Master CSI 1 — 2018-2019

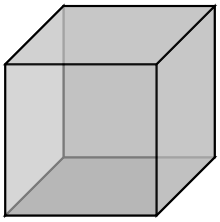
Théorie de l'information : DS du 24 octobre 2018

Master Sciences et Technologies, mention Mathématiques ou Informatique, parcours Cryptologie et Sécurité informatique

Responsable : Gilles Zémor

Durée : 1h30. Sans document. Les exercices sont indépendants.

– Exercice 1.



On tire au hasard avec la loi uniforme un sommet X d'un cube. Puis on choisit aléatoirement un sommet voisin Y de X, uniformément parmi ses trois voisins possibles. Que vaut l'information mutuelle $I(X,Y)$?

– **Solution.** On a : $H(X) = \log_2 8 = 3 = H(Y)$, $H(Y|X) = \log_2 3$, d'où $I(X,Y) = H(Y) - H(Y|X) = 3 - \log_2 3 \approx 1.42$ shannons.

– Exercice 2. On suppose que A et B sont deux événements d'un même espace probabilisé Ω, tels que $A \cap B = \emptyset$ et $P(A) = P(B) = 1/4$. Soient X et Y les variables aléatoires $\Omega \to \{0, 1, -1\}$ définies par :

$$X(\omega) = \begin{cases} 1 & \omega \in A \\ -1 & \omega \in B \\ 0 & \text{sinon} \end{cases} \qquad \text{et} \qquad Y(\omega) = \begin{cases} -1 & \omega \in A \\ 1 & \omega \in B \\ 0 & \text{sinon} \end{cases}$$

a) Les variables X et Y sont-elles indépendantes ?

b) Montrer que $H(X) = H(Y) = H(X,Y) = I(X,Y)$. Que vaut cette valeur commune ? Que valent $H(X|Y)$ et $H(Y|X)$?

c) Soit $Z = XY$. Montrer que $H(X|Z) = H((X,Y)|Z)$. Que vaut cette valeur ?

– **Solution.**

a) Non. On a $P(X = 1) = P(A) = 1/4$, de même $P(Y = 1) = P(B) = 1/4$, mais $P(X = 1, Y = 1) = P(A \cap B) = 0 \neq P(X = 1)P(Y = 1)$.

b) On a $P(X=1)=P(X=-1)=1/4$ et $P(X=0)=1/2$. D'où

$$H(X)=\frac{2}{4}\log_2 4+\frac{1}{2}\log_2 2=\frac{3}{2}.$$

Un calcul similaire donne $H(Y)=H(X)=\frac{3}{2}$.
Par ailleurs,

$$\begin{aligned}P(X=1,Y=-1)&=P(A)=1/4\\P(X=-1,Y=1)&=P(B)=1/4\\P(X=0,Y=0)&=P((A\cup B)^c)=1/2.\end{aligned}$$

Le couple (X,Y) suit donc la même loi $(1/4,1/4,1/2)$ que X et Y et $H(X,Y)=H(X)=H(Y)$. Donc $I(X,Y)=H(X)+H(Y)-H(X,Y)=H(X)$. La formule $H(X,Y)=H(X)+H(Y|X)=H(Y)+H(X|Y)$ donne donc $H(Y|X)=H(X|Y)=0$.

c) On a :

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

et comme la valeur de Y est entièrement déterminée par celle de X, on a $H(X,Z)=H((X,Y),Z)$. Donc $H(X|Z)=H((X,Y)|Z)$.
La loi de Z est $(1/2,1/2)$, donc $H(Z)=1$ et comme Z est entièrement déterminée par X on a $H(X,Z)=H(X)$. Donc

$$H(X|Z)=H(X)-H(Z)=\frac{3}{2}-1=\frac{1}{2}.$$

– Exercice 3. Soient X et Y deux variables indépendantes, à valeurs entières, la variable X prenant ses valeurs dans un ensemble de n entiers dénoté A, et la variable Y prenant ses valeurs dans un ensemble de n entiers également, dénoté B. Ceci sous-entend que $P(X=a)$ et $P(Y=b)$ sont toutes les deux non nulles pour tout $a\in A$ et $b\in B$. Si on suppose que $H(X+Y)=H(X)+H(Y)$, quel est le cardinal de l'ensemble $A+B=\{a+b, a\in A, b\in B\}$?

– **Solution.** Comme X et Y sont supposées indépendantes, on a $H(X,Y)=H(X)+H(Y)=H(X+Y)$. Donc $H((X,Y)|X+Y)=0$, ce qui implique qu'il n'y a jamais deux couples (a,b) et (a',b') distincts dont la somme vaut la même valeur, sinon on ne pourrait pas toujours déterminer (X,Y) à partir de $X+Y$. Donc $|A+B|=|A\times B|=|A||B|$.

– Exercice 4. Le code $\{10, 00, 11, 110\}$ est-il uniquement déchiffrable ?

– **Solution.** Oui. On peut décomposer une plage de 1 consécutifs suivie d'un 0 , soit $1 \cdots 10$, ainsi : si le nombre de 1 est supérieur ou égal à trois, la plage ne peut que commencer par le symbole 11. Ensuite il reste soit 110 soit 10 suivant la parité de nombre de 1 dans la plage. Une fois tous les 1 traités, il ne reste plus qu'à intercaler les symboles 00 ce qui se fait clairement sans ambiguité.

– Exercice 5.

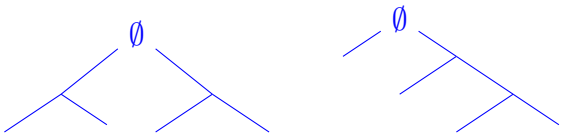
a) Soit C un code de Huffman à quatre mots. Quelles sont les possibles distributions des longueurs $\ell_1 \leqslant \ell_2 \leqslant \ell_3 \leqslant \ell_4$ de C ?

b) Donner un exemple de loi de probabilité $p_1 \geqslant p_2 \geqslant p_3 \geqslant p_4$ pour laquelle toutes les distributions de longueurs ci-dessus correspondent à un code de Huffman pour cette loi.

c) Caractériser l'ensemble des lois avec cette propriété.

– **Solution.**

a) Il n'y en n'a que deux possibles, $(2, 2, 2, 2)$ et $(1, 2, 3, 3)$ qui correspondent aux deux arbres binaires saturés à quatre feuilles.



b) $(p_1, p_2, p_3, p_4) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6})$. Il y a deux choix après la première étape de l'algorithme de Huffman, qui mènent aux deux arbres distincts.

c) Posons $p'_3 = p_3 + p_4$. Pour que les deux choix soient possibles après la première étape, c'est-à-dire associer p_2 et p'_3 ou bien associer p_1 et p_2, il faut que $p'_3 = p_1$. Il s'agit donc de caractériser les lois (p_1, p_2, p_3, p_4) telles que $p_1 \geqslant p_2 \geqslant p_3 \geqslant p_4$ et telles que $p'_3 = p_1$. Comme $p_1 + p_2 + p'_3 = 1$ on en déduit $p_2 = 1 - 2p_1$, donc, en écrivant $\frac{1}{2}p'_3 \leqslant p_2 \leqslant p_1$,

$$\frac{1}{2}p_1 \leqslant 1 - 2p_1 \leqslant p_1$$

soit

$$\frac{1}{3} \leqslant p_1 \leqslant \frac{2}{5}.$$

On trouve donc l'ensemble de lois :

$$(p_1, p_2, p_3, p_4) = (p_1, 1 - 2p_1, \frac{p_1}{2} + t, \frac{p_1}{2} - t)$$

avec

$$\frac{1}{3} \leqslant p_1 \leqslant \frac{2}{5}$$
$$0 \leqslant t \leqslant 1 - \frac{5}{2}p_1.$$

– Exercice 6. On considère une source $X_1, X_2, \ldots, X_k \ldots$ constituée de variables indépendantes de même loi. Cette loi est celle d'une variable prenant quatre valeurs avec probabilités $1/2, 1/4, 1/8, 1/8$.

a) Quelle est la longueur moyenne d'un encodage optimal, en bits, de la suite X_i. Quelle est la particularité de cette longueur moyenne ?

b) On considère maintenant la suite des bits encodés $b_1, b_2, \ldots, b_n \ldots$. Évaluer de manière approchée la quantité

$$\frac{H(b_1, b_2, \ldots, b_n)}{n}$$

et sa limite quand n tend vers l'infini.

c) En déduire $H(b_n)$, pour n grand, ainsi que la loi de b_n.

– **Solution.**

a) $\overline{\ell} = 14/8 = 7/4$. Elle est égale à l'entropie de X_i.

b) Comme $(b_1, \ldots, b_n)$ permet de reconstituer la source, on a $H(b_1, b_2, \ldots, b_n) = H(X_1, \ldots, X_k)$ où $(b_1, b_2, \ldots, b_n)$ encode $(X_1, \ldots, X_k)$. On a $n \approx k\overline{\ell}$, donc

$$\frac{H(b_1, b_2, \ldots, b_n)}{n} \approx \frac{H(X_1, \ldots, X_k)}{k\overline{\ell}}.$$

Comme les X_i sont indépendantes on a $H(X_1, \ldots, X_k) = kH(X_1)$, et comme $\overline{\ell} = H(X_1)$ on en déduit :

$$\frac{H(b_1, b_2, \ldots, b_n)}{n} \approx 1.$$

c) Comme $H(b_1, b_2, \ldots, b_n) \leqslant H(b_1) + \cdots + H(b_n)$ et que, les variables b_i étant binaires, $H(b_i) \leqslant 1$, on déduit de la question précédente qu'on doit avoir, pour n grand $H(b_n) \approx 1$. La loi de b_n ne peut donc que tendre vers la loi uniforme $(1/2, 1/2)$.