TD 1

Exercice 1. Dans la vallée de la mort :
- il pleut en moyenne 1 jour sur 100.
- la météo prédit 3 jours de pluie sur 100.
- chaque fois qu'il pleut, la météo l'a prévu.
- Monsieur Sûr-de-lui prévoit qu'il ne pleut jamais.

Est-il justifié de payer cher des investissements météo, alors que Monsieur Sûr-de-lui, qui ne coûte rien, se trompe moins souvent que la météo ?

Solution Intuitivement, alors que la météo est corrélée à la réalité, les prédictions de Monsieur Sûr-de-lui sont indépendantes du temps qu'il va faire. On s'attend donc à ce que l'information mutuelle entre le temps et les prédictions de Monsieur Sûr-de-lui soit nulle, alors que l'information mutuelle entre le temps et les prévisions météo soit positive.

Formellement : soient
X le temps (0 s'il fait beau, 1 s'il pleut),
Y la prévision météo(","),
Z la prévision de Monsieur Sûr-de-lui.
On donne les lois jointes de (X, Y) et celles de (X, Z) dans les deux tableaux suivants Elles

TABLE 1 – La loi jointe de (X, Y), à la case de "coordonnées" (a, b) on lit $p(X = a, Y = b)$.

$X \backslash Y$	0	1
0	0,97	0,02
1	0	0,01

TABLE 2 – La loi jointe de (X, Z), à la case de "coordonnées" (a, b) on lit $p(X = a, Z = b)$.

$X \backslash Z$	0	1
0	0,99	0
1	0,01	0

sont obtenues par le raisonnement suivant :

Loi de X : $p(X = 0) = 0,99$, $P(X = 1) = 0,01$

Loi de Y : $p(Y = 0) = 0,97$, $P(Y = 1) = 0,03$

On note d'après le troisième renseignement que : $P(Y = 1|X = 1) = 1$ et $P(Y = 0|X = 1) = 0$. Donc :

$$P(X = 1, Y = 1) = P(Y = 1|X = 1)P(X = 1) = 0,01$$

et

$$P(X = 1, Y = 0) = 0.$$

Par ailleurs

$$\begin{aligned} P(X = 0, Y = 1) &= P(Y = 1) - P(X = 1, Y = 1) = 0,03 - 0,01 = 0,02 \\ P(X = 0, Y = 0) &= P(X = 0) - P(X = 0, Y = 1) = 0,99 - 0,02 = 0,97 \end{aligned}$$

La deuxième loi est obtenue de manière similaire en démarrant le calcul en remarquant que

$$P(X=0,Z=1)=P(X=1,Z=1)=0.$$

Donner $I(X;Y), I(X,Z)$: une simple application numérique conduit à

$$\begin{aligned} I(X;Y) &\approx 0.05324 \\ I(X;Z) &= 0. \end{aligned}$$

Ce résultat s'interprête de la manière suivante. La météo apporte une information assez conséquente sur le temps, cela peut se quantifier par le rapport $\frac{I(X;Y)}{H(X)}$ qui est une quantité comprise entre 0 et 1, la valeur 0 indiquant que X et Y sont indépendants et une valeur de 1 le fait que X est une fonction déterministe de Y (voir exercice 4). Cette quantité vaut

$$\frac{I(X;Y)}{H(X)} \approx 0.65902$$

Dans le deuxième cas, le rapport $\frac{I(X;Z)}{H(X)}$ est nul. X est Z sont par conséquent deux variables aléatoires indépendantes et Z n'apporte aucune information sur X.

Exercice 2. Soit un vecteur $(X_1, X_2, \ldots, X_N)$ de N variables aléatoires, son entropie est par définition :

$$H(X_1X_2\ldots X_N) = -\sum p(x_1,\ldots,x_N)\log p(x_1,\ldots,x_N)$$

1. (Cas de l'indépendance) Montrer que si $X_1,\ldots,X_n$ sont indépendantes :

$$H(X_1X_2\ldots X_N) = \sum_{i=1}^{N} H(X_i)$$

2. (Cas général) Montrer que (« règle de chaînage pour l'entropie »)

$$H(X_1X_2\ldots X_N) = H(X_N|X_1,\ldots X_{N-1}) + H(X_{N-1}|X_1\ldots X_{N-2}) + \cdots + H(X_2|X_1) + H(X_1)$$

Solution

1. Par récurrence. La formule est clairement vraie pour $N=1$. Supposons la vérifiée pour $N=k$. Posons $X=(X_1,\ldots,X_k)$ et $Y=X_{k+1}$. Notons que X et Y sont indépendants. On obtient ainsi

$$\begin{aligned} H(X_1,\ldots,X_{k+1}) &= H(X,Y) \\ &= H(X)+H(Y)-I(X;Y) \\ &= H(X)+H(Y) \text{ (car } X \text{ et } Y \text{ sont indépendants)} \\ &= H(X_1,\ldots,X_k)+H(X_{k+1}) \\ &= H(X_1)+\cdots+H(X_k)+H(X_{k+1}) \text{ (par hyp. de récurrence)} \end{aligned}$$

2. Par récurrence à nouveau. C'est vrai pour $N=1$. Posons à nouveau $X=(X_1,\ldots,X_k)$ et $Y=X_{k+1}$.

$$\begin{aligned} H(X_1,\ldots,X_{k+1}) &= H(X,Y) \\ &= H(X)+H(Y|X) \\ &= H(X_1,\ldots,X_k)+H(X_{k+1}|X_1\ldots X_k) \\ &= H(X_1)+\cdots+H(X_k|X_1\ldots X_{k-1})+H(X_{k+1}|X_1\ldots X_k) \text{ (par hyp. de récurrence)} \end{aligned}$$

Exercice 3. Soit le vecteur aléatoire (X, Y, Z), tel que

$$\begin{aligned} p_{XYZ}(0,0,0) &= \frac{1}{4} \\ p_{XYZ}(0,1,0) &= \frac{1}{4} \\ p_{XYZ}(1,0,0) &= \frac{1}{4} \\ p_{XYZ}(1,0,1) &= \frac{1}{4} \end{aligned}$$

On rappelle la formule

$$H(Y|X) = \sum_x p(x) H(Y|X = x)$$

Donner les valeurs de $H(X)$, $H(Y|X)$, $H(Z|X, Y)$. Trouver $H(X, Y, Z)$ de deux manières : par la règle de chaînage, et directement.

On rappelle que $h(1/4) \approx 0.811$. Calculer $H(Y)$, Vérifier que $H(Y|X) \leq H(Y)$. Quelle information apporte X sur Y, Y sur X ?

Solution

- On a $p(X = 0) = p(X = 1) = 1/2$. Donc $H(X) = 1$.
- Ensuite $p(Y = 0|X = 0) = 1/2$ et $p(Y = 1|X = 0) = 1/2$. Donc $H(Y|X = 0) = 1$. De plus $p(Y = 1|X = 1) = 0$, et $p(Y = 0|X = 1) = 1$, donc $H(Y|X = 1) = 0$. Finalement

$$H(Y|X) = 1/2 H(Y|X = 0) + 1/2 H(Y|X = 1) = 1/2.$$

- $p(Z = 0|X = 0, Y = 0) = 1$, $p(Z = 1|X = 0, Y = 0) = 0$. Donc $H(Z|X = 0, Y = 0) = 0$. De même $p(Z = 0|X = 0, Y = 1) = 1$ $p(Z = 1|X = 0, Y = 1) = 0$. Donc $H(Z|X = 0, Y = 1) = 0$. Enfin $H(Z|X = 1, Y = 0) = 1$. Il vient

$$H(Z|X, Y) = 1/4.0 + 1/4.0 + 1/2 H(Z|X = 1, Y = 0) = 1/2.$$

- $H(X, Y, Z) = H(Z|X, Y) + H(Y|X) + H(X) = 1/2 + 1/2 + 1 = 2$.
- Directement : la variable aléatoire prend 4 valeurs équiprobables, donc $H = \log_2 4 = 2$.
- On a $p(Y = 0) = \frac{3}{4}$ et $p(Y = 1) = \frac{1}{4}$. Par conséquent $H(Y) = h(1/4) \approx 0.811$. $H(Y|X)$ vaut $\frac{1}{2}$. Cela est bien inférieur à $H(Y)$. L'information qu'apporte X sur Y vaut $I(X; Y)$ soit $H(Y) - H(Y|X) \approx 0.311$. L'information qu'apporte Y sur X vaut également $I(X; Y)$.

Exercice 4. Montrer que $H(Y|X) = 0$ si et seulement si Y est une fonction de X, c'est à dire que pour tout x, tel que $p(x) > 0$ il existe un y tel que $p(y|x) = 1$. On utilise ici la notation (ou plutôt l'abus de notation...) $p(y|x) = p(Y = y|X = x)$.

Solution On a $H(Y|X) = \sum_x p(x) H(Y|X = x)$. On doit donc avoir, pour tout x tel que $p(x) > 0$, $H(Y|X = x) = 0$. On doit donc avoir $p(y|x) = 1$ pour un certain y, et donc $p(y|x) = 0$ pour les autres.

Exercice 5. Soit X une variable aléatoire et $g(x)$ une fonction. En utilisant le règle de chaînage de deux manières différentes, montrer que

$$H(g(X)) \leq H(X).$$

Solution On écrit

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

Mais aussi

$$\begin{aligned} H(X, g(X)) &= H(X|g(X)) + H(g(X)) \\ &\geq H(g(X)) \end{aligned}$$

Exercice 6. Soit X une variable aléatoire réelle discrète d'entropie $H(X)$. Donner le lien général entre $H(Y)$ et $H(X)$

- quand $Y = 2^X$
- quand $Y = \cos X$.

Solution

1. Y est une fonction de X donc $H(Y) \leq H(X)$ d'après l'exercice précédent. Par ailleurs, X est une fonction de Y, puisque $X = \log_2 Y$. Par conséquent, $H(X) \leq H(Y)$. Par conséquent, $H(X) = H(Y)$. Cette technique de preuve permet de montrer que $H(X) = H(f(X))$ quand f est bijective.
2. On a pour les mêmes raisons $H(Y) \leq H(X)$. Cependant la fonction cosinus n'étant pas bijective, on n'a pas nécessairement $H(X) = H(Y)$. Ainsi, pour X une variable aléatoire à valeurs dans $\{-1, 1\}$ donnée par $p(X = -1) = p(X = 1) = \frac{1}{2}$, on a $H(X) = 1$. En revanche, ici $p(Y = \cos 1) = 1$. Par conséquent $H(Y) = 0$.

Exercice 7. On suppose que l'on a une balance à deux plateaux (qui permet donc juste de comparer les poids qui sont mis sur les deux plateaux) et de n pièces. Quand une pièce est authentique, elle a un certain poids t (qui est inconnu). Quand elle est frauduleuse elle est d'un poids différent de t (elle peut aussi bien être plus légère que plus lourde). On sait que parmi ces n pièces, une pièce au plus est frauduleuse.

1. Trouver une borne inférieure sur le nombre de pesées à effectuer de manière à être sur de détecter la pièce frauduleuse et de pouvoir dire si elle est plus légère ou plus lourde. Indication : utiliser l'exercice 5...
2. On suppose que l'on dispose également d'un tas infini de pièces authentiques. Donner la stratégie optimale pour détecter une pièce frauduleuse et dire si elle est plus légère ou plus lourde pour un tas de 13 pièces. Généraliser au cas $n = \frac{3^k-1}{2}$.

Solution

1. On numérote les pièces de 1 à n. On code toutes les possibilités par une variable entière X comprise entre $-n$ et n. X vaut 0 si et seulement il n'y aucune pièce frauduleuse. Dans le cas contraire, si X vaut i, cela veut dire que $|i|$ est la pièce frauduleuse. $X > 0$ signifie que la pièce frauduleuse est plus lourde et $X < 0$ signifie que la pièce frauduleuse est plus légère. Soit k un nombre de pesées pour lequel on est sur de trouver la valeur de X. Soit $(Y_1, Y_2, \ldots, Y_k) \in \{-1, 0, 1\}^k$, les résultats donnés par ces pesées. $Y_i = -1$ a pour signification qu'à la i-ème pesée le plateau de gauche soit plus léger que le plateau de droite, $Y_i = 0$ signifie que les deux plateaux ont même poids et finalement $Y_i = 1$ signifie que le plateau de gauche est le plus lourd. Notons que par hypothèse, X est une fonction de $(Y_1, Y_2, \ldots, Y_k)$. Par conséquent (exercice 5)
$$H(X) \leq H(Y_1, \ldots, Y_k).$$
Par ailleurs,
$$\begin{aligned} H(Y_1, \ldots, Y_k) &= H(Y_1) + H(Y_2|Y_1) + \cdots + H(Y_k|Y_1 \ldots Y_{k-1}) \text{ (chaînage)} \\ &\leq H(Y_1) + H(Y_2) + \cdots + H(Y_k) \text{ (le conditionnement ne peut que réduire l'entropie)} \\ &\leq k \log_2 3 \text{ ($\log_2 3$ est l'entropie maximale d'une variable aléatoire ternaire)} \end{aligned}$$

Par conséquent

$$k \geq \frac{H(X)}{\log_2 3}.$$

On optimise cette borne inférieure en choisissant de manière adaptée les configurations (en les prenant équiprobables). Dans ce cas

$$H(X) = \log_2(2n+1).$$

On obtient finalement

$$k \geq \frac{\log_2(2n+1)}{\log_2 3} = \log_3(2n+1).$$

2. La borne précédente s'applique également au cas où l'on dispose d'un tas auxiliaire de pièces authentiques. Il faut donc au moins 3 pesées dans ce cas. Cela peut être réalisé de la manière suivante. On prend neuf pièces du tas à tester, que l'on compare à 9 pièces du tas de pièces authentiques. Deux cas sont possibles.

   Soit les 9 pièces du tas à tester sont d'un poids différent. La pesée précédente a indiqué si la pièce frauduleuse est plus lourde ou plus légère. On peut supposer sans perdre en généralité qu'elle est plus légère. Dans ce cas, on trouve la pièce incriminée en deux pesées. On fait en effet trois tas de 3 pièces. On compare le premier tas avec le second. S'ils ont même poids, c'est que la pièce se trouve dans le dernier tas. Sinon, la pièce incriminée se trouve dans le tas le plus léger. Dans les deux cas, on est réduit à étudier un tas de 3 pièces. Il suffit de prendre deux pièces du tas et de les comparer pour déterminer quelle pièce est la plus légère des trois.

   Dans le deuxième cas, les 9 pièces ont même poids que les 9 pièces du tas auxiliaire. Elles sont donc toutes authentiques et l'on est réduit à étudier le tas des 4 pièces restantes. On prend 3 pièces de ce tas que l'on compare à 3 pièces du tas auxiliaire. A nouveau deux cas se présentent. Soit les deux tas ont même poids. Dans ce cas, il suffit de comparer la dernière pièce avec une pièce étalon du tas auxiliaire et le tour est joué. Soit on dispose d'un tas de 3 pièces dont on sait qu'elle est frauduleuse et si elle est plus lourde ou plus légère. On termine le travail comme précédemment en une pesée en comparant deux pièces de ce dernier tas.

   L'algorithme se généralise facilement à tout ensemble de $\frac{3^k-1}{2}$ pièces que l'on peut tester en k pesées. On peut le montrer par récurrence sur k en utilisant le résultat auxiliaire suivant (que l'on montre lui aussi par récurrence). Soit un tas de 3^k pièces dont l'une est frauduleuse et qui est plus lourde que les autres. On la trouve en k pesées. Il suffit en effet pour la première pesée de faire deux tas distincts de 3^{k-1} pièces chacune, puis de comparer le poids de ces 2 tas pour se réduire avec une pesée à un tas de 3^{k-1} pièces résiduelles.

   Voilà comme on procède pour le tas de $\frac{3^k-1}{2}$ pièces. On fait un tas comprenant 3^{k-1} pièces. On compare ce tas avec un tas formé de 3^{k-1} pièces du tas auxiliaire. Si les deux tas sont de poids différents il faut résoudre le problème susmentionné, ce que l'on fait donc en $k-1$ pesées. Sinon, on est ramené au problème de départ avec un tas de taille $\frac{3^k-1}{2} - 3^{k-1} = \frac{3^{k-1}-1}{2}$.

Exercice 8 (Le problème du mot de passe)**.** Un individu (probablement mal intentionné) cherche à accéder à un service protégé par un mot de passe qu'il ne connaît pas. Soit $\mathcal{M} = \{0,1\}^m$ l'ensemble des mots de passe possibles. Nous supposons que le système d'authentification est parfait et que la seule possibilité d'action pour l'attaquant consiste à essayer les mots de passe un par un.

On suppose ensuite que le mot de passe est choisi dans $\mathcal{M}$ selon une loi d'entropie $h \leq m$. Nous notons p_i les probabilités des lettres de $\mathcal{M}$ dans l'ordre décroissant (le mot le plus probable a pour probabilité p_1, le suivant pour probabilité p_2, ...).

1. Montrer que la meilleure stratégie consiste à tester les mots dans l'ordre des probabilités décroissantes. Exprimez le nombre moyen d'essais, $\mathcal{N}(p)$, en fonction des p_i.
2. Soient deux lois de probabilité $p = (p_i)_{i\geq 1}$ et $q = (q_i)_{i\geq 1}$ telles que les suites p_i et q_i soient décroissantes avec $q_i > 0$ pour tout $i \geq 1$ (en revanche p_i peut être nul à partir d'un certain

rang). On rappelle que la distance de Kullback de p par rapport à q est donnée par

$$D(p \parallel q) = \sum_{i \geq 1} p_i \log \frac{p_i}{q_i}. \tag{1}$$

Nous posons $q_i = (1-\alpha)\alpha^{i-1}$ pour un certain réel $0 < \alpha < 1$. On suppose que les entropies de $H(p)$ et $H(q)$ de p et q sont bien définies (c'est à dire que les sommes $-\sum_{i\geq 1} p_i \log p_i$ et $-\sum_{i\geq 1} q_i \log q_i$ sont bien définies). Montrer que si $H(p) = H(q)$, alors $\sum_{i\geq 1} ip_i \geq \sum_{i\geq 1} iq_i$. *Indication : on pourra tirer profit de la positivité de la distance de Kullback $D(p \parallel q) \geq 0$.*

3. Calculer l'entropie $H(q)$ de la loi q en fonction de α. Nous noterons H_α cette quantité. On rappelle les identités $\sum_{i\geq 1} \alpha^{i-1} = 1/(1-\alpha)$ et $\sum_{i\geq 1} i\alpha^{i-1} = 1/(1-\alpha)^2$
4. En déduire que pour tout réel $0 < \alpha < 1$ nous avons $1 < (1-\alpha)2^{H_\alpha} < e$, où e est la base du logarithme népérien.
5. Déduire du résultat précédent que que $\mathcal{N}(p) > c_1 2^h$ (on s'efforcera de donner une valeur à c_1). Interprétez le résultat.

Solution

1. On teste les mots de passe suivant un certain ordre, disons que l'on choisit une permutation π des entiers de 1 à 2^m et l'on teste d'abord le mot de passe d'indice $\pi(1)$, puis celui d'indice $\pi(2)$ et ainsi de suite. Le nombre moyen d'essais $N(\pi)$ est par conséquent égal à

$$N(\pi) = \sum_{i=1}^{2^m} ipi_{\pi(i)}$$

Supposons qu'il existe un couple (i,j) tel que $i < j$ et $\pi(i) > \pi(j)$. On peut diminuer le nombre moyen d'essais en choisissant un ordre (i.e. une permutation π') tel que $\pi'(k) = \pi(k)$ pour tout k différent de i et j et $\pi'(i) = \pi(j)$, $\pi'(j) = \pi(i)$. On a en effet :

$$\begin{aligned} N(\pi') - N(\pi) &= i(p_{\pi(j)} - p_{\pi(i)}) + j(p_{\pi(i)} - p_{\pi(i)}) \\ &= (j-i)(p_{\pi(i)} - p_{\pi(j)}) \\ &< 0. \end{aligned}$$

Cela implique que l'ordre trivial $\pi(1) = 1, \pi(2) = 2, \ldots, \pi(2^m) = 2^m$ permet de minimiser ce nombre d'essais.

2. On remarque d'abord que

$$\log q_i = \log(1-\alpha) + (i-1)\log\alpha$$

Par conséquent :

$$\begin{aligned} i &= \frac{\log q_i - \log(1-\alpha) + \log\alpha}{\log\alpha} \\ &= \frac{\log q_i + \log\frac{\alpha}{1-\alpha}}{\log\alpha} \end{aligned}$$

On utilise cette expression de i dans le calcul suivant

$$\begin{aligned}
\sum_i i p_i &= \sum_i p_i \frac{\log q_i + \log \frac{\alpha}{1-\alpha}}{\log \alpha} \\
&= \frac{1}{\log \alpha} \sum p_i \log q_i + \frac{\log \frac{\alpha}{1-\alpha}}{\log \alpha} \\
&= \frac{1}{\log \alpha} \sum p_i \log \frac{q_i}{p_i} p_i + \frac{\log \frac{\alpha}{1-\alpha}}{\log \alpha} \\
&= -\frac{1}{\log \alpha} D(p \parallel q) - \frac{1}{\alpha} H(p) + \frac{\log \frac{\alpha}{1-\alpha}}{\log \alpha} \\
&\leq -\frac{1}{\log \alpha} H(q) + \frac{\log \frac{\alpha}{1-\alpha}}{\log \alpha} \\
&= \frac{\log \frac{\alpha}{1-\alpha}}{\log \alpha} + \frac{1}{\log \alpha} \sum_i q_i \log q_i \\
&= \frac{\log \frac{\alpha}{1-\alpha}}{\log \alpha} + \frac{1}{\log \alpha} \sum_i q_i \left(\log(1-\alpha) + (i-1)\log \alpha\right) \\
&= \sum_i i q_i
\end{aligned}$$

3. Le calcul de l'entropie de q donne

$$\begin{aligned}
H_\alpha &= -\sum_i q_i \log q_i \\
&= -\sum_i (1-\alpha)\alpha^{i-1} \left(\log(1-\alpha) + (i-1)\log \alpha\right) \\
&= -\log(1-\alpha) + \log \alpha - \frac{\log \alpha}{1-\alpha} \\
&= \log \frac{\alpha}{1-\alpha} - \frac{\log \alpha}{1-\alpha}
\end{aligned}$$

4. On a

$$\begin{aligned}
(1-\alpha)2^{H_\alpha} &= (1-\alpha)\frac{\alpha}{1-\alpha} 2^{-\frac{\log \alpha}{1-\alpha}} \\
&= \alpha \alpha^{-\frac{1}{1-\alpha}} \\
&= \alpha^{-\frac{\alpha}{1-\alpha}}
\end{aligned}$$

Il est facile de vérifier que cette quantité est toujours comprise entre 1 et e.

5. Il est à noter que le calcul fournit par ailleurs

$$\sum_i i q_i = \frac{1}{1-\alpha}.$$

De tout ceci on déduit que

$$\mathcal{N}(p) \geq \mathcal{N}(q) = \frac{1}{1-\alpha}.$$

avec α qui est tel que $h = H_\alpha$. Notons que l'inégalité du point précédent implique que

$$2^{-H_\alpha} \leq 1 - \alpha \leq e 2^{-H_\alpha}.$$

Par conséquent :

$$\frac{1}{e} 2^{H_\alpha} \leq \frac{1}{1-\alpha} \leq 2^{H_\alpha},$$

c'est à dire

$$\frac{1}{e}2^h \leq \frac{1}{1-\alpha} \leq 2^h.$$

Cela implique que

$$\mathcal{N}(p) \geq \frac{1}{e}2^h.$$

Exercice 9. Montrer que

$$\binom{n}{t} \leq 2^{nh(t/n)}$$

Solution Soit $(X_1, X_2, \ldots, X_n)$ une variable aléatoire à valeurs dans $\{0,1\}^n$ qui est équidistribuée parmi l'ensemble des mots binaires de poids t. On a

$$H(X_1, X_2, \ldots, X_n) = \log_2 \binom{n}{t} \tag{2}$$

Par ailleurs,

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &\leq H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_1 \ldots Y_{n-1}) \text{ (chaînage)} \\ &\leq H(X_1) + H(X_2) + \cdots + H(X_n) \text{ (le conditionnement ne peut que réduire l'entropie)} \\ &\leq nh(t/n) \text{ (car } p(X_i = 1) = t/n \text{ et donc } H(X_i) = h(t/n).) \end{aligned}$$

Exercice 10. [Lemme de Fano- lien entre la probabilité d'erreur d'un estimateur et l'entropie conditionnelle] Soit X et Y deux variables aléatoires (avec X prenant ses valeurs dans un alphabet de taille a). On estime X par une certaine fonction $\hat{X}$ de Y. On note P_e la probabilité de l'estimateur, c'est à dire $P_e = p(\hat{X} \neq X)$. Montrer que

$$h(P_e) + P_e \log_2(a-1) \geq H(X|Y).$$

Indication : introduire une variable aléatoire E définie par

$$E = \begin{cases} 1 & \text{si } \hat{X} \neq X \\ 0 & \text{si } \hat{X} = X \end{cases}$$

puis écrire $H(E, X|Y)$ de deux manières différentes.

Solution Ecrivons maintenant l'entropie $H(E, X|Y)$ de deux manières différentes

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

Il est à noter d'une part que $H(E|X, Y) = 0$ (puisque E est une fonction de X et Y). Par ailleurs, concernant la seconde expression pour $H(E, X|Y)$ on a $H(E|Y) \leq H(E)$ (puisque le conditionnement ne peut que réduire l'entropie). Comme E est une variable aléatoire binaire on a

$$H(E) = -p(E=0)\log(p(E=0)) - p(E=1)\log(p(E=1)) = h(P_e).$$

Par ailleurs,

$$\begin{aligned} H(X|E, Y) &= p(E=0)H(X|Y, E=0) + p(E=1)H(X|Y, E=1) \\ &\leq (1-P_e)0 + P_e \log_2(a-1) \end{aligned}$$

puisque lorsque $E = 0$, X est une fonction de Y et donc $H(X|Y, E=0) = 0$ et que l'on peut toujours borner l'entropie conditionnelle de X connaissant Y et que $E = 1$ par l'entropie d'une variable aléatoire uniformément répartie parmi toutes les valeurs possibles de X à l'exception de $\hat{X}$. En rassemblant les deux égalités précédentes et cette nouvelle inégalité, on obtient le lemme annoncé.

Exercice 11. Soient X et Y deux variables aléatoires à valeurs dans un groupe $(G, +)$. Soit la variable aléatoire $Z = X + Y$.

1. Montrer que $H(Z|X) = H(Y|X)$.
2. Montrer que si X et Y sont indépendantes $H(Y) \leq H(Z)$ et $H(X) \leq H(Z)$ (utiliser la positivité de l'information mutuelle).
3. Donner un exemple de deux variables aléatoires X et Y telles que $H(X) > H(Z)$ et $H(Y) > H(Z)$.

Solution

1. On écrit de deux manières

$$\begin{aligned} H(Z, X, Y) &= H(Z|X, Y) + H(Y|X) + H(X) \\ &= 0 + H(Y|X) + H(X) \\ &= H(Y|X, Z) + H(Z|X) + H(X) \\ &= 0 + H(Z|X) + H(X) \end{aligned}$$

2. On écrit $I(Z; X) \geq 0$, puis

$$\begin{aligned} I(Z; X) &= H(Z) - H(Z|X) \\ &= H(Z) - H(Y|X) \\ &= H(Z) - H(Y) \text{ (car } X \text{ et } Y \text{ sont indépendantes)} \end{aligned}$$

   Donc $H(Z) \geq H(Y)$

3. On donne $X \in \{0, 1\}$ avec n'importe quelle proba, et $Y = 1 - X$. Alors $Z = 0$, et $H(Z) = 0$.