5 Codage de source

5.1 Mauvais code source

Lesquels des codes suivants ne peuvent pas être des codes de Huffman ?

1. Code composé des 3 mots $0, 10, 11$.
2. Code composé des 4 mots $00, 01, 10, 110$.
3. Code composé des 2 mots $01, 10$

5.2 Codage de source binaire et ternaire

Soit une source simple (sans mémoire) S sur un alphabet A_S composé de 5 lettres,
$A_S = \{s_1, s_2, s_3, s_4, s_5\}$.

1. Soit le code binaire $\{00, 11, 010, 111, 1010\}$:
   (a) Est-il à décodage unique (déchiffrable) ?
   (b) Est-il instantané ?
2. Soit le code binaire $\{00, 11, 010, 011, 101\}$:
   (a) Est-il instantané ?
   (b) Existe-t-il un code binaire instantané plus court ? (Justifier)
3. On s'intéresse maintenant à un codage ternaire de la source. On suppose que le jeu de probabilités des symboles de S est $P_S = \{1/3, 1/3, 1/9, 1/9, 1/9\}$.
   (a) Pour un codage ternaire à mots de longueur fixe l, quel serait le choix de l ?
   (b) Calculer l'entropie et la redondance de la source.
   (c) En déduire la longueur moyenne minimale d'un code source ternaire instantané.
   (d) Pour la source S :
      i. Proposer un code de Huffman (donner l'arbre).
      ii. Calculer sa longueur moyenne et son efficacité. Commenter les résultats obtenus.
      iii. Peut-on en déduire une propriété de la suite des symboles après codage ?

5.3 Longueurs des mots d'un code de Huffman

Soit une v.a. X qui prend 4 valeurs $\{1; 2; 3; 4\}$ avec les probabilités $\left\{\frac{1}{3}; \frac{1}{3}; \frac{1}{4}; \frac{1}{12}\right\}$.

1. Donner un code de Huffman pour cette v.a..
2. Montrer qu'il existe deux ensembles de longueurs de mots de code optimaux qui sont : $(1, 2, 3, 3)$ et $(2, 2, 2, 2)$.
3. En conclure qu'il existe des codes optimaux avec des longueurs de mots de code qui dépasse la longueur de code de Shannon $\left\lceil \log_2 \frac{1}{p(x)} \right\rceil$ pour certains symboles.