Université Paris 8 J. Lavauzelle

Master 1 Mathématiques et applications — année 2023–24

Théorie de l'Information – Feuille de TD 3

02/10/2023

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

`https://lvzl.fr/teaching/2023-24/ti.html`

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

Exercice 1. ($\star$) 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 2. ($\star$) 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 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 gamma.

Dans cet exercice, on souhaite coder efficacement une source à valeur dans l'ensemble des entiers naturels, sans avoir d'information préalable sur les entiers émis par la source.

Usuellement, pour coder un entier $n \in \mathbb{N}$ sous forme de chaîne de bits, on décompose l'entier en base 2 :

$$n = \sum_{i=0}^{k} n_i 2^i\,,$$

et on retourne la séquence $B(n) = (n_0, \ldots, n_k) \in \{0,1\}^{k+1}$ de longueur $k+1$, où $k = \lfloor \log_2(n) \rfloor$ (excepté pour $n = 0$, pour lequel on a $k = 0$).

Question 1.– Le code $B : \mathbb{N} \to \{0,1\}^+$ est-il préfixe ? Est-il uniquement décodable ? Pourquoi ?

Le code Gamma, introduit par Elias, propose une solution au problème précédent. Avant d'encoder l'entier n sous la forme de sa décomposition en base 2, on précise à l'aide d'un code unaire la longueur de n. Ainsi, le code $\Gamma : \mathbb{N} \to \{0,1\}^+$ est défini par :

$$\Gamma(n) = \underbrace{0 \;\cdots\cdots\cdots\; 0}_{\leftarrow\, 1+\lfloor \log_2(n) \rfloor\, \rightarrow} 1\, B(n) \quad \text{ pour } n \geq 1,$$

et $\Gamma(0) = 10$.

Question 2.– Le code Γ est-il préfixe ?

Question 3.– Quelle est la longueur de $\Gamma(n)$ pour $n \in \mathbb{N}$? Donner une condition sur la distribution (p_n) de la source pour que la longueur moyenne du code Γ soit finie.

Question 4.– 💻 Calculer numériquement des valeurs approchées de la longueur moyenne du code Γ lorsque :

- X suit une loi uniforme sur $\{0, 1, \ldots, N\}$ (et $p_X(n)$ est nulle pour $n \geq N$),
- X suit une loi géométrique,
- X suit une loi de Poisson.

Question 5.– ($\star\star\star$) Le code Gamma peut être vu comme le codage unaire de la taille en bits de n, concaténé avec la représentation binaire de n. Peut-on itérer ce procédé pour obtenir un code encore plus court ? Quelle est la longueur du codage de l'entier n obtenu ?