Master CSI 1 2009-2010

Théorie de l'information, MHT 813 : Examen du 23 avril 2010

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

Responsable : Gilles Zémor

Durée : 3h. Sans document. Les exercices sont indépendants.

– EXERCICE 1. Soit une loi de probabilité $(p_1, p_2, \ldots, p_n)$ où l'on a ordonné les p_i par ordre décroissant, soit $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n$.

a) On suppose que $p_1 > 2/5$. Démontrer que l'algorithme de Huffman binaire attribue toujours au symbole de probabilité p_1 un mot binaire de longueur au plus 2.

b) Examiner la situation plus finement et montrer que l'algorithme de Huffman attribue toujours au symbole de probabilité p_1 un mot binaire de longueur 1.

c) On suppose maintenant que $p_1 < 1/3$. Démontrer que que l'algorithme de Huffman binaire attribue toujours au symbole de probabilité p_1 un mot binaire de longueur au moins 2.

– EXERCICE 2. Supposons que la variable aléatoire X prend ses valeurs dans l'ensemble $\mathcal{X} = \{1, 2, 3, 4, 5\}$. On considère deux lois de probabilités $p = (p_x)_{x\in\mathcal{X}}$ et $q = (q_x)_{x\in\mathcal{X}}$ données par le tableau suivant :

Symbole	p_x	q_x	$C_1(x)$	$C_2(x)$
1	$\frac{1}{2}$	$\frac{1}{2}$	0	0
2	$\frac{1}{4}$	$\frac{1}{8}$	10	100
3	$\frac{1}{8}$	$\frac{1}{8}$	110	101
4	$\frac{1}{16}$	$\frac{1}{8}$	1110	110
5	$\frac{1}{16}$	$\frac{1}{8}$	1111	111

Les deux dernières colonnes du tableau représentent deux codes binaires C_1 et C_2

a) Calculer $H(p), H(q), D(p||q)$ et $D(q||p)$.

b) Les codes C_1 et C_2 sont-ils uniquement déchiffrables ? Montrer que C_1 est un code optimal pour la loi p et que C_2 est un code optimal pour la loi q.

c) Quelle est la longueur moyenne du codage par C_2 lorsque la loi est p ? Quelle est la longueur moyenne du codage par C_1 lorsque la loi est q ? Que vaut l'excès par rapport aux longueurs optimales ? Quel est le rapport avec la question **a)** ?

– Exercice 3.

a) Un émetteur transmet toutes les secondes à un destinataire un symbole binaire à travers un canal binaire symétrique de probabilité de transition p. Combien de shannons par seconde reçoit le destinataire ?

b) L'émetteur décide maintenant d'utiliser un code de Hamming binaire de paramètres $[7, 4, 3]$, c'est-à-dire qu'il continue d'émettre un symbole binaire par seconde, mais sous la forme d'un mot du code de Hamming toutes les sept secondes. Le destinataire décode chaque mot du code de Hamming au maximum de vraisemblance. Calculer, en fonction de p, la probabilité qu'un mot reçu soit correctement décodé.

c) Calculer la capacité d'un canal q-aire symétrique où $q = 16$ et de probabilité de transition P, c'est-à-dire d'un canal avec 16 entrées, 16 sorties, et tel que $P(Y = x|X = x) = 1 - P$ et $P(Y = y|X = x) = P/15$ pour tous x, y, $y \neq x$.

d) Combien de shannons par seconde reçoit le destinataire avec le procédé de codage de la question **b)** ? Comparer avec **a)** : pourquoi la perte était-elle inévitable ?

– Exercice 4. X est une variable aléatoire qui prend ses valeurs dans $\{0, 1, 2\}$. Un canal discret sans mémoire prend la variable X en entrée et rend en sortie la variable $Y = X + Z$ où la variable Z est indépendante de X et est uniformément distribuée dans l'ensemble $\{a, b\}$ où a et b sont des entiers distincts. Suivant les valeurs de a, b, calculer

a) la capacité maximale de ce canal : on donnera des valeurs de a, b qui atteignent ce maximum, ainsi que la loi de X associée.

b) la capacité minimale de ce canal : donner de même des valeurs de a, b qui atteignent ce minimum, ainsi que la loi de X associée.

– Exercice 5. Soit C_h le code de Hamming de matrice génératrice :

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

$\mathbb{F}_2$ désigne le corps à deux éléments. Soit $V = \mathbb{F}_2^7 \times \mathbb{F}_2^7$ l'espace vectoriel des matrices binaires 7×7. On considère le sous-espace vectoriel de V constitué des matrices dont toutes les lignes et toutes les colonnes sont des mots de C_h. On note cet espace vectoriel $C = C_h \otimes C_h$ et on considère C comme un code linéaire, de longueur donc égale à $7 \times 7 = 49$.

a) si x et y sont des vecteurs de $\mathbb{F}_2^7$, on note $x \otimes y$ la matrice dont la coordonnée (i, j) vaut $x_i y_j$. Montrer que si x et y sont des mots de C_h alors $x \otimes y$ est un mot de C.

b) Montrer que si c est un mot de C et que si on a $c_{ij} = 0$ pour tout $(i, j) \in \{1, 2, 3, 4\}^2$, alors c est la matrice nulle.

c) Montrer que si a et b sont des lignes de $\mathbf{G}$, alors $a \otimes b$ est une matrice dont toutes les coordonnées $(i, j) \in \{1, 2, 3, 4\}^2$ sont nulles, sauf une.

d) Trouver une base de C et en déduire la dimension du code C.

e) Trouver la distance minimale d de C.

f) Montrer que les mots de C de poids minimum sont forcément de la forme $x \otimes y$ où x et y sont des mots de C_h de poids minimum.

g) Combien y a-t-il de mots de poids d dans C ?

h) On considère l'algorithme de décodage suivant.
- Décoder en parallèle toutes les lignes de la matrice reçue, en appliquant l'algorithme de décodage standard du code C_h.
- puis décoder en parallèle toutes les colonnes de la matrice issue de l'étape de décodage précédente en appliquant l'algorithme de décodage standard du code C_h.

Montrer que cet algorithme décode correctement n'importe quelle configuration de 3 erreurs.

i) Montrer que l'algorithme de décodage de la question précédente ne décode pas toutes les configurations de 4 erreurs.