

Séance 1 : compression sans perte (par groupe de 5-6)

Exercice 1 :

On considère une source S pouvant émettre 5 symboles, dont la probabilité p_i de chaque symbole figure dans le tableau ci-dessous. Ce dernier fournit également deux codages binaires possibles C_1 et C_2 de S . Indiquer si ces codes sont uniquement déchiffrables et préfixés. Calculer la longueur moyenne \bar{n}_1 et \bar{n}_2 de leurs mots. Comparer ces valeurs à la longueur moyenne minimum \bar{n}_{\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
C_1	0	10	11	101	1001
C_2	00	10	11	010	011

Exercice 2 :

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

1. 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 quelle que soit la distribution de probabilité.
2. On donne les probabilités suivantes $\{0.45; 0.20; 0.16; 0.14; 0.05\}$ 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.

Exercice 3 :

Il arrive souvent qu'un même entier x (codé sur 1 octet) apparaisse plusieurs fois consécutivement, disons m fois, dans un flux de données. Dans ce cas, il est codé sous la forme (x,m) . Par exemple, on remplace 10 occurrences du caractère 5 par la séquence $(5,10)$. On parle de codage RLE (Run Length Encoding).

1. Coder la séquence avec la méthode RLE :

22222555777777788888811111

2. En pratique, le couple (x,m) est codé sous la forme xm où m est un entier codé sur 5 bits. Dans le meilleur des cas, une séquence de n octets peut se retrouver compressée en combien d'octets ? Dans le pire cas, une séquence de n octets peut se retrouver compressée en combien d'octets ?

3. Nous proposons d'utiliser un mécanisme de quantification. Chaque entier x est d'abord divisé par un entier non nul Q , puis x est ensuite remplacé par l'entier y qui correspond à la partie entière du ratio x/Q . Appliquer cette quantification sur la séquence d'entiers suivante :

23227253575757787828552131211

avec $Q=4$, puis avec $Q=8$.

4. Proposer un mécanisme simple pour estimer la valeur d'origine de x à partir de y . Appliquer ce mécanisme de décodage aux séquences précédentes. Calculer l'erreur quadratique moyenne de la séquence reconstruite par rapport à la séquence initiale pour chaque cas : $Q=4$, puis $Q=8$.
5. Appliquer le codage RLE sur chacune des nouvelles séquences obtenues. Que constate-t-on ?