Master Mathématiques et Applications, 1-ière année, Aix-Marseille Université

Modélisation en Traitement du Signal, Fiche d'exercices 3

Année 2015-16

---

1 Code de Huffman

On considère l'alphabet $\mathcal{A}$ équipé de la distribution de probabilités donnée dans le tableau ci-dessous

Symbole	a	b	c	d	e	f
Probabilité	0,4	0,2	0,15	0,1	0,8	0,7

1. Appliquer l'algorithme de Huffman, et en déduire un code pour cet alphabet.
2. Calculer l'entropie de Shannon de la distribution de probabilités P, ainsi que la longueur moyenne des mots. Donner l'efficacité du code.

2 Code de Huffman

On considère l'alphabet A équipé de la distribution de probabilités donnée dans le tableau ci-dessous

Symbole	a	b	c	d	e	f	g	h	i	j	k	l	m
Probabilité	5/24	3/24	2/24	2/24	2/24	2/24	2/24	1/24	1/24	1/24	1/24	1/24	1/24

1. Appliquer l'algorithme de Huffman, et en déduire un code pour cet alphabet.
2. Calculer l'entropie de Shannon de la distribution de probabilités P, ainsi que la longueur moyenne des mots. Donner l'efficacité du code.

3 Code de Shannon-Fano

Le code de Shannon-Fano, utilisé dans les années 50, est le premier code à avoir exploité la redondance d'une source. Le principe est basé sur une décomposition récursive de l'alphabet en deux parties telles que leur probabilité soit sensiblement égale.

On considère un alphabet $A = \{a\}$, et une distribution de probabilités $P = \{p(a), a \in A\}$. L'algorithme de Shannon-Fano comporte les étapes suivantes :

— Les probabilités d'apparition de chaque symbole sont placées dans un tableau trié par ordre décroissant de probabilités .

— L'alphabet est coupé en deux groupes de symboles A_0 et A_1 dont la somme des probabilités de chaque groupe avoisine 0.5.

— Le groupe A_0 est codé par un "0" et A_1 par un "1".

— Si un groupe A_i n'a qu'un seul élement, c'est une feuille terminale, sinon la procédure reprend récursivement à l'étape 2 sur le groupe A_i.

1. Appliquer l'algorithme de Shannon-Fano ci-dessus à l'alphabet de la section 2, et en déduire un code pour cet alphabet.
2. Calculer l'entropie de Shannon de la distribution de probabilités P, ainsi que la longueur moyenne des mots. Donner l'efficacité du code. Comparer les résultats aux résultats obtenus avec le code de Huffman

4 Théorème de Shannon

Dans cet exercice, on considère un alphabet $\mathcal{A} = \{a_0, \dots a_{M-1}\}$, muni d'une distribution de probabilités $P = \{p(a), a \in \mathcal{A}\}$, et on se propose de démontrer les principaux éléments liés au théorème de Shannon-Fano. On considère donc un codeur binaire $\alpha : a \in \mathcal{A} \to \alpha(a)$, et pour tout $a \in \mathcal{A}$ on note $\ell(a)$ la longueur du mot binaire $\alpha(a)$. On note $\overline{\ell} = \sum_{a\in\mathcal{A}} p(a)\ell(a)$ la longueur moyenne des mots associée à α.

1. On considère l'inégalité de Kraft
$$\sum_{a\in\mathcal{A}} 2^{-\ell(a)} \leq 1 \ .$$
Montrer que si α est un code de longueur constante (c'est à dire $\ell(a) = \ell$ pour tout $a \in \mathcal{A}$), alors $\ell \geq \mathsf{Card}(\mathcal{A})$.
2. Soit maintenant α un code de longueur variable (CLV). On suppose que le codeur satisfait l'inégalité de Kraft, et on veut montrer l'inégalité de Shannon $\overline{\ell} \geq H(P)$, où $H(P)$ est l'entropie de Shannon :
$$H(P) = -\sum_{a\in\mathcal{A}} p(a)\log_2(p(a))$$
   a) En utilisant la définition de ℓ, un logarithme en base 2 et l'inégalité de Kraft, montrer que
$$\overline{\ell} \geq -\sum_{a\in\mathcal{A}} p(a)\log_2(q(a)) \qquad \text{où} \qquad q(a) = \frac{2^{-\ell(a)}}{\sum_{b\in\mathcal{A}} 2^{-\ell(b)}}$$
   b) En utilisant la concavité du logarithme (inégalité de Jensen), montrer que
$$\sum_{a\in\mathcal{A}} p(a)\log_2(q(a)) \leq \sum_{a\in\mathcal{A}} p(a)\log_2(p(a));$$
   déduire du résultat précédent que
$$\overline{\ell} \geq H(P) \ .$$
   c) Sous quelles conditions l'inégalité peut elle être une égalité ?

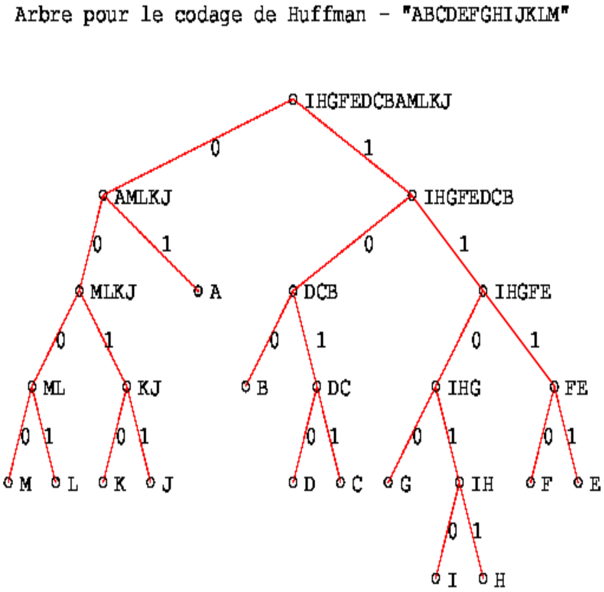


FIGURE 1 – Arbre binaire associé au code de Huffman

Corrections

1 Code de Huffman

1. Pour construire le code, il faut commencer par remplir la table

a (0,4)	a (0,4)	a (0,4)	a (0,4)	defcb (0,6)	adefcb (1)
b (0,2)	b (0,2)	def (0,25)	cb (0,35)	a (0,4)	
c (0,15)	c (0,15)	b (0,2)	def (0,25)		
d (0,1)	ef (0,15)	c (0,15)			
e (0,8)	d (0,1)				
f (0,7)					

On en déduit l'arbre de Huffman, puis le code

lettre	code	longueur
a	1	1
b	010	3
c	011	3
d	001	3
e	0000	4
f	0001	4

2. La longueur moyenne des mots vaut $\overline{\ell} = 2.35$, et l'entropie vaut $H = 2.296$. L'efficacité vaut donc $H/\overline{\ell} = 97\%$.

2 Code de Huffman

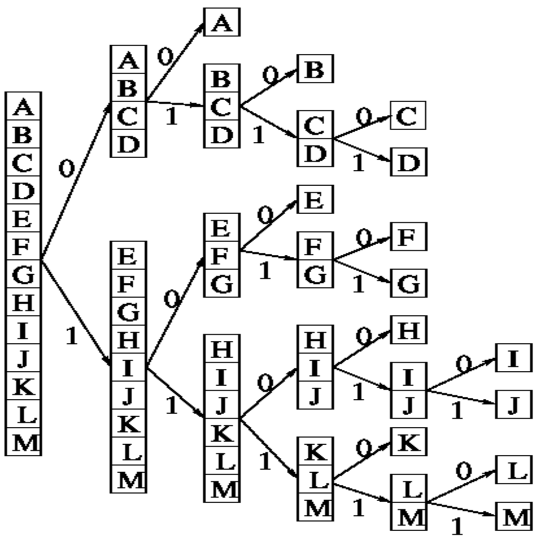
1. L'encodage est décrit dans la figure **??**, et le code se trouve dans la table **??** (tableau de gauche). L'arbre se trouve ci-dessous :

   et le code correspondant est là :
2. L'entropie vaut $H(P) \approx 3.4864$. La longueur moyenne pour un code de longueur constante serait de 4 bits/symbole, la longueur pour le code de Huffman vaut $\overline{\ell}_H \approx 3.5417$.

symbole	code	longueur
A	01	2
B	100	3
C	1011	4
D	1010	4
E	1111	4
F	1110	4
G	1100	4
H	11011	5
I	11010	5
J	0011	4
K	0010	4
L	0001	4
M	0000	4

Table 1 – Arbre binaire associé au code de Huffman

3 Code de Shannon-Fano



1.

symbole	code	longueur
A	00	2
B	010	3
C	0110	4
D	0111	4
E	100	3
F	1010	4
G	1011	4
H	1100	4
I	11010	5
J	11011	5
K	1110	4
L	11110	5
M	11111	5

2. L'entropie vaut $H(P) \approx 3.4864$. La longueur moyenne pour un code de longueur constante serait de 4 bits/symbole, la longueur pour le code de Huffman vaut $\overline{\ell}_{SF} \approx 3.5417$ (même résultat que pour le code de Huffman).

4 Théorème de Shannon

1. Prenons l'inégalité de Kraft dans le cas d'un code de longueur constante L :

$$\sum_{a\in\mathcal{A}} 2^{-L} = \mathsf{Card}(\mathcal{A})2^{-L} \leq 1 \ ,$$

on en déduit donc

$$2^L \geq \mathsf{Card}(\mathcal{A}) \quad \text{soit} \quad L \geq \log_2(\mathsf{Card}(\mathcal{A})) \ .$$

L devant être un entier, il est nécessairement supérieur ou égal à l'entier ls plus proche par valeur supérieure de $\log_2(\mathsf{Card}(\mathcal{A}))$. Par exemple, si $\mathsf{Card}(\mathcal{A}) = 10$, alors $L \geq 4$.

2. a) Calculons

$$\bar{\ell} = \sum_{a\in\mathcal{A}} p(a)\ell(a) = -\sum_{a\in\mathcal{A}} p(a)\log_2(2^{-\ell(a)}) \geq -\sum_{a\in\mathcal{A}} p(a)\log_2\left(\frac{2^{-\ell(a)}}{\sum_{b\in\mathcal{A}} 2^{-\ell(b)}}\right) = -\sum_{a\in\mathcal{A}} p(a)\log_2(q(a)) \ .$$

On a ici utilisé l'inégalité de Kraft, qui donne $2^{-\ell(a)} \geq 2^{-\ell(a)}/\sum_{b\in\mathcal{A}} 2^{-\ell(b)}$.

b) L'inégalité de Jensen donne

$$\sum_{a\in\mathcal{A}} p(a)\log_2\left(\frac{q(a)}{p(a)}\right) \leq \log_2\left(\sum_{a\in\mathcal{A}} p(a)\frac{q(a)}{p(a)}\right) = \log_2(\sum_{a\in\mathcal{A}} q(a)) = 0$$

car $Q = \{q(a)\}$ est une distribution de probabilités. On en déduit donc

$$\bar{\ell} \geq -\sum_{a\in\mathcal{A}} p(a)\log_2(q(a)) \geq -\sum_{a\in\mathcal{A}} p(a)\log_2(p(a)) = H(P) \ .$$