

CORRECTION DM6 FACULTATIF - MATHS 2 ECE 2019

Partie I - Entropie différentielle d'une variable à densité

1. (a) Soit $(x, y) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$. On a :

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

- (b) Soit $\alpha \in \mathbb{R}$. On a :

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

- (c) La fonction \log_2 est deux fois dérivable par opérations usuelles. De plus, on a pour $x \in \mathbb{R}_+^*$:

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

et :

$$\log_2''(x) = \underbrace{\frac{1}{\ln(2)}}_{>0} \times \underbrace{\frac{-1}{x^2}}_{<0}.$$

Donc pour tout $x \in \mathbb{R}_+^*$, $\log_2''(x) < 0$ et donc la fonction \log_2 est concave sur \mathbb{R}_+^* .

2. Comme \log_2 est continue et comme $f(x)$ est nulle en dehors de I , sous réserve de convergence absolue, on a d'après le théorème de transfert :

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

3. (a) i. Y est une transformée affine de X . On a directement pour tout $y \in \mathbb{R}$:

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

où f_Y désigne une densité de Y .

- ii. Faisons le changement de variable affine $x = y - c$.

L'application $y \mapsto y - c$ est strictement croissante et \mathcal{C}^1 . En posant $x = y - c$, on a $dx = dy$ et donc les intégrales suivantes ont la même nature et sont égales en cas de convergence :

$$\int_a^b f(x) \log_2(f(x))dx \quad \text{et} \quad \int_{a+c}^{b+c} f(y-x) \log_2(f(y-c))dy.$$

Or la première intégrale converge puisque X admet une entropie différentielle. Donc :

$$\begin{aligned} h(X) &= - \int_a^b f(x) \log_2(f(x))dx = - \int_{a+c}^{b+c} f(y-x) \log_2(f(y-c))dy \\ &= - \int_{a+c}^{b+c} f_Y(y) \log_2(f_Y(y))dy = h(Y). \end{aligned}$$

On reconnaît bien l'intégrale définissant l'entropie différentielle de Y (qui converge bien et donc qui existe) et on a $h(X) = h(Y)$.

- (b) i. Z est une transformée affine de X avec coefficient strictement positif. On a directement pour tout $z \in \mathbb{R}$:

$$f_Z(z) = \frac{1}{\alpha} f\left(\frac{z}{\alpha}\right)$$

où f_Z désigne une densité de Y .

ii. Faisons le changement de variable affine $x = \frac{z}{\alpha}$.

L'application $z \mapsto z/\alpha$ est strictement croissante et \mathcal{C}^1 . En posant $x = \frac{z}{\alpha}$, on a $dx = \frac{dz}{\alpha}$ et donc les intégrales suivantes ont la même nature et sont égales en cas de convergence :

$$\int_a^b f(x) \log_2(f(x)) dx \quad \text{et} \quad \int_{\alpha a}^{\alpha c} f\left(\frac{z}{\alpha}\right) \log_2\left(\left(\frac{z}{\alpha}\right)\right) \frac{dz}{\alpha}.$$

Or la première intégrale converge puisque X admet une entropie différentielle. Donc :

$$\begin{aligned} h(X) &= - \int_a^b f(x) \log_2(f(x)) dx = - \int_{\alpha a}^{\alpha c} f\left(\frac{z}{\alpha}\right) \log_2\left(\left(\frac{z}{\alpha}\right)\right) \frac{dz}{\alpha} \\ &= - \int_{\alpha a}^{\alpha c} \alpha f_Z(z) \log_2(\alpha f_Z(z)) \frac{dz}{\alpha} = - \int_{\alpha a}^{\alpha c} f_Z(z) (\log_2(\alpha) + \log_2(f_Z(z))) dz \\ &= - \int_{\alpha a}^{\alpha c} f_Z(z) \log_2(\alpha) dz - \int_{\alpha a}^{\alpha c} f_Z(z) \log_2(f_Z(z)) dz = - \log_2(\alpha) + h(Z). \end{aligned}$$

On reconnaît bien l'intégrale définissant l'entropie différentielle de Y (qui converge bien et donc qui existe) ainsi que l'espérance de $\log_2(\alpha)$ et on a donc $\boxed{h(Z) = h(X) + \log_2(\alpha)}$.

4. (a) i. Une densité de X est donnée par :

$$f_X : \begin{cases} \mathbb{R} & \rightarrow \mathbb{R}_+ \\ x & \mapsto \begin{cases} \frac{1}{a} & \text{si } x \in [0, a] \\ 0 & \text{si } x \notin [0, a] \end{cases} \end{cases}.$$

ii. L'intégrale $-\int_0^a f_X(x) \log_2(f_X(x)) dx$ converge puisque c'est une intégrale sur un segment d'une fonction continue. Donc $\boxed{X \text{ admet une entropie différentielle.}}$ De plus :

$$\begin{aligned} h(X) &= - \int_0^a f_X(x) \log_2(f_X(x)) dx \\ &= - \int_0^a \frac{1}{a} \log_2\left(\frac{1}{a}\right) dx \\ &= \frac{\log_2(a)}{a} \int_0^a 1 dx = \boxed{\log_2(a)}. \end{aligned}$$

iii. Soit $a > 0$. On a :

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

(b) Une densité de Y est donnée pour $y \in \mathbb{R}$ par :

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

On considère donc l'intégrale :

$$- \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \log_2\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\right) dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \left[\log_2(\sqrt{2\pi}) + \frac{y^2}{2 \ln(2)} \right] dy$$

Or $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$ et $\int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$ converge (le premier car f_Y est une densité et le second car

c'est le moment d'ordre 2). Par linéarité de l'intégrale, on a donc convergence et :

$$\begin{aligned} - \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \log_2 \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \right) dy &= \log_2(\sqrt{2\pi}) \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy}_{=1} + \frac{1}{2 \ln(2)} \underbrace{\int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy}_{=1} \\ &= \log_2(\sqrt{2\pi}) + \frac{1}{2 \ln(2)} = \frac{1}{2} \log_2(2\pi) + \frac{1}{2} \times \underbrace{\frac{\ln(e)}{\ln(2)}}_{=\log_2(e)} \\ &= \boxed{\frac{1}{2} \log_2(2\pi e)}. \end{aligned}$$

Donc Y admet une entropie différentielle et $h(Y) = \frac{1}{2} \log_2(2\pi e)$.

(c) Une densité de Z est donnée pour $z \in \mathbb{R}$ par :

$$f_Z(z) = \begin{cases} 0 & \text{si } z < 0 \\ \lambda e^{-\lambda z} & \text{si } z \geq 0 \end{cases}.$$

On considère donc l'intégrale :

$$- \int_0^{+\infty} \lambda e^{-\lambda z} \log_2(\lambda e^{-\lambda z}) dz = \int_0^{+\infty} \lambda e^{-\lambda z} \left[-\log_2(\lambda) + \frac{\lambda z}{\ln(2)} \right] dz$$

Comme dans la question précédente, on voit deux intégrales connues et convergente (intégrale de la densité et espérance). Donc par linéarité :

$$\begin{aligned} - \int_0^{+\infty} \lambda e^{-\lambda z} \log_2(\lambda e^{-\lambda z}) dz &= -\log_2(\lambda) \underbrace{\int_0^{+\infty} \lambda e^{-\lambda z} dz}_{=1} + \frac{\lambda}{\ln(2)} \underbrace{\int_0^{+\infty} z \lambda e^{-\lambda z} dz}_{=1/\lambda} \\ &= \log_2\left(\frac{1}{\lambda}\right) + \log_2(e) = \boxed{\log_2\left(\frac{e}{\lambda}\right)}. \end{aligned}$$

Et donc Z admet une entropie différentielle et $h(Z) = \log_2\left(\frac{e}{\lambda}\right)$.

(d) i. Vérifions que f est une densité.

- f est clairement positive sur \mathbb{R} .
- f est également continue par opérations usuelles (même la valeur absolue est bien continue).
- On a :

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

est bien convergente (c'est une des intégrales de références). On a même :

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

De même :

$$\int_{-\infty}^0 f(x) dx = \int_{-\infty}^0 \frac{1}{2} \lambda e^{-\lambda|x|} dx = \int_{-\infty}^0 \frac{1}{2} \lambda e^{\lambda x} dx = \frac{1}{2}.$$

On peut le montrer explicitement soit en passant par une primitive, soit par changement de variable $u = -x$.

Donc $\int_{-\infty}^{+\infty} f(x) dx$ converge et :

$$\boxed{\int_{-\infty}^{+\infty} f(x) dx = \frac{1}{2} + \frac{1}{2} = 1.}$$

Donc f est une densité de probabilité.

- ii. Le calcul est très similaire à la question précédente avec la loi exponentielle. On doit considérer l'intégrale :

$$-\int_{-\infty}^{+\infty} \frac{1}{2} \lambda e^{-\lambda|x|} \log_2 \left(\frac{1}{2} \lambda e^{-\lambda|x|} \right) dx$$

L'intégrande est visiblement paire. Il faut et il suffit donc de vérifier la convergence de $\int_0^{+\infty} \frac{1}{2} \lambda e^{-\lambda x} \log_2 \left(\frac{1}{2} \lambda e^{-\lambda x} \right) dx$. Or dans ce cas :

$$\begin{aligned} \int_0^{+\infty} \frac{1}{2} \lambda e^{-\lambda|x|} \log_2 \left(\frac{1}{2} \lambda e^{-\lambda|x|} \right) dx &= \int_0^{+\infty} \frac{1}{2} \lambda e^{-\lambda x} \log_2 \left(\lambda e^{-\lambda x} \right) dx + \int_0^{+\infty} \frac{1}{2} \lambda e^{-\lambda x} \log_2 \left(\frac{1}{2} \right) dx \\ &= -\frac{1}{2} \log_2 \frac{e}{\lambda} + \frac{1}{2} \log_2 \frac{1}{2} \end{aligned}$$

où le calcul est issu de la question précédente.

Par parité, W a donc bien une entropie différentielle et :

$$h(W) = 2 \times \left(\frac{1}{2} \log_2 \frac{e}{\lambda} - \frac{1}{2} \log_2 \frac{1}{2} \right) = \log_2 \left(\frac{e}{\lambda} \right) + 1.$$

5. (a) $\alpha X + \beta Y$ est une variable de loi normale centrée pour tout $\alpha, \beta \in \mathbb{R}$. En particulier, pour $\alpha = 1$ et $\beta = 0$, X est une variable de loi normale centrée.
- (b) Avec les notations de l'énoncé, on a $X = \sigma Z$ où Z est une variable de loi normale centrée réduite (transformation affine d'une loi normale). On a donc :

$$h(X) = h(Z) + \log_2(\sigma)$$

d'après une des questions précédentes. De plus, on a déjà calculer que $h(Z) = \frac{1}{2} \log_2(2\pi e)$. Donc :

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

- (c) i. Montrons l'existence par domination.

On a $(|X| - |Y|)^2 \geq 0$. Or $(|X| - |Y|)^2 = X^2 - 2|XY| + Y^2$. Donc :

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

Comme X et Y suivent des lois normales, elles admettent toutes deux des moments d'ordre 2 et donc $E\left(\frac{1}{2}(X^2 + Y^2)\right)$ existe. Par domination $E(XY)$ existe.

- ii. On procède de manière similaire, on a pour $\lambda \in \mathbb{R}$:

$$\begin{aligned} \lambda^2 E(Y^2) + 2\lambda E(XY) + E(X^2) &= E(\lambda^2 Y^2 + 2\lambda XY + X^2) \\ &= E((\lambda Y + X)^2) \end{aligned}$$

Or $(\lambda Y + X)^2 \geq 0$ donc :

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

- iii. Le polynôme précédent est un polynôme d'ordre 2 toujours positif. Nécessairement, il a donc au plus 1 racine (sinon, il serait strictement négatif entre ses racines).

Donc le discriminant associé est négatif ou nul. Calculons ce discriminant :

$$\Delta = (2E(XY))^2 - 4E(Y^2)E(X^2)$$

et donc on a $E(XY)^2 - E(X^2)E(Y^2) \leq 0$ c'est-à-dire :

$$E(XY)^2 \leq E(X^2)E(Y^2).$$

- iv. Comme X et Y sont centrées, on a $E(X) = E(Y) = 0$. D'après la formule de Huygens, on a ainsi $V(X) = E(X^2)$ et $V(Y) = E(Y^2)$. Or $V(X) = V(Y) = \sigma^2$ par hypothèse.

Donc :

$$E(XY)^2 \leq \sigma^4.$$

En passant à la racine, on obtient : $|E(XY)| \leq \sigma^2$ puis :

$$\boxed{|\rho| = \left| \frac{E(XY)}{\sigma^2} \right| \leq 1.}$$

$\boxed{\text{On a donc bien } \rho \in [-1, 1].}$

- v. Si X et Y sont indépendantes, on a $E(XY) = E(X)E(Y) = 0$. Et donc $\boxed{\rho = 0.}$

(d) i. On a :

$$\begin{aligned} h(X, Y) = 0 &\Leftrightarrow \log_2(2\pi e\sigma^2\sqrt{1-\rho^2}) = 0 \\ &\Leftrightarrow 2\pi e\sigma^2\sqrt{1-\rho^2} = 1 \\ &\Leftrightarrow \sqrt{1-\rho^2} = \frac{1}{2\pi e\sigma^2} \\ &\quad (\text{équivalent car } 1-\rho^2 > 0) \\ &\Leftrightarrow 1-\rho^2 = \frac{1}{4\pi^2 e^2 \sigma^4} \\ &\Leftrightarrow \rho^2 = 1 - \frac{1}{4\pi^2 e^2 \sigma^4} \\ &\Leftrightarrow \boxed{|\rho| = \sqrt{1 - \frac{1}{4\pi^2 e^2 \sigma^4}}.} \end{aligned}$$

ii. On a :

$$\begin{aligned} I(X, Y) &= h(X) + h(Y) - h(X, Y) \\ &= \frac{1}{2} \log_2(2\pi e\sigma^2) + \frac{1}{2} \log_2(2\pi e\sigma^2) - \log_2(2\pi e\sigma^2\sqrt{1-\rho^2}) \\ &= \log_2(2\pi e\sigma^2) - \log_2(2\pi e\sigma^2\sqrt{1-\rho^2}) \\ &= \log_2\left(\frac{2\pi e\sigma^2}{2\pi e\sigma^2\sqrt{1-\rho^2}}\right) \boxed{= -\log_2 \sqrt{1-\rho^2}.} \end{aligned}$$

- iii. Comme $\rho \in]-1, 1[$, on a $1-\rho^2 \in]0, 1[$. Ainsi, on a également $\sqrt{1-\rho^2} \in]0, 1[$ D'où $\log_2 \sqrt{1-\rho^2} \leq 0$.
Et enfin :

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

iv. On a :

$$\boxed{\lim_{\rho \rightarrow 1} I(X, Y) = \lim_{\rho \rightarrow 1} -\log_2 \underbrace{\sqrt{1-\rho^2}}_{\rightarrow 0} = +\infty.}$$

Partie II - Généralités sur l'entropie des variables discrètes

6. (a) D'après le théorème de transfert (pas de problème de convergence car la somme est finie), on a :

$$\begin{aligned} E(g(x)) &= \sum_{k=0}^n g(k)P(X = k) \\ &= \sum_{k=0}^n \log_2(P(X = k))P(X = k) \\ &= \boxed{-H(X).} \end{aligned}$$

(b) Pour tout $k \in \llbracket 0, n \rrbracket$, on a $0 < P(X = k) \leq 1$. Donc pour tout $k \in \llbracket 0, n \rrbracket$, $g(k) = \log_2(P(X = k)) \leq 0$.

Ainsi :

$$H(X) = - \sum_{k=0}^n \underbrace{P(X = k)}_{>0} \underbrace{\log_2(P(X = k))}_{\leq 0} \geq 0.$$

(c) i. On a :

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

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 ψ'' est négative et donc ψ est concave sur $]0, 1[$.

iii. On commence par résoudre pour $p \in]0, 1[$:

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

Comme ψ est concave, ψ' est décroissante et donc est positive avant $\frac{1}{2}$ et négative après.

En conséquence, ψ atteint un maximum local en $p_0 = \frac{1}{2}$. Et comme ψ est concave ce maximum est en fait global.

(d) On a :

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

7.

```
1 def Entropie(P):
    h = 0
    for k in range(len(P)):
        h = h - P[k]*np.log(P[k])/np.log(2)
5 return h
```

8. (a) Procédons par l'absurde. On suppose, quitte à renuméroter, que $P(X = x_1) = 1$. Montrons qu'il y a une absurdité.

Comme $N \geq 2$, on a $[X = x_2] \subset [X \neq x_1]$ donc :

$$P(X = x_2) \leq P(X \neq x_1) = 1 - P(X = x_1) = 0.$$

Or $P(X = x_2) \geq 0$ donc $P(X = x_2) = 0$. Mais c'est absurde car on a supposé que la loi de X était à support sur $\{x_1, x_2, \dots, x_N\}$ et donc $P(X = x_2) \neq 0$.

D'où $\forall i \in [1, N]$, $P(X = x_i) \neq 1$ et donc $P(X = x_i) < 1$.

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

$$\begin{aligned} E(\varphi(X)) &= p_1\varphi(x_1) + p_2\varphi(x_2) \\ &= p_1\varphi(x_1) + (1 - p_1)\varphi(x_2) \\ &\geq \varphi(p_1x_1 + (1 - p_1)x_2) \\ &\quad \text{(inégalité de convexité : le graphe de } \varphi \text{ est sous ses cordes)} \\ &\geq \varphi(E(X)). \end{aligned}$$

- (c) i. On a :

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

De plus, on sait que $p_i \neq 0$ pour tout $i \in [1, N - 1]$. Donc on a également $p'_i \neq 0$.

En outre, on a $1 - p_N = \sum_{j=1}^{N-1} p_j > p_i$ puisque $p_j > 0$ pour tout $j \neq i$. Donc $p'_i \frac{p_i}{1 - p_N} < 1$.

D'où :

$$0 < p'_i < 1.$$

- ii. D'après le théorème de transfert, on a $\sum_{i=1}^{N-1} p'_i\varphi(x_i) = E(\varphi(Y))$.

D'après la question précédente, on peut appliquer $\mathcal{P}(N - 1)$ à Y et donc : $E(\varphi(Y)) \geq \varphi(E(Y))$. Et donc :

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

- iii. On a d'après le théorème de transfert :

$$E(\varphi(X)) = \sum_{i=1}^N p_i\varphi(x_i)$$

puis :

$$\begin{aligned} E(\varphi(X)) &= p_N\varphi(x_N) + \sum_{i=1}^{N-1} p_i\varphi(x_i) \\ &= p_N\varphi(x_N) + (1 - p_N) \sum_{i=1}^{N-1} \underbrace{\frac{p_i}{1 - p_N}}_{=p'_i} \varphi(x_i) \\ &\geq p_N\varphi(x_N) + (1 - p_N)\varphi\left(\sum_{i=1}^{N-1} p_ix_i\right) \\ &\quad \text{(cf. question précédente)} \\ &\geq \varphi\left(p_Nx_N + (1 - p_N) \sum_{i=1}^{N-1} p'_ix_i\right) \\ &\quad \text{(inégalité de convexité)} \\ &\geq \varphi\left(\sum_{i=1}^N p_ix_i\right) \end{aligned}$$

et donc avec une dernière application du théorème de transfert :

$$\boxed{E(\varphi(X)) \geq \varphi(E(X))}.$$

Cela conclut l'hérédité de la récurrence. La question 8b a fait l'initialisation, donc $\mathcal{P}(N)$ est vraie pour tout $N \geq 2$ (et techniquement trivialement vraie pour $N = 1$).

(d) Si φ est concave, alors $-\varphi$ est convexe et donc :

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

et donc par linéarité de l'espérance $-E(\varphi(X)) \geq -\varphi(E(X))$ et ainsi :

$$\boxed{E(\varphi(X)) \leq \varphi(E(X))}.$$

9. (a) D'après l'inégalité de Jensen que l'on vient de montrer, on a :

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

puisque \log_2 est concave. De plus :

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

Donc $\log_2 \left(\sum_{k=0}^n p_k \frac{1}{(n+1)p_k} \right) = 0$. D'où :

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

(b) On a :

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

(c) D'après la question précédente : $\log_2(n+1) - H(X) = \sum_{k=0}^n \log_2((n+1)p_k)$. Or :

$$\sum_{k=0}^n \log_2((n+1)p_k) = - \underbrace{\sum_{k=0}^n \log_2 \frac{1}{(n+1)p_k}}_{\leq 0} \geq 0.$$

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

(d) Si $X \hookrightarrow \mathcal{U}(\llbracket 0, n \rrbracket)$, on a :

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

Cela prouve en particulier que l'entropie est maximale pour la loi uniforme.

10. (a) On utilise la formule des probabilités totales avec le système complet d'événements $\{[Y = k]\}_{k \in [0, n]}$:

$$\begin{aligned}
 P(X = Y) &= \sum_{k=0}^n P([X = Y] \cap [Y = k]) \\
 &= \sum_{k=0}^n P([X = k] \cap [Y = k]) \\
 &= \sum_{k=0}^n P(X = k) \underbrace{P(Y = k)}_{=P(X=k)} \text{ (indépendance)} \\
 &= \boxed{\sum_{k=0}^n (P(X = k))^2}.
 \end{aligned}$$

- (b) Posons $Y = \log_2(v(X))$. C'est bien une variable aléatoire puisque X est discrète. On pose $\varphi : t \mapsto 2^t$ qui est convexe sur \mathbb{R} . D'après l'inégalité de Jensen, on a :

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

Et donc :

$$\boxed{2^{E(\log_2(v(X)))} \leq E(\underbrace{2^{\log_2(v(X))}}_{=v(X)})}.$$

- (c) On a :

$$\begin{aligned}
 2^{-H(X)} &= 2^{E(\log_2(v(X)))} \\
 &\leq E(v(X)) \\
 &\leq \sum_{k=0}^n p_k v(k) \\
 &= \sum_{k=0}^n p_k^2 \boxed{= P(X = Y)}.
 \end{aligned}$$

- (d) Considérons X et Y indépendantes de loi $\mathcal{U}([0, n])$. On a :

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

Puis :

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

Ainsi, on a bien $\boxed{2^{-H(X)} = P(X = Y)}$.

Partie III - Entropie jointe et information mutuelle de deux variables discrètes

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

$$\begin{aligned}
 E(g(X, Y)) &= \sum_{k=0}^n \sum_{j=0}^n g(k, j) P([X = k] \cap [Y = j]) \\
 &= \sum_{k=0}^n \sum_{j=0}^n \log_2(P[X = k] \cap [Y = j]) P([X = k] \cap [Y = j]) \\
 &= \boxed{-H(X, Y)}.
 \end{aligned}$$

(b) On a :

$$\begin{aligned}
 H(X, Y) &= - \sum_{k=0}^n \sum_{j=0}^n \log_2(P[X = k] \cap [Y = j]) P([X = k] \cap [Y = j]) \\
 &= - \sum_{j'=0}^n \sum_{k'=0}^n \log_2(P[X = j'] \cap [Y = k']) P([X = j'] \cap [Y = k']) \quad (\text{changement } k = j' \text{ et } j = k') \\
 &= - \sum_{k'=0}^n \sum_{j'=0}^n \log_2(P[X = j'] \cap [Y = k']) P([X = j'] \cap [Y = k']) \quad (\text{interversion des sommes}) \\
 &= - \sum_{k'=0}^n \sum_{j'=0}^n \log_2([Y = k'] \cap P[X = j']) P([Y = k'] \cap [X = j']) \quad (\text{permutation des intersections}) \\
 &= \boxed{H(Y, X)}.
 \end{aligned}$$

(c) On a :

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

Concentrons-nous sur le terme dans la grande parenthèse. Pour $k \in \llbracket 0, n \rrbracket$, on a :

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

La dernière somme se simplifie :

$$\begin{aligned}
 \sum_{j=0}^n \frac{P([Y = j] \cap [X = k])}{P(X = k)} \log_2 P(X = k) &= \frac{\log_2 P(X = k)}{P(X = k)} \underbrace{\sum_{j=0}^n P([Y = j] \cap [X = k])}_{=P(X=k)} \\
 &= \log_2(P(X = k))
 \end{aligned}$$

où on a utilisé la formule des probabilités totales pour la simplification. D'où :

$$\begin{aligned}
 &\log_2(P(X = k)) + \sum_{j=0}^n \frac{P([Y = j] \cap [X = k])}{P(X = k)} \log_2 \frac{P([Y = j] \cap [X = k])}{P(X = k)} \\
 &= \sum_{j=0}^n \frac{P([Y = j] \cap [X = k])}{P(X = k)} \log_2 P([Y = j] \cap [X = k]).
 \end{aligned}$$

On a donc :

$$\begin{aligned}
 & H(X) + H(Y/X) \\
 = & - \sum_{k=0}^n P(X = k) \sum_{j=0}^n \frac{P([Y = j] \cap [X = k])}{P(X = k)} \log_2 P([Y = j] \cap [X = k]) \\
 = & - \sum_{k=0}^n \sum_{j=0}^n P(X = k) \frac{P([Y = j] \cap [X = k])}{P(X = k)} \log_2 P([Y = j] \cap [X = k]) \\
 = & - \sum_{k=0}^n \sum_{j=0}^n P([Y = j] \cap [X = k]) \log_2 P([Y = j] \cap [X = k]) \\
 = & \boxed{H(X, Y)}.
 \end{aligned}$$

(d) On montre de même que :

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

Donc :

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

En passant les termes des bons côtés, on obtient :

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

12. (a) La loi (marginale) de X est la donnée pour tout $k \in \llbracket 0, 3 \rrbracket$ de $P(X = k)$. Or on a :

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

en utilisant la formule des probabilités totales avec le système complet d'événements $\{[Y = j]\}_{j \in \llbracket 0, 3 \rrbracket}$.
On a donc :

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

On en déduit :

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

puisque c'est exactement le même calcul qu'à la question 6d.

(b) On a de manière similaire :

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

Y suit une loi uniforme, son entropie est donc maximale. On a $\boxed{H(Y) = \log_2(4) = 2}$.

(c) C'est un peu rébarbatif mais on le fait. On a d'abord :

$$\begin{aligned}
 H(X/Y=0) &= -\frac{4}{8} \log_2 \frac{4}{8} - \frac{4}{16} \log_2 \frac{4}{16} - \frac{4}{32} \log_2 \frac{4}{32} - \frac{4}{32} \log_2 \frac{4}{32} \boxed{= \frac{7}{4}}, \\
 H(X/Y=1) &= -\frac{4}{16} \log_2 \frac{4}{16} - \frac{4}{8} \log_2 \frac{4}{8} - \frac{4}{32} \log_2 \frac{4}{32} - \frac{4}{32} \log_2 \frac{4}{32} \boxed{= \frac{7}{4}}, \\
 H(X/Y=2) &= -\frac{4}{16} \log_2 \frac{4}{16} - \frac{4}{16} \log_2 \frac{4}{16} - \frac{4}{16} \log_2 \frac{4}{16} - \frac{4}{16} \log_2 \frac{4}{16} \boxed{= 2}, \\
 H(X/Y=3) &= -\frac{4}{4} \log_2 \frac{4}{4} + 0 + 0 + 0 \boxed{= 0}.
 \end{aligned}$$

Puis :

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

(d) On en déduit :

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

(e) Et on obtient également :

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

13. (a) On a :

$$\begin{aligned}
 I(X, Y) &= \sum_{k=0}^n \sum_{j=0}^n P([X = k] \cap [Y = j]) \log_2 \frac{P([X = k] \cap [Y = j])}{P(X = k)P(Y = j)} \\
 &= \sum_{j'=0}^n \sum_{k'=0}^n P([X = j'] \cap [Y = k']) \log_2 \frac{P([X = j'] \cap [Y = k'])}{P(X = j')P(Y = k')} \\
 &\quad \text{(changement } k = j' \text{ et } j = k') \\
 &= \sum_{k'=0}^n \sum_{j'=0}^n P([X = j'] \cap [Y = k']) \log_2 \frac{P([X = j'] \cap [Y = k'])}{P(X = j')P(Y = k')} \\
 &\quad \text{(permutation des sommes)} \\
 &= \sum_{k'=0}^n \sum_{j'=0}^n P([Y = k'] \cap [X = j']) \log_2 \frac{P([Y = k'] \cap [X = j'])}{P(Y = k')P(X = j')} \\
 &\quad \text{(permutation des intersections)} \\
 &= \boxed{I(Y, X)}.
 \end{aligned}$$

(b) On a :

$$\begin{aligned}
 &H(X) - H(X/Y) \\
 &= -\sum_{k=0}^n \underbrace{P(X = k)}_{=\sum_{j=0}^n P([X=k] \cap [Y=j])} \log_2(P(X = k)) + \sum_{j=0}^n P(Y = j) \sum_{k=0}^n P_{[Y=j]}(X = k) \log_2(P_{[Y=j]}(X = k)) \\
 &= -\sum_{k=0}^n \sum_{j=0}^n P([X = k] \cap [Y = j]) \log_2(P(X = k)) + \sum_{j=0}^n \sum_{k=0}^n \underbrace{P(Y = j)P_{[Y=j]}(X = k)}_{=P([X=k] \cap [Y=j])} \log_2(P_{[Y=j]}(X = k)) \\
 &= \sum_{k=0}^n \sum_{j=0}^n P([X = k] \cap [Y = j]) \log_2 \left(\frac{P_{[Y=j]}(X = k)}{P(X = k)} \right) \\
 &= \sum_{k=0}^n \sum_{j=0}^n P([X = k] \cap [Y = j]) \log_2 \left(\frac{P([Y = j] \cap [X = k])}{P(Y = j)P(X = k)} \right) \\
 &= \boxed{I(X, Y)}.
 \end{aligned}$$

- (c) Il suffit de remarquer que $P_{[X=k]}(X = j) = 0$ si $j \neq k$ et $P_{[X=k]}(X = j) = 1$ si $j = k$. Donc avec la convention $0 \log_2 0 = 0$, et comme $1 \log_2 1 = 0$, on a :

$$H(X/X = k) = - \sum_{j=0}^n P_{[X=k]}(X = j) \log_2(P_{[X=k]}(X = j)) = 0.$$

Donc $H(X/X) = \sum_{k=0}^n P(X = k)H(X/X = k) = 0$. Puis :

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

- (d) Si X et Y sont indépendantes, on a :

$$\log_2 \frac{P([X = k] \cap [Y = j])}{P(X = k)P(Y = j)} = \log_2 \underbrace{\frac{P(X = k)P(Y = j)}{P(X = k)P(Y = j)}}_{=1} = 0.$$

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

14. (a) Le système $\{[Y = j]\}_{j \in \llbracket 0, n \rrbracket}$ est un système complet d'événements. Donc, d'après la formule des probabilités totales pour tout $k \in \llbracket 0, n \rrbracket$:

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

d'où :

$$\begin{aligned} \sum_{j=0}^n p_j &= \sum_{j=0}^n \frac{P([X = k] \cap [Y = j])}{P(X = k)} \\ &= \frac{1}{P(X = k)} \underbrace{\sum_{j=0}^n P([X = k] \cap [Y = j])}_{=P(X=k)} \\ &= \boxed{1}. \end{aligned}$$

- (b) C'est une application de l'inégalité de Jensen. En effet, \log_2 est concave et donc :

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

Or on a :

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

Donc :

$$\boxed{E(\log_2(Z_k)) \leq \log_2(1) = 0}.$$

- (c) Commençons par remarquer que toute la construction précédente dépend d'un entier $k \in \llbracket 0, n \rrbracket$. On a donc en fait $n + 1$ inégalités $E(\log_2(Z_k)) \leq 0$.

Or les termes x_j qui forment le support de Z_k ressemblent fortement à l'argument du logarithme dans la formule de $I(X, Y)$. L'espérance de Z_k donne une somme sur j . Cela pousse donc à resommer sur k pour obtenir une formule apparentée à celle de $I(X, Y)$.

Après un peu de recherche au brouillon, on calcule :

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

Or $\sum_{k=0}^n \underbrace{P(X = k)}_{\geq 0} \underbrace{E(\log_2(Z_k))}_{\leq 0} \leq 0$. Donc :

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