Université de Lille — Faculté des Sciences et Technologies, Département Informatique

Compression, détection et correction d'erreurs

2025–2026
Licence mention informatique – Semestre 4

TD2 – Compression de données

---

1 Sources, quantité d'information et entropie

Exercice 1.1 : *Pour s'échauffer*

Q 1.1 Que dire de la longueur moyenne d'un codage optimal dans le cas où tous les symboles de la source sont équiprobables ?

Exercice 1.2 : *Caractérisation de codages optimaux*

Q 2.1 Combien de feuilles possède l'arbre binaire d'un codage binaire préfixe optimal de m symboles ?

Q 2.2 Quelle est la particularité des arbres de codages binaires préfixes optimaux ?

Q 2.3 Quels sont les arbres des codages préfixes binaires optimaux d'une source de trois symboles ? de quatre symboles ?

Q 2.4 Dans un codage binaire optimal peut-il y avoir exactement trois mots de longueur maximale ?

Q 2.5 La distribution des longueurs des mots d'un codage préfixe binaire optimal est-elle unique ?

Exercice 1.3 : *Entropie d'un message*

On considère un message M écrit avec les symboles d'un alphabet $\mathcal{S} = \{s_1, s_2, \ldots, s_m\}$. On désigne par n_i le nombre d'occurrences du symbole s_i dans le message M. La longueur du message est désignée par N. On a évidemment

$$N = \sum_{i=1}^{m} n_i.$$

Exprimez l'entropie de M en fonction des entiers N et n_i, $i = 1, \ldots, m$, de manière à ce que N n'apparaisse pas dans la somme.

Exercice 1.4 :

Imaginons un codage c pour lequel la longueur du mot associé à chaque symbole soit égale à la quantité d'information du symbole. Autrement dit pour chaque symbole $s \in \mathcal{S}$, $|c(s)| = I(s)$.

Q 4.1 Pourquoi un tel codage serait-il optimal ?

Q 4.2 Quelle contrainte a-t-on sur $I(s)$ et donc sur $p(s)$ pour qu'un tel codage existe ?

On considère une source de 8 symboles décrites par le tableau ci-dessous.

s	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
$p(s)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

Q 4.3 Déterminez un codage optimal pour cette source.

Q 4.4 Combien y a-t-il de codages binaires préfixes optimaux pour cette source ?

2 Algorithme de Huffman

Exercice 2.1 : *Codage de Huffman*

On considère la source d'information définie par l'alphabet $\mathcal{S} = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8\}$ avec la distribution de probabilités suivante f :

s	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
$p(s)$	0,4	0,18	0,1	0,1	0,07	0,06	0,05	0,04

Q 1.1 Calculez la quantité d'information contenue dans chacun des huit symboles, puis l'entropie de la source d'information $(\mathcal{S}, f)$.

Q 1.2 Calculez un codage de Huffman de cette source, puis la longueur moyenne de ce codage. Comparez cette longueur moyenne à l'entropie.

Pour approfondir

Exercice 2.2 :

Q 2.1 Même exercice pour la source suivante (les données ne sont pas des fréquences mais des effectifs)

x	a	b	c	d	e	f	g	h
$f(x)$	1	1	2	3	5	8	13	21

Q 2.2 Les effectifs donnés dans le tableau ci-dessous sont les premiers termes de la suite de Fibonacci. Peut-on généraliser le codage pour un alphabet source de n symboles dont les fréquences sont les n premiers termes de la suite de Fibonacci ?

Exercice 2.3 :

Q 3.1 Pour coder sept symboles, existe-t-il un codage binaire optimal comprenant

1. trois mots de longueur 2, deux mots de longueur 3 et deux mots de longueur 4 ?
2. un mot de longueur 2, cinq mots de longueur 3 et un mot de longueur 4 ?

Soit $\mathcal{S}$ la source munie de la distribution de fréquences f suivante :

x_1	x_2	x_3	x_4	x_5	x_6	x_7
1/8	1/32	1/4	1/8	1/16	5/32	1/4

Q 3.2 Construisez un codage binaire optimal pour $\mathcal{S}$.

Soient c_1 et c_2 les codages de $\mathcal{S}$ définis par

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
$c_1(x)$	0011	0000	10	01	0010	0001	11
$c_2(x)$	001	000000	1	0001	00001	000001	01

Q 3.3 Combien de bits faut-il en moyenne pour coder une séquence de 16 000 symboles à l'aide de ces deux codages ?

Q 3.4 Ces codages sont-ils optimaux ?

Pour approfondir

Exercice 2.4 : *Tailles d'un fichier*

L'analyse des caractères d'un fichier de programmes en PASCAL donne la répartition suivante (par ordre croissant des effectifs) :

W	$	Y	F	5	à	X	>	*	"
1	1	1	1	2	2	2	2	2	2
q	3	z	+	4	O	P	y	S	V
2	3	3	3	4	5	6	6	8	8
M	j	R	2	-	.	<	U	D	B
8	10	10	11	11	13	15	16	17	17
é	C	v	x	{	L	}	[	]	I
18	20	21	23	25	25	25	27	27	34
E	N	A	k	h	0	1	T	g	w
35	35	36	36	37	40	40	42	43	46
m	’	b	u	P	=	_	l	,	d
58	58	66	70	74	78	79	89	90	93
:	s	f	(	)	c	a	o	n	;
96	107	123	128	128	131	157	158	199	200
r	t	i	e	/	<ESP>				
216	273	278	283	304	1972				

<ESP> désigne le caractère espace.

Q 4.1 En supposant que chaque caractère est codé en code ASCII sur 8 bits, quelle est la taille de ce fichier (en bits) ?

Q 4.2 Donner une minoration de la taille de ce fichier s'il est codé par un codage préfixe.

Q 4.3 Quelle est la taille de ce fichier s'il est codé par un codage de Huffman ?

Exercice 2.5 : *Fichier aléatoire*

Supposons qu'un fichier comportant n symboles au total, composé de 256 symboles différents, chacun codé sur 8 bits, ait été généré avec un générateur pseudo-aléatoire. Dans ce cas, la fréquence de chaque symbole est comparable. On pourra d'ailleurs considérer que la fréquence minimale d'un symbole dans ce fichier est supérieure à la moitié de la fréquence maximale dans ce même fichier.

Q 5.1 Quelle sera la taille du fichier après utilisation de l'algorithme de Huffman ?

Exercice 2.6 : *Stratégie*

On a lancé deux fois une pièce de monnaie.

Q 6.1 Quelle est la quantité d'information contenue dans les affirmations suivantes :

1. PILE n'est pas apparu ?
2. PILE est apparu une seule fois ?
3. PILE est apparu deux fois ?

Q 6.2 Quelle est l'entropie de la source d'information indiquant le nombre de PILE ?

Q 6.3 Vous désirez connaître le nombre de PILE sortis lors de ces deux tirages. Pour cela vous pouvez poser autant de questions à réponses oui ou non que vous voulez. Quelle est la meilleure stratégie vous permettant de connaître ce nombre ? Par meilleure stratégie, on entend celle qui permet d'obtenir le nombre en un minimum de questions en moyenne.

Q 6.4 Reprendre le même exercice avec 3 lancers.