+1)}$$

Commentaire

- Remarquons qu'on a démontré plus tôt (en question ***9.c)***) :

$$H(X) \leqslant \log_2(n+1)$$

L'entropie $H(X)$ est donc majorée par la quantité $\log_2(n+1)$.

- Dans cette question ***9.d)***, on démontre que ce majorant est atteint lorsque la v.a.r. X suit la loi uniforme $\mathcal{U}(\llbracket 0, n \rrbracket)$. Notons que cela signifie que :
  - × la majoration obtenue en question précédente est *optimale* : il est impossible de trouver un meilleur majorant de l'entropie que $\log_2(n+1)$.
  - × la loi uniforme est une loi d'entropie maximale.
- Si l'on revient au contexte fourni en début d'énoncé, on vient de démontrer que l'intensité du désordre est maximale pour la loi uniforme. On peut comprendre ce résultat en se rappelant que les valeurs prises par la v.a.r. X de loi $\mathcal{U}(\llbracket 0, n \rrbracket)$ le sont avec même probabilité :

$$\forall k \in \llbracket 0, n \rrbracket,\ \mathbb{P}([X = k]) = \frac{1}{n+1}$$

Il n'y a donc aucune valeur parmi les entiers de 0 à n qui est « privilégiée » (au sens où aucune n'est prise avec une probabilité plus grande que les autres). Le désordre intervenant dans une expérience modélisée par une v.a.r. de loi uniforme est donc plus important que pour toute autre loi. □

10. Soient X et Y deux variables aléatoires *de même loi* à support $\{0, 1, \ldots, n\}$. On suppose en outre X et Y indépendantes.

a) Montrer : $\mathbb{P}([X = Y]) = \sum_{k=0}^{n} \big(\mathbb{P}([X = k])\big)^2$.

Démonstration.

- Tout d'abord : $\mathbb{P}([X = Y]) = \mathbb{P}([X - Y = 0])$

Commentaire

Remarquons qu'on se ramène ici à un cas particulier de loi d'une somme $X - Y$. Il faut donc se préparer à utiliser les méthodes usuelles pour la détermination de ce type de loi : la formule des probabilités totales.

- La famille $([X = k])_{k \in \llbracket 0, n \rrbracket}$ forme un système complet d'événements.
  Ainsi, par formule des probabilités totales :

$$\begin{aligned}
\mathbb{P}([X - Y = 0]) &= \sum_{k=0}^{n} \mathbb{P}([X = k] \cap [X - Y = 0]) \\
&= \sum_{k=0}^{n} \mathbb{P}([X = k] \cap [Y = k]) \\
&= \sum_{k=0}^{n} \mathbb{P}([X = k]) \times \mathbb{P}([Y = k]) && \textit{(car } X \textit{ et } Y \textit{ sont indépendantes)} \\
&= \sum_{k=0}^{n} \big(\mathbb{P}([X = k])\big)^2 && \textit{(car } X \textit{ et } Y \textit{ ont même loi)}
\end{aligned}$$

Finalement : $\mathbb{P}([X = Y]) = \sum_{k=0}^{n} \big(\mathbb{P}([X = k])\big)^2$.

□

b) On pose $v(k) = \mathbb{P}([X = k])$ pour tout k élément de $\{0, 1, \ldots, n\}$. Montrer :

$$2^{\mathbb{E}\left(\log_2\left(v(X)\right)\right)} \leqslant \mathbb{E}\left(2^{\log_2\left(v(X)\right)}\right) = \mathbb{E}\big(v(X)\big)$$

Démonstration.

- Démontrons que la fonction $\varphi : x \mapsto 2^{-x}$ est convexe sur $\mathbb{R}_+$.
  - × La fonction $\varphi : x \mapsto 2^{-x} = \mathrm{e}^{-x\,\ln(2)}$ est de classe $\mathcal{C}^2$ sur $\mathbb{R}_+$ en tant que composée de fonctions de classe $\mathcal{C}^2$ sur les intervalles adéquats.

Commentaire

- Détaillons la démonstration de la régularité de φ.
  La fonction φ est de classe $\mathcal{C}^2$ sur $\mathbb{R}_+$ car elle est la composée $\varphi = h_2 \circ h_1$ où :
  - × $h_1 : x \mapsto -x\,\ln(2)$ est :
    - de classe $\mathcal{C}^2$ sur $\mathbb{R}_+$ en tant que fonction polynomiale,
    - telle que : $h_1([0, +\infty[) \subset \mathbb{R}$.
  - × $h_2 : x \mapsto \mathrm{e}^x$ est de classe $\mathcal{C}^2$ sur $\mathbb{R}$.
- Il serait plus naturel d'introduire la fonction $\psi : x \mapsto 2^x$.
  En considérant la v.a.r. $Z = \log_2\left(v(X)\right)$ on aurait alors, pour peu que toutes les contraintes de l'inégalité de Jensen soient vérifiées :

  $$2^{\mathbb{E}(Z)} \leqslant \mathbb{E}\left(2^Z\right)$$

  Cette fonction ψ est bien convexe sur $\mathbb{R}_+$. Par contre, la v.a.r. Z ainsi définie est à valeurs négatives. Pour se conformer aux contraintes de l'énoncé, on va donc introduire $Z = -\log_2\left(v(V)\right)$ et considérer la fonction φ plutôt que ψ.

Soit $x \in \mathbb{R}_+$.

× Tout d'abord : $\varphi'(x) = -\ln(2)\,\mathrm{e}^{-x\,\ln(2)}$.

× Et ensuite : $\varphi''(x) = \big(\ln(2)\big)^2\,\mathrm{e}^{-x\,\ln(2)} \geqslant 0$.

La fonction φ est donc convexe sur $\mathbb{R}_+$.

- On note de plus : $Z = -\log_2\left(v(X)\right)$.
  Démontrons que Z est une v.a.r. finie à valeurs dans $\mathbb{R}_+$.
  - × Tout d'abord, Z est une v.a.r. finie car X l'est. On a même :

    $$Z(\Omega) = \{-\log_2\left(v(x)\right) \mid x \in X(\omega)\} = \{-\log_2\left(v(k)\right) \mid k \in [\![0, n]\!]\}$$

  - × Soit $k \in [\![0, n]\!]$. Comme la loi de X est à support $[\![0, n]\!]$:

    $$0 < \mathbb{P}([X = k]) \leqslant 1$$

donc	$0 < v(k) \leqslant 1$	
d'où	$\ln\left(v(k)\right) \leqslant 0$	*(par croissance de* ln *sur* $]0, +\infty[$*)*
puis	$\dfrac{\ln\left(v(k)\right)}{\ln(2)} \leqslant 0$	*(car* $\ln(2) > 0$*)*
ainsi	$-\log_2\left(v(k)\right) \geqslant 0$	

  On en déduit : $Z(\Omega) \subset \mathbb{R}_+$.

On a ainsi démontré que la v.a.r. Z est une v.a.r. finie et de loi à support $A \subset \mathbb{R}_+$.

- On obtient alors :
  - × la fonction φ est convexe sur $\mathbb{R}_+$,
  - × la v.a.r. Z est une v.a.r. finie de loi à support $A \subset \mathbb{R}_+$.

  Par inégalité de Jensen :

$$\begin{array}{ccccc} \mathbb{E}\big(\varphi(Z)\big) & & \geqslant & & \varphi\big(\mathbb{E}(Z)\big) \\ \shortparallel & & & & \shortparallel \\ \mathbb{E}\left(2^{\log_2\left(v(X)\right)}\right) & = & \mathbb{E}\left(2^{-\left(-\log_2\left(v(X)\right)\right)}\right) & & 2^{-\mathbb{E}\left(-\log_2\left(v(X)\right)\right)} \end{array}$$

- Enfin :
  - × par linéarité de l'espérance :

$$2^{-\mathbb{E}\left(-\log_2\left(v(X)\right)\right)} = 2^{-\left(-\mathbb{E}\left(\log_2\left(v(X)\right)\right)\right)} = 2^{\mathbb{E}\left(\log_2\left(v(X)\right)\right)}$$

$$\boxed{\text{Alors : } 2^{\mathbb{E}\left(\log_2\left(v(X)\right)\right)} \leqslant \mathbb{E}\left(2^{\log_2\left(v(X)\right)}\right).}$$

  - × par ailleurs :

$$2^{\log_2\left(v(X)\right)} = \exp\Big(\log_2\left(v(X)\right)\ln(2)\Big) = \exp\left(\frac{\ln\left(v(X)\right)}{\cancel{\ln(2)}}\cancel{\ln(2)}\right) = \exp\Big(\ln\left(v(X)\right)\Big) = v(X)$$

$$\boxed{\text{On en déduit : } \mathbb{E}\left(2^{\log_2\left(v(X)\right)}\right) = \mathbb{E}\big(v(X)\big).}$$

□

c) En déduire : $2^{-H(X)} \leqslant \mathbb{P}([X=Y])$.

Démonstration.

- D'après la question ***6.a)*** :

$$H(X) = -\mathbb{E}\big(g(X)\big)$$

où $g : \begin{array}{ccc} [\![0,n]\!] & \to & \mathbb{R} \\ k & \mapsto & \log_2\big(\mathbb{P}([X=k])\big) = \log_2\left(v(k)\right) \end{array}$

D'où : $H(X) = -\mathbb{E}\Big(\log_2\left(v(X)\right)\Big)$. Ainsi :

$$2^{\mathbb{E}\left(\log_2\left(v(X)\right)\right)} = 2^{-H(X)}$$

- De plus, par théorème de transfert :

$$\begin{aligned} \mathbb{E}\big(v(X)\big) &= \sum_{k=0}^{n} v(k)\,\mathbb{P}([X=k]) \\ &= \sum_{k=0}^{n} \big(\mathbb{P}([X=k])\big)^2 && \textit{(par définition de } v\textit{)} \\ &= \mathbb{P}([X=Y]) && \textit{(d'après } \textbf{\textit{10.a)}}\textit{)} \end{aligned}$$

$$\boxed{\text{Ainsi, d'après la question précédente : } 2^{-H(X)} \leqslant \mathbb{P}([X=Y]).}$$

□

d) Donner un exemple de loi où l'inégalité précédente est une égalité.

Démonstration.
On note X et Y deux v.a.r. aléatoires de même loi $\mathcal{U}(\llbracket 0,n \rrbracket)$, indépendantes.

- Tout d'abord :

$$\begin{aligned} 2^{-H(X)} &= \exp\big(-H(X)\ln(2)\big) \\ &= \exp\big(-\log_2(n+1)\ln(2)\big) \qquad \textit{(d'après \textbf{9.d)})} \\ &= \exp\left(\frac{\ln(n+1)}{\ln(2)}\ln(2)\right) \\ &= \frac{1}{\exp\big(\ln(n+1)\big)} = \frac{1}{n+1} \end{aligned}$$

- D'autre part, d'après ***10.a)*** :

$$\mathbb{P}([X=Y]) = \sum_{k=0}^{n}\big(\mathbb{P}([X=k])\big)^2 = \sum_{k=0}^{n}\left(\frac{1}{n+1}\right)^2 = \frac{1}{(n+1)^2}\sum_{k=0}^{n}1 = \frac{1}{(n+1)^2}\times(n+1)$$

Ainsi : $2^{-H(X)} = \dfrac{1}{n+1} = \mathbb{P}([X=Y])$.

> Pour X et Y deux v.a.r. de loi $\mathcal{U}(\llbracket 0,n \rrbracket)$, indépendantes, l'inégalité de ***10.c)*** est une égalité.

Commentaire

Comme pour la question ***9.***, on démontre dans cette question ***10.*** que :
- × la majoration obtenue en ***10.c)*** est optimale : il est impossible de trouver un meilleur majorant de $2^{-H(X)}$ que $\mathbb{P}([X=Y])$,
- × la loi uniforme est une loi pour lequel ce maximum est atteint.

□

Troisième partie : Entropie jointe et information mutuelle de deux variables discrètes

Soient X et Y deux variables aléatoires de lois à support $\{0,1,\dots,n\}$. On appelle **entropie jointe** de X et Y le réel :

$$H(X,Y) = -\sum_{k=0}^{n}\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[Y=j])\log_2\big(\mathbb{P}([X=k]\cap[Y=j])\big)$$

avec la convention : $0\times\log_2(0)=0$.

11. a) On définit la fonction $g : \{0,1,\dots,n\}^2 \to \mathbb{R}\cup\{-\infty\}$ en posant pour $(k,j)\in\{0,1,\dots,n\}^2$:

$$g(k,j) = \log_2\big(\mathbb{P}([X=k]\cap[Y=j])\big)$$

Montrer : $H(X,Y) = -\mathbb{E}\big(g(X,Y)\big)$.

Démonstration.
- Les v.a.r. X et Y sont finies, donc la v.a.r. $g(X,Y)$ l'est. Elle admet donc une espérance.

- Par théorème de transfert :

$$\begin{aligned}
\mathbb{E}\big(g(X,Y)\big) &= \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} g(x,y)\,\mathbb{P}([X=x]\cap[Y=y]) \\
&= \sum_{k=0}^{n}\sum_{j=0}^{n} g(k,j)\,\mathbb{P}([X=k]\cap[Y=j]) && \textit{(car } X(\Omega)\subset[\![0,n]\!] \textit{ et } Y(\Omega)\subset[\![0,n]\!]\textit{)} \\
&= \sum_{k=0}^{n}\sum_{j=0}^{n} \log_2\big(\mathbb{P}([X=k]\cap[Y=j])\big)\,\mathbb{P}([X=k]\cap[Y=j]) \\
&= -H(X,Y) && \textit{(par définition de } H(X,Y)\textit{)}
\end{aligned}$$

$$\boxed{H(X,Y) = -\mathbb{E}\big(g(X,Y)\big)}$$

□

b) Montrer : $H(X,Y) = H(Y,X)$.

Démonstration.
On a :

$$\begin{aligned}
H(X,Y) &= -\sum_{k=0}^{n}\left(\sum_{j=0}^{n} \log_2\big(\mathbb{P}([X=k]\cap[Y=j])\big)\,\mathbb{P}([X=k]\cap[Y=j])\right) \\
&= -\sum_{0\leqslant k,j\leqslant n} \log_2\big(\mathbb{P}([X=k]\cap[Y=j])\big)\,\mathbb{P}([X=k]\cap[Y=j]) \\
&= -\sum_{j=0}^{n}\left(\sum_{k=0}^{n} \log_2\big(\mathbb{P}([X=k]\cap[Y=j])\big)\,\mathbb{P}([X=k]\cap[Y=j])\right) && \textit{(en sommant suivant les colonnes)} \\
&= -\sum_{j=0}^{n}\left(\sum_{k=0}^{n} \log_2\big(\mathbb{P}([Y=j]\cap[X=k])\big)\,\mathbb{P}([Y=j]\cap[X=k])\right) && \textit{(car l'intersection est commutative)} \\
&= H(Y,X)
\end{aligned}$$

$$\boxed{H(X,Y) = H(Y,X)}$$

□

c) Pour tout k tel que $0 \leqslant k \leqslant n$, on pose :

$$H(Y\,|\,X=k) \;=\; -\sum_{j=0}^{n} \mathbb{P}_{[X=k]}([Y=j])\,\log_2\big(\mathbb{P}_{[X=k]}([Y=j])\big)$$

On appelle **entropie conditionnelle** de Y sachant X le réel :

$$H(Y\,|\,X) \;=\; \sum_{k=0}^{n} \mathbb{P}([X=k])\,H(Y\,|\,X=k)$$

Montrer : $H(X,Y) \;=\; H(X) + H(Y\,|\,X)$.

Démonstration.

- Soit $k \in [\![0, n]\!]$. Tout d'abord :

$$\begin{aligned}
& H(Y \mid X = k) \\
= \quad & -\sum_{j=0}^{n} \mathbb{P}_{[X=k]}([Y = j]) \log_2\left(\mathbb{P}_{[X=k]}([Y = j])\right) \\
= \quad & -\sum_{j=0}^{n} \frac{\mathbb{P}([X = k] \cap [Y = j])}{\mathbb{P}([X = k])} \log_2\left(\frac{\mathbb{P}([X = k] \cap [Y = j])}{\mathbb{P}([X = k])}\right) \\
= \quad & -\sum_{j=0}^{n} \frac{\mathbb{P}([X = k] \cap [Y = j])}{\mathbb{P}([X = k])} \left(\log_2\left(\mathbb{P}([X = k] \cap [Y = j])\right) - \log_2\left(\mathbb{P}([X = k])\right)\right) \\
= \quad & -\frac{1}{\mathbb{P}([X = k])} \left(\sum_{j=0}^{n} \mathbb{P}([X = k] \cap [Y = j]) \log_2\left(\mathbb{P}([X = k] \cap [Y = j])\right)\right. \\
& - \left.\sum_{j=0}^{n} \mathbb{P}([X = k] \cap [Y = j]) \log_2\left(\mathbb{P}([X = k])\right)\right) \\
= \quad & -\frac{1}{\mathbb{P}([X = k])} \left(\sum_{j=0}^{n} \mathbb{P}([X = k] \cap [Y = j]) \log_2\left(\mathbb{P}([X = k] \cap [Y = j])\right)\right. \\
& - \left.\log_2\left(\mathbb{P}([X = k])\right) \sum_{j=0}^{n} \mathbb{P}([X = k] \cap [Y = j])\right)
\end{aligned}$$

- Or, la famille $([Y = j])_{j \in [\![0,n]\!]}$ forme un système complet d'événements.
  Ainsi, par formule des probabilités totales :

$$\mathbb{P}([X = k]) = \sum_{j=0}^{n} \mathbb{P}([X = k] \cap [Y = j])$$

D'où :

$$\begin{aligned}
& H(Y \mid X = k) \\
= \quad & -\frac{1}{\mathbb{P}([X = k])} \left(\sum_{j=0}^{n} \mathbb{P}([X = k] \cap [Y = j]) \log_2\left(\mathbb{P}([X = k] \cap [Y = j])\right)\right. \\
& - \left.\log_2\left(\mathbb{P}([X = k])\right) \mathbb{P}([X = k])\right)
\end{aligned}$$

Ainsi :

$$\begin{aligned}
& \mathbb{P}([X = k])\ H(Y \mid X = k) \\
= \quad & -\left(\sum_{j=0}^{n} \mathbb{P}([X = k] \cap [Y = j]) \log_2\left(\mathbb{P}([X = k] \cap [Y = j])\right)\right. \\
& - \left.\log_2\left(\mathbb{P}([X = k])\right) \mathbb{P}([X = k])\right)
\end{aligned}$$

- On en déduit :

$$\begin{aligned}
& H(Y \mid X) \\
= \quad & \sum_{k=0}^{n} \mathbb{P}([X=k])\ H(Y \mid X=k) \\
= \quad & -\sum_{k=0}^{n} \left(\sum_{j=0}^{n} \mathbb{P}([X=k] \cap [Y=j])\ \log_2\left(\mathbb{P}([X=k] \cap [Y=j])\right) \right. \\
- \quad & \left. \log_2\left(\mathbb{P}([X=k])\right)\ \mathbb{P}([X=k]) \right) \\
= \quad & -\sum_{k=0}^{n} \left(\sum_{j=0}^{n} \mathbb{P}([X=k] \cap [Y=j])\ \log_2\left(\mathbb{P}([X=k] \cap [Y=j])\right) \right) \\
+ \quad & \sum_{k=0}^{n} \log_2\left(\mathbb{P}([X=k])\right)\ \mathbb{P}([X=k]) \\
= \quad & H(X,Y) - H(X)
\end{aligned}$$

Finalement : $H(X,Y) = H(X) + H(Y \mid X)$. □

d) Montrer que pour tout couple de variables aléatoires X et Y de lois à support $\{0,1,\ldots,n\}$, on a :

$$H(X) - H(X \mid Y) \ = \ H(Y) - H(Y \mid X)$$

Démonstration.
D'après la question ***11.b)*** :

$$H(X,Y) \ = \ H(Y,X)$$

D'où, d'après la question ***11.c)*** :

$$H(X) + H(Y \mid X) \ = \ H(Y) + H(X \mid Y)$$

Ainsi : $H(X) - H(X \mid Y) \ = \ H(Y) - H(Y \mid X)$. □

12. On considère dans cette question deux variables aléatoires de lois à support $\{0,1,2,3\}$. On suppose que la loi conjointe de (X,Y) est donnée par le tableau suivant :

j \ k	0	1	2	3
0	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
1	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
3	$\frac{1}{4}$	0	0	0

(on lit dans la $k^{\text{ème}}$ colonne et la $j^{\text{ème}}$ ligne la valeur de $\mathbb{P}([X=k] \cap [Y=j])$)

a) Déterminer la loi de X et montrer : $H(X) = \dfrac{7}{4}$.

Démonstration.

- Tout d'abord : $X(\Omega) \subset [\![0, 3]\!]$.
- Soit $k \in [\![0, 3]\!]$.
  La famille $([Y = j])_{j \in [\![0,3]\!]}$ forme un système complet d'événements.
  Ainsi, par formule des probabilités totales :

$$\mathbb{P}([X = k]) \;=\; \sum_{j=0}^{3} \mathbb{P}([X = k] \cap [Y = j])$$

On en déduit :

$\times$ $\mathbb{P}([X = 0]) \;=\; \dfrac{1}{8} + \dfrac{1}{16} + \dfrac{1}{16} + \dfrac{1}{4} \;=\; \dfrac{1}{2}$

$\times$ $\mathbb{P}([X = 1]) \;=\; \dfrac{1}{16} + \dfrac{1}{8} + \dfrac{1}{16} \;=\; \dfrac{1}{4}$

$\times$ $\mathbb{P}([X = 2]) \;=\; \dfrac{1}{32} + \dfrac{1}{32} + \dfrac{1}{16} \;=\; \dfrac{1}{8}$

$\times$ $\mathbb{P}([X = 3]) \;=\; \dfrac{1}{32} + \dfrac{1}{32} + \dfrac{1}{16} \;=\; \dfrac{1}{8}$

On peut récapituler ces résultats dans le tableau suivant :

k	0	1	2	3
$\mathbb{P}([X = k])$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

Commentaire

- On peut vérifier :

$$\sum_{k=0}^{3} \mathbb{P}([X = k]) \;=\; \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} \;=\; 1$$

  Cette propriété est vérifiée car la famille $([X = k])_{k \in [\![0,3]\!]}$ forme un système complet d'événements.
- On pourra s'en servir pour vérifier que les probabilités $(\mathbb{P}([X = k]))_{k \in [\![0,3]\!]}$ obtenues définissent bien une loi de probabilité pour X.

D'après ***6.d)***, $H(X) = \dfrac{7}{4}$

□

b) Déterminer la loi de Y et calculer $H(Y)$.

Démonstration.

- Tout d'abord : $Y(\Omega) \subset [\![0,3]\!]$.
- Soit $j \in [\![0,3]\!]$.
  La famille $([X=k])_{k\in[\![0,3]\!]}$ forme un système complet d'événements.
  Ainsi, par formule des probabilités totales :

$$\mathbb{P}([Y=j]) \;=\; \sum_{k=0}^{3} \mathbb{P}([X=k]\cap[Y=j])$$

On en déduit le tableau suivant :

j	0	1	2	3
$\mathbb{P}([Y=j])$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Ainsi : $Y \hookrightarrow \mathcal{U}([\![0,3]\!])$.

- D'après ***9.d)*** : $H(Y) = \log_2(3+1) = \log_2(2^2) = 2$.
  La dernière égalité est vraie d'après la question ***1.b)***.

$H(Y) = 2$

□

c) Montrer : $H(X\,|\,Y) = \dfrac{11}{8}$.

Démonstration.

- Soit $k \in [\![0,3]\!]$.

$$\mathbb{P}_{[Y=j]}([X=k]) \;=\; \frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([Y=j])}$$

Avec le tableau fourni par l'énoncé et le tableau de la question précédente, on obtient le tableau suivant :

j \ k	0	1	2	3
0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
3	1	0	0	0

(on lit dans la $k^{\text{ème}}$ colonne et la $j^{\text{ème}}$ ligne la valeur de $\mathbb{P}_{[Y=j]}([X=k])$)

- On en déduit :

$$\begin{aligned} H(X\,|\,Y=0) &= -\sum_{k=0}^{3} \mathbb{P}_{[Y=0]}([X=k])\,\log_2\left(\mathbb{P}_{[Y=0]}([X=k])\right) \\ &= -\left(\frac{1}{2}\,\log_2\left(\frac{1}{2}\right)+\frac{1}{4}\,\log_2\left(\frac{1}{4}\right)+\frac{1}{8}\,\log_2\left(\frac{1}{8}\right)+\frac{1}{8}\,\log_2\left(\frac{1}{8}\right)\right) \\ &= \frac{7}{4} \qquad \textit{(calcul déjà effectué en } \textbf{\textit{6.d)}}\textit{)} \end{aligned}$$

  Avec le même calcul : $H(X\,|\,Y=1)=\dfrac{7}{4}$.

- Ensuite :

$$\begin{aligned} H(X\,|\,Y=2) &= -\sum_{k=0}^{3} \mathbb{P}_{[Y=2]}([X=k])\,\log_2\left(\mathbb{P}_{[Y=2]}([X=k])\right) \\ &= -\left(\frac{1}{4}\,\log_2\left(\frac{1}{4}\right)+\frac{1}{4}\,\log_2\left(\frac{1}{4}\right)+\frac{1}{4}\,\log_2\left(\frac{1}{4}\right)+\frac{1}{4}\,\log_2\left(\frac{1}{4}\right)\right) \\ &= \log_2(4) \;=\; \log_2(2^2) \;=\; 2 \end{aligned}$$

- Enfin :

$$\begin{aligned} H(X\,|\,Y=3) &= -\sum_{k=0}^{3} \mathbb{P}_{[Y=3]}([X=k])\,\log_2\left(\mathbb{P}_{[Y=3]}([X=k])\right) \\ &= -\big(1\times\log_2(1)+0\times\log_2(0)+0\times\log_2(0)+0\times\log_2(0)\big) \\ &= \log_2(1) \qquad \textit{(par convention, d'après l'énoncé)} \\ &= 0 \end{aligned}$$

- On en déduit :

$$\begin{aligned} H(X\,|\,Y) &= \sum_{j=0}^{3} \mathbb{P}([Y=j])\,H(X\,|\,Y=j) \\ &= \frac{1}{4}\times\frac{7}{4}+\frac{1}{4}\times\frac{7}{4}+\frac{1}{4}\times 2+\frac{1}{4}\times 0 \\ &= \frac{22}{16} \end{aligned}$$

$$\boxed{\text{Finalement : } H(X\,|\,Y)=\frac{11}{8}.}$$

□

d) Que vaut $H(Y\,|\,X)$?

Démonstration.
D'après la question ***11.d)*** :

$$\begin{aligned} H(Y\,|\,X) &= H(Y)-\big(H(X)-H(X\,|\,Y)\big) \\ &= 2-\frac{7}{4}+\frac{11}{8} \qquad \textit{(d'après les questions } \textbf{\textit{12.a)}}\textit{, } \textbf{\textit{b)}} \textit{ et } \textbf{\textit{c)}}\textit{)} \end{aligned}$$

$$\boxed{\text{D'où : } H(Y\,|\,X)=\frac{13}{8}.}$$

□

e) Calculer $H(X,Y)$.

Démonstration.
D'après la question ***11.c)*** :

$$H(X,Y) \;=\; H(X) + H(Y \,|\, X) \;=\; \frac{7}{4} + \frac{13}{8}$$

$$\boxed{\text{Finalement : } H(X,Y) = \frac{27}{8}.}$$

Commentaire

- Dans cette question ***12.***, l'objectif est de calculer $H(X,Y)$ à l'aide des entropies conditionnelles et des formules déterminées en question ***11.***.
- Cet objectif est un peu *ad hoc* puisque, les v.a.r. X et Y pouvant prendre peu de valeurs, il aurait été plus rapide de commencer par la détermination de $H(X,Y)$ pour en déduire les valeurs des entropies conditionnelles $H(X \,|\, Y)$ et $H(Y \,|\, X)$ à l'aide de $H(X)$, $H(Y)$ et des formules de la question ***11.***.
  Il n'est cependant pas conseillé de traiter les questions de l'énoncé dans un autre sens : on passera à coup sûr à côté de nombreux points de barème.

□

13. Soient X et Y deux variables aléatoires de lois à support $\{0,1,\dots,n\}$. On appelle **information mutuelle** de X et de Y le réel :

$$I(X,Y) \;=\; \sum_{k=0}^{n} \sum_{j=0}^{n} \mathbb{P}([X=k] \cap [Y=j]) \, \log_2 \left(\frac{\mathbb{P}([X=k] \cap [Y=j])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])} \right)$$

a) Montrer : $I(X,Y) = I(Y,X)$.

Démonstration.
On a :

$$\begin{aligned}
I(X,Y) &= \sum_{k=0}^{n} \left(\sum_{j=0}^{n} \mathbb{P}([X=k] \cap [Y=j]) \, \log_2 \left(\frac{\mathbb{P}([X=k] \cap [Y=j])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])} \right) \right) \\
&= \sum_{0 \leqslant k,j \leqslant n} \mathbb{P}([X=k] \cap [Y=j]) \, \log_2 \left(\frac{\mathbb{P}([X=k] \cap [Y=j])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])} \right) \\
&= \sum_{j=0}^{n} \left(\sum_{k=0}^{n} \mathbb{P}([X=k] \cap [Y=j]) \, \log_2 \left(\frac{\mathbb{P}([X=k] \cap [Y=j])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])} \right) \right) && \textit{(en sommant suivant les colonnes)} \\
&= \sum_{j=0}^{n} \left(\sum_{k=0}^{n} \mathbb{P}([Y=j] \cap [X=k]) \, \log_2 \left(\frac{\mathbb{P}([Y=j] \cap [X=k])}{\mathbb{P}([Y=j])\,\mathbb{P}([X=k])} \right) \right) && \textit{(par commutativité de l'intersection et du produit)} \\
&= I(Y,X)
\end{aligned}$$

$$\boxed{I(X,Y) = I(Y,X)}$$

□

b) Montrer : $I(X,Y) = H(X) - H(X \mid Y)$.

Démonstration.

- Soit $(k,j) \in [\![0,n]\!]^2$.

$$\begin{aligned} & \log_2\left(\frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])}\right) \\ = \;& \log_2\left(\frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([Y=j])}\right) - \log_2\big(\mathbb{P}([X=k])\big) \\ = \;& \log_2\big(\mathbb{P}_{[Y=j]}([X=k])\big) - \log_2\big(\mathbb{P}([X=k])\big) \end{aligned}$$

- On en déduit :

$$\begin{aligned} & I(X,Y) \\ = \;& \sum_{k=0}^{n}\left(\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[Y=j])\,\log_2\left(\frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])}\right)\right) \\ = \;& \sum_{k=0}^{n}\left(\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[Y=j])\left(\log_2\big(\mathbb{P}_{[Y=j]}([X=k])\big) - \log_2\big(\mathbb{P}([X=k])\big)\right)\right) \\ = \;& \sum_{k=0}^{n}\left(\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[Y=j])\,\log_2\big(\mathbb{P}_{[Y=j]}([X=k])\big)\right) \qquad (*) \\ & - \sum_{k=0}^{n}\left(\log_2\big(\mathbb{P}([X=k])\big)\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[Y=j])\right) \qquad (**) \end{aligned}$$

- Commençons par simplifier $(*)$.

$$\begin{aligned} & \sum_{k=0}^{n}\left(\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[Y=j])\,\log_2\big(\mathbb{P}_{[Y=j]}([X=k])\big)\right) \\ = \;& \sum_{k=0}^{n}\left(\sum_{j=0}^{n}\mathbb{P}([Y=j])\;\mathbb{P}_{[X=k]}([Y=j])\,\log_2\big(\mathbb{P}_{[Y=j]}([X=k])\big)\right) \\ = \;& \sum_{j=0}^{n}\left(\sum_{k=0}^{n}\mathbb{P}([Y=j])\;\mathbb{P}_{[X=k]}([Y=j])\,\log_2\big(\mathbb{P}_{[Y=j]}([X=k])\big)\right) \\ = \;& \sum_{j=0}^{n}\left(\mathbb{P}([Y=j])\sum_{k=0}^{n}\mathbb{P}_{[X=k]}([Y=j])\,\log_2\big(\mathbb{P}_{[Y=j]}([X=k])\big)\right) \\ = \;& \sum_{j=0}^{n}\mathbb{P}([Y=j])\big(-H(X\mid Y=j)\big) \\ = \;& -H(X\mid Y) \end{aligned}$$

- Simplifions ensuite $(**)$.
  La famille $([Y=j])_{j\in[\![0,n]\!]}$ forme un système complet d'événements.
  Ainsi, par formule des probabilités totales :

$$\mathbb{P}([X=k]) \;=\; \sum_{j=0}^{n} \mathbb{P}([X=k]\cap[Y=j])$$

  D'où :

$$-\sum_{k=0}^{n}\left(\log_2\big(\mathbb{P}([X=k])\big)\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[Y=j])\right)$$

$$= \;-\sum_{k=0}^{n}\log_2\big(\mathbb{P}([X=k])\big)\,\mathbb{P}([X=k])$$

$$= \;H(X)$$

  On en déduit : $I(X,Y) = H(X) - H(X\,|\,Y)$.

□

c) Montrer : $I(X,X) = H(X)$.

Démonstration.

- Tout d'abord :

$$I(X,X) \;=\; \sum_{k=0}^{n}\left(\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[X=j])\,\log_2\left(\frac{\mathbb{P}([X=k]\cap[X=j])}{\mathbb{P}([X=k])\,\mathbb{P}([X=j])}\right)\right)$$

- Soit $k\in[\![0,n]\!]$.

$$\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[X=j])\,\log_2\left(\frac{\mathbb{P}([X=k]\cap[X=j])}{\mathbb{P}([X=k])\,\mathbb{P}([X=j])}\right)$$

$$= \sum_{\substack{j=0\\ j\neq k}}^{n}\mathbb{P}([X=k]\cap[X=j])\,\log_2\left(\frac{\mathbb{P}([X=k]\cap[X=j])}{\mathbb{P}([X=k])\,\mathbb{P}([X=j])}\right)$$

$$+ \sum_{\substack{j=0\\ j=k}}^{n}\mathbb{P}([X=k]\cap[X=j])\,\log_2\left(\frac{\mathbb{P}([X=k]\cap[X=j])}{\mathbb{P}([X=k])\,\mathbb{P}([X=j])}\right)$$

  En effet, pour tout $j\neq k$:

$$\mathbb{P}([X=k]\cap[Y=j]) \;=\; \mathbb{P}(\varnothing) \;=\; 0$$

  Or, par convention : $0\times\log_2(0)=0$. Ainsi :

$$\sum_{\substack{j=0\\ j\neq k}}^{n}\mathbb{P}([X=k]\cap[X=j])\,\log_2\left(\frac{\mathbb{P}([X=k]\cap[X=j])}{\mathbb{P}([X=k])\,\mathbb{P}([X=j])}\right) \;=\; 0$$

De plus :

$$\sum_{j=0}^{n} \mathbb{P}([X=k]\cap[X=j]) \log_2\left(\frac{\mathbb{P}([X=k]\cap[X=j])}{\mathbb{P}([X=k])\,\mathbb{P}([X=j])}\right)$$

$$= \sum_{\substack{j=0\\ j=k}}^{n} \mathbb{P}([X=k]\cap[X=j]) \log_2\left(\frac{\mathbb{P}([X=k]\cap[X=j])}{\mathbb{P}([X=k])\,\mathbb{P}([X=j])}\right)$$

$$= \mathbb{P}([X=k]\cap[X=k]) \log_2\left(\frac{\mathbb{P}([X=k]\cap[X=k])}{\mathbb{P}([X=k])\,\mathbb{P}([X=k])}\right)$$

$$= \mathbb{P}([X=k]) \log_2\left(\frac{\cancel{\mathbb{P}([X=k])}}{\cancel{\mathbb{P}([X=k])}\,\mathbb{P}([X=k])}\right)$$

$$= -\mathbb{P}([X=k]) \log_2\big(\mathbb{P}([X=k])\big)$$

- On en conclut :

$$I(X,X) = \sum_{k=0}^{n}\Big(-\mathbb{P}([X=k]) \log_2\big(\mathbb{P}([X=k])\big)\Big) = H(X)$$

$$\boxed{I(X,X) = H(X)}$$

□

d) Que vaut $I(X,Y)$ si X et Y sont indépendantes ?

Démonstration.
Supposons que X et Y sont indépendantes.

- Soit $(k,j) \in [\![0,n]\!]^2$.
  Comme X et Y sont indépendantes :

$$\mathbb{P}([X=k]\cap[Y=j]) = \mathbb{P}([X=k])\,\mathbb{P}([Y=j])$$

On en déduit :

$$\log_2\left(\frac{\mathbb{P}([X=k]\cap[Y=k])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])}\right) = \log_2\left(\frac{\mathbb{P}([X=k])\,\mathbb{P}([Y=k])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])}\right) = \log_2(1) = 0$$

- Ainsi :

$$I(X,Y) = \sum_{k=0}^{n}\left(\sum_{j=0}^{n} \mathbb{P}([X=k]\cap[Y=j]) \log_2\left(\frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])}\right)\right)$$

$$= \sum_{k=0}^{n}\left(\sum_{j=0}^{n} \mathbb{P}([X=k]\cap[Y=j]) \times 0\right)$$

$$= 0$$

$$\boxed{\text{Si } X \text{ et } Y \text{ sont indépendantes, alors : } I(X,Y) = 0.}$$

□

14. Soient X et Y deux variables aléatoires de lois à support $\{0,1,\dots,n\}$. On fixe $0 \leqslant k \leqslant n$. Pour $0 \leqslant j \leqslant n$, on pose : $p_j = \dfrac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])}$. On suppose que $p_j > 0$ pour tout $0 \leqslant j \leqslant n$ et on pose : $x_j = \dfrac{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])}{\mathbb{P}([X=k]\cap[Y=j])}$.

a) Montrer : $\sum_{j=0}^{n} p_j = 1$.

Démonstration.

- Tout d'abord :

$$\sum_{j=0}^{n} p_j \;=\; \sum_{j=0}^{n} \frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])} \;=\; \frac{\sum_{j=0}^{n} \mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])}$$

- Or, on a déjà démontré (en question ***11.c)*** par exemple) :

$$\sum_{j=0}^{n} \mathbb{P}([X=k]\cap[Y=j]) \;=\; \mathbb{P}([X=k])$$

On en déduit :

$$\sum_{j=0}^{n} p_j \;=\; \frac{\mathbb{P}([X=k])}{\mathbb{P}([X=k])} \;=\; 1$$

$$\boxed{\sum_{j=0}^{n} p_j = 1}$$

□

b) Soit Z_k une variable aléatoire de loi à support $\{x_0,\dots,x_n\}$ dont la loi est donnée par $\mathbb{P}([Z_k = x_j]) = p_j$ pour $0 \leqslant j \leqslant n$. Montrer :

$$\mathbb{E}\big(\log_2(Z_k)\big) \;\leqslant\; 0$$

Démonstration.

- La v.a.r. Z_k est bien définie car :
  - × $\forall j \in [\![0,n]\!]$, $p_j \geqslant 0$,
  - × $\sum_{j=0}^{n} p_j = 1$.
- Démontrons : $Z_k(\Omega) \subset \mathbb{R}_+$.
  - × Tout d'abord, d'après l'énoncé : $Z_k(\Omega) \subset \{x_0,\dots,x_n\}$.
  - × Or, pour tout $j \in [\![0,n]\!]$: $x_j = \dfrac{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])}{\mathbb{P}([X=k]\cap[Y=j])} \geqslant 0$.

$$\boxed{Z_k(\Omega) \subset \mathbb{R}_+}$$

- On remarque alors que :
  - × la fonction $\log_2$ est concave sur $\mathbb{R}_+^*$ (d'après ***1.c)***),
  - × la v.a.r. Z_k est de loi à support $A = \{x_0,\dots,x_n\} \subset \mathbb{R}_+$, avec $\mathrm{Card}(A) = n+1$.

  Par inégalité de Jensen (question ***8.d)***) :

$$\mathbb{E}\big(\log_2(Z_k)\big) \;\leqslant\; \log_2\big(\mathbb{E}(Z_k)\big)$$

- Or, par définition de Z_k :

$$\begin{aligned}
\mathbb{E}(Z_k) &= \sum_{j=0}^{n} x_j\, \mathbb{P}([Z_k = x_j]) \\
&= \sum_{j=0}^{n} x_j\, p_j \\
&= \sum_{j=0}^{n} \frac{\cancel{\mathbb{P}([X=k])}\ \mathbb{P}([Y=j])}{\cancel{\mathbb{P}([X=k]\cap[Y=j])}}\ \frac{\cancel{\mathbb{P}([X=k]\cap[Y=j])}}{\cancel{\mathbb{P}([X=k])}} \\
&= \sum_{j=0}^{n} \mathbb{P}([Y=j]) \\
&= 1 \qquad \textit{(car } ([Y=j])_{j\in[\![0,n]\!]} \textit{ est un système complet d'événements)}
\end{aligned}$$

D'où : $\log_2\big(\mathbb{E}(Z_k)\big) = \log_2(1) = 0$.

Finalement : $\mathbb{E}\big(\log_2(Z_k)\big) \leqslant 0$.

□

c) En déduire : $I(X,Y) \geqslant 0$.

Démonstration.

- Par théorème de transfert :

$$\begin{aligned}
&\mathbb{E}\big(\log_2(Z_k)\big) \\
&= \sum_{j=0}^{n} \log_2(x_j)\, \mathbb{P}([Z_k = x_j]) \\
&= \sum_{j=0}^{n} \log_2(x_j)\, p_j \\
&= \sum_{j=0}^{n} \log_2\left(\frac{\mathbb{P}([X=k])\ \mathbb{P}([Y=j])}{\mathbb{P}([X=k]\cap[Y=j])}\right)\ \frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])} \\
&= -\frac{1}{\mathbb{P}([X=k])} \sum_{j=0}^{n} \mathbb{P}([X=k]\cap[Y=j])\, \log_2\left(\frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])\ \mathbb{P}([Y=j])}\right)
\end{aligned}$$

- De plus, d'après la question précédente :

$$\mathbb{E}\big(\log_2(Z_k)\big) \leqslant 0$$

D'où :

$$-\frac{1}{\mathbb{P}([X=k])} \sum_{j=0}^{n} \mathbb{P}([X=k]\cap[Y=j])\, \log_2\left(\frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])\ \mathbb{P}([Y=j])}\right) \leqslant 0$$

Comme $\mathbb{P}([X=k]) \geqslant 0$:

$$\sum_{j=0}^{n} \mathbb{P}([X=k]\cap[Y=j])\, \log_2\left(\frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])\ \mathbb{P}([Y=j])}\right) \geqslant 0$$

- En sommant ces inégalités pour k variant de 0 à n :

$$\sum_{k=0}^{n}\sum_{j=0}^{n}\mathbb{P}([X=k]\cap[Y=j])\,\log_2\left(\frac{\mathbb{P}([X=k]\cap[Y=j])}{\mathbb{P}([X=k])\,\mathbb{P}([Y=j])}\right) \geqslant 0$$

$$\shortparallel$$

$$I(X,Y)$$

$$\boxed{I(X,Y)\geqslant 0}$$

Commentaire

- On a donc démontré dans cette question que l'information mutuelle de X et de Y, $I(X,Y)$ est minorée par 0.
- Remarquons qu'on a démontré plus tôt (en question ***13.d)***) :

$$X \text{ et } Y \text{ indépendantes} \quad \Rightarrow \quad I(X,Y)=0$$

  Le minorant trouvé dans cette question ***14.c)*** est donc atteint si X et Y sont indépendantes. Cela signifie en particulier que la minoration trouvée est optimale : il est impossible de trouver un meilleur minorant de l'information mutuelle que 0.
- Il n'est pas si surprenant de trouver que l'information mutuelle est minimale dans le cas de 2 v.a.r. X et Y indépendantes puisque l'énoncé nous introduit en toute première page l'information mutuelle $I(X,Y)$ comme une mesure de l'information apportée par les v.a.r. X et Y l'une sur l'autre.
- Dans tout ce problème, on introduit et étudie quelques notions de statistiques, plus particulièrement de la théorie de l'information développée par Fisher et Shannon. Pour d'autres exemples d'utilisation de cette théorie, on pourra se référer au sujet `ESSEC-II 2009`.

□