Travaux dirigés n°2
Codage de la source (1)

Exercice n°1 : Soit le langage représentant les quatre symboles A, C, G, T. On considère le codage suivant :

Symbole	Code
A	0
C	01
G	001
T	101

a) Quels sont les symboles de code ? Quelle est la valence du codage ?
b) Quels sont les valeurs L1, L2, L3, L4 ? Est-ce un codage déchiffrable, préfixe, vérifie t'il l'égalité de Kraft-McMillan ?
c) Soit le codage représentant les quatre symboles A, C, G, T où A est représenté par le mot de code « 0 », C par « 100 », G par « 101 », T par « 111 ».

Quels sont les valeurs L1, L2, L3, L4 ? Est-ce un codage déchiffrable?

Exercice n°2 : Soit X={1,2,3,4,5}. Considérez les deux lois de probabilité p et q définies sur un alphabet, et les deux codes C1 et C2 dans le tableau suivant :

$x \in X$	P(x)	Q(x)	C1	C2
1	0.5	0.5	0	0
2	0.25	0.125	10	100
3	0.125	0.125	110	101
4	0.0625	0.125	1110	110
5	0.0625	0.125	1111	111

a) Calculer H(p), H(q)
b) Vérifier que C1 est optimal pour p et que C2 est optimal pour q (calculer leur longueur moyenne)
c) On suppose qu'on utilise C2 pour coder une source $X \sim p$. Quelle est la longueur moyenne des mots de code. De combien elle dépasse l'entropie ?
d) Quelle est la pénalité si on utilise C1 pour une source $X \sim q$

Exercice n°3 : Soit une source binaire sans mémoire X telle que P(0) = 0.9 et P(1) =0.1.

a) Calculer l'entropie de X.
b) On groupe maintenant les éléments binaires de X par mots de 3. Les mots de la nouvelle source ainsi constituée X^3 sont codés selon le tableau ci-dessous.

Mot	000	001	010	011	100	101	110	111
Code	0	11	101	10011	1000	100101	1001000	1001001

– Calculer la probabilité de chaque mot.
– Quelle est l'entropie de X^3 ?
– Montrer que le code est à préfixes différents.
– Calculer la longueur moyenne des mots de code.
– En déduire l'efficacité.