Université Paris 8 — Année 2022–2023

M1 Mathématiques et applications, parcours ACC

Théorie de l'information – Feuille de TD 3

05/10/2022

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

`www.math.univ-paris13.fr/~lavauzelle/teaching/2022-23/theorie-information.html`

($\star$) exercice fondamental ($\star\star$) pour s'entraîner ($\star\star\star$) pour aller plus loin 💻 sur machine

Exercice 1. ($\star$) Arbre binaire 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.– Décoder le texte encodé suivant, tout en essayant de bien comprendre comment vous arrivez à « séparer » les lettres :

$$010011110011111100.$$

Question 3.– Construire l'arbre binaire correspondant à ce code.

Question 4.– Est-il possible de rendre ce code plus efficace, quelle que soit la distribution de la source X sur $\mathcal{X}$? Si oui, que deviendrait le texte encodé de la question 2 ?

Exercice 2. ($\star$) 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. ($\star\star$) 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 :

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

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. ($\star\star$) 💻 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 XIXème siècle.

L'alphabet du code Morse est retranscrit avec deux symboles : l'impulsion courte « $\cdot$ » et l'impulsion longue « $-$ ».

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

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

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.

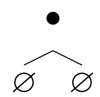
Exercice 5. ($\star\star\star$) Formalisation des arbres binaires.

Cet exercice a pour but de démontrer quelques propriétés sur les arbres binaires, qui peuvent être utiles au cours.

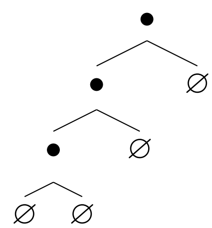
Formellement, on peut donner la définition suivante d'un arbre binaire :

— l'arbre vide (sans racine), noté $\varnothing$, est un arbre binaire,
— si G et D sont deux arbres binaires, alors le couple (G, D) est un arbre binaire.

Par exemple, l'arbre binaire $(\varnothing, \varnothing)$ est l'arbre réduit à une racine. C'est l'arbre :



Autre exemple, l'arbre $(((\varnothing, \varnothing), \varnothing), \varnothing)$ peut-être représenté ainsi :



Descendants. Si $T = (G, D)$ est un arbre binaire non-vide, on appelle G le descendant gauche de T et D son descendant droit. Si $G = D = \varnothing$, alors on dit que T est une feuille. Sinon, c'est un nœud interne. On note $n(T)$ son nombre de nœuds internes d'un arbre T et $f(T)$ son nombre de feuilles. Exprimé plus formellement :

— si $T = \varnothing$, alors $f(T) = 0$ et $n(T) = 0$
— si $T = (\varnothing, \varnothing)$, alors $f(T) = 1$ et $n(T) = 0$
— sinon, pour $T = (G, D)$ avec G ou D non-vide, alors on définit par induction $f(T) = f(G) + f(D)$ et $n(T) = 1 + n(G) + n(D)$.

Enfin, on note $v(T) = n(T) + f(T)$ le nombre total de nœuds de T.

Hauteur. La hauteur d'un arbre binaire est définie par une relation de récurrence :

— par convention, l'arbre binaire vide a hauteur $h(\varnothing) = -1$;
— si $T = (G, D)$ est non-vide, alors $h(T) = 1 + \max\{h(G), \max(D)\}$.

Pour revenir aux exemples précédents, l'arbre $T = (\varnothing, \varnothing)$ réduit à une racine possède $f(T) = 1$ feuille, aucun nœud interne ($n(T) = 0$) et sa hauteur est $h(T) = 0$. D'autre part, l'arbre $T = (((\varnothing, \varnothing), \varnothing), \varnothing)$ possède $f(T) = 1$ feuille, $n(T) = 2$ nœuds internes et sa hauteur est $h(T) = 2$.

Question 1.– Vérifier que dans le cas général, le nombre de nœuds total $v(T)$ satisfait :

— si $T = \varnothing$, alors $v(T) = 0$;
— si $T = (G, D)$, alors $v(T) = 1 + v(G) + v(D)$.

Question 2.– Démontrer par récurrence que $h(T) \leq n(T) + 1 \leq 2^{h(T)}$. Donner la forme des arbres pour les cas d'égalité.

Question 3.– Démontrer par récurrence que $1 \leq f(T) \leq 2^{h(T)}$. Donner la forme des arbres pour les cas d'égalité.

On note maintenant $n_1(T)$ le nombre de nœuds internes ayant 1 descendant, et $n_2(T)$ le nombre de nœuds internes ayant 2 descendants.

Question 4.–

1. Quelle est la relation entre $n(T)$, $n_1(T)$ et $n_2(T)$?
2. Donner un exemple d'arbre T avec $n_1(T) = 0$ et $n(T) = 3$.
3. Donner un exemple d'arbre T avec $n_2(T) = 0$ et $n(T) = 3$.

Question 5.– Démontrer par récurrence que $f(T) = n_2(T) + 1$.