Master CSI 1 — 2022-2023

Théorie de l'information : DS du 7 mars 2023

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.

a) Soit $X = (X_0, X_1, X_2, X_3, X_4)$ la variable aléatoire de loi uniforme et prenant ses valeurs parmi les cinq décalés circulaires du quintuplet binaire (10100). Calculer $H(X_0)$ et $H(X_i|X_0)$ pour $i \neq 0$.

b) Que vaut $H(X_iX_{i+1}X_{i+2})$ (où la somme des indices s'entend modulo 5) ? En déduire, sans faire de calcul supplémentaire, la valeur de $H(X_{i+2}|X_iX_{i+1})$.

– EXERCICE 2. Soit la suite de variables aléatoires de Bernoulli $X_0, X_1, \ldots, X_i$ construite de la manière suivante : X_0 est une variable de loi $P(X_0 = 1) = 1/3, P(X_0 = 0) = 2/3$. Pour définir les variables $X_i, i \geqslant 1$, on utilise une suite auxiliaire $Z_1, Z_2, \ldots, Z_i, \ldots$ où les Z_i sont des variables de Bernoulli indépendantes et uniformes, donc telles que $P(Z_i = 1) = P(Z_i = 0) = 1/2$. Pour $i \geqslant 1$, la variable X_i est définie ainsi :

— si $X_{i-1} = 1$ alors $X_i = 0$
— si $X_{i-1} = 0$, alors on pose $X_i = Z_i$.

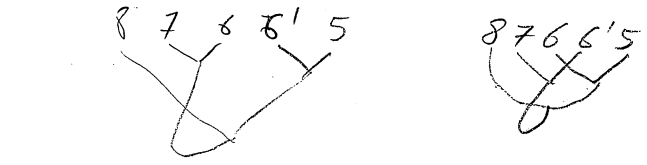
a) Calculer la loi de $X_1, X_2, \ldots$ et en déduire $H(X_i)$ pour tout i.

b) Pour $i \geqslant 1$, calculer $H(X_0, X_1, \ldots, X_i)$. Quelle est la limite de $\frac{1}{i}H(X_0, \ldots X_{i-1})$ lorsque $i \to \infty$?

c) Montrer que $(X_i, X_{i+1}, X_{i+2}, X_{i+3})$ suit la même loi quel que soit i. Si $n = 4k$, que devient la longueur moyenne de $X_0, X_1, \ldots, X_{n-1}$ si on coupe la suite en k blocs de taille 4, $(X_0, X_1, X_2, X_3), (X_4, X_5, X_6, X_7) \ldots$ et si on applique un codage de Huffman sur chaque bloc ?

– EXERCICE 3. On considère une variable aléatoire X prenant ses valeurs dans l'ensemble $\{a, b, c, d, e, f, g\}$ avec la loi de probabilité

$$p = (p_a = \frac{1}{4}, p_b = \frac{7}{32}, p_c = \frac{3}{16}, p_d = \frac{5}{32}, p_e = \frac{3}{32}, p_f = \frac{1}{16}, p_g = \frac{1}{32}).$$



Quelles sont les valeurs possibles de la distribution des longueurs

$$(\ell_a, \ell_b, \ell_c, \ell_d, \ell_e, \ell_f, \ell_g)$$

d'un code de Huffman pour cette loi ? Donner un code de Huffman correspondant pour chacun de ces cas. Quelle en est la longueur moyenne ?

– EXERCICE 4. On considère l'ensemble des lois $p = (p_1, p_2, p_3, p_4)$ vérifiant la propriété $p_1 \geqslant 2p_2 \geqslant 2p_3 \geqslant 2p_4$, que l'on appelera $(\star)$.

a) Montrer que $p_1 \geqslant 2/5$.

b) A quelle loi π pensez-vous pour maximiser la valeur de son entropie parmi les lois vérifiant la propriété $(\star)$?

c) Montrer que votre hypothèse est vérifiée en écrivant que $H(\pi) - H(p)$ est supérieure à une divergence de Kullback pour tout p vérifiant $(\star)$.

– EXERCICE 5. Soit le quadruplet $X = (X_1, X_2, X_3, X_4)$ de variables aléatoires choisi uniformément dans l'ensemble des 24 permutations de $(1, 2, 3, 4)$. On se propose de déterminer X (en d'autres termes de trier les quatre entiers) en n'effectuant que des comparaisons deux à deux. On cherche le nombre moyen de comparaisons d'un algorithme optimal (qui minimise donc le nombre moyen de comparaisons).

a) On commence par supposer $X_1 \leqslant X_2$ et $X_3 \leqslant X_4$. Faire la liste des 6 possibilités et proposer un algorithme de tri sous forme d'un arbre binaire qui est égal à un arbre de Huffman pour la loi uniforme sur un ensemble à six éléments. Quel est le nombre moyen de comparaisons de votre algorithme ?

b) En déduire le nombre moyen de comparaisons d'un algorithme optimal de tri de quatre entiers.