Master CSI 1 2010-2011

Théorie de l'information : DS du 18 février 2011

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

– Exercice 1. Soit X_1 une variable aléatoire prenant ses valeurs dans l'ensemble $\{0, 1, \dots, m\}$ et soit X_2 une variable aléatoire prenant ses valeurs dans $\{m+1, m+2, \dots, n\}$. Soit X une variable aléatoire qui vaut X_1 avec probabilité α et X_2 avec probabilité $1-\alpha$. Exprimer $H(X)$ en fonction de $H(X_1)$, $H(X_2)$ et α.

– **Solution.** Définissons la variable de Bernoulli Y qui vaut 0 si $X = X_1$ et 1 si $X = X_2$. On a :

$$\begin{aligned} H(X) &= H(Y) + H(X|Y) = H(Y) + \alpha H(X|Y=0) + (1-\alpha)H(X|Y=1) \\ &= h(\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2). \end{aligned}$$

– Exercice 2. Vous jetez un dé à six faces. Quelle est l'information mutuelle entre la face du dessus et la face latérale la plus proche de vous.

– **Solution.** Soient X la face du dessus et Y la face latérale la plus proche. On a $I(X,Y) = H(X) - H(X|Y)$. La loi de X est uniforme et X prend six valeurs. Connaissant la valeur de Y la loi de X est uniforme mais parmi quatre valeurs seulement. Donc

$$I(X,Y) = \log_2 6 - \log_2 4 = \log_2 3 - 1.$$

– Exercice 3. Soit X une variable aléatoire prenant ses valeurs dans l'ensemble $\{1, 2, \dots, m\}$. Soit T une permutation aléatoire de l'ensemble $\{1, 2, \dots, m\}$. On suppose T indépendante de X. Démontrer que, quelles que soient les lois de X et T :

$$H(T(X)) \geqslant H(X).$$

(On pourra utiliser plusieurs expressions de $H(T(X), T)$).

– **Solution.** Comme T est inversible, la connaissance de T et de $T(X)$ détermine X, donc :

$$H(T(X), T) = H(X, T, T(X)).$$

De même, la connaissance de X et T détermine $T(X)$, donc

$$H(X, T, T(X)) = H(X, T) = H(X) + H(T)$$

puisque X et T sont indépendantes. Par ailleurs

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

d'où l'on déduit :

$$\begin{aligned} H(T(X)) &= H(X) + H(T) - H(T|H(T(X)) \\ &= H(X) + I(T, T(X)) \\ &\geqslant H(X). \end{aligned}$$

– Exercice 4. Soit une variable aléatoire X prenant m valeurs avec une loi $p = (p_1, \ldots, p_m)$. On code cette variable avec un code C préfixe de distribution des longueurs $\ell_1, \ell_2, \ldots, \ell_m$. On suppose $\sum_{i=1}^{m} 2^{-\ell_i} = 1$ et on pose $q = (q_1, q_2, \ldots, q_m)$ avec $q_i = 2^{-\ell_i}$. Montrer que la longueur moyenne du codage de X par le code C vaut :

$$\bar{\ell} = H(p) + \mathrm{D}(p\,||\,q).$$

– **Solution.**

$$\begin{aligned} \bar{\ell} &= \sum_i p_i \ell_i = \sum_i p_i \log_2 \frac{1}{q_i} \\ &= \sum_i p_i \log_2 \frac{p_i}{p_i q_i} \\ &= \sum_i p_i \log_2 \frac{1}{p_i} + \sum_i p_i \log_2 \frac{p_i}{q_i} \\ &= H(p) + \mathrm{D}(p\,||\,q). \end{aligned}$$

– Exercice 5. Parmi les codes suivants

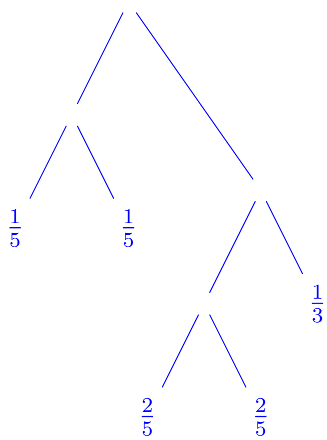
$$\{0, 10, 11\}, \{00, 01, 10, 111\}, \{000, 011, 010, 100, 110, 001\},$$

lesquels ne peuvent être produits par l'algorithme de Huffman quelle que soit la loi de probabilité associée ? Pourquoi ?

– **Solution.** Les trois codes sont des codes préfixes. Le premier code est un code de Huffman. Les deux autres ne le sont pas car les codes de Huffman sont optimaux, or quelle que soit la loi associée aux deux autres codes, il est possible de diminuer la longueur moyenne sans détruire la propriété préfixe en remplaçant le mot 111 par 11 pour le deuxième code et le mot 110 par 11 pour le troisième code

– Exercice 6. Trouver l'arbre de Huffman binaire pour une source X de loi $p = (\frac{1}{3}, \frac{1}{5}, \frac{1}{5}, \frac{2}{15}, \frac{2}{15})$. Montrer pourquoi le code que vous trouvez est également optimal pour la loi $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$. Quelles sont les longueurs moyennes de ce code pour ces deux lois ?

– **Solution.** L'arbre de Huffman est celui-ci.



La longueur moyenne pour cette loi est $34/15$. Pour la loi uniforme la longueur moyenne est $12/5$.