Université de Caen Normandie, UFR des Sciences
Licence 3

Module " Algorithmique, structures informatiques et cryptologie"
Théorie de l'information

Exercice 1 : Source, entropie et codes

Considérons la source sans mémoire $\mathcal{S}$ sur l'alphabet $\{A, C, G, T\}$ dont les fréquences sont données par

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(C) = \mathbb{P}(G) = \mathbb{P}(T) = \frac{1}{6}.$$

1. Calculez l'entropie de $\mathcal{S}$.

Considérons les quatre codes suivants de la source $\mathcal{S}$:

Symbole	A	C	G	T
Code 1	00	01	10	11
Code 2	0	10	110	111
Code 3	1	01	001	0001
Code 4	0	01	10	11

2. Calculez la longueur moyenne de ces codes vis à vis de la source $\mathcal{S}$. Quel code choisiriez-vous en pratique ?
3. Construisez les arbres binaires associés à chaque code.
4. Quels sont les codes préfixes ? Quels sont les codes décodables ?

Exercice 2 : Codes de Huffman (exam 2017)

On considère la variable aléatoire X sur l'alphabet $\{1, 2, \ldots, 6\}$ dot la distribution est donnée par

$$P(X = 1) = P(X = 2) = 0.25, \qquad P(X = 3) = 0.20, \qquad P(X = 4) = P(X = 5) = P(X = 6) = 0.1.$$

1. Donnez le code de Huffman associé à X.
2. Quelle est la longueur moyenne de ce code ?

Exercice 3 : Session 2 (2017)

On considère une source à deux symboles $S = \{a, b\}$ avec pour probabilité $p_a = 0.1$ et $p_b = 0.9$.

1. Calculer l'entropie de la source et trouver un code optimal pour cette source dont vous donnerez la longueur moyenne.
2. On considère maintenant la source $S' = S \times S \times S$ formée de triplets de symbole de S Les probabilités sont données par $p_{xyz} = p_x p_y p_z$ pour $(x, y, z) \in S^3$.
   Calculer les probabilités de chaque triplet et en déduire l'entropie de la source.
3. Appliquer l'algorithme de Huffman pour déterminer un code optimal pour S'. Vous déterminerez sa longueur moyenne. Quel code est le plus efficace en comparant avec le code de la question 1.

Exercice 4 : Codes de Huffman et canal

On considère une source sans mémoire qui produit un bit par unité de temps selon la distribution $p_0 = 3/4$ et $p_1 = 1/4$. On souhaite faire passer ce flot de bit à travers un canal qui ne peut faire passer que 0.82 bits par unité de temps (ou de manière équivalente, 82 bits toutes les 100 unités de temps).

Trouvez un code de Huffman qui permet de compresser le flot de la source afin qu'il puisse passer par le canal (il faudra considérer des blocs de bits).

Exercice 5 : Canal binaire symétrique

Soit X une source sans mémoire sur l'alphabet binaire $\{0, 1\}$ donnée par la distribution :

$$\mathbb{P}[X = 0] = \frac{1}{8}, \qquad \mathbb{P}[X = 1] = \frac{7}{8}.$$

Considérons le canal symétrique binaire $\mathcal{C}$ de paramètre $p = \frac{1}{16}$.

1. Si X est la source en entrée du canal $\mathcal{C}$, et Y la source en sortie du canal, donnez la loi conjointe de (X, Y).
2. En déduire la loi de Y. Est-ce la même que celle de X ?
3. Quelle est l'entropie de X et de Y ? Quelle est l'entropie conjointe de (X, Y) ?
4. En déduire l'information mutuelle $I(X, Y)$ et les entropies conditionnelles $H(X|Y)$ et $H(Y|X)$.
5. Donnez la capacité du canal $\mathcal{C}$.

Exercice 6 : Canal symétrique

Soit X une source sans mémoire sur l'alphabet ternaire $\{0, 1, 2\}$ donnée par la distribution :

$$\mathbb{P}[X = 0] = \frac{1}{8}, \qquad \mathbb{P}[X = 1] = \frac{3}{8}, \qquad \mathbb{P}[X = 2] = \frac{4}{8}.$$

Considérons le canal ternaire $\mathcal{C}$ dont la matrice de transition est donnée par

$$\mathcal{M} = \begin{pmatrix} 3/4 & 1/8 & 1/8 \\ 1/8 & 3/4 & 1/8 \\ 1/8 & 1/8 & 3/4 \end{pmatrix}.$$

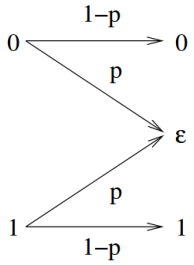
1. Si X est la source en entrée du canal $\mathcal{C}$, et Y la source en sortie du canal, donnez la loi conjointe de (X, Y).
2. En déduire la loi de Y. Est-ce la même que celle de X ?
3. Quelle est l'entropie de X et de Y ? Quelle est l'entropie conjointe de (X, Y) ?
4. En déduire l'information mutuelle $I(X, Y)$ et les entropies conditionnelles $H(X|Y)$ et $H(Y|X)$.
5. Soit X' une source sans mémoire sur l'alphabet $\{0, 1, 2\}$. Montrez que l'information mutuelle $I(X', Y')$ avec Y' la source à la sortie du canal vérifie

$$I(X', Y') = H(Y') - H(\mathcal{M}) \quad \text{avec} \quad H(\mathcal{M}) = -\frac{3}{4}\log_2\frac{3}{4} - 2 \times \frac{1}{8}\log_2\frac{1}{8}.$$

6. Donnez la capacité du canal $\mathcal{C}$.

Exercice 7 : exam 2018 (5pts)

Considérons le canal binaire à effacement $\mathcal{C}$ suivant, de paramètre $p \in [0,1]$:



1. Donnez la matrice de transition du canal $\mathcal{C}$. (0,5pt)

Soit X une source sans mémoire sur l'alphabet $\{0,1\}$ telle que

$$\text{Prob}\,(X=1) = 1 - \text{Prob}\,(X=0) = a.$$

Soit Y la source résultant du passage de la source X à travers le canal $\mathcal{C}$.

2. Calculez en fonction de a et p, les entropies $H(X)$, $H(Y)$, $H(X,Y)$ et l'information mutuelle $I(X,Y)$.
   Pour simplifier les formules, nous posons $h(t) = -t\log_2 t - (1-t)\log_2(1-t)$. (3pts)
3. En déduire la capacité du canal $\mathcal{C}$. (1,5pt)

Exercice 8 : Code de Hamming (7,4)

L'alphabet d'entrée du code de Hamming(7,4) est $\{0,1\}^4$ et l'alphabet de sortie est $\{0,1\}^7$. Si les quatre bits d'entrée sont donnés par $m = (m_1, m_2, m_3, m_4)$, alors les sept bits en sortie $y = (y_1, y_2, \ldots, y_7)$ sont définis par :

$$y_7 = m_4, \quad y_6 = m_3, \quad y_5 = m_2, \quad y_3 = m_1$$

$$y_1 = y_3 + y_5 + y_7 \mod 2, \quad y_2 = y_3 + y_6 + y_7 \mod 2, \quad y_4 = y_5 + y_6 + y_7 \mod 2.$$

Rappelons que les sommes de contrôle précédentes permettent de corriger jusqu'à une erreur. La première (resp. deuxième/troisième) somme de contrôle indique une erreur potentielle à une position d'indice paire (resp. dont le deuxième/troisième bit est à 1) si la somme de contrôle n'est pas vérifiée.

1. Quel est le code associé aux mots suivants :

$$0000, \quad 0001, \quad 0010, \quad 0100, \quad 1000, \quad 1001, \quad 0110$$

2. Y a-t'il une erreur dans les mots de code suivants ? Si oui, corriger la.

$$y = (1,0,1,0,1,1,0), \qquad y = (1,1,1,0,1,1,1), \qquad y = (0,0,1,0,0,1,0).$$

3. La matrice génératrice d'un code linéaire (ce qu'est le code de Hamming) est une matrice G telle que si $m = (m_1, m_2, m_3, m_4)$ est encodée en $y = (y_1, \ldots, y_7)$, alors $G \cdot {}^t m = {}^t y$. Trouvez une expression de G.

4. La matrice de contrôle est une matrice H tel que pour tout mot de code $y = (y_1, \ldots, y_7)$ sans erreur, on a $H \cdot {}^t y = {}^t(0, 0, 0)$. Trouvez une expression de H
5. Calculez $H \cdot G$.
6. Le syndrome d'un mot de code y (avec ou sans erreur) est défini par $s = H \cdot {}^t y$. Donnez le syndrome des y de la question 2.

Exercice 9 : Code de Hamming généraux

En s'inspirant du code de Hamming $(7, 4)$, il est possible de construire des codes de Hamming dont :

— l'alphabet de sortie sont les blocs de $2^n - 1$ bits,
— l'alphabet d'entrée sont les blocs de $m = (2^n - 1) - n$ bits,
— dont les sommes de contrôles C_i sont en position 2^i pour $i = 0, 1, 2, \ldots, n - 1$.

Questions :

1. Construisez le code de Hamming $(15, 11)$ (qui correspond à $n = 4$).
2. Considérons le message codé $y = (1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1)$. Décodez ce message.
3. Quelle est l'efficacité des codes de Hamming $(7, 4)$, $(15, 11)$ et $(31, 26)$? Que se passe-t'il lorsque n tend vers l'infini ?