ENSEIRB **Janvier 2009**
Informatique 1ère année

Théorie de l'information

Documents autorisés : une feuille A4 manuscrite recto-verso.
Calculatrices autorisées. Les exercices sont indépendants.

Exercice 1

On considère une source binaire S, émettant les symboles 0 et 1 avec probabilités respectives p et $1-p$ $(0 < p < 0.5)$. On note S^k la source d'ordre k, émettant des k-uplets successifs de symboles de S ; ainsi, S^2 emet les symboles 00, 01, 10 et 11 avec probabilités respectives p^2, $p(1-p)$, $p(1-p)$ et $(1-p)^2$.

1. Donner, en fonction de p, les entropies de S, S^2 et S^3.
2. Justifier l'existence d'un code uniquement déchiffrable pour S^2, d'assortiment de longueurs $\{1, 2, 3, 3\}$. Si un tel code est optimal, quels symboles peuvent être codés avec longueur 1 ? avec longueur 3 ?
3. En déduire la longueur moyenne d'un tel code, et déterminer pour quelles valeurs de p, un tel code est plus efficace qu'un code de longueur uniforme égale à 2.
4. Pour $p = 1/3$, déterminer un codage binaire optimal pour chacune des sources S, S^2 et S^3, et calculer son efficacité.
5. Justifier, sans calcul, qu'un code binaire optimal pour S^4 a une efficacité au moins égale à celle d'un code binaire optimal pour S^2.

Exercice 2

On considère une source S qui émet les symboles a, b, c, d, e, f avec probabilités respectives 0.4, 0.1, 0.06, 0.1, 0.3 et 0.04.

1. Quelle est l'entropie de S ? Quelle est l'efficacité maximale d'un code binaire de longueur constante pour S ?
2. Alice prétend avoir construit pour S un code binaire préfixe de longueur moyenne 2.1. Est-ce possible ? Et si le code est ternaire, est-ce possible ?
3. Bob propose le code suivant pour S : $C(a) = 0$, $C(b) = 100$, $C(c) = 1100$, $C(d) = 101$, $C(e) = 111$, $C(f) = 1101$. Ce code est-il uniquement déchiffrable ? Décoder 1101011000101111. Ce code est-il plus efficace qu'un code binaire de longueur constante le plus efficace possible ?
4. On souhaite coder S (de manière uniquement déchiffrable) avec des mots ayant tous une longueur d'au plus 3.

(a) un mot d'un tel code peut-il avoir longueur 1 ?

(b) en déduire l'assortiment des longueurs que doit avoir un tel code pour être le plus efficace possible. Donner un tel code ; est-il plus efficace que celui de Bob ?

5. Construire un code de Huffmann pour S, et donner son efficacité.

Exercice 3

On rappelle qu'un mot v est un *suffixe* d'un mot w, s'il existe un mot u tel que l'on ait $w = u.v$ (le produit étant un produit de concaténation). On note $\tilde{u}$ le *miroir* d'un mot u, c'est-à-dire le mot u "lu à l'envers" : si $u = u_1 u_2 \dots u_\ell$ (où les u_i sont des lettres de l'alphabet), $\tilde{u} = u_\ell u_{\ell-1} \dots u_1$.

Un code est dit *suffixe* si, parmi les mots du code, il n'en existe pas deux codant des lettres distinctes et dont l'un soit un suffixe de l'autre.

1. Rappeler la définition d'un code préfixe.
2. Pour tout code C, on note $\tilde{C}$ le code formé des miroirs des mots de code de C. Montrer que $\tilde{C}$ est un code suffixe si et seulement si C est un code préfixe.
3. Une source S admet-elle toujours un code suffixe qui soit optimal (parmi tous les codes uniquement déchiffrables) ?
4. Donner une raison pour laquelle les codes suffixes sont moins utilisés que les codes préfixes (on pourra penser au code ternaire défini par $C(a) = 0$, $C(b) = 01$, $C(c) = 11$ et au décodage des mots 01^{1000} et 01^{1001}).