Université Paris 8 Année 2021–2022

M1 Mathématiques et applications, parcours ACC et CSSD

Théorie de l'information – Feuille de TD 4

14/10/2021

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

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

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

Exercice 1. ($\star$) Codes de Huffman et de Shannon–Fano.

Soit X une variable aléatoire donnée par la distribution de probabilité $(\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12})$.

Question 1.– Quel est le code de Shannon–Fano associé à cette variable aléatoire ?

Question 2.– Trouver deux codes de Huffman distincts (c'est-à-dire, avec des longueurs de mots différentes) pour la source X. Comparer leur longueur moyenne à celle du code de Shannon–Fano.

Exercice 2. ($\star\star$) Code sur la loi conjointe.

Soit $\mathcal{X} = \{a, b\}$ et X, Y deux variables indépendantes sur $\mathcal{X}$ de même loi de Bernoulli de paramètre λ. On note $Z = (X, Y)$ la variable produit, définie sur $\mathcal{X}^2$, de loi conjointe $p_{X,Y}$.

Question 1.– Calculer $\mathbb{P}(Z = z)$ pour tout $z = (x, y) \in \mathcal{X}^2$.

Question 2.– Quelle est l'entropie de Z ?

Question 3.– Décrire le code de Huffman de source Z. Selon la valeur de λ, on pourra distinguer plusieurs formes pour l'arbre binaire associé au code.

Question 4.– Tracer le graphe de la longueur moyenne du code de Huffman en fonction de λ. Sous quelle condition sur λ le code de Huffman est-il strictement meilleur que le code de longueur fixe égale à 2 ?

Question 5.– Décrire le code de Shannon–Fano de source Z. On donnera les longueurs des mots en fonction de λ.

Question 6.– Sous quelle condition sur λ le code de Shannon–Fano est-il strictement meilleur que le code de longueur fixe égale à 2 ? On pourra s'aider d'un logiciel pour les résolutions numériques.

Exercice 3. ($\star\star$) Mots prépondérants dans un code de Huffman.

Considérons le code de Huffman sur une source X de distribution $p_1 \geq \cdots \geq p_m$.

Question 1.– Démontrer que si la probabilité d'occurence la plus forte vérifie $p_1 > 2/5$, alors le symbole associé à cette probabilité est encodé par un mot de longueur 1.

Question 2.– Démontrer que s'il existe un mot de longueur 1, alors la probabilité $p_1 \geq 1/3$.

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 :

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

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.– ($\star\star\star$)

1. 💻 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.
2. 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 lentier n obtenu ?

Exercice 5. ($\star\star\star$) Borne entropique et codes optimaux.

Question 1.– Pour quelle valeur de λ la variable de Bernoulli de paramètre λ donne un code de Shannon–Fano optimal ?

Question 2.– Soit m un entier ≥ 3 quelconque. Trouver une variable aléatoire sur $\{x_1, \ldots, x_m\}$ telle que le code de Shannon–Fano associé est optimal.

Question 3.– Soit $\epsilon > 0$. Déterminer une distribution sur une variable aléatoire X telle que, pour tout code C sur X, la longueur moyenne $\overline{\ell}(C) \geq H(X) + 1 - \epsilon$.

Question 4.– Soit $\epsilon > 0$. Déterminer une distribution sur une variable aléatoire X telle que le code de Shannon–Fano sur X a une longueur moyenne $\overline{\ell}(C_{SF}) \geq H(X) + 1 - \epsilon$, et le code de Huffman sur X a une longueur moyenne $\overline{\ell}(C_H) \leq H(X) + \epsilon$.