TP N°01 - TICOD
Notion d'Entropie et Compression de l'Information Par Codage d'Huffman

HALLA Senia - IGE 43 Groupe 02

10 November 2021

1 Mesure de l'information :

1.1 Calcul l'Entropie de la séquence X :

C'est une mesure pour décrire le désordre d'un système physique, dans le cas d'un signal on parle alors de la quantité d'information, on la calcule avec la formule suivante :

$$H(x) = -\sum_{i=1}^{M} p_i.log_m(p_i)$$

Alphabet	A	T	S	U	I	O	N	D	H(X)
Probabilités	$\frac{1}{33}$	$\frac{1}{33}$	$\frac{2}{33}$	$\frac{3}{33}$	$\frac{4}{33}$	$\frac{5}{33}$	$\frac{6}{33}$	$\frac{11}{33}$	2.62

Résultat de l'Entropie H(X) = 2.62 bits/message

2 Observation d'une Densité de Probabilité :

2.1 La fonction Histogram(x,h):

Elle donne l'histogramme de la distribution des symboles de la source "x", avec le nombre bins spécifié par "h" .

2.2 La Sortie du programme pour L = 1, 2, 3 et 4:

Voir Figure 1 - Partie 1 Code

2.3 La séquence de symbole :

La nature de la séquence de symbole est une suite de nombre qui suivent une distribution uniforme

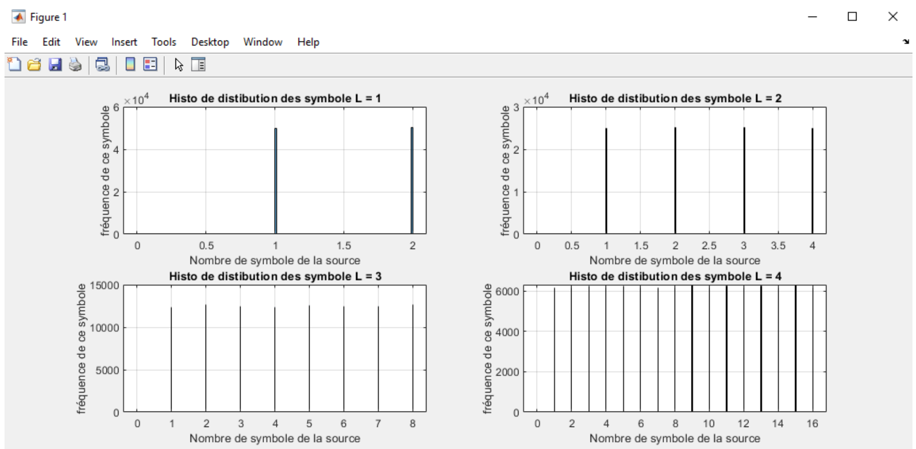

Figure 1: Histogramme Pour différentes valeur de Longeur

3 Mesure d'Entropie :

3.1 Le contenu de P :

Le vecteur P contient la probabilité de l'émission de chaque symbole.

3.2 La quantité d'information I et l'entropie H :

Voir le fichier Matlab.

3.3 Utilisation de Logarithme Base 2 :

On utilise la base 2 pour le calcul de logarithme, parce qu'on cherche le nombre des bits (qui est à la base binaire 2) dans un message.

3.4 Les valeurs de H pour différente valeurs de L :

L = 1 , H(X) = 1.000000 bits/message
L = 2 , H(X) = 1.999955 bits/message
L = 3 , H(X) = 2.999965 bits/message
L = 4 , H(X) = 3.999911 bits/message

4 Programmation d'une Source :

4.1 Comparaison P et P_z :

Le vecteur de probabilités P est presque identique au vecteur de probabilités P_z.

4.2 Source générés :

Une source Uniforme (Gaussienne)

4.3 Comparaison entre les deux Sources x et z :

Pour L = 3 :
H(X) = 2.999965 bits/message
$H_z(X) = 2.061676 bits/message$

On en déduit qu'avec la même source, on a réduit l'entropie de 3 à 2.062 bits/message cela revient à la distribution normale des probabilités.

5 Codage et Compression:

5.1 Commentaires :

Voir le code Matlab

5.2 Les Symboles et leur Codes :

Symbole	Code	Longueur du Code	Probabilité P_i
3	1	1	0.4713
2	01	2	0.2539
4	001	3	0.1250
1	0001	4	0.0624
5	00000	5	0.060
6	000010	6	0.0156
0	0000110	7	0.0078
7	0000111	7	0.0039

$$LMoyz = \sum_{i=1}^{8} p_i.n_i = 2.07960$$

5.3 Le degré de compression :

$\frac{H(Z)}{LMoyz} = 0.99$

5.4 L'arborescencce de Huffman :

Voir Figure 2 :

5.5 La longeur Moyenne du mot-code :

Pour la source z, la langueur moyenne du mot code après le codage Huffman est LMoyz=3 bits/message. (Ainsi que pour la source x)

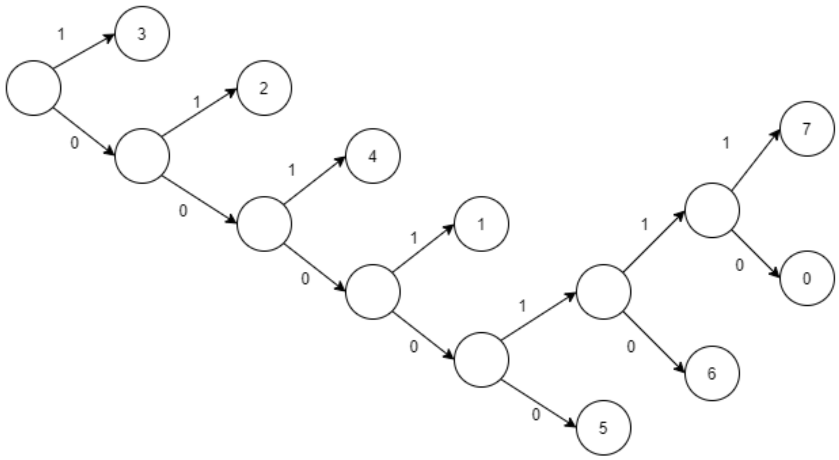

Figure 2:

5.6 L'intérêt de ce type de codage :

Pour une source équi-répartie, il n'y a pas un intérêt d'utiliser ce type de codage puisque on trouvera la même longueur moyenne si on utilise un codage de langueur fixe.