Théorie de l'information - TD2

Exercices à rendre pour le 20/10 : 7 et 8.

Exercice 1 (Codes préfixes et uniquement décodables)**.** *Parmi les codes binaires suivants, dire lesquels sont préfixes et lesquels sont uniquement décodables. Ici, on assimile un code C sur un ensemble $\{a_1, \dots, a_n\}$ au n-uplet $(C(a_1), \dots, C(a_n))$ (à permutation près).*

1. $(00, 01, 10, 11)$
2. $(0, 01, 001)$
3. $(000, 01, 001)$
4. $(00, 10, 11, 011, 010)$
5. $(0, 01)$

Exercice 2 (Écriture binaire)**.**

1. *Le code binaire sur $\{0, 1, \dots, 9\}$ qui à un entier associe son écriture en base 2 est-il préfixe ? Uniquement décodable ?*
2. *Le code binaire sur $\mathbb{N}$ qui à un entier associe son écriture en base 2 est-il préfixe ? Uniquement décodable ?*
3. *Le codage sur $\{0, 1, \dots, 9\}^\star$ qui à l'écriture d'un entier en base 10 associe l'écriture de ce même entier en base 2 est-il uniquement décodable ?*

Exercice 3 (Étude d'un code)**.** *On définit le code thermomètre (ou code unaire) sur l'alphabet $\mathbb{N}$ par:* $T(0) = 1, T(1) = 01, T(2) = 001, T(3) = 0001 \dots$

1. *Montrer que T est uniquement décodable.*
2. *Soit X une source sur $\mathbb{N}_{>0}$. Montrer que la longueur moyenne de T relativement à X est finie si et seulement si X admet une espérance finie. Dans ce cas, calculer la longueur moyenne de T.*
3. *Montrer que si X suit la loi géométrique de paramètre $1/2$, le code T est optimal.*
4. *Donner une autre loi pour laquelle T est optimal.*

Exercice 4 (Divergence et changement de distribution)**.** *Soit A un alphabet, et X une source à valeurs dans A de distribution p. Soit q une autre distribution sur A. On rappelle que la divergence de Kullback-Leibler de p par rapport à q est définie par*

$$D_{\mathrm{KL}}(p \| q) = \sum_{\substack{x \in A \\ p(x) > 0}} p(x) \log_2 \left(\frac{p(x)}{q(x)} \right).$$

1. *Montrer qu'il existe un code préfixe C sur A tel que la longueur de $C(a)$ soit $\lceil -\log_2 q(a) \rceil$.*
2. *Pour un tel C, montrer que sa longueur moyenne ℓ_C vérifie*
$$H(X) + D_{\mathrm{KL}}(p \| q) \leqslant \ell_C \leqslant H(X) + D_{\mathrm{KL}}(p \| q) + 1.$$
3. *Que représente alors $D_{\mathrm{KL}}(p \| q)$?*

Exercice 5 (Code de Huffman)**.** *Soit X une variable aléatoire à valeurs dans $\{0, 1, 2, 3, 4, 5\}$ de loi p_X définie par $p_X(0) = p_X(1) = p_X(2) = 1/8$, $p_X(3) = 1/2$, $p_X(4) = p_X(5) = 1/16$.*

1. *Construire un code binaire de Huffman H associé à X.*
2. *Coder le message 024322 à l'aide de H.*
3. *Calculer la longueur moyenne et l'efficacité de H relativement à X.*

Exercice 6 (Code de Shannon-Fano). *On considère la source X de l'exercice précédent.*

1. *Construire un code binaire de Shannon-Fano SF associé à X.*
2. *Coder le message* 024322 *à l'aide de SF.*
3. *Calculer la longueur moyenne et l'efficacité de SF relativement à X.*

Exercice 7 (Code de Gray). *Soit X une variable aléatoire discrète à valeurs dans $\{0, 1, 2, 3\}$ de loi p_X telle que $p_X(0) = 1/6$, $p_X(1) = 1/3$, $p_X(2) = p_X(3) = 1/4$. On considère le code de Gray $G\colon 0 \mapsto 00$, $1 \mapsto 01$, $2 \mapsto 11$, $3 \mapsto 10$. Sa particularité est que les mots de code associés à deux entiers successifs diffèrent seulement d'un bit.*

1. *Calculer l'entropie de X.*
2. *Coder le message* 301223 *à l'aide de G.*
3. *Calculer la longueur moyenne et l'efficacité de G relativement à X.*
4. *Construire le code de Huffman associé à la source X. Calculer sa longueur moyenne et son efficacité relativement à X.*
5. *Conclure.*

Exercice 8 (Codes de longueurs données). *Pour quelles valeurs de $n \geqslant 4$ existe-t-il un code binaire C uniquement décodable sur $\{1 \dots n\}$ dont les longueurs des mots de code vérifient :*
$\ell(C(1)) = \ell(C(2)) = 2$, $\ell(C(3)) = 3$, et $\ell(C(i)) = 5$ pour tout $i \in \{4 \dots n\}$?

Exercice 9 (Compression à perte). *Soit X une variable aléatoire à valeurs dans un ensemble fini S. Soit T une partie non vide de S.*

1. *Soit C un code binaire uniquement décodable sur T. Proposer un code binaire sur S de même longueur que C, avec une fonction de décodage de probabilité d'erreur $p(X \notin T)$.*
2. *On suppose dans cette question que S est un ensemble à 8 éléments, et la distribution de probabilités de X est $(\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16})$. Proposer un code à perte pour X avec une probabilité d'erreur $1/4$, puis un code avec une probabilité d'erreur $1/2$.*

Exercice 10 ($\star$ Symbole de poids fort dans le codage de Huffman). *On considère une variable aléatoire X à valeurs dans un alphabet $A = \{a_1, \dots, a_n\}$, dont la loi p_X vérifie $p_X(a_1) \geqslant p_X(a_2) \geqslant \cdots \geqslant p_X(a_n)$. Soit H le code de Huffman associé à X.*

1. *Montrer que si $p_X(a_1) > 2/5$ alors $H(a_1)$ est de longueur 1.*
2. *Montrer que s'il existe i tel que la longueur de $H(a_i)$ soit 1 alors $p_X(a_1) \geqslant 1/3$.*
3. *Que se passe-t-il entre $1/3$ et $2/5$?*