Université de Caen Normandie, UFR des Sciences
Licence 3

**Module "Protection de données"**
**TD2 - Codage de sources ; durée : 2h30**

### 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 Shannon-Fano et 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 Shannon-Fano associé à $X$. (2pts)
2. Quelle est la longueur moyenne de ce code ? (0,5pt)
3. Donnez le code de Huffman associé à $X$. (2pts)
4. Quelle est la longueur moyenne de ce code ? (0,5pt)

### 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**

Voici un texte sur 6 lettres :

`ABBCADEBBFCAFFECBBBAAFED`

1. Donnez un code de longueur fixe pour ce texte. Quelle est sa longueur moyenne ?
2. Donnez le code de Huffman associé à l'extrait. Quelle est sa longueur moyenne ?
3. Quelle est l'entropie du texte ? Comparez avec les différentes longueurs moyennes trouvées.

**Exercice 5 : 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 6 : Code binaire de longueur fixe**

Voici un extrait de "Le Cid" de Pierre Corneille (1606-1684) :

`o rage o desespoir`

1. Combien de symboles différents y-a-t'il (on considère les espaces) ?
2. Combien de bits sont nécessaires pour encoder un symbole avec un code binaire de longueur fixe ?
3. Donnez un code binaire possible. Donnez le codage de la citation.
4. Quelle est l'entropie $h_1$ de la citation si l'on considère Pierre Corneille comme une source sans-mémoire ?
5. Nous supposons maintenant les symboles émis par couple de lettres. Reprenez les questions 1 à 4 avec cette nouvelle supposition.
6. Notons $h_2$ la nouvelle entropie obtenue. Si Pierre Corneille avait été sans mémoire, quel lien y aurait-il entre $h_1$ et $h_2$ ?
7. Que constatez-vous en terme de compression ?