


Théorie de l'information – Feuille de TD 3

07/10/2021

Le corrigé de certains exercices sera disponible à l'adresse suivante :

www.math.univ-paris13.fr/~lavauzelle/teaching/2021-22/theorie-information.html

(★) exercice fondamental (★★) pour s'entraîner (★★★) pour aller plus loin  sur machine

Exercice 1. (★) Arbre d'un code.

On considère le code suivant sur l'alphabet $\mathcal{X} = \{a, b, c, d, e, f, g\}$

a	b	c	d	e	f	g
00	01	100	101	1100	1101	1111

Question 1.– Ce code est-il préfixe ? Est-il uniquement décodable ? Vérifie-t'il l'inégalité de Kraft ?

Question 2.– Construire l'arbre binaire correspondant à ce code. Est-il possible de rendre ce code plus efficace, quelle que soit la distribution de la source X sur \mathcal{X} ?

Exercice 2. (★) Codes préfixes et uniquement décodables.

Parmi les codes suivants, lesquels sont préfixes ? Lesquels sont uniquement décodables ?

1. $\{0, 10, 11\}$
2. $\{0, 01, 11\}$
3. $\{0, 01, 10\}$
4. $\{0, 101, 111, 100\}$
5. $\{00, 10, 100, 11, 110\}$

Exercice 3. (★★) Code unaire.

Considérons le code suivant sur l'ensemble discret \mathbb{N} . À tout entier $n \in \mathbb{N}$, on associe la séquence de bits :

$$U(n) = \underbrace{0 \cdots \cdots 0}_n 1 \in \{0, 1\}^{n+1}.$$

Question 1.– Démontrer que le code U est préfixe.

Question 2.– Sous quelle condition sur la source X , à valeurs dans \mathbb{N} , la longueur moyenne du code est-elle finie ?

Question 3.– Déterminer une source d'entropie finie pour laquelle le code U est optimal.

Exercice 4. (**) Code Morse.

Le code Morse international permet de transcrire un texte sous forme d'impulsions courtes ou longues. Il a été utilisé pour la télégraphie au XIX^{ème} siècle.

L'alphabet du code Morse est retranscrit avec deux symboles : l'impulsion courte « · » et l'impulsion longue « – ».

Voici un dictionnaire des encodages des lettres de l'alphabet courant :

A	· –	B	– · · ·	C	– · · ·	D	– · ·	E	·
F	· · – ·	G	– – ·	H	· · · ·	I	· ·	J	· – – –
K	– · – ·	L	· – · ·	M	– –	N	– ·	O	– – –
P	· – – ·	Q	– – · –	R	· – ·	S	· · ·	T	–
U	· · –	V	· · · –	W	· – –	X	– · · –	Y	– · – –
Z	– – · ·								

Question 1.– Le code Morse est-il uniquement décodable ?

Question 2.– Calculer la longueur moyenne du code Morse en supposant que la loi de distribution des lettres est uniforme.

En français, après l'analyse de plusieurs œuvres littéraires, on estime que la distribution des lettres dans un texte suit approximativement la loi dont les probabilités sont données dans la table suivante :

A	8,15%	B	0,97%	C	3,15%	D	3,73%	E	17,39%
F	1,12%	G	0,97%	H	0,85%	I	7,31%	J	0,45%
K	0,02%	L	5,69%	M	2,87%	N	7,12%	O	5,28%
P	2,80%	Q	1,21%	R	6,64%	S	8,14%	T	7,22%
U	6,38%	V	1,64%	W	0,03%	X	0,41%	Y	0,28%
Z	0,15%								

Question 3.–

1. Calculer l'entropie de la langue française.
2. Calculer la longueur moyenne du code Morse pour un texte en français. Commenter.

En anglais, une analyse identique donne la distribution suivante :

A	8,08%	B	1,67%	C	3,18%	D	3,99%	E	12,56%
F	2,17%	G	1,80%	H	5,27%	I	7,24%	J	0,14%
K	0,63%	L	4,04%	M	2,60%	N	7,38%	O	7,47%
P	1,91%	Q	0,09%	R	6,42%	S	6,59%	T	9,15%
U	2,79%	V	1,00%	W	1,89%	X	0,21%	Y	1,65%
Z	0,07%								

Question 4.–

1. Calculer l'entropie de la langue anglaise.
2. Calculer la longueur moyenne du code Morse pour un texte en français. Commenter.