Codage de source

Problème 1

Pour une région donnée, les prévisions d'un météorologiste se répartissent selon les fréquences relatives données par le tableau ci-dessous. Les colonnes correspondent au temps effectif, que l'on représente par la variable aléatoire T, prenant pour valeur 0 ou 1 selon qu'il fait *mauvais temps* (0) ou *beau temps* (1). Les lignes correspondent à la prévision du météorologiste, que l'on identifie par la variable aléatoire M, également à valeurs dans $\{0, 1\}$ selon qu'il avait prévu du *mauvais temps* (0) ou du *beau temps* (1).

$P(M = i, T = j)$	beau temps ($T = 1$)	mauvais temps ($T = 0$)
beau temps ($M = 1$)	5/8	1/16
mauvais temps ($M = 0$)	3/16	1/8

1. Calculer les probabilités marginales $P(M = i)$ et $P(T = j)$, avec $i, j \in \{0, 1\}$.

2. Montrer que le météorologiste se trompe 1 fois sur 4.

3. Un étudiant éclairé annonce qu'en prévoyant toujours du beau temps, il se trompe moins souvent que le météorologiste. Vérifier cette assertion.

4. On souhaite archiver le temps T en adoptant un codage binaire. En utilisant le premier théorème de Shannon, donner la place mémoire moyenne minimale qu'il faut prévoir, en bits par réalisation de T.

5. Refaire le calcul précédent dans le cas du stockage de M. Quelle place mémoire totale minimale représente le stockage de M et de T, en bits par réalisation du couple (M, T) ?

6. On souhaite à présent coder *conjointement* M et T. Calculer $H(M;T)$. En déduire la place mémoire moyenne minimale qu'il faut prévoir, en bits par réalisation du couple (M, T).

7. Interpréter la différence obtenue entre les résultats des 2 questions précédentes.

8. Proposer un codage de Huffman pour coder les couples (M, T).

9. Calculer la longueur moyenne $\overline{n}$ des mots du code binaire conçu à la question précédente. Comme on pouvait s'y attendre, quelle double inégalité vérifie $\overline{n}$?

Exercice 1

Indiquer pour chacun des codes suivants s'il est régulier, déchiffrable, instantané et complet : $\mathcal{C}_1 = \{00, 01, 10, 11\}$, $\mathcal{C}_2 = \{0, 01, 11\}$, $\mathcal{C}_3 = \{0, 10, 11\}$, $\mathcal{C}_4 = \{0, 11, 111\}$.

Exercice 2

On considère une source S pouvant émettre 5 symboles, dont la probabilité p_i de chacun figure dans le tableau ci-dessous. Ce dernier fournit également deux codages binaires possibles $\mathcal{C}_1$ et $\mathcal{C}_2$ de S. Indiquer si ces codes sont déchiffrables et instantanés. Calculer la longueur moyenne $\overline{n}_1$ et $\overline{n}_2$ de leurs mots. Comparer ces valeurs à la longueur moyenne minimum $\overline{n}_{\text{min}}$ des mots de tout codage binaire de S.

s_i	s_1	s_2	s_3	s_4	s_5
p_i	0.50	0.18	0.14	0.12	0.06
$\mathcal{C}_1$	0	10	11	101	1001
$\mathcal{C}_2$	00	10	11	010	011

Exercice 3

Une table traçante utilise les commandes suivantes

lever la plume (LP)
baisser la plume (BP)
transfert avec incrémentation à gauche (–X)
transfert avec incrémentation à droite (+X)
transfert avec incrémentation en haut (+Y)
transfert avec incrémentation en bas (–Y).

Quel est le nombre moyen minimum de bits requis pour ce jeu de commandes, sachant que les probabilités respectives des différents états sont données par

$$P_{\mathrm{LP}} = P_{\mathrm{BP}} = P_{-\mathrm{X}} = 0.1 \qquad P_{+\mathrm{X}} = 0.3 \qquad P_{+\mathrm{Y}} = P_{-\mathrm{Y}} = 0.2$$

Construire un code binaire de Shannon. Utiliser la technique de Huffman pour élaborer un autre code binaire. Comparer les deux solutions obtenues.

Exercice 4

Un lycée doit communiquer par voie télématique une liste de résultats au baccalauréat concernant 2500 étudiants. Ces résultats sont les suivants : 250 mentions bien, 375 mentions assez-bien, 1125 mentions passable, 625 insuffisants et 125 absents. Établir un code de Huffman binaire pour compresser le fichier. Calculer la longueur moyenne des mots utilisés. Calculer la taille du fichier si l'on codait l'information de manière classique, à l'aide d'un code à longueur fixe comprenant 8 bits. Évaluer le gain en taille de fichier réalisé grâce au code de Huffman. Calculer le temps de transmission du fichier si l'on utilise un modem fonctionnant à 14400 bits/s.

Problème 2

On considère un code comprenant deux mots de longueur 2, deux mots de longueur 3 et un mot de longueur 4.

1. Par l'inégalité de McMillan, montrer qu'il existe un code binaire déchiffrable respectant ces longueurs de mots. Dessiner un arbre de codage possible. Modifier celui-ci de sorte à réduire la longueur moyenne des mots du code.

2. On donne les probabilités suivantes $\{0.50, 0.18, 0.14, 0.12, 0.06\}$ à chacun des 5 états d'une source. Associer ces probabilités aux mots du code proposé à la question précédente de sorte à minimiser la longueur moyenne de codage. Calculer celle-ci et montrer qu'il existe des codes binaires plus performants.

3. Proposer un code binaire à l'aide de la méthode de Huffman. Comparer la longueur moyenne de ses mots à celle obtenue à la question précédente.