Théorie de l'Information

Résolution du Devoir # 1

SIC-SICOM

Maria-João Rendas

1. Soit X une variable aléatoire discrète. Montrer que l'entropie d'une fonction de X est inférieure ou égale à l'entropie de X en justifiant les pas suivants:

$$\begin{aligned} H(X,g(X)) &\overset{(a)}{=} H(X)+H(g(X)|X) \\ &\overset{(b)}{=} H(X) \\ H(X,g(X)) &\overset{(c)}{=} H(g(X))+H(X|g(X)) \\ &\overset{(d)}{\geq} H(g(X)) \end{aligned}$$

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

<u>résolution</u>

- (a) est une instance de la règle de la chaîne pour l'entropie conjointe:

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

avec $Y = g(X)$.

- (b) découle du fait que $H(g(X)|X) = 0$. Par définition

$$H(g(X)|X) = \mathrm{E}_X\left[H(g(X)|X=x)\right] = \mathrm{E}_X\left[\mathrm{E}_{y\in g(\mathcal{X})}\left(p(y|X=x)\log\frac{1}{p(y|X=x)}\right)\right].$$

Mais, comme Y est une fonction déterministe de X,

$$p(y|X=x) = \begin{cases} 1, & y = g(x) \\ 0, & y \neq g(x) \end{cases}$$

d'où

$$p(y|X=x)\log\frac{1}{p(y|X=x)} = 0, \forall y \in g(\mathcal{X})$$

et donc

$$H(g(X)|X) = \mathrm{E}_X\left[\mathrm{E}_{y\in g(\mathcal{X})}0\right] = 0.$$

- (c) règle de la chaîne (comme en (a)).

- (d) est une conséquence immédiate de la non-négativité de l'entropie conditionnelle: $H(X|Y) \geq 0$, avec $Y = g(X)$.

2. **résolution**
Nous venons de voir (dans le problème précédant) qu'en général

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

Pour avoir égalité en (d), il faut que

$$H(X|g(X)) = 0.$$

C'est à dire,

$$H(g(X)) = H(X) \Leftrightarrow H(X|g(X)) = 0.$$

Cette dernière équation implique (nous l'avons vu en cours) que la connaissance de $g(X)$ fixe d'une façon unique la valeur de X, ce qui est équivalent à dire que $g(X)$ est une fonction *inversible*.

(a) Soit $Y = X^5$, où X est une variable aléatoire qui prend des valeurs entières positives et négatives. Quelle est la relation entre $H(X)$ et $H(Y)$?
Comme x^5 est une fonction inversible (voir Figure 1), nous avons dans ce cas

$$H(X^5) = H(X).$$

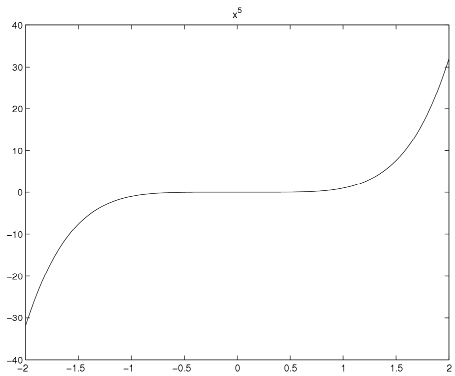

Figure 1: graphe de x^5.

(b) Et si $Y = X^2$?
Cette fonction, voir Figure 2, n'est pas en général inversible sur tout son domaine (qui comprend des valeurs positives et négatives), et donc $H(X|Y) > 0$, et par conséquence

$$H(Y = X^2) < H(X).$$

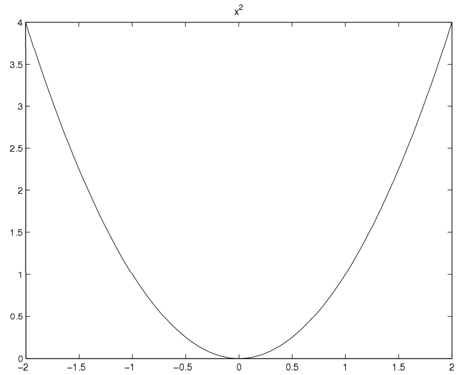

Figure 2: graphe de x^2.

Nous nottons que si le domaine de X, $\mathcal{X}$, est tel que

$$\text{Si } x \in \mathcal{X} \Rightarrow -x \notin \mathcal{X}$$

alors nous aurons $H(Y = x^2) = H(X)$.

(c) Et si $Y = \tan X$?
La Figure 3 montre que cette fonction n'est pas inversible sur $\mathbf{R}$, et donc, en général

$$H(Y = \tan X) < H(X).$$

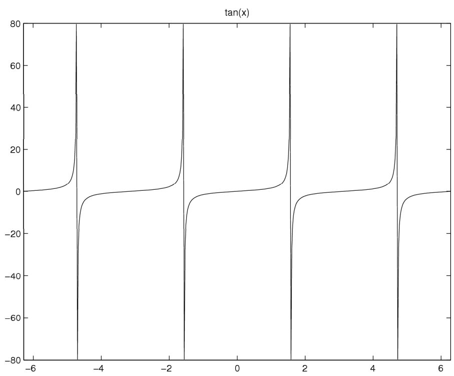

Figure 3: graphe de $\tan x$.

3. (a) Considérez qu'une pièce équilibrée est lancée. Quelle est l'information mutuelle entre la face en haut et celle qui est en bas ?
résolution
Nottons $X \in \mathcal{A} = \{\text{face}, \text{pile}\}$ la variable aléatoire qui décrit la face en

haut, et $Y \in \mathcal{A}$ une autre variable aléatoire qui décrit la face en bas de la pièce. La loi de probabilité conjointe est (en supposant que la pièce est équilibrée)

$$p_{X,Y}(x,y) = \begin{cases} 1/2, & x \neq y \\ 0, & x = y \end{cases},$$

car il faut bien que la face en haut (X) soit diférente de celle en bas (Y). De cette loi nous obtenons directement la loi de probabilité conditionnelle

$$p_{Y|X=x}(y|x) = \begin{cases} 1, & y \neq x \\ 0, & y = x \end{cases}, \forall x \in \mathcal{A},$$

el les lois marginales

$$p_X(x) = \frac{1}{2}, \forall x \in \mathcal{A}, \qquad p_Y(y) = \frac{1}{2}, \forall y \in \mathcal{A}.$$

L'information mutuelle est par définition

$$I(X;Y) = H(X) - H(Y|X) = H(Y) - H(Y|X)$$

Les entropies de X et de Y sont l'entropie d'une loi uniforme dans un alphabet de dimension 2, et donc

$$H(X) = H(Y) = \log 2 = 1 \text{ bit}.$$

L'entropie conditionnelle de Y sachant X (égale à celle de X sachant Y, car le problème est complètement symétrique) est

$$H(Y|X) = \mathrm{E}_X\left[H(Y|X=x)\right] = \mathrm{E}_X\left[0\right] = 0,$$

car une fois X observée, la valeur de Y est déterminée avec certitude.
Nous obtenous donc

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) = 1 \ bit.$$

(b) Un dé équilibré est lancé. Quelle est l'information mutuelle entre la face en haut et celle en bas ?
Comme la face en bas (X) est une fonction déterministe de celle qui est en haut (Y), le même raisonnement que celui effectué pour l'alinéa préc édante conduit à

$$I(X;Y) = H(X) - H(X|Y) = H(X) = \log 6,$$

car toutes les six faces sont équiprobables (le dé est *équilibré*).

(c) Quelle est l'information mutuelle entre la face en haut et celle qui est tournée vers le joueur ?
Soit $X \in \mathcal{B} = \{1,2,3,4,5,6\}$ la face en haut, et $Y \in \mathcal{B}$ la face en bas

(dans le même lancement d'un dé équilibré). Nous rappellons la d éfinition d'information mutuelle:

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

Dans notre cas, nous pouvons constanter facilement que $H(X) = H(Y) = \log 6$; ce résultat a déjà été utilisé dans l'alinéa précédante. Calculons maintenant $H(Y|X)$. Une fois que la valeur de la face en haut est fixée, la face en bas est aussi uniquement déterminée, et toutes les autres *quatre* faces peuvent être, avec égale probabilité, tournées vers le joueur. Les lois conditionnelles $P(Y|X = x)$ sont donc des lois uniformes dans un alphabet de dimension quatre et par conséquence

$$H(Y|X = x) = \log 4, \forall x \in \mathcal{B}.$$

L'entropie conditionnelle est la moyenne sur X de cette valeur (constante), et donc

$$H(Y|X) = \mathrm{E}_X [\log 4] = \log 4 = 2 \text{ bits}.$$

Finalement, l'information mutuelle est

$$I(X;Y) = H(Y) - H(Y|X) = \log 6 - \log 4 = \log \frac{2}{3} = 0.585 \ bits.$$

4. Supposez qu'on dispose de n pièces qui peuvent ou pas comprendre une pièce fausse. La pièce fausse peut être plus lourde ou plus légère que les vraies. On dispose d'une balance pour peser les pièces.

   (a) Comptez le nombre d'états dans lesquels l'ensemble de n pièces peut être, et le nombre de possibles résultats des k pesages. En les comparant, déterminez une borne supérieure au nombre n de pièces parmis lesquelles une suite de k pesages permettent de déterminer quelle est la pièce fausse (si elle existe) et si elle elle plus plégère ou plus lourde que les autres.
   résolution
   Les états possibles de l'ensemble de n pièces sont

   - Toutes les pièces sont similaires (1 état)
   - Il y a une des n pièces qui est plus lourde (n états, chacun correspondant à l'identité de la pièce plus lourde)
   - Il y a une des pièces qui est plus légère (n états).

   Le nombre total d'états est donc

   $$N = 2n + 1.$$

   à chaque pesée, il y a trois résultats possibles

   - équilibre des plats de la balance ($A = B$)
   - le plat gauche est plus lourd que celui de droite ($A > B$)
   - le plat de gauche est plus léger que celui de droite ($A < B$).

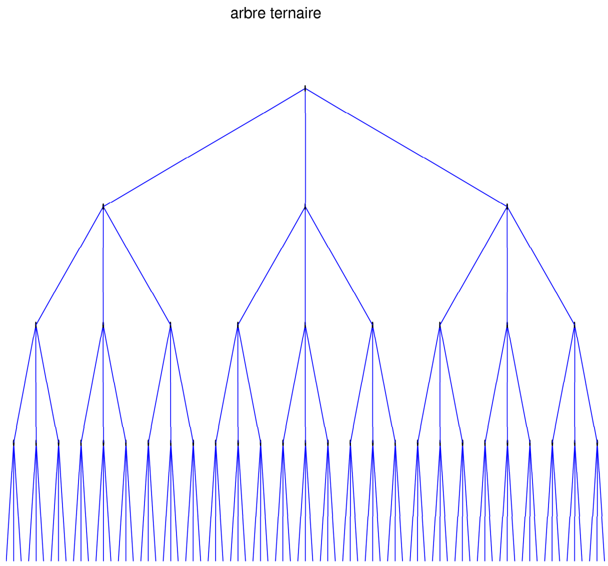

Figure 4: Arbre ternaire.

Le nombre de situations possibles qui peuvent être discriminées avec k pesages est donc 3^k, illustrées par les feuilles de l'arbre ternaire de la Figure 4, pour $k = 4$. Pour pouvoir détecter dans lequel, parmis les $2n + 1$ états possibles, est l'ensemble de n pièces, nous devons donc avoir

$$3^k \geq 2n + 1 \Leftrightarrow n \leq (3^k - 1)/2.$$

5. Une pièce est lancée jusqu'à l'occurence d'une face. Soit X le nombre de lancements.

   (a) Déterminez l'entropie $H(X)$ en bits. L'expression suivante peut être utile :

   $$\sum_{n=1}^{\infty} r^n = \frac{r}{1-r}, \qquad \sum_{n=1}^{\infty} n r^n = \frac{r}{(1-r)^2}.$$

   Déterminez une séquence de questions binaires (réponses oui/non) de la forme " $X \in S$?", où S est un ensemble, Comparez $H(X)$ avec le nombre moyen de questions nécessaires pour déterminer X.

   (b) Soit Y le nombre de lancements jusquà l'occurence d'une deuxième face. Par exemple, $Y = 5$ si la deuxième face est obtenue au cinquième lancement. Argumentez que $H(Y) = H(X_1 + X_2) < H(X_1, X_2) = 2H(X)$, et interpretez cette relation.

résolution

(a) Soit p la probabilité d'une face dans chaque lancement de la pièce ($p = 1/2$ pour une pièce équilibrée). La loi du nombre de lancements jusquà l'occurrence de la première face est la *loi géométrique* :

$$p(X = n) = p(1-p)^{n-1}, n = 1, 2, \ldots$$

L'entropie de cette loi est, par définition

$$H(X) = -\sum_{n=1}^{\infty} p(1-p)^{n-1} \log\left(p(1-p)^{n-1}\right) = -\sum_{n=0}^{\infty} p(1-p)^{n} \log\left(p(1-p)^{n}\right)$$

Cette expression peut être écrite comme

$$\begin{aligned} H(X) &= -p\sum_{n=0}^{\infty}(1-p)^n \log\left(p(1-p)^n\right) \\ &= -p\left[\sum_{n=0}^{\infty}(1-p)^n \log p + \sum_{n=0}^{\infty}(1-p)^n (n)\log(1-p)\right]. \end{aligned}$$

Finalement, l'entropie est:

$$H(X) = -p\left[\log p\frac{1}{p} + \log(1-p)\frac{1-p}{p^2}\right] = \frac{-p\log p - (1-p)\log(1-p)}{p} = \frac{H(p)}{p}.$$

Ce résultat peut être interprété de la façon suivante. Chaque lancement a une entropie $H(p)$ associée. Comme en moyenne nous devons faire $\frac{1}{p}$ lancements pour obtenir une face, l'entropie totale doit être $\frac{H(p)}{p}$.
Une bonne séquence de questions est celle où la quantité d'information associée à chaque question, sachant les réponses aux questions préc édantes, est maximale (1 bit, pour des questions "oui/non"), c'est à dire, dont les deux réponses possibles sont également probables. Considérons que la *pièce est équilibrée*, $p = 1/2$. Dans ce cas, la suite de questions

"$X = 1$" ?

"$X = 2$" ?

$\vdots$

"$X = n$" ?

$\vdots$

a effectivement une entropie maximale, car sachant que $x \notin [1, n-1]$ les probabilités pour que $X = n$ et $X > n$ sont les mêmes (1/2).
Le nombre de questions nécessaires pour déterminer la valeur de $X = x$ est $k = x$, et donc le nombre moyen de questions nécessaires est $E[k] = E[X] = \frac{1}{p} = 2$. Nous pouvons constater que cette valeur coincide avec l'entropie $H(X) = H(p)/p = 2$.

(b) La variable aléatoire Y est la somme du nombre de lancements jusqu'à l'obtention d'une face X_1 plus le nombre de lancements *additionnels* n écessaires pour avoir la deuxième face. Comme le résultat de chaque lancement est statistiquement indépendant des autres, X_1 etX_2 sont des variables aléatoires statistiquement indépendantes, et donc

$$H(X_1, X_2) = H(X_1) + H(X_2|X_1) = H(X_1) + H(X_2) = 2H(X)$$

où nous avons utilisé le fait que X_1 et X_2 sont identiquement distribuées. Par la règle de chaîne pour l'entropie conjointe

$$\begin{aligned} H(Y) &= H(Y, X_1) - H(X_1|Y) \\ &\stackrel{(a)}{<} H(Y, X_1) = H(X_1) + H(Y|X_1) = H(X_1) + H(X_1 + X_2|X_1) \\ &\stackrel{(b)}{=} H(X_1) + H(X_2|X_1) \\ &\stackrel{(c)}{\leq} H(X_1) + H(X_2) = 2H(X) \end{aligned}$$

où les différents pas ont les justifications suivantes:

(a) car $H(X_1|Y) > 0$: la valeur de X_1 n'est pas complètement déterminée par Y (mais elle n'est pas indépendante non plus, et donc l'in égalité est stricte),

(b) car quand X_1 est connue, il ne reste que l'incertitude associée à la valeur de X_2,

(c) car X_1 et X_2 sont statistiquement indépendantes,

ce qui vérifie l'ennoncé du problème.

6. Un tournoi consiste dans une séquence de trois combats et termine aussitôt un des joueurs gagne deux combats. Soit X la variable aléatoire qui représente le résultat d'un tournoi entre les joueurs A et B, par exemple AA, ou BAB. Soit Y le nombre de combats effectués. Y peut prendre les valeurs 2 et 3.

   (a) Sous l'hypothèse que les joueurs A et B sont équilibrés et que les combats sont indépendants, calculez $H(X)$, $H(Y)$, et $H(X|Y)$.

   (b) Soit Z l'équipe gagnante. Déterminez $H(X|Z)$. Comparez avec $H(X)$. Déterminez $H(Z|X)$.

   (c) Déterminez $I(Y;Z)$.

<u>résolution</u>

La Figure 5 représente l'arbre binaire qui décrit les séquences posssibles dans ce problème (les feuilles de l'arbre). Pour chaque feuille, nous indiquons les valeurs correspondantes des variables X, Y et Z.

(a) Si les joueurs sont équilibrés, la probabilité de chaque branche est 1/2. Nous pouvons donc établir la probabilité conjointe de X et Y, résum ée par le tableau suivant

$Y:X$	AA	BB	ABA	ABB	BAA	BAB
2	1/4	1/4	0	0	0	0
3	0	0	1/8	1/8	1/8	1/8

Par un simple calcul de loi marginale, nous obtenons facilement que Y est une variable uniforme binaire, $Y \in \{2, 3\}$, que la loi de X est la suivante

X	AA	BB	ABA	ABB	BAA	BAB
p(X)	1/4	1/4	1/8	1/8	1/8	1/8

et que $p(X|Y = 2)$ est uniforme dans $\{AA, BB\}$ et $p(Y|X = 3)$ est uniforme dans $\{ABA, ABB, BAA, BAB\}$. Par application directe de la définition d'entropie, nous obtenons

$$H(Y) = \log 2 = 1 bit,$$

et

$$H(X) = 2\frac{1}{4}\log 4 + 4\frac{1}{8}\log 8 = \frac{1}{2}2 + \frac{1}{2}3 = 2.5 \text{ bits}.$$

Nous pouvons encore écrire

$$H(X|Y = 2) = \log 2 = 1 \text{ bit} \qquad H(X|Y = 3) = \log 4 = 2\text{bits}.$$

Nous avons donc

$$H(X|Y) = \frac{1}{2}H(X|Y = 2) + \frac{1}{2}H(X|Y = 3) = \frac{1}{2}(1 + 2) = 1.5 \text{ bits}$$

(Vous pouvez obtenir le même résultat à partir de la règle de chaîne pour le pair (X, Y).)

(b) La variable Z est aussi indiquée dans la Figure 5. La loi conjointe de Z et X est

$Z:X$	AA	BB	ABA	ABB	BAA	BAB
A	1/4	0	1/8	0	1/8	0
B	0	1/4	0	1/8	0	1/8

Il est évident que Z est une variable uniforme dans l'ensemble $\{A, B\}$, et donc son entropie est $H(Z) = 1$ bit.
Conditionnellement à la valeur de $Z = z$, la séquence X peut prendre trois valeurs; une correspondant à la suite $X = zz$ de longeur 2 (avec probabilit é 1/2), et les deux autres correspondant à la victoire de l'autre équipe dans le premier ou le deuxième jeu ($\overline{z}z$ et $z\overline{z}$), chacune avec probabilité 1/4. Les entropies conditionnelles sont donc indépendantes de la valeur de z:

$$H(X|Z = z) = \frac{1}{2}\log 2 + \frac{2}{4}\log 4 = 1.5 \text{ bits}.$$

Ce résultat peut être obtenu de la règle de chaîne:

$$H(X, Z) = H(X) + H(Z|X) = H(Z) + H(X|Z)$$

et donc

$$H(X|Z) = H(X) + H(Z|X) - H(Z)$$

Comme Z est une fonction déterministe de X, $H(Z|X) = 0$:

$$H(X|Z) = H(X) - H(Z).$$

En utilisant les valeurs de $H(X)$ de l'alinéa précédante, et la valeur de $H(Z)$ déterminée dans cette alinéa, nous obtenons

$$H(X|Z) = 2.5 - 1 = 1,5 \text{ bits.}$$

(c) L'information mutuelle est

$$I(Y;Z) = H(Z) - H(Z|Y) = H(Y) - H(Z|Y).$$

Nous avons déjà vu que $H(Z) = H(Y) = 1$ bit. La loi conjointe de Z et Y est représentée dans le tableau suivant .

$Z : Y$	2	3
A	1/4	1/4
B	1/4	1/4

D'où nous concluons imédiatement que les deux variables sont statistiquement indépendantes:

$$P_{Z,Y}(z,y) = P_Z(z)p_Y(y),$$

et donc les entropies conditionnelles sont égales à l'entropie non-conditionnelles:

$$H(Z|Y) = H(Z), \qquad H(Y|Z) = H(Y).$$

L'information mutuelle est donc nulle, comme une analyse intuitive du problème sugère (si les équipes sont équilibrées, savoir le nombre de jeux n'apporte pas d'information sur quelle équipe a gagné) :

$$I(Z;Y) = 0.$$

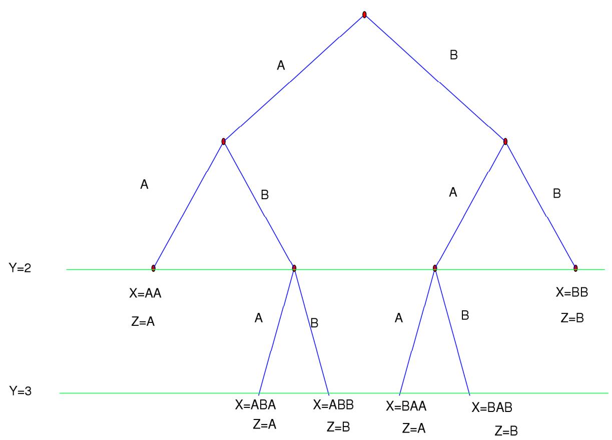

Figure 5: Représentation du tournoi par un arbre binaire.