MA331
Théorie de l'information et Codage

Nicolas Barbot

9 Janvier 2020

Le barème est donné à titre très indicatif

1 Entropie

Une source produit un caractère x à partir d'un alphabet $\mathcal{A} = \{0, 1, \ldots, 9, a, b, \ldots, z\}$ avec :
— une probabilité 1/3 si x est une voyelle (a, e, i, o, u et y) ;
— une probabilité 1/3 si x est un chiffre ;
— une probabilité 1/3 si x est une des 21 consonnes ;
Pour chaque ensemble (voyelle, chiffre ou consonne), la probabilité de sélectionner un caractére donné est uniforme.

1. (1 point) Donner la probabilité de l'événement a, b et 4.
2. (2 points) Déterminer l'entropie de la source.
3. (1 point) Cette source produit une séquence de 1 millon de caractères. Déterminer la taille en octet du fichier généré par un codeur de source idéal.

2 Capacité

L'entrée X d'un canal est une entree binaire dont les symboles sont dans $\{0, 1\}$. La sortie Y est ternaire et ses symboles sont dans $\{1, e, 0\}$. La distribution de probabilite de l'entrée X est $\{p(X = 0) = \alpha, p(X = 1) = 1 - \alpha\}$. La matrice de transition du canal est :

$$P = \begin{pmatrix} 1 - \beta - \epsilon & \epsilon \\ \beta & \beta \\ \epsilon & 1 - \beta - \epsilon \end{pmatrix}$$

1. (2 points) Déterminer la distribution de probabilité de la sortie Y en fonction de α, β et ϵ (c'est à dire $P(Y = 0)$, $P(y = e)$ et $P(Y = 1)$). Calculer l'entropie de la sortie $H(Y)$.
2. (1 point) Calculer $H(Y|X)$ et en déduire l'information mutuelle transmise par ce canal.
3. (1 point) Montrer que $H(Y)$ est maximale lorsque $\alpha = 1/2$. En déduire la capacite du canal.
4. (½ point) Calculer la capacité du canal lorsque $\beta = 0$.
5. (½ point) Calculer la capacité du canal lorsque $\epsilon = 0$.

3 Codage de Huffman

Une source X binaire produit des symboles dans l'alphabet $\mathcal{A}_X = \{0; 1\}$ avec les probabilités $\mathcal{P}_X = \{0.9; 0.1\}$.

1. (2 points) Déterminer le code de Huffman de l'extension à l'ordre 2 et 3 de cette source (en considérant des symboles de longueur 2 et 3).
2. (2 points) Déterminer l'entropie de la source, $H(X)$, ainsi que l'entropie des sources étendues $H(X^2)$ et $H(X^3)$.
3. (1 point) Discuter la performance des l'algorithmes utilisés. Comparer aux autres solutions de codage de source que vous connaissez.

4 Codes linéaires

Soit $\mathcal{A} = \{0, 1\}$ Soit le code linéaire binaire $C(15, 4)$ défini par la fonction d'encodage suivante

$$\begin{array}{cccc} \Phi : & \mathcal{A}^4 & \to & \mathcal{A}^{15} \\ & c_1c_2c_3c_4 & \mapsto & c_1c_2c_3c_4\ r_1...r_{11} \end{array}$$

où :

$$\begin{array}{ccc} r_1 = c_3 + c_4 & r_5 = c_1 + c_4 & r_9 = c_1 + c_2 + c_4 \\ r_2 = c_2 + c_4 & r_6 = c_1 + c_3 & r_{10} = c_1 + c_2 + c_3 \\ r_3 = c_2 + c_3 & r_7 = c_1 + c_3 + c_4 & r_{11} = c_1 + c_2 + c_3 + c_4 \\ r_4 = c_2 + c_3 + c_4 & r_8 = c_1 + c_2 & \end{array}$$

On admettra que la distance minimale de $\mathcal{C}$ est $\delta(\mathcal{C}) = 6$

1. (1 point) Combien d'erreurs peut-on détecter, et combien peut-on en corriger ?
2. (1 point) Ce code est-il systématique ?
3. (1 point) Ce code est-il parfait ?
4. (2 points) Donner une matrice de contrôle de $\mathcal{C}$.
5. (2 points) On reçoit les messages :

$$m_r = 1111\ 00010010110$$

$$m'_r = 0001\ 01101100110$$

   m_r et m'_r appartiennent-ils à $\mathcal{C}$, si non les corriger.
6. (2 points (bonus)) Dans le cas d'un canal binaire symétrique de probabilité d'erreur binaire $p_e = 10^{-1}$, calculer la probabilité d'erreur mot (WER) et la probabilité d'erreur bit (BER).
   On prendra comme approximation de $(1 - p_e)^n \simeq 1 - np_e$.
   Est-il préférable d'utiliser un code de Hamming $C(7, 4)$ dans ce cas ?