Master CSI 1 — 2008-2009

Théorie de l'information, MHT 813 : Examen du 24 avril 2009

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. Quels sont les arbres qui sont associés à un code de Huffman binaire ?

– Exercice 2. On prend le n-uple ordonné $(1, 2, \ldots, n)$ et on le perturbe aléatoirement en tirant un numéro au hasard et en le réinsérant au hasard dans la suite. Par exemple, pour $n = 10$, on produit $(1, 2, 3, 7, 4, 5, 6, 8, 9, 10)$ en retirant 7 de sa place initiale entre 6 et 8 et en l'insérant entre 3 et 4.

Quelle est l'entropie du n-uple résultant ?

– **Solution.** Soit X la variable aléatoire décrivant le n-uple produit par la procédure. Il s'agit de décrire l'ensemble N des n-uples perturbés possibles, de trouver la loi de X, i.e. la valeur de $p_x = P(X = x)$ pour tout $x \in N$, puis d'appliquer la formule $H(X) = \sum_x p_x \log 1/p_x$.

La perturbation consiste à choisir un couple (i, j) d'entiers entre 1 et n, où i décrit le numéro tiré au hasard et j décrit l'emplacement où le numéro est replacé. Par exemple pour $n = 5$, le choix $(i, j) = (2, 1)$ produit le 5-uple $(2, 1, 3, 4, 5)$: le choix $(i, j) = (5, 3)$ produit le 5-uple $(1, 2, 5, 3, 4)$, et $(1, 5)$ donne $(2, 3, 4, 5, 1)$. Le choix de (i, j) est uniforme parmi les n^2 choix possibles.

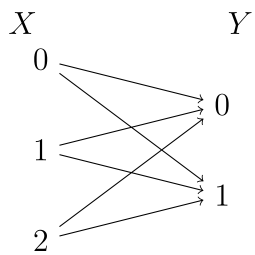
Il faut réaliser que
- tous les choix (i, i) produisent le même n-uple $(1, 2, \ldots, n)$.
- les choix $(i, i + 1)$ et $(i + 1, i)$ produisent le même n-uple.
- n'importe quel autre choix (i, j) produit un n-uple unique.

On en déduit que
- un n-uple apparait avec probabilité $n/n^2 = 1/n$,
- $n - 1$ n-uples apparaissent avec probabilité $2/n^2$,
- $n^2 - n - 2(n - 1) = (n - 1)(n - 2)$ n-uples apparaissent avec probabilité $1/n^2$.

Donc l'entropie du n-uple résultant vaut :

$$
\begin{aligned}
H(X) &= \frac{1}{n}\log_2 n + (n-1)\frac{2}{n^2}\log_2\frac{n^2}{2} + (n-1)(n-2)\frac{1}{n^2}\log_2 n^2 \\
&= (2-\frac{1}{n})\log_2 n - \frac{2}{n} + \frac{2}{n^2}.
\end{aligned}
$$

– Exercice 3. On considère le canal discret sans mémoire :

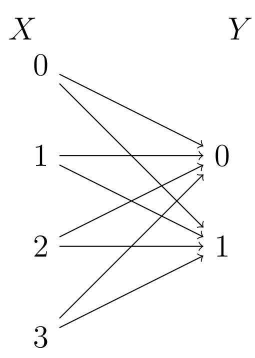


où les probabilités de transition sont données par

$$
\begin{aligned}
P(Y=1|X=0) = P(Y=1|X=1) = P(Y=0|X=2) = p \\
P(Y=0|X=0) = P(Y=0|X=1) = P(Y=1|X=2) = 1-p
\end{aligned}
$$

pour un certain paramètre p. Calculer la capacité de ce canal.

– Exercice 4. On considère le canal discret sans mémoire :



où les probabilités de transition sont données par

$$
\begin{aligned}
P(Y=1|X=0) &= p & \qquad P(Y=0|X=0) &= 1-p \\
P(Y=0|X=1) &= p & \qquad P(Y=1|X=1) &= 1-p \\
P(Y=1|X=2) &= p & \qquad P(Y=0|X=2) &= 1-p \\
P(Y=0|X=3) &= p & \qquad P(Y=1|X=3) &= 1-p
\end{aligned}
$$

pour un certain paramètre p.

a) Calculer, en fonction de p, la capacité de ce canal.

b) En déduire, dans le cas où la loi de X est uniforme, la valeur de $H(X|Y)$.

– EXERCICE 5. Soit C un code linéaire binaire défini par la matrice de parité $\mathbf{H}$ suivante :

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

a) Quels sont les paramètres $[n, k, d]$ (longueur, dimension, distance minimale) de ce code ?

b) Démontrer qu'un quelconque vecteur de $\{0,1\}^{16}$ de poids 3 se transforme de manière unique en un mot de code de poids 4 en changeant un «0» en un «1». En déduire le nombre de mots de poids 4 du code C.

c) Par une démarche analogue, trouver le nombre de mots de poids 6 de ce code.

d) Montrer que n'importe quel vecteur de $\{0,1\}^{16}$ est à distance de Hamming au plus 2 d'un mot de C.

e) Combien y a-t-il de vecteurs de $\{0,1\}^{16}$ qui ne sont ni des mots de code ni à distance de Hamming 1 d'un mot de C ?

f) Soit $\mathbf{x}$ un vecteur de $\{0,1\}^{16}$ qui n'est ni un mot de C, ni à distance de Hamming 1 d'un mot de C. Montrer qu'il existe 8 mots de C à distance de Hamming 2 de $\mathbf{x}$.

g) Quels sont les paramètres du code dual $C^{\perp}$ de C ?

h) On reçoit le vecteur suivant avec cinq coordonnées effacées :

$$[????01011101110?].$$

Montrer que le mot du code C coïncidant avec les coordonnées non effacées est unique et le trouver.

i) On efface aléatoirement et avec une loi uniforme quatre coordonnées d'un mot $\mathbf{c}$ du code C. Calculer la probabilité qu'il soit possible de décoder et de retrouver $\mathbf{c}$ sans ambiguïté.

j) Montrer que le code C peut corriger simultanément une erreur et un effacement dans n'importe quelle paire $\{i, j\}$ de positions.

k) Soit σ la fonction syndrome associée à $\mathbf{H}$,

$$\begin{array}{rcl} \{0,1\}^{16} & \longrightarrow & \{0,1\}^{5} \\ \mathbf{x} & \mapsto & \mathbf{H}\,{}^t\mathbf{x}. \end{array}$$

Soit $\mathbf{x} = [x_1 \dots x_{16}]$ un vecteur aléatoire uniforme de $\{0,1\}^{16}$. Quel est le nombre minimum de coordonnées x_i qu'il faut connaître pour avoir un bit d'information (un shannon) sur la valeur du syndrome $\sigma(\mathbf{x})$? Trouver un ensemble minimal de coordonnées x_i dont la connaissance procure deux bits d'information sur la valeur de $\sigma(\mathbf{x})$.

– **Solution.**

a) $[16, 11, 4]$.

b) un mot de poids 3 a pour syndrome la somme de trois colonnes de $\mathbf{H}$: le syndrome a donc un 1 (somme de trois 1) en dernière position. On en déduit qu'il s'agit d'une colonne de $\mathbf{H}$. Il suffit de modifier la coordonnée correspondante du mot de poids 3 pour obtenir un mot de code. Cette coordonnée est forcément une coordonnée qui passe de «0» à «1» sinon on obtiendrait un mot de poids 2 or $d = 4$.

On en déduit que le nombre de mots de poids 4 égale $\frac{1}{4}\binom{16}{3} = 140$.

c) De même que précédemment, un mot de poids 5 est exactement à distance 1 d'un mot de code. Si on enlève tous les 140×12 mots de poids 5 qui sont à distance 1 d'un mot de code de poids 4, il reste $\binom{16}{5} - 140.12 = 2688$ mots de poids 5 qui sont tous à distance 1 d'un mot de code de poids 6. Il y a donc $2688/6 = 448$ mots de code de poids 6.

d) Il suffit de voir que n'importe quel vecteur non nul s de $\{0,1\}^5$ est somme d'au plus 2 colonnes de $\mathbf{H}$. Ceci est manifestement le cas, si le dernier bit de s est «1», c'est déjà une colonne de $\mathbf{H}$, sinon on lui ajoute la dernière colonne de $\mathbf{H}$ et on trouve une colonne de $\mathbf{H}$.

e) Comme $d = 4$, les boules de rayon 1 centrées autour des mots de code sont disjointes. Leur réunion contient $2^{11}(1 + 16)$ mots. La réponse est donc $2^{16} - 2^{15} - 2^{11} = 2^{15} - 2^{11} = 30720$.

f) Un mot qui n'est ni un mot de code ni à distance 1 d'un mot de code est à distance 2 d'un mot de code et a pour syndrome s un vecteur de $\{0,1\}^5$ qui se termine par un 0. La modification d'une coordonnée quelconque donne un syndrome qui se termine par un 1 : un tel syndrome est une colonne de $\mathbf{H}$ qui détermine donc de manière unique la deuxième coordonnée à modifier pour obtenir un mot de code. Il y a 16 manières de procéder ainsi, mais modifier la coordonnée i puis la coordonnée j donne le même mot que si l'on modifie la coordonnée j puis la coordonnée i. On tombe donc sur $16/2 = 8$ mots de code possibles.

g) $[16, 5, 8]$.

h) $[0011010111011101]$.

i) Il s'agit de la probabilité de ne pas tomber sur le support d'un mot de code de poids 4, c'est-à-dire d'après la question **b)**

$$1 - \frac{140}{\binom{16}{4}} = \frac{12}{13}.$$

j) Le code raccourci, c'est-à-dire déduit de C en «oubliant» une coordonnée est manifestement un code linéaire de distance minimale $4 - 1 = 3$. Le code raccourci où on a oublié la coordonnée effacée peut donc corriger une erreur. Une fois l'erreur corrigée on peut corriger l'effacement sans ambiguïté.

k) Il faut intercepter le support d'un mot du code engendré par les lignes de $\mathbf{H}$, soit au minimum 8 coordonnées x_i. Si on veut deux bits d'information il faut intercepter la réunion du support de deux mots de code non nuls, soit au minimum 12 symboles. Par exemple tous les x_i sauf x_3, x_4, x_{15}, x_{16}.