Master CSI 1 — 2017-2018

Théorie de l'information : DS du 7 novembre 2017

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 à pile ou face trois fois de suite. Quelle est l'information mutuelle $I(X,Y)$ entre le nombre X de «face» et le nombre Y de paires de «face» consécutifs (Y prend donc ses valeurs dans l'ensemble $\{0,1,2\}$).

– EXERCICE 2. Soit $\mathcal{X} = \{1,2,\dots,n\}$ et soit X une variable aléatoire à valeurs dans $\mathcal{X}$. Soient f et g deux permutations de $\mathcal{X}$, i.e. deux bijections de $\mathcal{X}$ dans $\mathcal{X}$. Soit Y une variable de Bernoulli telle que $P(Y=0) = P(Y=1) = 1/2$ et supposée indépendante de X. Soit Z la variable définie par $Z = f(X)$ si $Y = 0$ et $Z = g(X)$ si $Y = 1$.

Montrer que $H(X) \leqslant H(Z)$. Donner un exemple de variable X et de fonctions f et g pour lesquelles $H(Z) = H(X) + 1$.

– EXERCICE 3.

a) Soient $(a_i)_{1\leqslant i\leqslant m}$ et $(b_i)_{1\leqslant i\leqslant m}$ deux suites finies de m entiers strictement positifs. On pose

$$a = \sum_{i=1}^{m} a_i \quad \text{et} \quad b = \sum_{i=1}^{m} b_i.$$

Utiliser la formule $\mathrm{D}(p\,||\,q) \geqslant 0$ pour montrer l'inégalité

$$\sum_{i=0}^{m} a_i \log \frac{a_i}{b_i} \geqslant a \log \frac{a}{b}.$$

b) Soit $p = (p_1,\dots,p_n)$ et $q = (q_1,\dots,q_n)$ deux lois de probabilité définies sur $\mathcal{X} = \{1,2,\dots,n\}$. Soit $A = (A_j)_{1\leqslant j\leqslant k}$ une partition de $\mathcal{X}$, c'est-à-dire que l'on a

$$\mathcal{X} = A_1 \cup A_2 \cup \cdots A_k$$

pour des A_j disjoints. On définit les lois de probabilité p_A et q_A sur $\{1,2,\dots,k\}$ par

$$p_A(j) = \sum_{i\in A_j} p_i$$

$$q_A(j) = \sum_{i\in A_j} q_i.$$

Montrer que $\mathrm{D}(p\,||\,q) \geqslant \mathrm{D}(p_A\,||\,q_A)$. A quelle condition a-t-on égalité ?

c) Si $p = (p_1, \ldots, p_n)$ et $q = (q_1, \ldots, q_n)$ sont deux lois définies sur $\mathcal{X} = \{1, 2, \ldots, n\}$, on définit la distance statistique $d(p, q)$ entre p et q par

$$d(p,q) = \sum_{i=1}^{n} |p_i - q_i|.$$

Soit $A = (A_1, A_2)$ où $A_1 = \{i \in \mathcal{X}, p_i \geqslant q_i\}$ et $A_2 = \{i \in \mathcal{X}, p_i < q_i\}$. Montrer que $d(p, q) = d(p_A, q_A)$.

d) L'inégalité de Pinsker affirme que pour toutes lois p et q on a :

$$\mathrm{D}(p \,||\, q) \geqslant \frac{1}{2} d^2(p, q).$$

Montrer que pour démontrer l'inégalité de Pinsker, il suffit de la démontrer pour des lois de Bernoulli.

– EXERCICE 4. Soit π la loi de probabilité $\pi = (1/3, 1/3, 1/9, 1/9, 1/27, 1/27, 1/27)$. Pouvez-vous donner un code préfixe *ternaire*, i.e. constitué de mots sur l'alphabet $\{0, 1, 2\}$, dont la longueur moyenne associée à une source X de loi π égale l'entropie ternaire $H_3(X) = H(X)/\log 3$?

– EXERCICE 5. Existe-t-il un code préfixe ternaire constitué de deux mots de longueur 1, d'un mot de longueur 2, et de 7 mots de longueur 3 ?

– EXERCICE 6. On dispose d'un lot de 12 pièces numérotées de 1 à 12 d'apparences identiques. Toutes ont le même poids connu, soit de 5g (disons), sauf une pièce défectueuse dont le poids est légèrement différent. On dispose d'une balance moderne qui donne le poids avec précision de n'importe quelle quantité posée dessus. On nous donne l'information supplémentaire que la pièce défectueuse est, avec probabilité 1/2, dans l'ensemble $\{1, 2, 3, 4\}$.

Quel est le minimum de l'espérance du nombre de pesées nécessaire pour déterminer la pièce défectueuse ? Comment faut-il procéder pour minimiser ce nombre moyen de pesées ? On pourra donner la procédure sous forme d'un arbre.