Théorie de l'Information

Devoir # 3

25 octore 2006

SIC-SICOM

Maria-João Rendas

1. Lequel de ces codes ne peut pas être un code de Huffman?

   (a) $\{1, 01, 00\}$

   (b) $\{00, 01, 10, 110\}$

   (c) $\{01, 10\}$

   Justifiez toutes vos réponses.

2. Considérez la variable aléatoire avec la loi de probabilité dans le tableau suivant:

$x \in \mathcal{X}$	p(x)
x_1	0.5
x_2	0.26
x_3	0.11
x_4	0.04
x_5	0.04
x_6	0.03
x_7	0.02

   (a) Construisez un code de Huffman pour X.

   (b) Calculez la longueur moyenne du code obtenu.

   (c) Construisez un code ternaire pour X.

3. Des mots comme "Stop" ou "Feu" sont petits, pas car leur utilisation est fréquente, mais peut-être car on souhaite miniser le temps nécessaire pour les dire. Supposez que $X = i$ avec probabilité $p_i, i = 1, \ldots, m$. Soit ℓ_i le nombre de bits néessaires pour coder $X = i$, et c_i le coût par lettre du mot X_i. Le coût moyen du code est donc

$$C = \sum_{i=1}^{m} p_i \ell_i c_i$$

   (a) Minimisez C défini par l'équation précédante sur toutes les longueurs $\{\ell_i\}_{i=1}^{m}$, telles que l'inégalité de Kraft est satisfaite:

$$\sum_{i=1}^{m} 2^{-\ell_i} \leq 1.$$

Ingnorez, pour cette minimisation, les contraintes qui imposent que les ℓ_i soient des entiers. Calculez les longueurs optimales $\ell_i^\star$, et le coût minimal qui leur est associé, $C^\star$.

(b) Comment pouvez-vous utiliser l'algorithme de Huffmann pour minimiser C sur tous les codes (bianires) uniquement décodables ? Soit $C_{\textit{Huffmann}}$ le coût de ce code optimal.

(c) Montrez que

$$C^\star \leq C_{\textit{Huffmann}} \leq C^\star + \sum p_i c_i.$$

4. Soit $\mathcal{X} = \{1, 2, 3, 4, 5\}$. Considérez les deux lois de probabilité p et q définies sur cet alphabet, et les deux codes C_1 et C_2 dans le tableau suivant

$x \in \mathcal{X}$	$p(x)$			
1	0.5	0.5	0	0
2	0.25	0.125	10	100
3	0.125	0.125	110	101
4	0.0625	0.125	1110	110
5	0.0625	0.125	1111	111

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

(b) Vérifiez que C_1 est optimal pour p et que C_2 est optimal pour q (calculez leurs longueurs moyennes).

(c) Admettez que l'on utilise C_2 pour une source $X \sim p$. Quelle est la longueur moyenne des mots de code? De combien elle dépasse l'entropie $H(p)$?

(d) Quelle est la pénalité si nous utilisons C_1 pour une source $X \sim q$?

5. On nous donne 6 bouteilles de vin. On sait qu'une des bouteilles est mauvaise. L'analyse visuelle des bouteilles permet de déterminer la probabilité pour que chacune d'entre elles soit celle qui est mauvaise:

bouteille # i	p_i
1	7/26
2	5/26
3	4/26
4	4/26
5	3/26
6	3/26

Nous souhaitons déterminer la mauvaise bouteille.
Supposez que l'on goute les vins un à un. Choisissez l'ordre par laquelle ils doivent être goutés pour que le nombre d'essais soit minimisé. Rappelez-vous que si les premières 5 bouteilles sont bonnes, vous ne devez pas gouter la dernière.

(a) Quel est le nombre moyen d'essais qui doivent être réalisés?

(b) Quelle bouteille doit être testée en premier?

Admettez maintenant que vous pouvez gouter le mélange du contenu de plusieurs bouteilles, au lieu de les gouter une à une.

(c) Quelle est la valeur minimale du nombre moyen de tests qui doivent être faits pour déterminer la mauvaise bouteille?

(d) Quel mélange doit être testé en premier ?