ESSEC II 2019 - voie E

Correction proposée par G. Dupont.

Première partie : Entropie différentielle d'une variable aléatoire à densité.

1. (a) Soient $x, y \in \mathbf{R}_+^*$. Alors

$$\log_2(xy) = \frac{\ln(xy)}{\ln(2)} = \frac{\ln(x) + \ln(y)}{\ln(2)} = \frac{\ln(x)}{\ln(2)} + \frac{\ln(y)}{\ln(2)} = \log_2(x) + \log_2(y).$$

On a donc bien

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

(b) Soit $\alpha \in \mathbf{R}$. On a

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

On a donc bien

$$\boxed{\forall \alpha \in \mathbf{R}, \quad \log_2(2^\alpha) = \alpha.}$$

(c) $\log_2 : x \mapsto \dfrac{\ln(x)}{\ln(2)}$ est $\mathcal{C}^\infty$ sur $\mathbf{R}_+^*$ car ln l'est. En outre, pour tout $x \in \mathbf{R}_+^*$, on a

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

de sorte que

$$\boxed{\log_2 \text{ est concave.}}$$

2. D'après le théorème de transfert, sous réserve de convergence absolue des intégrales considérées, on a (avec la convention $u \ln(u) = 0$ quand $u = 0$),

$$\begin{aligned}
\mathbf{E}\left[\log_2(f(X))\right] &= \int_{-\infty}^{+\infty} \log_2(f(x)) f(x) \, dx \\
&= \int_a^b \log_2(f(x)) f(x) \, dx \\
&= -h(X).
\end{aligned}$$

Ainsi, on a

$$\boxed{h(X) = -\int_a^b \log_2(f(x)) f(x) \, dx.}$$

3. (a) i. Soit $x \in \mathbf{R}$. On a

$$\begin{aligned}
F_Y(x) &= \mathbf{P}\left(Y \leq x\right) \\
&= \mathbf{P}\left(X + c \leq x\right) \\
&= \mathbf{P}\left(X \leq x - c\right) \\
&= F_X(x - c).
\end{aligned}$$

Puisque X admet une densité, F_X est continue sur $\mathbf{R}$ et $\mathbf{C}^1$ sur $\mathbf{R}$ sauf éventuellement en un nombre fini de points. Par composition avec la fonction affine $x \mapsto x - c$, il en est de même de F_Y et donc Y admet une densité f_Y. En tout point x où F_Y est $\mathcal{C}^1$, on a

$$\boxed{f_Y(x) = F_Y'(x) = F_X'(x-c) = f(x-c).}$$

ii. On a

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

Puisque X admet une entropie différentielle, $\log_2(f(X))$ admet une espérance, donc $\log_2(f_Y(Y))$ admet une espérance et donc Y admet une entropie différentielle. En outre,

$$h(Y) = -\mathbf{E}\left[\log_2(f_Y(Y))\right] = -\mathbf{E}\left[\log_2(f(X))\right] = h(X).$$

En conclusion,

$$\boxed{\forall c \in \mathbf{R}, \quad h(X+c) = h(X).}$$

(b) i. Pour tout $x \in \mathbf{R}$, on a

$$\begin{aligned} F_Z(x) &= \mathbf{P}\left(Z \leq x\right) \\ &= \mathbf{P}\left(\alpha X \leq x\right) \\ &= \mathbf{P}\left(X \leq \frac{x}{\alpha}\right) \\ &= F_X\left(\frac{x}{\alpha}\right). \end{aligned}$$

Comme précédemment, Z admet une densité définie pour $x \in \mathbf{R}$ par

$$\boxed{f_Z(x) = \frac{1}{\alpha} f\left(\frac{x}{\alpha}\right).}$$

ii. On a

$$\log_2\left(f_Z(Z)\right) = \log_2\left(\frac{1}{\alpha} f\left(\frac{Z}{\alpha}\right)\right) = \log_2\left(\frac{1}{\alpha}\right) + \log_2\left(f(X)\right) = -\log_2(\alpha) + \log_2\left(f(X)\right).$$

Puisque $\log_2(f(X))$ admet une espérance, il s'ensuit que $\log_2(f_Z(Z))$ en admet aussi une et donc Z admet une entropie différentielle. Alors

$$\begin{aligned} h(Z) &= -\mathbf{E}\left[\log_2(f_Z(Z))\right] \\ &= -\mathbf{E}\left[-\log_2(\alpha) + \log_2\left(f(X)\right)\right] \\ &= \log_2(\alpha) + h(X). \end{aligned}$$

En conclusion,

$$\boxed{\forall \alpha > 0, \quad h(\alpha X) = h(X) + \log_2(\alpha).}$$

4. (a) Soit $a > 0$ et $X \rightsquigarrow \mathcal{U}([0,a])$.

i. Une densité de X est

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

ii. L'intégrale $\displaystyle\int_0^a \frac{1}{a} \log_2\left(\frac{1}{a}\right)\, dx$ converge absolument donc X admet une entropie différentielle et

$$\begin{aligned} h(X) &= -\mathbf{E}\left[\log_2(f_X(X))\right] \\ &= -\int_0^a \frac{1}{a} \log_2\left(\frac{1}{a}\right)\, dx \\ &= \log_2(a). \end{aligned}$$

Ainsi,

$$\boxed{h(X) = \log_2(a).}$$

iii. On a la chaîne d'équivalences :

$$\begin{aligned} h(X) > 0 \quad &\Leftrightarrow \quad \log_2(a) > 0 \\ &\Leftrightarrow \quad \frac{\ln(a)}{\ln(2)} > 0 \\ &\Leftrightarrow \quad \ln(a) > 0 \\ &\Leftrightarrow \quad a > 1. \end{aligned}$$

Ainsi,

$$\boxed{h(X) > 0 \quad \Leftrightarrow \quad a > 1.}$$

(b) Soit $Y \rightsquigarrow \mathcal{N}(0,1)$. Sous réserve de convergence, on a

$$\begin{aligned} h(Y) &= -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \log_2\left(\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right) \, dx \\ &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} \left(-\frac{1}{2}\log_2(2\pi) + \log_2(e^{-x^2/2})\right) \, dx \\ &= \frac{1}{2}\log_2(2\pi) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{x^2/2} \, dx + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{x^2}{2\ln(2)} e^{x^2/2} \, dx \\ &= \frac{1}{2}\log_2(2\pi) + \frac{1}{2\ln(2)} \mathbf{E}\left[Y^2\right] \\ &= \frac{1}{2}\left[\log_2(2\pi) + \frac{1}{\ln(2)}\right] \\ &= \frac{1}{2}\left[\log_2(2\pi) + \log_2(e)\right] \\ &= \frac{1}{2}\log_2(2\pi e). \end{aligned}$$

Les intégrales étant bien convergentes, Y admet une entropie différentielle et

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

(c) Soit $Z \rightsquigarrow \mathcal{E}(\lambda)$, avec $\lambda > 0$. Sous réserve de convergence, on a

$$\begin{aligned} h(Z) &= -\int_0^{+\infty} \lambda e^{-\lambda x} \log_2(\lambda e^{-\lambda x}) \, dx \\ &= -\int_0^{+\infty} \lambda e^{-\lambda x} \left[\log_2(\lambda) - \frac{\lambda x}{\ln 2}\right] \, dx \\ &= -\log_2(\lambda) \int_0^{+\infty} \lambda e^{-\lambda x} \, dx + \frac{\lambda}{\ln 2} \int_0^{+\infty} x\lambda e^{-\lambda x} \, dx \\ &= -\log_2(\lambda) + \frac{\lambda}{\ln 2} \mathbf{E}\left[Z\right] \\ &= -\log_2(\lambda) + \frac{1}{\ln 2} \\ &= \log_2\left(\frac{e}{\lambda}\right). \end{aligned}$$

Toutes les intégrales considérées étant convergentes, Z admet une entropie différentielle et

$$\boxed{h(Z) = \log_2\left(\frac{e}{\lambda}\right).}$$

(d) i. $f : x \mapsto \frac{1}{2}\lambda e^{-\lambda|x|}$ est continue sur $\mathbf{R}$ comme composée de fonctions continues sur $\mathbf{R}$ et à valeurs positives. En outre, f est une fonction paire donc, sous réserve de convergence,

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

la dernière égalité provenant du fait que l'on intègre la densité d'une loi exponentielle de paramètre λ sur $\mathbf{R}$.

Ainsi,

$$\boxed{f \text{ est une densité de probabilité sur } \mathbf{R}.}$$

Remarque : La loi correspondante est appelée *loi de Laplace* de paramètre $(0,1)$. Pour une étude générale des lois de Laplace, on pourra consulter le sujet d'ESSEC 2017 de la voie E.

ii. Sous réserve de convergence, on a

$$\begin{aligned} h(W) &= -\int_{-\infty}^{+\infty} \frac{1}{2}\lambda e^{-\lambda|x|}\log_2\left(\frac{1}{2}\lambda e^{-\lambda|x|}\right)\ dx \\ &= -\int_0^{+\infty} \lambda e^{-\lambda x}\log_2\left(\frac{1}{2}\lambda e^{-\lambda x}\right)\ dx \quad \text{(par parité)} \\ &= -\int_0^{+\infty} \lambda e^{-\lambda x}\log_2(\lambda e^{-\lambda x}) + \log_2(2)\int_0^{+\infty} \lambda e^{-\lambda x}\ dx \\ &= h(Z) + 1 \\ &= \log_2\left(\frac{e}{\lambda}\right) + 1 \end{aligned}$$

Toutes les intégrales considérées étant convergentes, W admet une entropie différentielle et

$$\boxed{h(W) = \log_2\left(\frac{e}{\lambda}\right) + 1.}$$

5. (a) On a $X = 1 \times X + 0 \times Y$ donc, le couple (X,Y) étant gaussien centré, X suit une loi normale centrée de variance égale à σ^2. Ainsi,

$$\boxed{X \rightsquigarrow \mathcal{N}(0,\sigma^2).}$$

(b) Si $Z \rightsquigarrow \mathcal{N}(0,1)$, X a la même loi que σZ. Ainsi,

$$\begin{aligned} h(X) &= h(\sigma Z) \\ &= h(Z) + \log_2(\sigma) \quad \text{(d'après 3.b)} \\ &= \frac{1}{2}\log_2(2\pi e) + \log_2(\sigma) \\ &= \frac{1}{2}\log_2(2\pi e\sigma^2). \end{aligned}$$

Ainsi,

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

(c) i. Pour tous $x, y \in \mathbf{R}$, on a

$$|xy| \leq \frac{x^2+y^2}{2}$$

donc $0 \leq |XY| \leq \frac{1}{2}(X^2+Y^2)$.

Puisque X^2 et Y^2 admettent une espérance, il s'ensuit que $|XY|$ admet une espérance et donc

$$\boxed{\mathbf{E}[XY] \text{ existe.}}$$

ii. Pour tout $\lambda \in \mathbf{R}$, $X + \lambda Y$ suit une loi normale centrée donc admet un moment d'ordre 2. En outre

$$\mathbf{E}\left[(X+\lambda Y)^2\right] = \mathbf{E}\left[X^2\right] + 2\lambda \mathbf{E}[XY] + \lambda^2 \mathbf{E}\left[Y^2\right]$$

et cette quantité est positive comme espérance d'une variable aléatoire positive.
Ainsi,

$$\boxed{\forall \lambda \in \mathbf{R}, \quad 0 \leq \mathbf{E}\left[X^2\right] + 2\lambda \mathbf{E}[XY] + \lambda^2 \mathbf{E}\left[Y^2\right].}$$

iii. X et Y étant fixées, considérons l'application définie sur $\mathbf{R}$ par :

$$P : \lambda \mapsto \mathbf{E}\left[X^2\right] + 2\lambda \mathbf{E}[XY] + \lambda^2 \mathbf{E}\left[Y^2\right].$$

C'est une fonction polynomiale de degré 2 en λ, de signe constant (positif) d'après la question précédente. Son discriminant $\Delta = 4\mathbf{E}[XY]^2 - 4\mathbf{E}\left[X^2\right]\mathbf{E}\left[Y^2\right]$ est donc négatif. Autrement dit,

$$4\mathbf{E}[XY]^2 - 4\mathbf{E}\left[X^2\right]\mathbf{E}\left[Y^2\right] \leq 0,$$

c'est-à-dire,

$$\boxed{\mathbf{E}[XY]^2 \leq \mathbf{E}\left[X^2\right]\mathbf{E}\left[Y^2\right].}$$

iv. Les variables aléatoires X et Y étant centrées, on a $\mathbf{E}\left[X^2\right] = \mathbf{V}(X) = \sigma^2 = \mathbf{V}(Y) = \mathbf{E}\left[Y^2\right]$ donc

$$\rho^2 = \frac{\mathbf{E}[XY]^2}{\sigma^4} = \frac{\mathbf{E}[XY]^2}{\mathbf{V}(X)\mathbf{V}(Y)} = \frac{\mathbf{E}[XY]^2}{\mathbf{E}[X^2]\mathbf{E}[Y^2]} \leq 1$$

et on a bien

$$\boxed{-1 \leq \rho \leq 1.}$$

v. Si X et Y sont indépendantes, on a $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$. Or X est centrée donc $\mathbf{E}[X] = 0$. Ainsi,

$$\boxed{\text{Si } X \text{ et } Y \text{ sont indépendantes, alors } \rho = 0.}$$

Remarque : Le candidat sagace aura reconnu le *coefficient de corrélation* qui est au programme d'ECE pour les couples de variables aléatoires *discrètes*. Ses propriétés se généralisent en fait aux variables aléatoires réelles quelconques.

(d) Supposons $|\rho| < 1$, où ρ est le nombre défini 5.c.iv.

i. On a

$$\begin{aligned} h(X,Y) = 0 \quad &\Leftrightarrow \quad \log_2(2\pi e\sigma^2\sqrt{1-\rho^2}) = 0 \\ &\Leftrightarrow \quad 2\pi e\sigma^2\sqrt{1-\rho^2} = 1. \end{aligned}$$

Ainsi,

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

ii. On a

$$\begin{aligned} I(X,Y) &= h(X) + h(Y) - h(X,Y) \\ &= \frac{1}{2}\log_2(2\pi e\sigma^2) + \frac{1}{2}\log_2(2\pi e\sigma^2) - \log_2(2\pi e\sigma^2\sqrt{1-\rho^2}) \quad \text{(d'après 5.b)} \\ &= \log_2(2\pi e\sigma^2) - \log_2(2\pi e\sigma^2\sqrt{1-\rho^2}) \end{aligned}$$

$$= \log_2\left(\frac{2\pi e\sigma^2}{2\pi e\sigma^2\sqrt{1-\rho^2}}\right)$$
$$= \log_2\left(\frac{1}{\sqrt{1-\rho^2}}\right).$$

Ainsi,

$$\boxed{I(X,Y) = \log_2\left(\frac{1}{\sqrt{1-\rho^2}}\right).}$$

iii. On a

$$\begin{aligned} |\rho| < 1 \quad &\Rightarrow \quad 0 \leq \rho^2 < 1 \\ &\Rightarrow \quad 0 < 1-\rho^2 \leq 1 \\ &\Rightarrow \quad 0 < \sqrt{1-\rho^2} \leq 1 \\ &\Rightarrow \quad \frac{1}{\sqrt{1-\rho^2}} \geq 1 \\ &\Rightarrow \quad \log_2\left(\frac{1}{\sqrt{1-\rho^2}}\right) \geq 0 \\ &\Rightarrow \quad I(X,Y) \geq 0. \end{aligned}$$

On a donc bien

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

iv. On a

$$\sqrt{1-\rho^2} \xrightarrow[\rho\to 1]{} 0^+$$

donc

$$\frac{1}{\sqrt{1-\rho^2}} \xrightarrow[\rho\to 1]{} +\infty$$

et donc, en passant au $\log_2$, on a

$$\boxed{I(X,Y) = \log_2\left(\frac{1}{\sqrt{1-\rho^2}}\right) \xrightarrow[\rho\to 1]{} +\infty.}$$

Deuxième partie : Généralités sur l'entropie des variables discrètes

6. (a) Soit

$$g : \left\{ \begin{array}{ccc} [\![0,n]\!] & \longrightarrow & \mathbf{R} \\ k & \mapsto & \log_2(\mathbf{P}(X=k)). \end{array} \right.$$

D'après le théorème de transfert, on a

$$\begin{aligned} \mathbf{E}[g(X)] &= \sum_{k=0}^{n} g(k)\mathbf{P}(X=k) \\ &= \sum_{k=0}^{n} \log_2(\mathbf{P}(X=k))\mathbf{P}(X=k) \\ &= -H(X) \end{aligned}$$

et donc

$$\boxed{H(X) = -\mathbf{E}\left[g(X)\right].}$$

(b) Pour tout $k \in [\![0, n]\!]$, on a

$$\begin{aligned} 0 < \mathbf{P}\left(X = k\right) \leq 1 \quad &\Rightarrow \quad g(k) = \log_2(\mathbf{P}\left(X = k\right)) \leq 0 \\ &\Rightarrow \quad \mathbf{P}\left(X = k\right)\log_2(\mathbf{P}\left(X = k\right)) \leq 0 \\ &\Rightarrow \quad \sum_{k=0}^{n} \mathbf{P}\left(X = k\right)\log_2(\mathbf{P}\left(X = k\right)) \leq 0 \end{aligned}$$

et donc

$$\boxed{H(X) = -\sum_{k=0}^{n} \mathbf{P}\left(X = k\right)\log_2(\mathbf{P}\left(X = k\right)) \geq 0.}$$

(c) Soit $p \in]0,1[$ et $X \rightsquigarrow \mathcal{B}(p)$.

i. On a

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

ii. ψ est $\mathcal{C}^2$ sur $]0,1[$ par composition et, pour tout $p \in]0,1[$, on a

$$\psi'(p) = \log_2(1-p) - \log_2(p)$$

et donc

$$\begin{aligned} \psi''(p) &= -\frac{1}{(1-p)\ln(2)} - \frac{1}{p\ln(2)} \\ &= -\frac{1}{\ln 2}\frac{p + (1-p)}{p(1-p)} \\ &= -\frac{1}{p(1-p)\ln 2} < 0. \end{aligned}$$

Ainsi, $\psi'' < 0$ et donc

$$\boxed{\psi \text{ est (strictement) concave sur }]0,1[.}$$

iii. ψ étant concave sur l'ouvert $]0,1[$, elle admettra un maximum en tout point où sa dérivée s'annule. Or, pour tout $p \in]0,1[$, on a

$$\begin{aligned} \psi'(p) = 0 \quad &\Leftrightarrow \quad \frac{1}{\ln(2)}\ln\left(\frac{1-p}{p}\right) = 0 \\ &\Leftrightarrow \quad \frac{1-p}{p} = 1 \\ &\Leftrightarrow \quad 1 - p = p \\ &\Leftrightarrow \quad p = \frac{1}{2}. \end{aligned}$$

Ainsi,

$$\boxed{\psi \text{ atteint son maximum en } p_0 = \frac{1}{2}.}$$

(d) On a

$$H(X) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{2}{8}\log_2\left(\frac{1}{8}\right)$$

$$
\begin{aligned}
&= -\frac{1}{2}\log_2(2^{-1}) - \frac{1}{4}\log_2(2^{-2}) - \frac{2}{8}\log_2(2^{-3}) \\
&= \frac{1}{2} + \frac{2}{4} + \frac{6}{8} \\
&= \frac{14}{8} \\
&= \frac{7}{4}.
\end{aligned}
$$

Ainsi,

$$\boxed{H(X) = \frac{7}{4}.}$$

7. On propose le programme suivant :

```
function h = Entropie(P)
        n = length(P)
        h = 0
        for k = 1:n
                h = h - P(k)*log(P(k))/log(2)
        end
endfunction
```

8. (a) Soit $N \geq 2$ et $i \in [\![1, N]\!]$. On a

$$\sum_{j=1}^{n} p_j = 1 \quad \Rightarrow \quad p_i = 1 - \sum_{\substack{j=1\\ j\neq i}}^{n} p_j$$

mais il existe au moins un $j \in [\![1, N]\!] \setminus \{i\}$ car $N \geq 2$ et chaque p_j est strictement positif par définition du support. Ainsi, $\displaystyle\sum_{\substack{j=1\\ j\neq i}}^{n} p_j > 0$ et donc $p_i < 1$.

Il s'ensuit que

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

(b) Si $N = 2$, on a

$$
\begin{aligned}
\mathbf{E}\left[\varphi(X)\right] &= p_1\varphi(x_1) + p_2\varphi(x_2) \\
&= p_1\varphi(x_1) + (1-p_1)\varphi(x_2) \\
&\geq \varphi(p_1x_1 + (1-p_1)x_2) \quad \text{(inégalité de convexité)} \\
&= \varphi(\mathbf{E}\left[X\right]).
\end{aligned}
$$

Ainsi,

$$\boxed{\mathcal{P}(2) \text{ est vraie.}}$$

(c) Soit $N \geq 3$ tel que $\mathcal{P}(N-1)$ est vraie. Pour chaque $i \in [\![1, N-1]\!]$, on pose $p'_i = \dfrac{p_i}{1-p_N}$.

i. On a

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

Par ailleurs, pour tout $i \in [\![1, N-1]\!]$, $p_i > 0$ et $1 - p_N = \sum_{j=1}^{N-1} p_j > p_i$ donc

$$\boxed{0 < p'_i < 1.}$$

ii. D'après le théorème de transfert et $\mathcal{P}(N-1)$ appliquée à Y, on a

$$\begin{aligned}\sum_{i=1}^{N-1} p'_i \varphi(x_i) &= \mathbf{E}[\varphi(Y)] \\ &\geq \varphi(\mathbf{E}[Y]) \\ &= \varphi\left(\sum_{i=1}^{N-1} p'_i x_i\right).\end{aligned}$$

Ainsi,

$$\boxed{\sum_{i=1}^{N-1} p'_i \varphi(x_i) \geq \varphi\left(\sum_{i=1}^{N-1} p'_i x_i\right).}$$

iii. On a

$$\begin{aligned}\mathbf{E}[\varphi(X)] &= \sum_{i=1}^{N} p_i \varphi(x_i) \\ &= p_N \varphi(x_N) + \sum_{i=1}^{N-1} p_i \varphi(x_i) \\ &= p_N \varphi(x_N) + (1-p_N) \sum_{i=1}^{N-1} p'_i \varphi(x_i) \\ &\geq p_N \varphi(x_N) + (1-p_N)\varphi\left(\sum_{i=1}^{N-1} p'_i x_i\right) \quad \text{(d'après 8.c.ii)} \\ &\geq \varphi\left(p_N x_N + (1-p_N)\sum_{i=1}^{N-1} p'_i x_i\right) \quad \text{(inégalité de convexité)} \\ &= \varphi\left(\sum_{i=1}^{N} p_i x_i\right) \\ &= \varphi(\mathbf{E}[X]).\end{aligned}$$

Ainsi,

$$\boxed{\varphi \text{ convexe} \quad \Rightarrow \quad \mathbf{E}[\varphi(X)] \geq \varphi(\mathbf{E}[X]).}$$

(d) Si φ est concave, $-\varphi$ est convexe et donc, d'après la propriété établie précédemment

$$\mathbf{E}[-\varphi(X)] \geq -\varphi(\mathbf{E}[X]),$$

soit, par linéarité de l'espérance,

$$-\mathbf{E}[\varphi(X)] \geq -\varphi(\mathbf{E}[X])$$

et donc

$$\boxed{\varphi \text{ concave} \quad \Rightarrow \quad \mathbf{E}[\varphi(X)] \leq \varphi(\mathbf{E}[X]).}$$

9. (a) D'après l'inégalité de Jensen et la concavité de $\log_2$, on a

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

mais

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

donc

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

et

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

(b) On a

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

On a donc bien

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

(c) On a

$$\begin{aligned}\log_2(n+1) - H(X) &= \sum_{k=0}^{n} p_k \log_2((n+1)p_k)\\ &= -\sum_{k=0}^{n} p_k \log_2\left(\frac{1}{(n+1)p_k}\right) \geq 0 \quad \text{(d'après 9.a)}\end{aligned}$$

donc

$$\boxed{\log_2(n+1) \geq H(X).}$$

(d) Si $X \rightsquigarrow \mathcal{U}([\![0,n]\!])$, on a

$$\boxed{H(X) = -\sum_{k=0}^{n} \frac{1}{n+1}\log_2\left(\frac{1}{n+1}\right) = \log_2(n+1).}$$

Remarque : Ceci montre que, parmi les lois supportées sur $[\![0,n]\!]$, la loi uniforme est celle qui a l'entropie maximale. En d'autres termes, c'est dans le cas de la loi uniforme que l'imprévisibilité est maximale.

10. Soit X, Y indépendantes et de même loi supportées sur $[\![0,n]\!]$.

(a) On a

$$(X = Y) = \bigcup_{k=0}^{n} (X = k, Y = k)$$

donc

$$\begin{aligned}
\mathbf{P}(X = Y) &= \mathbf{P}\left(\bigcup_{k=0}^{n} (X = k, Y = k)\right) \\
&= \sum_{k=0}^{n} \mathbf{P}(X = k, Y = k) \quad \text{(incompatibilité)} \\
&= \sum_{k=0}^{n} \mathbf{P}(X = k)\,\mathbf{P}(Y = k) \quad \text{(indépendance)} \\
&= \sum_{k=0}^{n} \mathbf{P}(X = k)^2 \quad \text{(même loi).}
\end{aligned}$$

Ainsi,

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

(b) Posons $Y = \log_2(v(X))$ et $\varphi : t \mapsto 2^t$ qui est une fonction convexe sur $\mathbf{R}$. Alors d'après l'inégalité de Jensen, on a

$$\varphi(\mathbf{E}[Y]) \leq \mathbf{E}[\varphi(Y)].$$

Autrement dit,

$$\boxed{2^{\mathbf{E}[\log_2(v(X))]} \leq \mathbf{E}\left[2^{\log_2(v(X))}\right] = \mathbf{E}[v(X)].}$$

(c) On a

$$\begin{aligned}
2^{-H(X)} &= 2^{\mathbf{E}[\log_2(v(X))]} \quad \text{(d'après 6.a)} \\
&\leq \mathbf{E}[v(X)] \quad \text{(d'après 10.b)} \\
&= \sum_{k=0}^{n} p_k v(k) \\
&= \sum_{k=0}^{n} p_k^2 \\
&= \mathbf{P}(X = Y) \quad \text{(d'après 10.a).}
\end{aligned}$$

Ainsi,

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

(d) Si on considère X, Y indépendantes de loi $\mathcal{U}([\![0, n]\!])$, on a d'après 9.d,

$$2^{-H(X)} = 2^{-\log_2(n+1)} = \frac{1}{n+1}$$

et

$$\mathbf{P}(X = Y) = \sum_{k=0}^{n} \left(\frac{1}{n+1}\right)^2 = \frac{1}{n+1}.$$

Ainsi,

$$\boxed{\text{Si } X, Y \rightsquigarrow \mathcal{U}([\![0, n]\!]) \text{ sont indépendantes, alors } 2^{-H(X)} = \mathbf{P}(X = Y).}$$

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

11. (a) D'après le théorème de transfert pour les couples des variables aléatoires discrètes, on a

$$\mathbf{E}\left[g(X,Y)\right] = \sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}\left(X=k, Y=j\right)\log_2(\mathbf{P}\left(X=k,Y=j\right)) = -H(X,Y).$$

Ainsi,

$$\boxed{H(X,Y) = -\mathbf{E}\left[g(X,Y)\right].}$$

(b) On a

$$\begin{aligned}
H(Y,X) &= -\sum_{j=0}^{n}\sum_{k=0}^{n} \mathbf{P}\left(Y=j, X=k\right)\log_2(\mathbf{P}\left(Y=j,X=k\right)) \\
&= -\sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}\left(X=k, Y=j\right)\log_2(\mathbf{P}\left(X=k,Y=j\right)) \\
&= H(X,Y).
\end{aligned}$$

Ainsi,

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

(c) On a

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

On a donc bien

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

(d) On a

$$\begin{aligned} H(Y) - H(Y|X) &= H(Y) - (H(X,Y) - H(X)) \\ &= H(Y) + H(X) - H(X,Y). \end{aligned}$$

Et, par symétrie du rôle de X et de Y et par symétrie de H, il vient

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

Ainsi,

$$\boxed{H(X) - H(X|Y) = H(Y) - H(Y|X).}$$

12. (a) Le calcul de la loi marginale se fait classiquement à l'aide de la formule des probabilités totales et on obtient

$$\begin{aligned} \mathbf{P}(X=0) &= \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = \frac{1}{2} \\ \mathbf{P}(X=1) &= \frac{1}{16} + \frac{1}{8} + \frac{1}{16} = \frac{1}{4} \\ \mathbf{P}(X=2) &= \frac{1}{32} + \frac{1}{32} + \frac{1}{16} = \frac{1}{8} \\ \mathbf{P}(X=3) &= \frac{1}{32} + \frac{1}{32} + \frac{1}{16} = \frac{1}{8}. \end{aligned}$$

Il suit alors des calculs effectués en 6.d que

$$\boxed{H(X) = \frac{7}{4}.}$$

(b) De la même manière, on obtient

$$\begin{aligned} \mathbf{P}(Y=0) &= \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{32} = \frac{1}{4} \\ \mathbf{P}(Y=1) &= \frac{1}{16} + \frac{1}{8} + \frac{1}{32} + \frac{1}{32} = \frac{1}{4} \\ \mathbf{P}(Y=2) &= \frac{4}{16} = \frac{1}{4} \\ \mathbf{P}(Y=3) &= \frac{1}{4}. \end{aligned}$$

Ainsi, $Y \rightsquigarrow \mathcal{U}([\![0,3]\!])$ et il suit alors de 9.d que

$$\boxed{H(Y) = \log_2(4) = 2.}$$

(c) Le calcul direct de $H(X|Y)$ est possible mais nous apparaît trop long. Nous nous proposons donc de passer par la formule établie en 11.c en calculant $H(X,Y)$. On a

$$\begin{aligned} H(X,Y) &= -\frac{1}{4}\log_2\left(\frac{1}{4}\right) - 2 \times \frac{1}{8}\log_2\left(\frac{1}{8}\right) - 6 \times \frac{1}{16}\log_2\left(\frac{1}{16}\right) - 4 \times \frac{1}{32}\log_2\left(\frac{1}{32}\right) \\ &= \frac{2}{4} + \frac{6}{8} + \frac{24}{16} + \frac{20}{32} \\ &= \frac{27}{8}. \end{aligned}$$

Ainsi,

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

(d) D'après 11.c, on a

$$H(Y|X) = H(X,Y) - H(X) = \frac{27}{8} - \frac{7}{4} = \frac{13}{8}.$$

Ainsi,

$$\boxed{H(Y|X) = \frac{13}{8}.}$$

(e) On a déjà vu en 12.c que

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

13. (a) On a

$$\begin{aligned}
I(Y,X) &= \sum_{j=0}^{n}\sum_{k=0}^{n} \mathbf{P}(Y=j, X=k)\log_2\left(\frac{\mathbf{P}(Y=j, X=k)}{\mathbf{P}(Y=j)\,\mathbf{P}(X=k)}\right)\\
&= \sum_{j=0}^{n}\sum_{k=0}^{n} \mathbf{P}(X=k, Y=j)\log_2\left(\frac{\mathbf{P}(X=k, Y=j)}{\mathbf{P}(X=k)\,\mathbf{P}(Y=j)}\right)\\
&= \sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}(X=k, Y=j)\log_2\left(\frac{\mathbf{P}(X=k, Y=j)}{\mathbf{P}(X=k)\,\mathbf{P}(Y=j)}\right)\\
&= I(X,Y).
\end{aligned}$$

On a donc bien

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

(b) On a

$$\begin{aligned}
H(X) - H(X|Y) &= -\sum_{k=0}^{n} \mathbf{P}(X=k)\log_2(\mathbf{P}(X=k))\\
&\quad + \sum_{j=0}^{n} \mathbf{P}(Y=j)\sum_{k=0}^{n} \mathbf{P}_{(Y=j)}(X=k)\log_2(\mathbf{P}_{(Y=j)}(X=k))\\
&= -\sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}(X=k, Y=j)\log_2(\mathbf{P}(X=k))\\
&\quad + \sum_{j=0}^{n}\sum_{k=0}^{n} \mathbf{P}(X=k, Y=j)\log_2(\mathbf{P}_{(Y=j)}(X=k))\\
&= \sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}(X=k, Y=j)\log_2\left(\frac{\mathbf{P}_{(Y=j)}(X=k)}{\mathbf{P}(X=k)}\right)\\
&= \sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}(X=k, Y=j)\log_2\left(\frac{\mathbf{P}(X=k, Y=j)}{\mathbf{P}(X=k)\,\mathbf{P}(Y=j)}\right)\\
&= I(X,Y).
\end{aligned}$$

On a donc bien

$$\boxed{H(X) - H(X|Y) = I(X,Y).}$$

(c) En appliquant la question précédente à X et X, on a

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

mais

$$\begin{aligned}
H(X|X) &= -\sum_{k=0}^{n} \mathbf{P}\left(X=k\right) \sum_{j=0}^{n} \mathbf{P}_{(X=k)}(X=j) \log_2(\mathbf{P}_{(X=k)}(X=j)) \\
&= -\sum_{k=0}^{n} \mathbf{P}\left(X=k\right) \mathbf{P}_{(X=k)}(X=k) \log_2(\mathbf{P}_{(X=k)}(X=k)) \\
&= -\sum_{k=0}^{n} \mathbf{P}\left(X=k\right) \log_2(1) \\
&= 0.
\end{aligned}$$

Ainsi,

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

(d) Si X et Y sont indépendantes, pour tous $j,k \in [\![0,n]\!]$, on a $\mathbf{P}\left(X=k, Y=j\right) = \mathbf{P}\left(X=k\right)\mathbf{P}\left(Y=j\right)$ donc

$$\log_2\left(\frac{\mathbf{P}\left(X=k, Y=j\right)}{\mathbf{P}\left(X=k\right)\mathbf{P}\left(Y=j\right)}\right) = 0$$

et donc

$$I(X,Y) = \sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}\left(X=k, Y=j\right) \log_2\left(\frac{\mathbf{P}\left(X=k, Y=j\right)}{\mathbf{P}\left(X=k\right)\mathbf{P}\left(Y=j\right)}\right) = 0.$$

En conclusion :

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

14. (a) $k \in [\![0,n]\!]$ étant fixé, on pose pour tout $j \in [\![0,n]\!]$,

$$p_j = \frac{\mathbf{P}\left(X=k, Y=j\right)}{\mathbf{P}\left(X=k\right)} > 0 \quad \text{et} \quad x_j = \frac{\mathbf{P}\left(X=k\right)\mathbf{P}\left(Y=j\right)}{\mathbf{P}\left(X=k, Y=j\right)}.$$

i. On a

$$\begin{aligned}
\sum_{j=0}^{n} p_j &= \sum_{j=0}^{n} \frac{\mathbf{P}\left(X=k, Y=j\right)}{\mathbf{P}\left(X=k\right)} \\
&= \sum_{j=0}^{n} \mathbf{P}_{(X=k)}(Y=j) \\
&= \sum_{j=0}^{n} \mathbf{P}\left(Y_{(X=k)} = j\right) \\
&= 1 \quad (\text{car } Y_{(X=k)}(\Omega) \subset [\![0,n]\!]),
\end{aligned}$$

où $Y_{(X=k)}$ désigne la variable aléatoire Y conditionnelle à l'événement $(X=k)$.
On a donc bien

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

ii. La fonction $\log_2$ est concave donc, d'après 8.d, on a

$$\mathbf{E}\left[\log_2(Z_k)\right] \leq \log_2(\mathbf{E}\left[Z_k\right]).$$

Or

$$\begin{aligned}
\mathbf{E}\left[Z_k\right] &= \sum_{j=0}^{n} x_j p_j \\
&= \sum_{j=0}^{n} \frac{\mathbf{P}\left(X=k\right)\mathbf{P}\left(Y=j\right)}{\mathbf{P}\left(X=k, Y=j\right)} \times \frac{\mathbf{P}\left(X=k, Y=j\right)}{\mathbf{P}\left(X=k\right)} \\
&= \sum_{j=0}^{n} \mathbf{P}\left(Y=j\right) \\
&= 1.
\end{aligned}$$

Ainsi,

$$\boxed{\mathbf{E}\left[\log_2(Z_k)\right] \leq \log_2(1) = 0.}$$

iii. On a

$$\begin{aligned}
\forall k \in [\![0,n]\!],\ \mathbf{E}\left[\log_2(Z_k)\right] \leq 0 \quad &\Rightarrow \quad \forall k \in [\![0,n]\!],\ \mathbf{P}\left(X=k\right)\mathbf{E}\left[\log_2(Z_k)\right] \leq 0 \\
&\Rightarrow \quad \sum_{k=0}^{n} \mathbf{P}\left(X=k\right)\mathbf{E}\left[\log_2(Z_k)\right] \leq 0 \\
&\Leftrightarrow \quad \sum_{k=0}^{n} \mathbf{P}\left(X=k\right) \sum_{j=0}^{n} \log_2(x_j)\mathbf{P}\left(Z_k = x_j\right) \leq 0 \\
&\Leftrightarrow \quad \sum_{k=0}^{n} \mathbf{P}\left(X=k\right) \sum_{j=0}^{n} \log_2(x_j) p_j \leq 0 \\
&\Leftrightarrow \quad \sum_{k=0}^{n} \mathbf{P}\left(X=k\right) \sum_{j=0}^{n} \log_2(x_j) \frac{\mathbf{P}\left(X=k, Y=j\right)}{\mathbf{P}\left(X=k\right)} \leq 0 \\
&\Leftrightarrow \quad \sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}\left(X=k, Y=j\right)\log_2(x_j) \leq 0 \\
&\Leftrightarrow \quad \sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}\left(X=k, Y=j\right)\log_2\left(\frac{\mathbf{P}\left(X=k\right)\mathbf{P}\left(Y=j\right)}{\mathbf{P}\left(X=k, Y=j\right)}\right) \leq 0 \\
&\Leftrightarrow \quad \sum_{k=0}^{n}\sum_{j=0}^{n} \mathbf{P}\left(X=k, Y=j\right)\log_2\left(\frac{\mathbf{P}\left(X=k, Y=j\right)}{\mathbf{P}\left(X=k\right)\mathbf{P}\left(Y=j\right)}\right) \geq 0 \\
&\Leftrightarrow \quad I(X,Y) \geq 0.
\end{aligned}$$

On a donc bien,

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

Remarque : En conclusion, pour des variables aléatoires à valeurs dans $[\![0,n]\!]$, on a montré que l'information mutuelle était toujours positive. Elle est donc minimale quand les deux variables aléatoires sont indépendantes (12.d).

Commentaire : Les notions étudiées dans ce problème relèvent de la *théorie de l'information* développée indépendamment par Ronald Aymler Fisher et Claude Shannon. Cette théorie, qui joue un rôle important en statistiques, fait régulièrement l'objet de problèmes de concours. Le candidat intéressé pourra par exemple se pencher sur le sujet de 2009 de l'épreuve d'ESSEC II de la voie E.