Master CSI 1 2013-2014

Théorie de l'information : DS du 22 octobre 2013

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

Responsable : Gilles Zémor

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

– Exercice 1. Soient X et Y deux variables indépendantes, toutes deux de loi uniforme dans $\mathcal{X} = \{0, 1, 2, 3\}$. Soit Z la variable de Bernoulli qui vaut 0 si $X > Y$ et 1 sinon.

a) Calculer $H(Z|X)$, $H(Z|Y)$ et $H(Z)$.

b) En déduire $H(X|Z)$ et $H(Y|Z)$.

– **Solution.**

a) On a :

$$P(Z = 0|X = 0) = 0$$
$$P(Z = 0|X = 1) = \frac{1}{4}$$
$$P(Z = 0|X = 2) = \frac{2}{4}$$
$$P(Z = 0|X = 3) = \frac{3}{4}$$

D'où l'on déduit :

$$\begin{aligned} H(Z|X) &= \sum_{i=0}^{3} P(X = i)H(Z|X = i) \\ &= \frac{1}{4}h\left(\frac{1}{4}\right) + \frac{1}{4}h\left(\frac{1}{2}\right) + \frac{1}{4}h\left(\frac{3}{4}\right) \\ &= \frac{1}{4} + \frac{1}{2}h\left(\frac{1}{4}\right) = \frac{5}{4} - \frac{3}{8}\log_2 3 \approx 0.66. \end{aligned}$$

Un calcul similaire donne

$$H(Z|Y) = H(Z|X) = \frac{1}{4} + \frac{1}{2}h\left(\frac{1}{4}\right).$$

Enfin, on a $P(Z = 0) = 6/16 = 3/8$, donc

$$H(Z) = h\left(\frac{3}{8}\right) \approx 0.95.$$

b) On a

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

D'où

$$\begin{aligned} H(X|Z) = H(X) + H(Z|X) - H(Z) &= \log_2 4 + \frac{1}{4} + \frac{1}{2} h\left(\frac{1}{4}\right) - h\left(\frac{3}{8}\right) \\ &= \frac{1}{4} + \frac{5}{8}\log_2 5 \approx 1.7. \end{aligned}$$

On trouve de même $H(Y|Z) = H(X|Z)$.

– Exercice 2. Démontrer que quelles que soient les variables X, Y, Z on a :

$$H(X,Y) + H(Y,Z) \geqslant H(X,Y,Z) + H(Y).$$

On pourra utiliser la relation $H(X|Y) \geqslant H(X|(Y,Z))$.

– **Solution.** On a $H(X,Y,Z) = H(Y,Z) + H(X|(Y,Z)) \leqslant H(Y,Z) + H(X|Y)$, d'où

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

– Exercice 3. Soit $\pi = (p_1, \dots, p_m)$ une loi de probabilité. Soit π' la loi obtenue à partir de π en sélectionant deux coordonnées i et j dans $\{1, 2, \dots, m\}$ et en remplaçant les probabilités p_i et p_j par $(p_i + p_j)/2$ et $(p_i + p_j)/2$.

Montrer que $H(\pi) \leqslant H(\pi')$. On pourra introduire des variables X et X' à valeurs dans $\{1, 2, \dots, m\}$ et de lois π et π', ainsi qu'une variable Y qui vaut 1 si $X, X' \in \{i, j\}$ et 0 sinon, et s'intéresser à $H(X|Y)$ et $H(X'|Y)$.

– **Solution.** Comme $H(\pi) = H(X) = H(Y,X) = H(Y) + H(X|Y)$ et $H(\pi') = H(Y) + H(X'|Y)$, il suffit de démontrer $H(X|Y) \leqslant H(X'|Y)$.

Or,

$$\begin{aligned} P(X = i|Y = 1) &= \frac{P(X = i)}{P(Y = 1)} = \frac{p_i}{p_i + p_j} \quad \text{et} \quad & P(X = i|Y = 0) &= 0 \\ P(X' = i|Y = 1) &= \frac{(p_i + p_j)/2}{p_i + p_j} = \frac{1}{2} \quad \text{et} \quad & P(X' = i|Y = 0) &= 0 \end{aligned}$$

D'où

$$H(X|Y) = P(Y=1)H(X|Y=1) = (p_i+p_j)h\left(\frac{p_i}{p_i+p_j}\right)$$
$$\leqslant (p_i+p_j)h\left(\frac{1}{2}\right) = H(X'|Y).$$

– Exercice 4. Soit une variable aléatoire X prenant les valeurs a, b, c, d, e, f, g de loi de probabilité $(P(X=a), P(X=b), \ldots, P(X=g))$ égale à :

$$\pi = (0.35,\ 0.3,\ 0,2,\ 0.05,\ 0.05,\ 0.03,\ 0.02).$$

a) Trouver un code de Huffman binaire pour X.

b) Trouver la longueur moyenne $\overline{\ell}$ pour ce code et cette loi de probabilité.

c) Quelles sont toutes les distributions des longueurs

$$\ell_1 \geqslant \ell_2 \geqslant \ell_3 \geqslant \ell_4 \geqslant \ell_5 \geqslant \ell_6 \geqslant \ell_7$$

possibles pour tous les codes de Huffman associés à la loi π ?

– **Solution.**

a) On a par exemple, $a \mapsto 0$, $b \mapsto 10$, $c \mapsto 110$, $d \mapsto 1110$, $e \mapsto 11110$, $f \mapsto 111110$, $g \mapsto 111111$.

b) $\overline{\ell} = 0.35 + 0.3 \times 2 + 0.2 \times 3 + 0.05 \times 4 + 0.05 \times 5 + 0.03 \times 6 + 0.02 \times 6 = 2.3$

c) $(6,6,5,4,3,2,1)$, $(5,5,4,3,2,2,2)$, $(4,4,4,4,2,2,2)$, $(5,5,5,5,3,2,1)$.

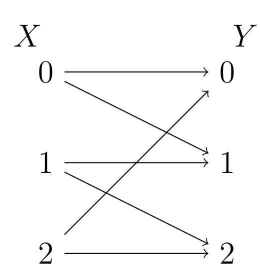
– Exercice 5. On considère les trois codes suivants :

$$C_1 = \{00, 10, 001, 101, 011, 111\}, C_2 = \{00, 10, 01, 101, 110, 011\}, C_3 = \{01, 10, 01110\}.$$

Lesquels sont uniquement déchiffrables ? Pourquoi ?

– **Solution.** Le premier est UD car suffixe (préfixe si l'on lit de droite à gauche). Le deuxième n'est pas UD car ne vérifie pas l'inégalité de Huffman. Le troisième n'est pas préfixe ni suffixe mais est tout de même UD : on peut déchiffrer en isolant d'abord les blocs de trois «1» consécutifs, ce qui identifie le sous-mot 01110, ensuite la lecture se fait de manière unique car il ne survit que des mots de longueur 2.

– Exercice 6. On considère le canal discret sans mémoire :



où les probabilités de transition sont données par :

$$P(Y=0|X=0) = P(Y=1|X=1) = P(Y=2|X=2) = 1-p$$
$$P(Y=1|X=0) = P(Y=2|X=1) = P(Y=0|X=2) = p.$$

Calculer la capacité de ce canal en fonction du paramètre p.

– **Solution.** Pour tout $i \in \{0,1,2\}$ on a $H(Y|X=i) = h_2(p)$, donc

$$H(Y|X) = h_2(p).$$

On a $I(X,Y) = H(Y) - H(Y|X) = H(Y) - h_2(p)$. On vérifie par ailleurs aisément que si X est uniforme alors Y l'est aussi, par conséquent

$$C = \max_{p(X)} (H(Y) - H(Y|X)) = \log_2 3 - h_2(p) \text{ shannons.}$$