Université Kasdi Merbah Ouargla
Département d'Informatique et des Technologies de l'Information

Master 1 Fondamental — Module TI
Mai 2022

Corrigé examen TI

Exercice 1 : Entropie d'une source et codage binaire (5 pts)

Soit une variable aléatoire $X = \{a, b, c, d\}$

1. Déterminer l'entropie $H(X_p)$ quand les probabilités associées à X sont $P(X) = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$
2. Déterminer l'entropie $H(X_q)$ quand les probabilités associées à X sont $Q(X) = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$
3. On définit le code binaire suivant

$$c(X = a) = 0$$
$$c(X = b) = 10$$
$$c(X = c) = 110$$
$$c(X = d) = 111$$

Ce code est-t-il un code sans préfixe ?

Calculer la longueur moyenne des mots de code (en bits/symbole) quand X est distribuée selon P(X) puis selon Q(X)

4. Comparer les résultats. Dans quel cas de distribution de probabilité ce code est absolument optimal ?

Réponses :

1. $H(X_p) = -\sum_{i=1}^{4} p_i \log_2 p_i = -\frac{1}{2} \times \log_2(\frac{1}{2}) - \frac{1}{4} \times \log_2(\frac{1}{4}) - 2 \times \frac{1}{8} \times \log_2(\frac{1}{8}) = \frac{1}{2} + \frac{1}{2} + \frac{3}{4} = \frac{2+2+3}{4} =$ *74=1.75 bits /symbole* **(1 pt)**
2. $H(X_q) = -\sum_{i=1}^{4} p_i \log_2 p_i = -4 \times \frac{1}{4} \times \log_2(\frac{1}{4}) = 2 bits/symboe$ *ou bien* $H(X_q) = \log_2(2^2) =$ *2 bits/symbole* car *une loi de probabilité* uniforme, son entropie est donc log*2n. Si n = 22, son entropie est* alors *2.* **(1 pt)**
3. Ce code est sans préfixe car aucun code ne commence par l'un des autres codes **(0.5 pt)**

Calcul de la longueur moyenne du code

a) Cas de la distribution $P(X)$

X	a	b	c	d
$P(X)$	½=0.5	¼=0.25	1/8=0.125	1/8=0.125
m_i	0	10	110	110
l_i	1	2	3	3

$\bar{L}_p = \sum_{i=1}^{6} p_i \times l_i = 1 \times (0.5) + 2 \times (0.25) + 3 \times (0.125 + 0.125) = 1.75 \ (bits/symbole)$ **(0.75 pt)**

b) Cas de la distribution $Q(X)$

X	a	b	c	d
$Q(X)$	1/4=0.25	¼=0.25	1/4=0.25	¼=0.25
m_i	0	10	110	110
l_i	1	2	3	3

$\bar{L}_q = \sum_{i=1}^{6} p_i \times l_i = 1 \times (0.25) + 2 \times (0.25) + 3 \times (0.25 + 0.25) = 0.25 + 0.5 + 1.5 = 2.25 \ (bits/symbole)$ **(0.75 pt)**

4. Comparaison des résultats
   Le meilleur code dépend de la loi de l'émission de la source **(0.25 pt)**
   $\bar{L}_p = H(X_p) = 1.75$
   $\bar{L}_q = 2.25 > H(X_q) = 2$ **(0.25 pt)**
   Le code associé à la distribution $P(X)$ est absolument optimal car $\bar{L}_p = H(X_p)$ **(0.5 pt)**

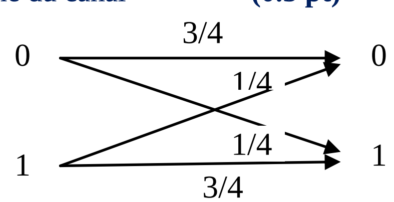
Exercice 2 : Modélisation mathématique d'un canal (5 pts)

Soit une source binaire X={0 , 1} de probabilité P(x=0)=0.5 et P(x=1)=0.5.
Soit un récepteur (canal) de même alphabet, pour chaque symbole émis, le canal de transmission à la probabilité ¼ de faire une erreur.

1. Représenter le canal par un graphe et en déduire la matrice de distribution des probabilités de réception P(Y|X)
2. Trouver les distributions conjointe P(X,Y) et marginale P(Y)
   Nous rappelons la formule : $p(y|x) = \frac{p(x,y)}{p(x)} \Longrightarrow p(x,y) = p(x) * p(y|x)$
3. Calculer les entropies H(X), H(Y), H(X,Y) et H(Y|X)
4. Calculer l'information mutuelle de X et Y : I(X,Y) et I(Y,X)

Réponses :

1. Graphe du canal **(0.5 pt)**



La matrice de distribution des probabilités de réception (probabilité conditionnelle) **(0.5 pt)**
P(Y|X)

X/Y	0	1
0	3/4	1/4
1	1/4	3/4

2. Les distributions conjointes et marginales **(1 pt)**

$$P(0,0) = p(0).p(0|0) = 1/2 * 3/4 = 3/8$$
$$P(0,1) = p(0).p(1|0) = 1/2 * 1/4 = 1/8$$
$$P(1,0) = p(1).p(0|1) = 1/2 * 1/4 = 1/8$$
$$P(1,1) = p(1).p(1|1) = 1/2 * 3/4 = 3/8$$

P(X,Y) : distribution conjointe , les distributions marginales P(X) et P(Y)

X/Y	0	1	P(X)
0	3/8	1/8	3/8+1/8=1/2
1	1/8	3/8	1/8+3/8=1/2
P(Y)	3/8+1/8=1/2	1/8+3/8=1/2	

3. Calcul des entropies :

$H(X) = \log_2(2^1) = 1\ bit/symbole$: Entropie de la source **(0.25 pt)**

$H(y) = \log_2(2^1) = 1\ bit/symbole$: Entropie de la réception **(0.25 pt)**

$H(X,Y) = -\sum_{i=1}^{4} p_i \log_2 p_i = -(2 \times \frac{1}{8} \times \log_2(\frac{1}{8}) -2 \times \frac{3}{8} \times \log_2(\frac{3}{8}) \approx 1.8 \frac{bits}{symbole}$: Entropie mutuelle **(0.25 pt)**

$H(X|Y) = -\sum_{i=1}^{4} p_i \log_2 p_i = -(2 \times \frac{1}{4} \times \log_2(\frac{1}{4}) -2 \times \frac{3}{4} \times \log_2(\frac{3}{4}) \approx 0.8\, bits/symbole$: Entropie conditionnelle **(0.25 pt)**

4. Information mutuelle moyenne de X et Y
   $I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = 1 - 0.8 = 0.2\, bit/symbole$ **(1 pt)**
   Ou bien:
   $I(X\,;Y) = H(X) + H(Y) - H(X,Y) = 1 + 1 - 1.8 = 0.2\, bit/symbole$

<u>Exercice 3 : Compression RLE et Huffman (10 pts)</u>

Considérant l'alphabet de la source **X= {a, b, c, d}** et la séquence de texte à compresser : **Seq= «abbcccdddd»**.

Méthode 1 – Codage de Huffman

1. Constituer la table des fréquences des symboles de l'alphabet X contenus dans la séquence donnée.
2. Calculer l'Entropie associée. Que dire de l'efficacité du codage qui associe 8 bits à chaque caractère ?
3. A partir de la table des fréquences, déterminer un arbre de Huffman et le code associé.
4. Calculer la taille du code de Huffman (en bits) de cette séquence. Quel est le taux de compression ?

Réponses :

1. Table des fréquences et Pi **(1 pt)**

Symbole	a	b	c	d
Nombre d'occurrences	1	2	3	4
Fréquences Pi	1/10=0.1	2/10=0.2	3/10=0.3	4/10=0.4

$Taille\ de\ \boldsymbol{Seq}\ \ en\ bits\ = 10\ x\ 8\ = 80\ bits$

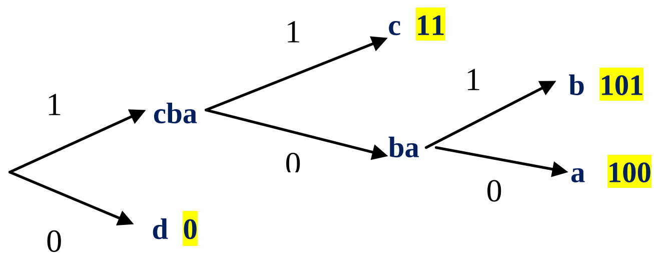
2. Entropie :
   $H = -\sum_{i=1}^{4} p_i log_2 p_i =$
   $-0.1 \times log_2(0.1) - 0.2 \times log_2(0.2) - 0.3 \times log_2(0.3) - 0.4 \times log_2(0.4) \approx 0.92\ bits / symbole$ **(0.5 pt)**
   L'entropie H ($\sim$0.92 $bits/symbole$) < codage ascii (8 bits par caractère) **(0.5 pt)**

   $\log_2(0.4) \approx -1.32$, $\log_2(0.3) \approx -1.74$, $\log_2(0.2) \approx -2.32$, $\log_2(0.1) \approx -3.32$

3. Arbre de Huffman

d	0.4	d	0.4	cba	0.6
c	0.3	c	0.3	d	0.4
b	0.2	ba	0.3		
a	0.1				

(0.5 pt)



$$Convention\ arbitraire : 1 \equiv symboles\ les\ plus\ probables, 0 \equiv\ symboles\ les\ moins\ probables$$

(0.25 pt)

Table de code associé **(0.25 pt)**

X	a	b	c	d
P(X)	0.1	0.2	0.3	0.4
m_i	100	101	11	0
l_i	3	3	2	1

4. $Taille\ du\ code\ de\ Huffman\ en\ bits = 1 \times 3 + 2 \times 3 + 3 \times 2 + 4 \times 1 = 3 + 6 + 6 + 4 = 19\ bits$
   $Le\ taux\ de\ compression = \frac{80-19}{80} = 1 - \frac{19}{80} = 76.25\%$ **(1 pt)**

Méthode 2 – Codage RLE (Run Lenght Encoding)

1. Constituer le code Rle correspondant à la séquence donnée Seq.
2. Chaque couple obtenue est encodée sur 16 bits : 8 bits pour le nombre de répétitions et 8 bits pour le caractère répété. Quelle est la longueur du (en bits) obtenu ?
3. Calculer le taux de compression

Réponses :

1. Code RLE : remplacer chaque caractère identique par un couple (nb, C) ou nb : nbr. de caractères (sur 8 bits), C : le caractère en répétition (sur 8 bits)
   Seq= «abbcccdddd».
   Le code RLE associé : (1,a), (2,b),(3,c), (4,d) **(1 pt)**
2. Longueur du code (en bits)
   $4\ couples \times 16\ bits = 64\ bits$: taille du code RLE de cette séquence **(1 pt)**
3. $Le\ taux\ de\ compression = \frac{80-64}{80} = 1 - \frac{64}{80} = 20\%$ **(1 pt)**

Méthode 3 – Codage RLE + Huffman

1. Appliquer le codage de Huffman au code RLE Trouvé et calculer le nouveau taux de compression
2. Comparer les performances des 3 méthodes (Huffman, RLE, RLE+Huffman)

Réponses :

1. Codage de Huffman appliqué au code RLE

X	(1,a)	(2,b)	(3,c)	(4,d)
Nombre d'occurrences	1	1	1	1
p_i	1/4=0.25	1/4=0.25	1/4=0.25	1/4=0.25

(0.25 pt)

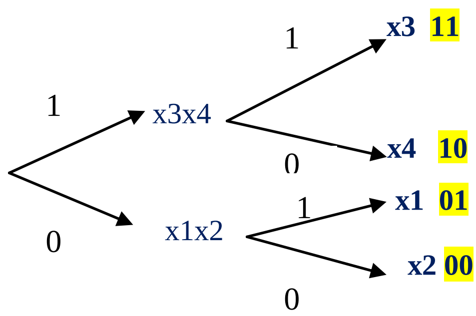
L'entropie : $H = \log_2(2^2) = 2\ bits/symbole$ < 16 bits par un couple RLE, donc une compression du code RLE est possible

Arbre de Huffman :

(1,a)=x1	0.25	x3x4	0.5	x3x4	0.5
(2,b)=x2	0.25	x1	0.25	x1x2	0.5
(3,c)=x3	0.25	x2	0.25		
(4,d)=x4	0.25				

(0.25 pt)

(0.25 pt)



X	(1,a)	(2,b)	(3,c)	(4,d)
Nombre d'occurrences	1	1	1	1
p_i	0.25	0.25	0.25	0.25
m_i	01	00	11	10
l_i	2	2	2	2

$Longueur\ du\ code\ (en\ bits)\ =\ 4\ \times\ (1 \times 2) =\ 8\ bits$ **(0.25 pt)**

$Le\ taux\ de\ compression\ = \frac{80-8}{80} = 1 - \frac{8}{80} = 90\%$ **(1 pt)**

2. $La\ méthode\ mixte\ (RLE + HUFFMAN)\ présente\ le\ taux\ de\ compression\ le\ plus\ élevé$ **(1 pt)**