



PRÉPASIMPLE

prepasimple.fr

ESSEC

Mathématiques 2 — Option économique

2019

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Corrigé

Première partie : Entropie différentielle d'une variable à densité

1. a) Pour $x, y > 0$, la propriété du logarithme népérien donne $\ln(xy) = \ln x + \ln y$. En divisant par $\ln 2 > 0$, on obtient

$$\log_2(xy) = \log_2 x + \log_2 y.$$

1. b) Par définition d'une puissance réelle, $2^\alpha = e^{\alpha \ln 2}$. Ainsi

$$\log_2(2^\alpha) = \frac{\ln(e^{\alpha \ln 2})}{\ln 2} = \alpha.$$

1. c) La fonction $\log_2$ est de classe C^2 sur $]0, +\infty[$ et ses dérivées sont

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

Le critère de concavité pour les fonctions de classe C^2 montre donc que $\log_2$ est strictement concave, et en particulier concave, sur $\mathbb{R}_+^*$.

2. Question adaptée au programme actuel.

La condition de convergence absolue assure l'existence de l'espérance demandée.

Posons $q(x) = \log_2(f(x))$ sur I et $q(x) = 0$ ailleurs. La fonction q est continue aux points où f est continue et strictement positive ; les exceptions admises sont celles du cadre des densités du programme. L'intégrale

$$\int_a^b |q(x)| f(x) dx = \int_a^b f(x) |\log_2(f(x))| dx$$

est finie par hypothèse. De plus, $f = 0$ hors de I , donc $\mathbb{P}(X \notin I) = 0$. Le théorème de transfert donne

$$E(q(X)) = \int_a^b f(x) \log_2(f(x)) dx.$$

Par conséquent, $h(X) = -E(\log_2(f(X)))$, où la valeur attribuée hors de I n'influence pas l'espérance. Une simple convergence conditionnelle de l'intégrale ne suffirait pas à

appliquer le transfert : c'est la raison de la précision de l'énoncé adapté.

3. a) i) Pour tout $y \in \mathbb{R}$, $F_Y(y) = \mathbb{P}(c + X \leq y) = F_X(y - c)$. En dérivant aux points de continuité de f , on obtient une densité

$$f_Y(y) = f(y - c).$$

Son support est $c + I$.

3. a) ii) Méthode 1 — Transfert et identité des variables

Avec probabilité 1, $f_Y(Y) = f(Y - c) = f(X) > 0$. Ainsi $\log_2(f_Y(Y)) = \log_2(f(X))$. Le membre de droite admet une espérance absolue par hypothèse. Il en est donc de même du membre de gauche, et la question 2 donne

$$h(Y) = -E(\log_2(f_Y(Y))) = -E(\log_2(f(X))) = h(X).$$

Méthode 2 — Translation dans l'intégrale

Sur un segment $[u, v] \subset I$, la substitution affine $y = x + c$ donne

$$\int_{u+c}^{v+c} f_Y(y) |\log_2(f_Y(y))| dy = \int_u^v f(x) |\log_2(f(x))| dx.$$

En faisant tendre les bornes vers celles de I , on obtient la convergence absolue pour Y . La même substitution, sans valeurs absolues, donne alors $h(Y) = h(X)$. La première méthode utilise directement la variable transformée ; la seconde montre que déplacer une densité ne modifie pas son intégrale d'entropie.

3. b) i) Comme $\alpha > 0$, $F_Z(z) = F_X(z/\alpha)$. La dérivation fournit

$$f_Z(z) = \frac{1}{\alpha} f(z/\alpha),$$

de support αI .

3. b) ii) Méthode 1 — Transformation du logarithme de la densité

Avec probabilité 1,

$$\log_2(f_Z(Z)) = \log_2(f(X)/\alpha) = \log_2(f(X)) - \log_2 \alpha.$$

Cette variable admet une espérance absolue puisque $\log_2(f(X))$ en admet une et que le second terme est constant. Donc

$$h(Z) = -E(\log_2(f(X))) + \log_2 \alpha = h(X) + \log_2 \alpha.$$

Méthode 2 — Changement d'échelle dans l'intégrale

Sur les segments correspondants, la substitution affine $z = \alpha x$ transforme l'intégrale d'entropie en

$$-\int_u^v f(x) \log_2 \left(\frac{f(x)}{\alpha} \right) dx = -\int_u^v f(x) \log_2(f(x)) dx + \log_2 \alpha \int_u^v f(x) dx.$$

La majoration

$$f(x) |\log_2(f(x)/\alpha)| \leq f(x) |\log_2(f(x))| + |\log_2 \alpha| f(x)$$

assure la convergence absolue. Le passage aux bornes de I est donc justifié; comme $\int_I f = 1$, on retrouve $h(Z) = h(X) + \log_2 \alpha$. Le changement d'échelle explique le terme supplémentaire, tandis que la première méthode évite de reprendre les bornes d'intégration.

4. a) i) On peut choisir

$$f_X(x) = \begin{cases} 1/a & \text{si } 0 \leq x \leq a, \\ 0 & \text{sinon.} \end{cases}$$

Les valeurs aux deux extrémités n'affectent pas la loi.

4. a) ii) Sur le support, $\log_2(f_X(x)) = -\log_2 a$ est constant. L'intégrale absolue vaut $|\log_2 a|$, donc elle est finie. Ainsi

$$h(X) = - \int_0^a \frac{1}{a} \log_2(1/a) dx = \log_2 a.$$

4. a) iii) La fonction $\log_2$ est strictement croissante et s'annule en 1. Par conséquent, $h(X) > 0$ si et seulement si $a > 1$. L'entropie différentielle peut donc être négative : elle l'est pour $0 < a < 1$.

4. b) Une densité de Y est $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$. Pour tout réel y ,

$$-\log_2(f_Y(y)) = \frac{1}{2} \log_2(2\pi) + \frac{y^2}{2 \ln 2}.$$

La valeur absolue de cette expression est majorée par une constante plus $y^2/(2 \ln 2)$. Or $E(Y^2) = V(Y) + E(Y)^2 = 1$. L'espérance absolue existe donc. Par transfert et linéarité,

$$h(Y) = \frac{1}{2} \log_2(2\pi) + \frac{1}{2 \ln 2} = \frac{1}{2} \log_2(2\pi e).$$

4. c) On choisit $f_Z(z) = \lambda e^{-\lambda z}$ pour $z \geq 0$, et 0 sinon. Sur $[0, +\infty[$,

$$-\log_2(f_Z(z)) = -\log_2 \lambda + \frac{\lambda z}{\ln 2}.$$

La valeur absolue est majorée par $|\log_2 \lambda| + \lambda z / \ln 2$. Comme $E(Z) = 1/\lambda$, cette expression est intégrable. On en déduit

$$h(Z) = -\log_2 \lambda + \frac{\lambda E(Z)}{\ln 2} = \log_2(e/\lambda).$$

4. d) i) La fonction f est positive et continue sur $\mathbb{R}$. Par parité, pour $R > 0$,

$$\int_{-R}^R f(x) dx = \lambda \int_0^R e^{-\lambda x} dx = 1 - e^{-\lambda R}.$$

En faisant tendre R vers $+\infty$, on obtient $\int_{-\infty}^{+\infty} f(x) dx = 1$. Donc f est bien une densité.

4. d) ii) On commence par justifier $E(|W|)$. La parité donne

$$E(|W|) = \lambda \int_0^{+\infty} x e^{-\lambda x} dx.$$

L'intégration par parties est effectuée d'abord sur $[0, R]$:

$$\lambda \int_0^R x e^{-\lambda x} dx = -R e^{-\lambda R} + \frac{1 - e^{-\lambda R}}{\lambda}.$$

Les croissances comparées donnent $R e^{-\lambda R} \rightarrow 0$. Ainsi $E(|W|) = 1/\lambda$. D'autre part,

$$-\log_2(f(W)) = \log_2(2/\lambda) + \frac{\lambda|W|}{\ln 2}.$$

L'espérance absolue existe grâce au moment précédent. Finalement,

$$h(W) = \log_2(2/\lambda) + \frac{1}{\ln 2} = \log_2(2e/\lambda).$$

5. a) Question adaptée au programme actuel : le cas $\gamma = 0$ de la définition est simplement explicité ; il n'intervient pas pour X , dont la variance est strictement positive.

Dans la définition du couple gaussien centré, on prend $(\alpha, \beta) = (1, 0)$. Il existe $\gamma \in \mathbb{R}$ et $Z \sim \mathcal{N}(0, 1)$ tels que X ait même loi que γZ . Sa variance est alors $\gamma^2 = \sigma^2 > 0$, donc $\gamma \neq 0$. La transformation affine d'une loi normale donne $X \sim \mathcal{N}(0, \sigma^2)$.

5. b) En posant $\sigma = \sqrt{\sigma^2} > 0$, X a même loi que σZ , avec $Z \sim \mathcal{N}(0, 1)$. D'après 3.b) et 4.b),

$$h(X) = \frac{1}{2} \log_2(2\pi e) + \log_2 \sigma = \frac{1}{2} \log_2(2\pi e \sigma^2).$$

L'existence est assurée par 3.b), puisque celle de $h(Z)$ a été établie.

5. c) i) L'inégalité $(|X| - |Y|)^2 \geq 0$ donne $2|XY| \leq X^2 + Y^2$. Les variables X et Y sont centrées et de variance σ^2 , donc $E(X^2) = E(Y^2) = \sigma^2 < +\infty$. Par comparaison, $E(|XY|) \leq \sigma^2 < +\infty$. Ainsi $E(XY)$ existe.

5. c) ii) Pour $\lambda \in \mathbb{R}$, la variable $(X + \lambda Y)^2$ admet une espérance, car les trois termes de son développement en admettent une. Sa positivité et la linéarité donnent

$$0 \leq E((X + \lambda Y)^2) = E(X^2) + 2\lambda E(XY) + \lambda^2 E(Y^2).$$

5. c) iii) Posons $c = E(XY)$ et prenons $\lambda = -c/\sigma^2$, ce qui est permis puisque $\sigma^2 > 0$.

L'inégalité précédente donne

$$0 \leq \sigma^2 - \frac{c^2}{\sigma^2}.$$

En multipliant par $\sigma^2 > 0$, on obtient $c^2 \leq \sigma^4 = E(X^2)E(Y^2)$. Aucune indépendance n'a été utilisée.

5. c) iv) L'inégalité précédente fournit $|E(XY)| \leq \sigma^2$. Par division par $\sigma^2 > 0$, $|\rho| \leq 1$, soit $\rho \in [-1, 1]$.

5. c) v) Si X et Y sont indépendantes, leur produit a pour espérance $E(X)E(Y) = 0$, puisque les deux variables sont centrées. Donc $\rho = 0$.

5. d) i) Sous l'hypothèse $|\rho| < 1$, l'argument du logarithme est strictement positif. Le logarithme de base 2 s'annule uniquement en 1, donc

$$h(X, Y) = 0 \iff 2\pi e\sigma^2\sqrt{1-\rho^2} = 1.$$

Pour une variance fixée, cette condition équivaut à $\sigma^2 \geq 1/(2\pi e)$ et $\rho^2 = 1 - (2\pi e\sigma^2)^{-2}$. Si $\sigma^2 < 1/(2\pi e)$, aucune valeur de ρ ne convient.

5. d) ii) Les deux entropies marginales sont égales à $\frac{1}{2} \log_2(2\pi e\sigma^2)$. Par les propriétés du logarithme,

$$I(X, Y) = \log_2(2\pi e\sigma^2) - \log_2(2\pi e\sigma^2\sqrt{1-\rho^2}) = -\frac{1}{2} \log_2(1-\rho^2).$$

La formule de $h(X, Y)$ est une définition donnée dans l'énoncé : aucun calcul de densité de couple n'est nécessaire.

5. d) iii) Puisque $0 < 1 - \rho^2 \leq 1$, on a $\log_2(1 - \rho^2) \leq 0$. Par conséquent, $I(X, Y) \geq 0$. L'égalité a lieu exactement lorsque $\rho = 0$.

5. d) iv) Lorsque $\rho \rightarrow 1$ avec $|\rho| < 1$, $1 - \rho^2 \rightarrow 0^+$. Le logarithme tend alors vers $-\infty$, donc $I(X, Y) \rightarrow +\infty$.

Deuxième partie : Généralités sur l'entropie des variables discrètes

6. a) La variable $g(X)$ est à support fini et prend uniquement des valeurs réelles, puisque toutes les probabilités marginales sont strictement positives. Elle admet donc une espérance et le transfert donne

$$E(g(X)) = \sum_{k=0}^n g(k) \mathbb{P}(X = k) = \sum_{k=0}^n \mathbb{P}(X = k) \log_2(\mathbb{P}(X = k)) = -H(X).$$

6. b) Pour tout k , $0 < \mathbb{P}(X = k) \leq 1$, donc $\log_2(\mathbb{P}(X = k)) \leq 0$. Chaque terme $-\mathbb{P}(X = k) \log_2(\mathbb{P}(X = k))$ est positif ou nul ; leur somme $H(X)$ l'est aussi. Pour $n = 0$, l'unique probabilité vaut 1 et l'entropie vaut 0.

6. c) i) Les probabilités de 0 et 1 sont respectivement $1 - p$ et p . Ainsi

$$\psi(p) = H(X) = -p \log_2 p - (1 - p) \log_2(1 - p).$$

6. c) ii) La fonction ψ est de classe C^2 sur $]0, 1[$. En dérivant les deux termes,

$$\psi'(p) = \frac{\ln(1 - p) - \ln p}{\ln 2}, \quad \psi''(p) = -\frac{1}{\ln 2} \left(\frac{1}{1 - p} + \frac{1}{p} \right) = -\frac{1}{p(1 - p) \ln 2} < 0.$$

Donc ψ est strictement concave sur $]0, 1[$.

6. c) iii) Méthode 1 — Étude de la dérivée

La dérivée est positive si $(1 - p)/p > 1$, soit $p < 1/2$, nulle en $1/2$, et négative si $p > 1/2$. Ainsi ψ atteint son unique maximum en $p = 1/2$, et ce maximum vaut 1.

Méthode 2 — Concavité et symétrie

L'expression de ψ donne $\psi(1-p) = \psi(p)$. Pour $p \neq 1/2$, la stricte concavité appliquée aux deux points distincts p et $1-p$ fournit

$$\psi(1/2) > \frac{1}{2}\psi(p) + \frac{1}{2}\psi(1-p) = \psi(p).$$

La valeur $\psi(1/2) = 1$ est donc l'unique maximum. La première méthode décrit toutes les variations ; la seconde exploite la symétrie de l'échange entre succès et échec.

6. d) Les logarithmes de $1/2$, $1/4$ et $1/8$ valent respectivement -1 , -2 et -3 . Par conséquent,

$$H(X) = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + 2 \times \frac{1}{8} \times 3 = \frac{7}{4}.$$

7. Code adapté en Python.

On additionne les contributions des $n+1$ valeurs possibles. Les probabilités sont strictement positives par l'hypothèse de support, donc chaque logarithme est défini.

Python

```
import numpy as np

def entropie(P):
    h = 0.0
    for k in range(len(P)):
        h = h - P[k] * np.log(P[k]) / np.log(2)
    return h
```

La variable h est la somme partielle des contributions. `range(len(P))` parcourt les indices de 0 à $\text{len}(P)-1$, soit exactement les indices des probabilités. La valeur renvoyée approche l'entropie en base 2, les erreurs éventuelles provenant des arrondis numériques. Pour la loi $(1/2, 1/4, 1/8, 1/8)$, le résultat vaut $1,75$.

8. a) Question adaptée au programme actuel : la définition de convexité utile à la récurrence est rappelée dans l'énoncé, sans supposer de régularité supplémentaire.

Comme la loi a pour support les N points donnés, $p_i > 0$. Puisque $N \geq 2$, il existe un autre indice $j \neq i$ de probabilité strictement positive. Dès lors $p_i \leq 1 - p_j < 1$. Ainsi $0 < p_i < 1$.

8. b) Pour une loi à deux valeurs, $p_1 + p_2 = 1$ et les deux coefficients sont dans $]0, 1[$. Par définition de la convexité,

$$\varphi(E(X)) = \varphi(p_1 x_1 + p_2 x_2) \leq p_1 \varphi(x_1) + p_2 \varphi(x_2) = E(\varphi(X)).$$

Toutes les espérances existent, puisque les variables prennent un nombre fini de valeurs réelles. Cela prouve $\mathcal{P}(2)$.

8. c) i) On a

$$1 - p_N = \sum_{i=1}^{N-1} p_i > 0,$$

donc les p'_i sont définis, positifs et de somme

$$\sum_{i=1}^{N-1} p'_i = \frac{1 - p_N}{1 - p_N} = 1.$$

Comme $N - 1 \geq 2$, chaque terme p_i est strictement inférieur à la somme des $N - 1$ termes ; par conséquent $p'_i < 1$.

8. c) ii) Les valeurs $x_1, \dots, x_{N-1}$ sont distinctes et positives ou nulles, et toutes leurs probabilités p'_i sont strictement positives. La variable Y satisfait donc les hypothèses de $\mathcal{P}(N - 1)$. Cette hypothèse donne

$$E(\varphi(Y)) \geq \varphi(E(Y)),$$

qui est précisément l'inégalité demandée après écriture des deux espérances comme sommes finies.

8. c) iii) Posons

$$m' = \sum_{i=1}^{N-1} p'_i x_i.$$

D'après la question précédente,

$$E(\varphi(X)) = (1 - p_N) \sum_{i=1}^{N-1} p'_i \varphi(x_i) + p_N \varphi(x_N) \geq (1 - p_N) \varphi(m') + p_N \varphi(x_N).$$

La convexité appliquée aux deux points m' et x_N , tous deux dans $\mathbb{R}_+$, donne

$$(1 - p_N) \varphi(m') + p_N \varphi(x_N) \geq \varphi((1 - p_N)m' + p_N x_N) = \varphi\left(\sum_{i=1}^N p_i x_i\right) = \varphi(E(X)).$$

Cela établit $\mathcal{P}(N)$. Par récurrence, la propriété est vraie pour tout support fini d'au moins deux valeurs. Pour un support réduit à un point, elle est une égalité immédiate.

8. d) Si φ est concave, $-\varphi$ est convexe. L'inégalité déjà prouvée donne

$$E(-\varphi(X)) \geq -\varphi(E(X)).$$

En multipliant par -1 , on obtient

$$E(\varphi(X)) \leq \varphi(E(X)).$$

Si des valeurs d'une variable auxiliaire coïncident, on additionne d'abord leurs probabilités; on retrouve ainsi une loi sur un support de valeurs distinctes, sans changer ses espérances.

9. a) Méthode 1 — Inégalité de Jensen démontrée à la question 8

On introduit la variable T qui prend la valeur $t_k = 1/((n+1)p_k)$ lorsque $X = k$. Elle a un support fini de valeurs strictement positives et

$$E(T) = \sum_{k=0}^n p_k \frac{1}{(n+1)p_k} = 1.$$

Pour appliquer rigoureusement 8.d), dont le domaine est $\mathbb{R}_+$, posons $a = \min_k t_k > 0$. La variable $T - a$ est positive ou nulle, et $u \mapsto \log_2(u + a)$ est concave sur $\mathbb{R}_+$. La question 8.d) donne alors

$$E(\log_2 T) \leq \log_2(E(T - a) + a) = \log_2(E(T)) = 0.$$

En développant $E(\log_2 T)$, on obtient exactement l'inégalité demandée, y compris lorsque certaines valeurs t_k coïncident. Le cas d'une variable constante est couvert par l'égalité signalée après 8.c).

Méthode 2 — Tangente au logarithme

La fonction $u \mapsto u - 1 - \ln u$, définie sur $]0, +\infty[$, a pour dérivée $1 - 1/u$. Elle décroît jusqu'à 1, puis croît, et vaut 0 en 1. Ainsi $\ln u \leq u - 1$ pour tout $u > 0$.

En appliquant cette inégalité à chaque t_k puis en multipliant par p_k et en sommant,

$$\sum_{k=0}^n p_k \log_2 t_k \leq \frac{1}{\ln 2} \left(\sum_{k=0}^n p_k t_k - \sum_{k=0}^n p_k \right) = 0.$$

La première méthode suit le résultat préparatoire de l'énoncé ; la seconde montre que la seule majoration par la tangente en 1 suffit ici.

9. b) Par additivité du logarithme et parce que

$$\sum p_k = 1,$$

$$\sum_{k=0}^n p_k \log_2((n+1)p_k) = \log_2(n+1) \sum_{k=0}^n p_k + \sum_{k=0}^n p_k \log_2 p_k = \log_2(n+1) - H(X).$$

9. c) Le membre de gauche de 9.a) est l'opposé du membre de gauche de 9.b). On a donc $H(X) - \log_2(n+1) \leq 0$, soit $H(X) \leq \log_2(n+1)$. Pour $n = 0$, les deux membres valent 0.

9. d) Pour la loi uniforme, $p_k = 1/(n+1)$ pour chaque k . Ainsi

$$H(X) = -(n+1) \frac{1}{n+1} \log_2 \left(\frac{1}{n+1} \right) = \log_2(n+1).$$

La borne précédente est donc atteinte par la loi uniforme.

10. a) Les événements $[X = k] \cap [Y = k]$, pour $0 \leq k \leq n$, sont deux à deux incompatibles et leur réunion est $[X = Y]$. Par indépendance et égalité des deux lois,

$$\mathbb{P}(X = Y) = \sum_{k=0}^n \mathbb{P}(X = k) \mathbb{P}(Y = k) = \sum_{k=0}^n p_k^2.$$

10. b) On doit tenir compte du fait que $\log_2(v(X))$ est négative ou nulle, alors que Jensen a été prouvée sur $\mathbb{R}_+$. Posons $L = \log_2(v(X))$ et $c = -\min_{0 \leq k \leq n} \log_2 p_k \geq 0$. Toutes les probabilités sont positives, donc c est fini et $L + c \geq 0$.

La fonction $\varphi(u) = 2^{u-c}$ est de classe C^2 sur $\mathbb{R}_+$ et sa dérivée seconde vaut $(\ln 2)^2 2^{u-c} > 0$: elle est convexe. En appliquant 8.c) à $L + c$, on obtient

$$2^{E(L+c)-c} = \varphi(E(L+c)) \leq E(\varphi(L+c)) = E(2^L).$$

Comme $2^{\log_2(v(X))} = v(X)$, cela donne $2^{E(\log_2(v(X)))} \leq E(v(X))$, comme demandé.

10. c) Par 6.a), $E(\log_2(v(X))) = -H(X)$. Par transfert,

$$E(v(X)) = \sum_{k=0}^n p_k v(k) = \sum_{k=0}^n p_k^2,$$

qui vaut $\mathbb{P}(X = Y)$ d'après 10.a). L'inégalité de 10.b) devient donc

$$2^{-H(X)} \leq \mathbb{P}(X = Y).$$

- 10. d)** Si les deux variables indépendantes suivent la loi uniforme sur $\{0, 1, \dots, n\}$, alors $H(X) = \log_2(n+1)$ et $\mathbb{P}(X = Y) = (n+1)/(n+1)^2 = 1/(n+1)$. Les deux membres de l'inégalité valent $1/(n+1)$: c'est un exemple d'égalité, valable aussi pour $n = 0$.

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

Pour cette partie, on utilise les notations $q_{kj} = \mathbb{P}([X = k] \cap [Y = j])$, $a_k = \mathbb{P}(X = k)$ et $b_j = \mathbb{P}(Y = j)$. Les marginales sont strictement positives ; les probabilités jointes peuvent être nulles. Toutes les sommes sont finies.

- 11. a)** Question adaptée au programme actuel.

On définit $g(k, j) = 0$ lorsque $q_{kj} = 0$: la fonction g est ainsi à valeurs réelles, même si le tableau joint comporte des zéros. Le transfert pour un couple discret à support fini donne

$$E(g(X, Y)) = \sum_{k=0}^n \sum_{j=0}^n q_{kj} g(k, j) = \sum_{k=0}^n \sum_{j=0}^n q_{kj} \log_2 q_{kj} = -H(X, Y),$$

où les termes de probabilité nulle valent 0. La valeur choisie pour g sur ces cases n'influence pas l'espérance. Cela évite de faire intervenir une valeur $-\infty$ dans un transfert pour une fonction réelle.

- 11. b)** En échangeant les deux indices dans la somme finie, on échange

$q_{kj} = \mathbb{P}(X = k, Y = j)$ et $\mathbb{P}(Y = j, X = k)$, qui sont égaux. Chaque terme est donc inchangé et $H(X, Y) = H(Y, X)$.

- 11. c)** Comme $a_k > 0$, les probabilités conditionnelles q_{kj}/a_k sont définies. Sur les cases où $q_{kj} > 0$, on a $\log_2(q_{kj}/a_k) = \log_2 q_{kj} - \log_2 a_k$. Les cases nulles apportent 0. Par conséquent,

$$\begin{aligned}
 H(Y | X) &= - \sum_{k=0}^n \sum_{j: q_{kj} > 0} q_{kj} \log_2 \left(\frac{q_{kj}}{a_k} \right) \\
 &= - \sum_{k=0}^n \sum_{j=0}^n q_{kj} \log_2 q_{kj} + \sum_{k=0}^n \left(\sum_{j=0}^n q_{kj} \right) \log_2 a_k.
 \end{aligned}$$

Or

$$\sum_j q_{kj} = a_k.$$

Ainsi

$$H(Y | X) = H(X, Y) + \sum_k a_k \log_2 a_k = H(X, Y) - H(X),$$

soit

$$H(X, Y) = H(X) + H(Y | X).$$

Il s'agit d'une identité entre sommes explicitement définies, et non de l'emploi d'un résultat général d'espérance conditionnelle.

11. d) La question 11.c) et sa version avec les deux variables échangées donnent

$$H(X) + H(Y | X) = H(X, Y) = H(Y) + H(X | Y).$$

En regroupant les termes, on obtient $H(X) - H(X | Y) = H(Y) - H(Y | X)$.

12. a) La loi de X s'obtient en sommant les colonnes du tableau. Les quatre probabilités sont

$$a_0 = \frac{1}{8} + \frac{1}{16} + \frac{1}{16} + \frac{1}{4} = \frac{1}{2},$$

$$a_1 = \frac{1}{16} + \frac{1}{8} + \frac{1}{16} = \frac{1}{4},$$

$$a_2 = a_3 = \frac{1}{32} + \frac{1}{32} + \frac{1}{16} = \frac{1}{8}.$$

On retrouve la loi de 6.d), donc $H(X) = 7/4$.

12. b) Chacune des quatre lignes a une somme égale à $1/4$. Par exemple la première donne $1/8 + 1/16 + 1/32 + 1/32 = 1/4$; la deuxième a les mêmes termes, la troisième contient quatre fois $1/16$ et la quatrième contient $1/4$ seulement. Donc Y suit la loi uniforme sur $\{0, 1, 2, 3\}$ et $H(Y) = \log_2 4 = 2$.

12. c) Conditionnellement à $Y = j$, on divise la ligne j par $1/4$. Les quatre lois de X sachant $Y = j$ sont donc :

j	Loi sur $(0, 1, 2, 3)$	Entropie
0	$(1/2, 1/4, 1/8, 1/8)$	$7/4$
1	$(1/4, 1/2, 1/8, 1/8)$	$7/4$
2	$(1/4, 1/4, 1/4, 1/4)$	2
3	$(1, 0, 0, 0)$	0

La deuxième entropie est égale à la première par permutation des probabilités. Les deux dernières correspondent à une loi uniforme sur quatre valeurs et à une loi certaine.

D'après la définition,

$$H(X | Y) = \frac{1}{4} \left(\frac{7}{4} + \frac{7}{4} + 2 + 0 \right) = \frac{11}{8}.$$

12. d) L'identité 11.d) donne $H(Y | X) = H(Y) - H(X) + H(X | Y)$. Par conséquent,

$$H(Y | X) = 2 - \frac{7}{4} + \frac{11}{8} = \frac{13}{8}.$$

12. e) Méthode 1 — Décomposition de l'entropie jointe

La question 11.c), avec les variables échangées, donne

$$H(X, Y) = H(Y) + H(X | Y) = 2 + \frac{11}{8} = \frac{27}{8}.$$

Méthode 2 — Somme directe sur le tableau

Le tableau contient deux cases égales à $1/8$, six cases égales à $1/16$, quatre cases égales à $1/32$, une case égale à $1/4$ et trois cases nulles. Ces dernières ne contribuent pas. Ainsi

$$H(X, Y) = 2 \times \frac{1}{8} \times 3 + 6 \times \frac{1}{16} \times 4 + 4 \times \frac{1}{32} \times 5 + \frac{1}{4} \times 2 = \frac{27}{8}.$$

La première méthode réutilise les entropies déjà calculées ; la seconde donne un contrôle direct de leur cohérence avec les seize cases.

- 13.** a) Question adaptée au programme actuel : les termes de probabilité jointe nulle sont explicitement définis comme nuls dans l'information mutuelle ; cela rend notamment la question 13.c) bien définie.

Le produit $a_k b_j$ et la probabilité q_{kj} sont symétriques lorsque l'on échange le rôle de X, k et de Y, j . En échangeant les indices de la somme finie, on obtient donc $I(X, Y) = I(Y, X)$.

- 13.** b) Pour chaque case de probabilité strictement positive,

$$q_{kj} \log_2 \left(\frac{q_{kj}}{a_k b_j} \right) = q_{kj} \log_2 q_{kj} - q_{kj} \log_2 a_k - q_{kj} \log_2 b_j.$$

Les cases nulles ne contribuent à aucun des trois termes. Après sommation et utilisation des marginales,

$$I(X, Y) = -H(X, Y) + H(X) + H(Y).$$

Comme $H(X, Y) = H(Y) + H(X | Y)$, on en déduit

$$I(X, Y) = H(X) - H(X | Y).$$

13. c) Conditionnellement à $X = k$, la variable X est certaine, égale à k ; son entropie conditionnelle vaut 0. Donc $H(X | X) = 0$ et 13.b) donne $I(X, X) = H(X)$.

13. d) Si X et Y sont indépendantes, $q_{kj} = a_k b_j$ pour tous les indices. Le quotient dans chaque logarithme vaut 1, donc chaque terme de la somme est nul. Ainsi $I(X, Y) = 0$.

14. a) Puisque $a_k > 0$, la somme est bien définie et

$$\sum_{j=0}^n p_j = \frac{1}{a_k} \sum_{j=0}^n q_{kj} = \frac{a_k}{a_k} = 1.$$

Les p_j sont donc les probabilités de la loi de Y sachant $X = k$.

14. b) Question adaptée au programme actuel.

Les x_j sont définis pour $j \in J_k$, puisque $q_{kj} = a_k p_j > 0$ sur cet ensemble. Ils sont strictement positifs. La définition $Z_k = x_J$ conserve les bonnes probabilités même si plusieurs valeurs x_j coïncident : le transfert est effectué sur J .

On calcule alors

$$E(Z_k) = \sum_{j \in J_k} p_j x_j = \sum_{j \in J_k} \frac{q_{kj}}{a_k} \frac{a_k b_j}{q_{kj}} = \sum_{j \in J_k} b_j \leq \sum_{j=0}^n b_j = 1.$$

Cette espérance est strictement positive et finie. Pour employer la concavité du logarithme avec le domaine de 8.d), on peut poser $a = \min_{j \in J_k} x_j > 0$ et appliquer cette question à $Z_k - a$ et à la fonction $u \mapsto \log_2(u + a)$, concave sur $\mathbb{R}_+$. Si Z_k est constante, l'égalité de Jensen est immédiate. Dans tous les cas,

$$E(\log_2 Z_k) \leq \log_2(E(Z_k)) \leq 0.$$

Dans le cas particulier où tous les p_j sont strictement positifs, $J_k = \{0, \dots, n\}$ et $E(Z_k) = 1$. La majoration précédente conserve donc le raisonnement d'origine tout en permettant les cases jointes nulles, notamment celles de la question 12.

14. c) Pour chaque indice k , l'expression obtenue à la question précédente s'écrit

$$E(\log_2 Z_k) = \sum_{j \in J_k} \frac{q_{kj}}{a_k} \log_2 \left(\frac{a_k b_j}{q_{kj}} \right) \leq 0.$$

On multiplie par $a_k > 0$ puis par -1 , ce qui donne

$$\sum_{j \in J_k} q_{kj} \log_2 \left(\frac{q_{kj}}{a_k b_j} \right) \geq 0.$$

En sommant sur k , on retrouve la définition de $I(X, Y)$; les cases hors de J_k contribuent 0 par convention. Donc $I(X, Y) \geq 0$ pour tout couple de lois marginales à support $\{0, \dots, n\}$, même lorsque certaines probabilités jointes sont nulles.

Relevé des adaptations

- 2. La convergence absolue est explicitée dans la définition de l'entropie différentielle afin de justifier l'espérance par transfert. La fonction logarithme de la densité est prolongée arbitrairement sur l'ensemble de probabilité nulle où la densité s'annule.
- 5. Le cas $\gamma = 0$ de la définition déjà donnée est explicité comme variable constante nulle, afin de ne pas supposer une loi normale non dégénérée pour la combinaison nulle.
- 7. Code adapté en Python : indices commençant à zéro et documentation de `len` et du logarithme. Le calcul conserve la somme et la base du logarithme de l'énoncé.

- 8. La définition de convexité employée dans la démonstration par récurrence est rappelée, sans hypothèse de dérivabilité supplémentaire.
- 11.a). La valeur $-\infty$ de la source sur les cases jointes nulles est remplacée par 0 pour définir une fonction réelle et appliquer le transfert. Les entropies restent identiques.
- 13. La convention attribuant la valeur 0 aux termes de probabilité jointe nulle est précisée pour rendre la formule définie, y compris pour $I(X, X)$.
- 14.b). La variable est définie par $Z_k = x_j$, car les x_j ne sont pas nécessairement distincts. Les termes sont regroupés lorsque plusieurs valeurs coïncident. La restriction aux probabilités conditionnelles strictement positives et la majoration $E(Z_k) \leq 1$ permettent de traiter les cases nulles et de conclure 14.c) dans toute la généralité annoncée.