

M1 ACC - Théorie de l'information

Contrôle continu - corrigé

Notation Sur vos copies, chaque question est notée sur 1 point. La note finale (sur 24) est la partie entière supérieure de la moyenne pondérée de ces notes selon le barème indiqué ci-dessous.

Exercice 1 (Questions de cours). Soit $\mathcal{P} = (\Omega, p)$ un espace probabilisé fini. Soient X, Y deux variables aléatoires sur \mathcal{P} à valeurs dans un ensemble fini S . Soit $C: S \rightarrow \{0, 1\}^*$ un code binaire sur S .

1. (1 point) Définir l'entropie $H(X)$ de X , et l'entropie conditionnelle $H(X|Y)$ de X sachant Y .

Correction

$$H(X) = - \sum_{x \in S} p(X = x) \log_2(p(X = x))$$
$$H(X|Y) = - \sum_{x, y \in S} p(X = x, Y = y) \log_2(p(X = x|Y = y))$$

2. (1 point) Définir l'information mutuelle $I(X; Y)$ entre X et Y .

Correction

$$I(X; Y) = \sum_{x, y \in S} p(X = x, Y = y) \log_2 \left(\frac{p(X = x, Y = y)}{p(X = x)p(Y = y)} \right)$$

3. (2 points) Énoncer et démontrer la relation qui lie $I(X; Y)$, $H(X)$ et $H(X|Y)$.

Correction

$$I(X; Y) = H(X) - H(X|Y)$$

Preuve :

$$\begin{aligned} I(X; Y) &= \sum_{x, y \in S} p(X = x, Y = y) \log_2 \left(\frac{p(X = x, Y = y)}{p(X = x)p(Y = y)} \right) \\ &= \sum_{x, y \in S} p(X = x, Y = y) \log_2 \left(\frac{p(X = x, Y = y)}{p(Y = y)} \right) - \sum_{x, y \in S} p(X = x, Y = y) \log_2(p(X = x)) \\ &= -H(X|Y) - \sum_{x \in S} p(X = x) \log_2(p(X = x)) \\ &= H(X) - H(X|Y). \end{aligned}$$

4. (1 point) Définir la longueur moyenne et l'efficacité de C relativement à X .

Correction La longueur moyenne de C relativement à X est

$$\ell_X(C) = \sum_{x \in S} p(X = x) \ell(C(x)).$$

L'efficacité de C relativement à X est

$$\text{Eff}_X(C) = \frac{H(X)}{\ell_X(C)}.$$

5. (1 point) Qu'affirme le théorème de Shannon au sujet de cette efficacité ?

Correction Il affirme que $\text{Eff}_X(C) \leq 1$.

Exercice 2 (Codes préfixes et uniquement décodables).

1. (3 points) Les codes suivants sur $\{1 \dots 4\}$ sont-ils réguliers ? préfixes ? uniquement décodables ?

(a) (0, 01, 001, 0001)

(b) (0, 10, 110, 1110)

(c) (1, 01, 1, 00).

Correction

(a) Ce code est régulier, mais pas uniquement décodable car l'un des mots de code est la concaténation de deux autres : $0||01 = 001$. Il n'est donc pas non plus préfixe.

(b) Ce code est préfixe, donc régulier et uniquement décodable.

(c) Ce code n'est pas régulier puisque deux des mots de code sont égaux. Il n'est donc pas non plus préfixe, ni uniquement décodable.

2. (2 points) Donner un exemple de code uniquement décodable et non préfixe sur l'ensemble $\{1 \dots 5\}$.

Correction Prenons par exemple $C = (0, 01, 001, 0001, 00001)$. Il n'est pas préfixe car 0 est préfixe de tous les autres mots de code. Pour tout mot $x = b_1 \dots b_n \in \{0, 1\}^*$, notons $\bar{x} = b_n \dots b_1$ le mot obtenu à partir de x en lisant les bits de droite à gauche. Le code

$$D = (0, 10, 100, 1000, 10000) = (\overline{C(1)}, \overline{C(2)}, \overline{C(3)}, \overline{C(4)}, \overline{C(5)})$$

est préfixe, et donc uniquement décodable. Or pour tout $x = a_1 \dots a_n \in \{1 \dots 5\}^*$,

$$\overline{C^*(x)} = D(a_n) \dots D(a_1) = D^*(\bar{x}).$$

Par conséquent, pour tous $x, y \in \{1 \dots 5\}^*$, si $C^*(x) = C^*(y)$ alors $D(\bar{x}) = D(\bar{y})$ et $x = y$. Cela signifie que C est uniquement décodable.

3. (3 points) Soient n, p deux entiers naturels. On cherche à dénombrer les codes binaires C sur $\{1 \dots n\}$ tels que pour tout $i \in \{1 \dots n\}$, la longueur de $C(i)$ appartienne à $\{1 \dots p\}$.

(a) Combien y a-t-il de tels codes ?

Correction Il y en a autant que d'applications $\{1 \dots n\} \rightarrow S$, où

$$S = \bigcup_{0 < j \leq p} \{0, 1\}^j.$$

L'ensemble S a pour cardinal $\sum_{j=1}^p 2^j = 2^{p+1} - 2$. Le nombre de codes cherché est donc $(2^{p+1} - 2)^n$.

(b) Combien y a-t-il de tels codes réguliers ?

Correction *Le nombre de codes réguliers est le nombre d'injections $\{1 \dots n\} \rightarrow S$. Il correspond au nombre d'arrangements de n éléments de S , c'est-à-dire $n! \binom{2^{p+1}-2}{n}$.*

(c) *Combien y a-t-il de tels codes réguliers C vérifiant que $C(1)$ n'est un préfixe d'aucun autre mot de code ?*

Correction *Soit $\ell \in \{1 \dots p\}$. Soit $u \in \{0, 1\}^\ell$, et N_ℓ le nombre de codes adéquats tels que $C(1) = u$. Comme $|\{0, 1\}^\ell| = 2^\ell$, le nombre cherché est $\sum_{\ell=1}^p 2^\ell N_\ell$. De plus, N_ℓ est le nombre de codes réguliers sur $\{2 \dots n\}$ à valeurs dans $S_u = \{s \in S \mid u \text{ n'est pas un préfixe de } s\}$. Le cardinal de S_u est $|S| - 2^{p-\ell}$. Le nombre total cherché est donc*

$$(n-1)! \sum_{\ell=1}^p 2^\ell \binom{2^{p+1} - 2^{p-\ell} - 2}{n-1}.$$

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

Correction *D'après l'inégalité de Kraft-McMillan, il existe un tel code si et seulement si*

$$\frac{1}{2^2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{n-4}{2^5} \leq 1$$

c'est-à-dire

$$\frac{5}{8} + \frac{n-4}{32} \leq 1.$$

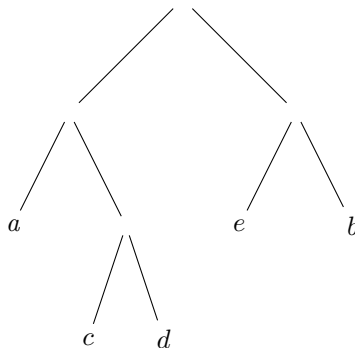
Ceci est équivalent à $n-4 \leq \frac{32-5}{8}$, i.e. $n \leq 16$.

Exercice 4 (Codes de Huffman et Shannon-Fano). *Soit X une variable aléatoire à valeurs dans $\{a, b, c, d, e\}$ de loi p_X définie par le tableau suivant.*

i	a	b	c	d	e
$p_X(i)$	$1/6$	$1/3$	$1/8$	$1/8$	$1/4$

1. (2 points) *Construire un code de Huffman C associé à X .*

Correction *L'algorithme vu en cours produit l'arbre suivant:*



Le code est donc donné par : $a \mapsto 00$, $b \mapsto 11$, $c \mapsto 010$, $d \mapsto 011$, $e \mapsto 10$.

2. (2 points) Construire un code de Shannon-Fano C' associé à X .

Correction On commence par construire le tableau des longueurs des mots de code ; la longueur de $C'(i)$ est $\lceil -\log_2(p(X=i)) \rceil$.

i	a	b	c	d	e
$\ell(C'(i))$	3	2	3	3	2

Les sommes croissantes des $1/2^{\lceil -\log_2(p(X=i)) \rceil}$ sont 0, 1/4, 1/2, 5/8, 3/4. On obtient donc le code suivant.

i	a	b	c	d	e
$C'(i)$	100	00	101	110	01

3. (1 point) Sans faire de calcul, donner des inégalités (strictes quand c'est possible) entre les longueurs moyennes $\ell_X(C)$, $\ell_X(C')$ et l'entropie $H(X)$.

Correction Comme les $p_X(i)$ ne sont pas tous des puissances de 2, il n'existe pas de code de longueur moyenne $H(X)$. Un code de Huffman étant toujours optimal, on en déduit que

$$H(X) < \ell_X(C) \leq \ell_X(C').$$

4. (2 points) Calculer les efficacités de C et C' relativement à X . Est-ce que C' est optimal ?

Correction L'entropie de X vaut :

$$H(X) = \frac{1}{6} \log(6) + \frac{1}{3} \log(3) + 2 \cdot \frac{1}{8} \log(8) + \frac{1}{4} \log(4) = \frac{1}{2} \log(3) + 17/12 \simeq 2.20.$$

La longueur moyenne de C est

$$\ell_X(C) = 2 \left(\frac{1}{6} + \frac{1}{3} + \frac{1}{4} \right) + 3 \frac{1}{4} = \frac{9}{4} = 2.25$$

et son efficacité vaut $H(X)/\ell_X(C) \simeq 0.98$. La longueur moyenne de C' est

$$\ell_X(C') = 2 \left(\frac{1}{3} + \frac{1}{4} \right) + 3 \left(\frac{1}{6} + \frac{1}{8} + \frac{1}{8} \right) = 29/12 \simeq 2.41$$

et son efficacité vaut $H(X)/\ell_X(C') \simeq 0.91$. Elle est strictement inférieure à celle de C , le code C' n'est donc pas optimal.