Théorie de l'Information – Solutions feuille de TD2

25/09/2023

Retrouvez le sujet du TD et d'autres exercices à l'adresse :

`https://lvzl.fr/teaching/2023-24/ti.html`

($\star$) exercice fondamental ($\star\star$) pour s'entraîner ($\star\star\star$) pour aller plus loin sur machine

Exercice 1. ($\star$) Information mutuelle.

Question 1.– Lors d'un lancer de pièce (supposée équilibrée), quelle est l'information mutuelle entre chacune des deux faces de la pièce ?

Question 2.– Lors d'un lancer de dé à 6 faces équilibré, quelle est l'information mutuelle entre deux faces opposées du dé ? Et entre deux faces adjacentes ?

Question 3.– Dans une urne contenant $n = 4$ boules noires et $r = 2$ boules rouges, on effectue deux tirages consécutifs et sans remise d'une boule. On note T_1 le premier tirage et T_2 le second tirage. Calculer l'information mutuelle entre ces deux tirages.

Solutions de l'Exercice 1.

Solution Q1. Soient X et Y les variables aléatoires correspondant aux deux faces d'une pièce. On a la loi conjointe suivante :

p_{XY}	$X = \text{pile}$	$X = \text{face}$
$Y = \text{pile}$	0	$1/2$
$Y = \text{face}$	$1/2$	0

Donc, $H(X) = H(Y) = H(XY) = h(1/2) = 1$. Par conséquent, $I(X;Y) = 1 = H(X)$, autrement dit $H(X \mid Y) = 0$. C'est compréhensible : la valeur de X dépend de manière déterministe de celle de Y. Si X est pile, alors Y est face, et inversement.

Solution Q2. Soit X la variable aléatoire correspondant à la face du dessus du dé, Y celle du dessous, et Z l'une des faces latérales (quelconque).

Sur un dé classique, on sait que X et Y sont liées par la formule $Y = 7 - X$. Ainsi, on a

$$I(X\,;Y) = H(X) - H(X \mid Y) = H(X) - 0 = \log_2(6)\,.$$

Concernant l'information mutuelle entre X et Z, une première méthode consiste à calculer les probabilité conjointes. Par symétrie du dé, on a :

$$\begin{aligned} I(X\,;Z) &= \sum_{x=1}^{6}\sum_{z=1}^{6} p_{X,Z}(x,z)\log_2\frac{p_{X,Z}(x,z)}{p_X(x)p_Z(z)} \\ &= 6 \times \sum_{z=1}^{6} p_{X,Z}(x=1,z)\log_2\frac{p_{X,Z}(x=1,z)}{p_X(1)p_Z(z)}\,. \end{aligned}$$

Or, $p(x = 1{,}z)$ est différent de zéro pour $z = 2{,}3{,}4{,}5$, et vaut dans ce cas $1/24$. Une nouvelle fois, par symétrie du dé on peut considérer ces 4 cas simultanément et on a donc :

$$I(X\,;Z) = 6 \times 4 \times \frac{1}{24} \times \log_2\frac{\frac{1}{24}}{\frac{1}{6}\cdot\frac{1}{6}} = \log_2\frac{3}{2} = \log_2(3) - 1\,.$$

Pour obtenir ce résultat sans calcul, on peut également observer que $Z \mid X$ prend n'importe quelle des 4 valeurs autres que X et Y, c'est-à-dire que $Z \mid X$ est tirée uniformément dans l'ensemble $\{1, \ldots, 6\} \setminus \{X, 7-X\}$. On a donc :

$$I(X\,;Z) = I(Z\,;X) = H(Z) - H(Z\,|\,X) = \log_2(6) - \log_2(4) = \log_2(3) - 1\,.$$

Solution Q3. Écrivons la table de probabilité :

	$T_1 = N$	$T_1 = R$
$T_2 = N$	$2/5$	$4/15$
$T_2 = R$	$4/15$	$1/15$

On a alors

$$H(T_1) = H(T_2) = h(2/3) = \frac{2}{3}\log_2\frac{3}{2} + \frac{1}{3}\log_2 3 = \log_2(3) - \frac{2}{3}\,.$$

où $h(\cdot)$ est la fonction d'entropie binaire. Par ailleurs

$$H(T_1, T_2) = \frac{2}{5}\log_2\frac{5}{2} + 2 \times \frac{4}{15}\log_2\frac{15}{4} + \frac{1}{15}\log_2 15 = \log_2(5) + \frac{3}{5}\log_2(3) - \frac{22}{15}\,.$$

Finalement,

$$I(T_1\,;T_2) = H(T_1) + H(T_2) - H(T_1, T_2) = -\log_2(5) + \frac{7}{5}\log_2(3) + \frac{2}{15}\,.$$

Exercice 2. ($\star$) Information mutuelle et non-création d'information.

Soient X, Y deux variables aléatoires à valeurs respectives dans $\mathcal{X}$ et $\mathcal{Y}$. On considère également $f : \mathcal{X} \to \mathcal{X}$ et $g : \mathcal{Y} \to \mathcal{Y}$ deux applications.

Question 1.– Démontrer que $I(X\,;f(X)) = H(f(X))$.

Question 2.– Démontrer que $I(X\,;Y) \leq I(X\,;g(Y))$.

Question 3.– En déduire que $I(f(X)\,;g(Y)) \leq I(X;Y)$.

Solutions de l'Exercice 2.

Solution Q1. Par le principe de non-création d'information, on a $H(f(X) \mid X) = 0$. L'information mutuelle est symétrique, par conséquent

$$I(X\,;\,f(X)) = I(f(X)\,;\,X) = H(f(X)) - H(f(X) \mid X) = H(f(X))\,.$$

Solution Q2. Observons d'abord que, d'après le principe de non-création d'information,

$$H(X \mid g(Y)) = H(X, g(Y)) - H(X) \leq H(X, Y) - H(X) = H(X \mid Y)$$

Ainsi, on obtient :

$$I(X\,;g(Y)) = H(X) - H(X \mid g(Y)) \geq H(X) - H(X \mid Y) = I(X\,,Y)\,.$$

Solution Q3. On utilise deux fois le résultat de la question précédente, et la symétrie de l'information mutuelle :

$$I(f(X)\,;g(Y)) \leq I(f(X)\,;Y) \leq I(X\,;Y)$$

Exercice 3. ($\star\star$) Entropie maximale à espérance fixée.

Soit Γ la loi géométrique de paramètre $\gamma > 0$. On rappelle que Γ est une variable aléatoire discrète (mais non finie) à valeurs entières, et qu'elle est donnée par

$$p(\Gamma = n) := (1-\gamma)\gamma^n, \quad n \in \mathbb{N}\,.$$

Question 1.– Calculer l'espérance de Γ.

Question 2.– Calculer $H(\Gamma)$.

Question 3.– Démontrer que toute variable aléatoire X, à valeurs dans $\mathbb{N}$ et d'espérance $\mu = \mathbb{E}(\Gamma)$, vérifie :

$$H(X) \leq H(\Gamma)\,.$$

Indication : on peut établir la relation $H(\Gamma) - H(X) = D_{\mathrm{KL}}(p_X \,\|\, p_\Gamma)$.

Solutions de l'Exercice 3.

Solution Q1. Ce calcul est classique. Par exemple, en voici un déroulement :

$$\begin{aligned}
\mathbb{E}(\Gamma) &= \sum_{n\geq 0} n(1-\gamma)\gamma^n \\
&= \sum_{n\geq 0} n\gamma^n - \sum_{n\geq 0} n\gamma^{n+1} \\
&= \sum_{n\geq 0} n\gamma^n - \sum_{n\geq 0} (n+1)\gamma^{n+1} + \sum_{n\geq 0} \gamma^{n+1} \\
&= \sum_{n\geq 0} n\gamma^n - \sum_{n\geq 1} n\gamma^n + \frac{1}{1-\gamma} - 1 \\
&= 0 + \frac{\gamma}{1-\gamma} = \frac{\gamma}{1-\gamma}
\end{aligned}$$

Remarque : on peut également s'intéresser à la série entière $f(\gamma) = \sum_{n\geq 0} \gamma^n = \frac{1}{1-\gamma}$. Son rayon de convergence est 1 ; en la dérivant on obtient $f'(\gamma) = \sum_{n\geq 1} n\gamma^{n-1}$ d'une part, et $f'(\gamma) = \frac{1}{(1-\gamma)^2}$ d'autre part. On en déduit que

$$\mathbb{E}(\Gamma) = \sum_{n\geq 0} n(1-\gamma)\gamma^n = (1-\gamma)\gamma f'(\gamma) = \frac{\gamma}{1-\gamma}\,.$$

Solution Q2. On a :

$$\begin{aligned}
H(\Gamma) &= -\sum_{n\in\mathbb{N}} (1-\gamma)\gamma^n \log_2((1-\gamma)\gamma^n) \\
&= (1-\gamma)\log_2\left(\frac{1}{1-\gamma}\right)\sum_{n\in\mathbb{N}} \gamma^n + \log_2\left(\frac{1}{\gamma}\right)\sum_{n\in\mathbb{N}} n(1-\gamma)\gamma^n \\
&= \log_2\left(\frac{1}{1-\gamma}\right) + \log_2\left(\frac{1}{\gamma}\right)\mathbb{E}(\Gamma)
\end{aligned}$$

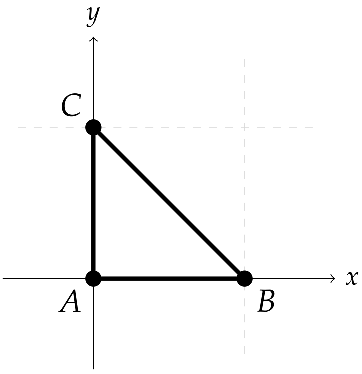
Solution Q3. Soit X une variable aléatoire à valeurs entières et d'espérance égale à $\mathbb{E}(\Gamma)$, et $(p_n)_{n\in\mathbb{N}}$ sa distribution. Alors :

$$\begin{aligned}
D_{\mathrm{KL}}(p_X \,\|\, p_\Gamma) &= \sum_{n\in\mathbb{N}} p_n \log_2\left(\frac{p_n}{(1-\gamma)\gamma^n}\right) \\
&= \sum_{n\in\mathbb{N}} p_n \left(\log_2\left(\frac{1}{1-\gamma}\right) + n\log_2\left(\frac{1}{\gamma}\right) + \log_2(p_n)\right) \\
&= \log_2\left(\frac{1}{1-\gamma}\right) \times \left(\sum_{n\in\mathbb{N}} p_n\right) + \log_2\left(\frac{1}{\gamma}\right) \times \left(\sum_{n\in\mathbb{N}} np_n\right) + \sum_{n\in\mathbb{N}} p_n \log_2 p_n \\
&= \log_2\left(\frac{1}{1-\gamma}\right) \times 1 + \log_2\left(\frac{1}{\gamma}\right) \times \mathbb{E}(X) + \sum_{n\in\mathbb{N}} p_n \log_2 p_n \\
&= \log_2\left(\frac{1}{1-\gamma}\right) + \log_2\left(\frac{1}{\gamma}\right)\mathbb{E}(\Gamma) + \sum_{n\in\mathbb{N}} p_n \log_2 p_n \\
&= H(\Gamma) - H(X)\,.
\end{aligned}$$

La divergence KL étant positive ou nulle, on obtient donc $H(\Gamma) \geq H(X)$. Ainsi, à espérance fixée, une variable qui suit une loi géométrique maximise son entropie.

Exercice 4. ($\star\star$) Tirage sur les sommets d'un triangle.

On considère un triangle (ABC) du plan dont les sommets ont pour coordonnées : $A = (0,0)$, $B = (1,0)$, $C = (0,1)$.



On effectue des tirages aléatoires parmi les trois sommets du triangle. On note X l'abscisse du point obtenu et Y son ordonnée.

Question 1.– Dans cette question, on suppose que le tirage est uniforme.

1. Déterminer les lois de X et Y.
2. Que peut-on dire de $H(X \mid Y = 1)$?
3. Calculer $H(X)$, $H(X \mid Y)$ et l'information mutuelle $I(X\,;Y)$.

Question 2.– Dans cette question, on fixe $\alpha \in [0,1]$ et on suppose que l'on tire A avec probabilité α, et B et C chacun avec probabilité $\frac{1-\alpha}{2}$.

1. Déterminer les lois de X et Y.
2. Calculer et tracer l'information mutuelle $I(X\,;Y)$ en fonction de α.

Solutions de l'Exercice 4.

Solution Q1.

1. On a $p(X = 0) = p(A) + p(C) = \frac{2}{3}$, donc $p(X = 1) = 1 - \frac{2}{3} = \frac{1}{3}$. De même, $p(Y = 0) = \frac{2}{3}$ et $p(Y = 1) = \frac{1}{3}$.
2. On a $H(X \mid Y = 1) = 0$. En effet, le seul point tel que $Y = 1$ est le point C, ce qui impose $X = 0$. La variable $X \mid Y = 1$ est donc déterministe.
3. Par conséquent $H(X) = H(Y) = \frac{1}{3}\log(3) + \frac{2}{3}\log(3/2) = \log(3) - \frac{2}{3}$. Comme $H(X,Y) = \log(3)$ (le tirage des points est uniforme sur 3 valeurs possible), on obtient $H(X \mid Y) = H(X,Y) - H(Y) = \frac{2}{3}$. Enfin, $I(X\,;Y) = H(X) - H(X \mid Y) = \log(3) - \frac{4}{3}$. Notons que ces deux dernières valeurs auraient pu être obtenues par le calcul :

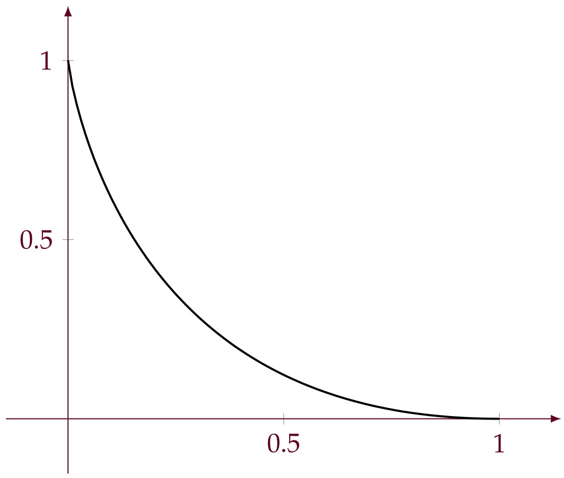
$$H(X \mid Y) = p(Y=0)H(X \mid Y=0) + p(Y=1)H(X \mid Y=1) = \tfrac{2}{3} \times 1 + \tfrac{1}{3} \times 0 = \tfrac{2}{3}$$

et

$$\begin{aligned}
I(X\,;Y) &= p(X=0,Y=0)\log\left(\tfrac{p(X=0,Y=0)}{p(X=0)p(Y=0)}\right) \\
&\quad + p(X=1,Y=0)\log\left(\tfrac{p(X=1,Y=0)}{p(X=1)p(Y=0)}\right) \\
&\quad + p(X=0,Y=1)\log\left(\tfrac{p(X=0,Y=1)}{p(X=0)p(Y=1)}\right) \\
&= \tfrac{1}{3}\log\left(\tfrac{1/3}{2/3\times 2/3}\right) + \tfrac{1}{3}\log\left(\tfrac{1/3}{1/3\times 2/3}\right) + \tfrac{1}{3}\log\left(\tfrac{1/3}{2/3\times 1/3}\right) \\
&= \tfrac{1}{3}\left(\log(3/4) + \log(3/2) + \log(3/2)\right) = \log(3) - \tfrac{4}{3}
\end{aligned}$$

Solution Q2.

1. On a $p(X = 0) = p(A) + p(C) = \frac{1+\alpha}{2}$ donc $p(X = 1) = \frac{1-\alpha}{2}$. La loi de X est donc une loi de Bernoulli de paramètre $\lambda = \frac{1+\alpha}{2}$. La variable Y suit la même loi.
2. Par conséquent
$$H(X) = H(Y) = h_2(\lambda) = \frac{1-\alpha}{2}\log\left(\frac{2}{1-\alpha}\right) + \frac{1+\alpha}{2}\log\left(\frac{2}{1+\alpha}\right)$$
Par ailleurs
$$\begin{aligned}
H(X \mid Y) &= H(X,Y) - H(Y) \\
&= 2 \times \frac{1-\alpha}{2}\log\left(\frac{2}{1-\alpha}\right) + \alpha\log\left(\frac{1}{\alpha}\right) - H(Y) \\
&= \frac{1-\alpha}{2}\log\left(\frac{2}{1-\alpha}\right) + \alpha\log\left(\frac{1}{\alpha}\right) - \frac{1+\alpha}{2}\log\left(\frac{2}{1+\alpha}\right)
\end{aligned}$$
donc
$$\begin{aligned}
I(X\,;Y) &= H(X) - H(X \mid Y) \\
&= (1+\alpha)\log\left(\frac{2}{1+\alpha}\right) - \alpha\log\left(\frac{1}{\alpha}\right)
\end{aligned}$$



Exercice 5. ($\star\star\star$) Estimation graphique de grandeurs informationnelles.

Dans les quatre graphes de la Figure 1, on représente le tirage de 100 points dans le plan selon quatre distributions distinctes.

Pour chacune de ces distributions, on note X la variable aléatoire représentant l'abscisse du point tiré, et Y son ordonnée. Le tirage d'un point correspond donc à la réalisation de la variable conjointe (X, Y). On note $H(X, Y)$ son entropie.

Pour simplifier l'étude, on pourra supposer que les tirages de certains points géométriques « extrémaux » (par exemple : les coins du carré de la distribution 4) sont de probabilité nulle.

Question 1.– Pour chacune des distributions, exprimer, si besoin en fonction de $H(X, Y)$:

1. l'entropie $H(X)$ de l'abscisse et celle $H(Y)$ de l'ordonnée ;
2. l'entropie $H(X|Y)$ de l'abscisse sachant l'ordonnée ;
3. l'information mutuelle $I(X;Y)$ entre l'abscisse et l'ordonnée.

Solutions de l'Exercice 5.

Solution Q1. Pour la distribution 1 :

1. $H(Y) = 0$ car Y est constante. Par conséquent $H(X) = H(X,Y)$.
2. On a donc $H(X \mid Y) = H(X) = H(X,Y)$.
3. On en déduit que $I(X;Y) = 0$.

Pour la distribution 2 :

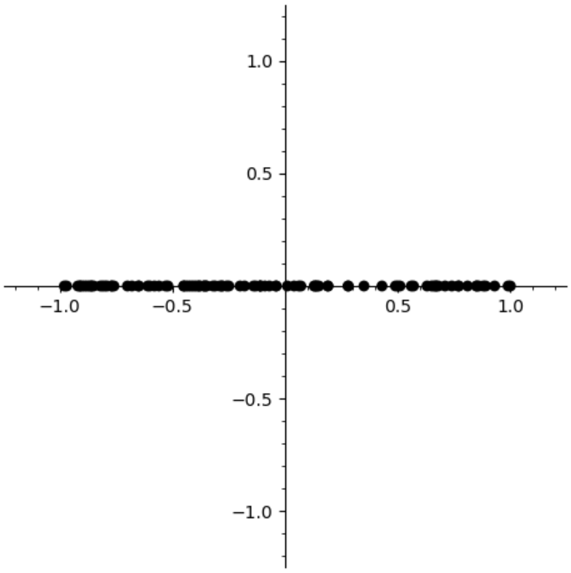
1. On observe que $X = Y$. Donc $H(X) = H(Y) = H(X,Y)$.
2. Comme $X = Y$, on a $H(X \mid Y) = 0$ (car Y détermine complèement X).
3. Par conséquent, $I(X;Y) = H(X,Y)$.

Pour la distribution 3 :

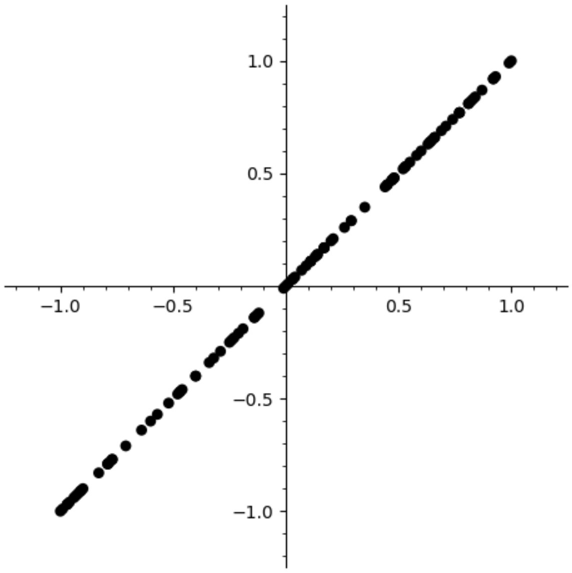
1. Par symétrie de la distribution, il est clair que $H(X) = H(Y)$. On pourra exprimer cette valeur en fonction de $H(X,Y)$, calculons $H(X \mid Y)$ dans le point suivant.
2. Pour chaque valeur $Y = y_i$, deux choix de X sont possibles et équiprobables. Par conséquent $H(X \mid Y = y_i) = 1$ pour tout i, donc $H(X \mid Y) = 1$. On en déduit que $H(X) = H(X,Y) - 1$.
3. On obtient $I(X;Y) = H(X) + H(Y) - H(X,Y) = H(X,Y) - 2$.

Pour la distribution 4 :

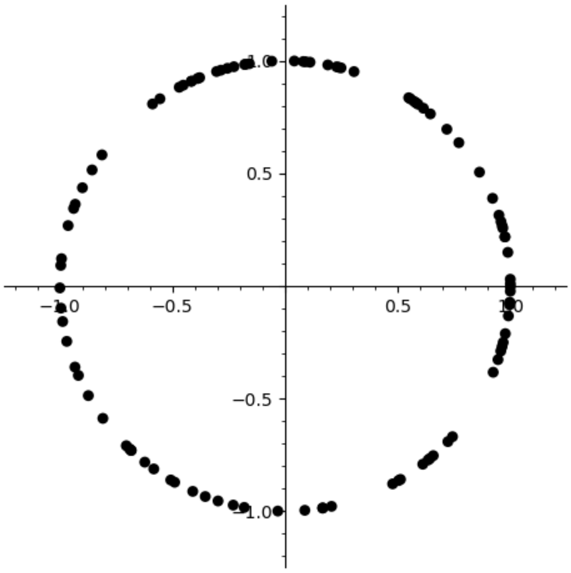
1. Par symétrie de la distribution, on a encore $H(X) = H(Y)$. Pour calculer exactement $H(X)$, observons que :
   — dans un cas sur 4, on a $X = 1$;
   — dans un cas sur 4, on a $X = -1$;
   — dans un cas sur deux, X est distribuée uniformément sur l'intervalle discretisé $]-1,1[$; pour ce cas de figure, on remarque que si $N = 2^{H(X,Y)}$ sont équiprobables pour (X,Y), alors par symétrie par rappor



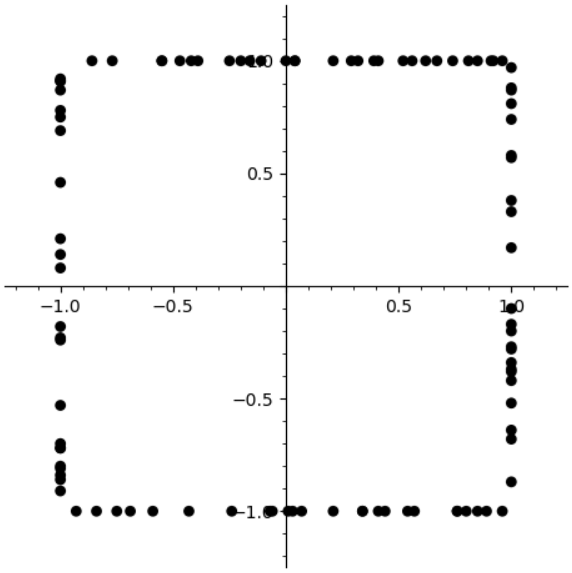
(a) Distribution 1



(b) Distribution 2



(c) Distribution 3



(d) Distribution 4

FIGURE 1 – Tirages selon quatre distributions dans le plan

à l'axe des abscisses, $N/2$ valeurs sont équiprobables pour X lorsque $X \notin \{-1,1\}$, ce qui mène à une entropie résiduele de $H(X,Y) - 1$.

Par conséquent $H(X) = 2 \times \frac{1}{4} \log_2(4) + \frac{1}{2}(H(X,Y) - 1) = \frac{H(X,Y)+1}{2}$.

2. Par conséquent $H(X \mid Y) = H(X,Y) - H(Y) = \frac{H(X,Y)-1}{2}$.
3. On a enfin $I(X\,;Y) = H(X) + H(Y) - H(X,Y) = 1$.