Source Coding

1. Codage de Huffman. [**CT 5.4**] La source X peut émettre 7 symboles différents, de probabilités $(0.49, 0.26, 0.12, 0.04, 0.04, 0.03, 0.02)$.

a) Construire un code de Huffman binaire pour cette source, et calculer sa longueur moyenne.

b) Construire un code de Huffman ternaire pour cette source, et calculer sa longueur moyenne.

2. Codage de Huffman (bis). [**CT 5.5**] Construire un code de Huffman binaire (lettres 0 et 1) pour une source X pouvant émettre 5 symboles différents $\{e_1, \ldots, e_5\}$, de probabilités respectives $(1/3, 1/5, 1/5, 2/15, 2/15)$. Expliquez pourquoi ce code est aussi optimal pour les probabilités $(1/5, 1/5, 1/5, 1/5, 1/5)$.

3. Mauvais codes. [**CT 5.6**] Lequel parmi ces codes ne peut pas être un code de Huffman, quelles que soient les probabilités des symboles codés ? Justifier votre réponse.

a) $\{0, 10, 11\}$

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

c) $\{01, 10\}$

4. La borne supérieure $H(X) + 1$ bits ne peut être améliorée si le codage est effectué symbole par symbole. Soit X une variable de Bernoulli de paramère p. Notons C_H le code de Huffman associé à cette source, et L la longueur moyenne de ce code. Soit $\epsilon > 0$. Montrez qu'il existe p tel que $L > H(X) + 1 - \epsilon$.
L'encadrement de la longueur moyenne d'un code (symbole par symbole) ne peut être amélioré. Les bornes dépendent de la loi de la variable aléatoire.

5. Codes ternaires atteignant $H_3(X)$. [**CT 5.10**] Une variable aléatoire X peut prendre m valeurs différentes. Son entropie en base 2 est $H(X)$, et donc son entropie en base 3 est donnée par $H_3(X) = H(X)/\log_2 3$ (rappel : cela signifie que le log de la formule donnant l'entropie est en base 3). On suppose que l'on a su construire un code ternaire pour X dont la longueur moyenne L atteint la borne inférieure : $L = H_3(X)$.

a) Montrez que toute valeur possible e_i pour X a une probabilité p_i du type 3^{-n_i}, où $n_i \in \mathbb{N}$. C'est à dire que X a une loi de probabilité triadique.

b) Montrez que cela n'est possible que pour m impair.

6. Comparaison de Shannon et de Huffman. [**CT 5.12**] Une variable aléatoire X peut prendre 4 valeurs, avec la loi de probabilité $(1/3, 1/3, 1/4, 1/12)$.

a) Contruisez un code de Huffman (binaire) pour X et calculez la longueur moyenne de ce code.

b) Soit un code à longueur fixe (tous les mots de code ont la même longueur) associé à la source X. Comparez la longueur moyenne du code à longueur fixe à celle du code de Huffman.

c) Quelles seraient les longueurs des mots d'un code de Shannon associé à la source X ? En conclure qu'un code de Huffman peut être plus long que le code de Shannon sur certains mots, et pourtant rester meilleur en moyenne.

7. Trous dans l'inégalité de Kraft. [**CT 5.3**] On suppose qu'un code préfixe a été construit sur le jeu de longueurs $l_1, \ldots, l_m$, et qu'il vérifie l'inégalité de Kraft au sens strict : $\sum D^{-l_i} < 1$. L'alphabet de ce code D-aire est $\mathcal{D} = \{0, 1, \ldots, D-1\}$. Montrez qu'il existe des éléments de $\mathcal{D}^*$ qui ne peuvent pas être décodés sous la forme de suites de mots de code .