Université de la Méditerranée
MAT16 *Pratique de la cryptographie*

Semestre 2 | **Exercices pour TD 1 (semaine 1)** | 2012–2013

1. **a.** Déterminer si le codage suivant est *i*) sans pertes, *ii*) instantané, et *iii*) uniquement décodable (avec explications) :

x	A	B	C	D
$C(x)$	000	001	01	1

b. Calculer l'entropie de $\{\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{D}\}$ avec

$$p(\mathtt{A}) = p(\mathtt{B}) = 0.125,\ p(\mathtt{C}) = 0.250,\ \text{et } p(\mathtt{D}) = 0.500,$$

et déterminer si C est optimal.

2. Soit $X = \{\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{D}, \mathtt{E}, \mathtt{F}, \mathtt{G}, \mathtt{H}\}$ l'espace de probabilité avec les probabilités suivantes :

x	A	B	C	D	E	F	G	H
$p(x)$	0.25	0.25	0.20	0.14	0.10	0.04	0.01	0.01

a. Calculer l'entropie $H(X)$.

b. Construire l'arbre de codage associé à un codage de Huffman C pour X, et donner les codes $C(x)$ pour chaque élément x.

c. Trouver l'espérance mathématique de la fonction longueur de $C(x)$, et vérifier le théorème *Codage source* de Shannon.

3. Soit $\mathcal{A} = \{\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{D}, \mathtt{E}\}$ un alphabet avec probabilités

$$p(\mathtt{A}) = 0.40,\ p(\mathtt{B}) = 0.1,\ p(\mathtt{C}) = p(\mathtt{D}) = 0.20,\ \text{et } p(\mathtt{E}) = 0.1,$$

et $M =$ `AAACDCBDCEAADADACEDAEADCABEDADDCECAAAAAD`.

a. Trouver un codage instantané $C : \mathcal{A} \longrightarrow \{0,1\}^*$, avec $E(\ell_C) \leq H(\mathcal{A}) + 1$.

b. Comparer la longueur de M multiplié par $\log_2(5)$ et la longueur de $C(M)$.

4. Soit $\mathcal{A} = \{\mathtt{A}, \mathtt{B}, \mathtt{C}\}$, et supposons que les digrammes suivent les probabilités :

		y		
	$p(xy)$	A	B	C
	A	0	4/15	1/15
x	B	8/27	8/27	0
	C	1/27	4/135	1/135

a. Calculer les probabilités $p(x)$ pour chaque lettre x dans $\mathcal{A}$.

b. Déterminer $H(\mathcal{A}^2)$, $H(\mathcal{A})$ et vérifier $H(\mathcal{A}^2) \leq 2H(\mathcal{A})$.

c. Trouver des codages $C_1 : \mathcal{A} \longrightarrow \{0,1\}^*$, et $C_2 : \mathcal{A}^2 \longrightarrow \{0,1\}^*$, qui sont uniquement décodable, avec l'espérance des longueurs bornée par $H(\mathcal{A}) + 1$ et $H(\mathcal{A}^2) + 1$.

d. Trouver les longueurs des codes $C_1(M)$ et $C_2(M)$ où

$M =$ `ABBABABABABABABBBABBBBBABABABABABBBACACABBABBBBABBABACBBBABA`.

5. On veut construire un code compresseur binaire sur une source S de longueur infinie et d'alphabet $\{0, 1\}$ telle que la probabilité d'émission de 0 est p et de 1 est $1-p$. On propose le code suivant : on compte le nombre d'occurences de 0 avant l'apparition d'un 1 et

- Si une chaîne de quatre 0 consécutifs apparaît on la code par 0.
- Si strictement moins de quatre 0 apparaissent avant un 1 on code la chaîne par le mot du code $1e_1e_2$ où e_1e_2 est la représentation binaire du nombre de 0 apparus avant le 1. Par exemple 001 est codé par 110 car deux 0 sont apparus avant le 1.

a. Expliciter les 5 mots du code. Le code est-il sans préfixe ?

b. On suppose que les probabilités d'apparition des symboles 0 et 1 sont indépendantes. Calculer la probabilité d'apparition d'une chaîne de k 0 suivis d'un 1 pour $k \in \mathbb{N}$.

c. Donner en fonction de p la longueur moyenne du code.

6. On considère un texte source formé à partir de 5 symboles distincts (a, b, c, d, r) avec les fréquences d ?apparition suivantes : $f(a) = 0,43$; $f(b) = 0,20$; $f(c) = 0,1$; $f(d) = 0,09$; $f(r) = 0,18$.

a. Générer un arbre de Huffman binaire et proposer le codage correspondant en affectant les 0 aux probabilités les plus grandes.

b. Coder le texte suivant et calculer le gain de compression par rapport à un code binaire de longueur fixe et minimale : abracadabrabracadabra

7. Les nombres de Fibonacci sont définis par : $F(0) = F(1) = 1$ et $F(i) = F(i-1) + F(i-2)$ pour $i \geq 2$. On considère l'ensemble des lettres $\{a, b, c, d, e, f, g, h\}$ affectées des poids basés sur les 8 premiers nombres de Fibonnacci : $poids(a) = 1$; $poids(b) = 1, \ldots$

a. Donner un codage, selon la méthode de Huffman, pour ces 8 lettres.

b. Généraliser la réponse pour trouver le codage lorsque les poids sont les n premiers nombres de Fibonacci. Quel est la longueur du codage de F_i pour $0 \leq i \leq n$?

c. Quelle est la longueur moyenne du code ?