

1. Soient \mathcal{A} , \mathcal{X} , et \mathcal{S} des espaces de probabilité

$$\mathcal{A} = \{A, B, C, D\}, \quad \mathcal{X} = \{U, V, X, Y, Z\}, \quad \mathcal{S} = \{I, S, N, E\},$$

avec les probabilités suivantes :

x	A	B	C	D	x	U	V	X	Y	Z	x	I	S	N	E
$p(x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	$p(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$p(x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{3}$

et soient $C_1 : \mathcal{A} \rightarrow \{0, 1\}^*$, $C_2 : \mathcal{X} \rightarrow \{0, 1, 2\}^*$ et $C_3 : \mathcal{S} \rightarrow \{0, 1\}^*$ les codages :

x	A	B	C	D	x	U	V	X	Y	Z	x	I	S	N	E
$C_1(x)$	111	110	10	0	$C_2(x)$	2	1	02	01	00	$C_3(x)$	10	01	00	1

- a. Calculez les entropies de \mathcal{A} , \mathcal{X} , et \mathcal{S} , et déterminez les espérances mathématiques des longueurs des mots dans $C_1(\mathcal{A})$, $C_2(\mathcal{X})$, et $C_3(\mathcal{S})$.
- b. Lesquels de ces codages sont uniquement décodables et lesquels sont optimaux ?
- c. Comparez les espérances mathématiques des longueurs des mots de code avec les entropies.
- d. Pour chacune de ces comparaisons, déterminez si elle est compatible avec le théorème *codage source* de Shannon, ou sinon, expliquez pourquoi.

Solution.

- a. Les entropies et espérances des longueurs sont :

X	$H(X)$	$E(S_{C_i})$
\mathcal{A}	$7/4$	$7/4$
\mathcal{X}	$4/3 \log_2(3) \doteq 2.113$	$4/3$
\mathcal{S}	$7/6 + 1/2 \log_2(3) \doteq 1.959$	$5/3$

- b. Les codes C_1 et C_2 sont des codages préfixes, alors uniquement décodables, mais $C_3(\text{IE}) = C_3(\text{ES})$ (ou $C_2(\text{ENE}) = C_3(\text{IS})$), alors C_3 n'est pas uniquement décodable. Parmi les codages uniquement décodables, C_1 et C_2 sont optimaux (seulement les codages uniquement décodable – tel que son extension $C : X^* \rightarrow \mathcal{B}^*$ est sans perte – peut être optimaux).

c. On a trouvé que les codages C_1 et C_2 sont optimaux avec les comparaisons :

$$\begin{aligned} H(\mathcal{A}) &= E(S_{C_1}) = 7/4 \\ H(\mathcal{X}) &= E(S_{C_2}) \log_2(3) = 4/3 \log_2(3) \\ H(\mathcal{S}) &> E(S_{C_3}), \end{aligned}$$

car $7/6 + 1/2 \log_2(3) > 7/6 + 1/2 = 5/3$. Le facteur $\log_2(3)$ est nécessaire parce que c'est un codage en $\{0, 1, 2\}^*$. Pour les codage binaires, ce facteur est $\log_2(2)$, qui est égal à 1.

d. La dernière comparaison semble être une contradiction au théorème de Shannon, mais la borne du théorème $H(\mathcal{S}) \leq E(S_{C_3})$ n'applique pas parce que C_3 n'est pas uniquement décodable.

Remarque. Pour un codage uniquement décodable $C : X \rightarrow \{0, 1, \dots, m\}^*$, la borne $H(X) \leq E(S_C) \log_2(m)$ est toujours vérifiée, mais selon le théorème de Shannon (codage source), il existe un tel codage qui aussi vérifié

$$E(S_C) \leq \frac{H(X)}{\log_2(m)} + 1.$$

2. Soit $\Gamma = \{\alpha, \beta, \gamma, \delta, \varepsilon\}$ un espace de probabilité avec des probabilités suivantes :

x	α	β	γ	δ	ε
$p(x)$	0.4	0.3	0.1	0.1	0.1

a. Trouvez un codage binaire préfixe B avec l'espérance mathématique de longueur borné par $H(\Gamma) + 1$.

b. Trouvez un codage préfixe $C : \Gamma \rightarrow \{0, 1, 2\}^*$ avec $E(S_C) \log_2(3) < E(S_B)$, où $S_B(x)$ et $S_C(x)$ sont les longueurs des mots $B(x)$ et $C(x)$.

Remarque : Vous pouvez utiliser l'estimation $\log_2(3) \doteq 1,585$.

Solution. Pour B nous prenons un codage de Huffman, mais les probabilités sont plus proches à des puissances de 3 (0.333, 0.111) que des puissances de 2 (0.50, 0.250, 0.125). Par conséquent, les valeurs de $-\log_3(p(x))$ sont plus proches à des entiers, et nous trouvons un codage préfixe C avec longueurs des mots de code qui sont proche à ces entiers.

x	$p(x)$	$-\log_2(p(x))$	$B(x)$	$S_B(x)$	$-\log_3(p(x))$	$C(x)$	$S_C(x)$
α	0.4	1.32	0	1	0.83	0	1
β	0.3	1.76	10	2	1.10	1	1
γ	0.1	3.32	110	3	2.09	20	2
δ	0.1	3.32	1110	4	2.09	21	2
ε	0.1	3.32	1111	4	2.09	22	2

a. Comme B est un codage de Huffman, il vérifié la borne $E(S_B) \leq H(\Gamma) + 1$.

b. On calcule $E(S_C) \log_2(3) = 1.3 \log_2(3) \leq 1.3 \cdot 1.6 = 2.08 < E(S_B) = 2.1$.

3. Soit $C : \mathcal{X} = \{a, b, c, d, e\} \rightarrow \{0, 1, 2\}^*$ le codage

$$C(a) = 0, \quad C(b) = 1, \quad C(c) = 20, \quad C(d) = 21, \quad C(e) = 22,$$

étendu a \mathcal{X}^* par $C(x_1 \dots x_n) = C(x_1) \dots C(x_n)$.

a. Trouver les mots de codes de $C(\mathcal{X}^*)$ dans $\{0, 1, 2\}^n$ pour $n = 1$ et $n = 2$. Quels mots de $\{0, 1, 2\}^n$ ne sont pas dans $C(\mathcal{X}^*) \cap \{0, 1, 2\}^n$?

b. Démontrez l'égalité

$$|C(\mathcal{X}^*) \cap \{0, 1, 2\}^n| \geq 2 \cdot 3^{n-1}. \quad [\text{corrigé}]$$

c. Si $T : \mathcal{X} \rightarrow \mathcal{A}$ est un codage, tel que pour un $\varepsilon > 0$, la borne

$$|T(\mathcal{X}^*) \cap \mathcal{A}^n| \geq \varepsilon |\mathcal{A}^n|$$

est vérifiée, trouvez le taux de transmission de T . Par conséquence, quel est le taux de transmission de C ?

Solution.

a. Par calcul explicite, on trouve

$$C(\mathcal{X}^*) \cap \{0, 1, 2\} = \{0, 1\} \text{ et } C(\mathcal{X}^*) \cap \{0, 1, 2\}^2 = \{00, 01, 10, 11, 20, 21, 22\}.$$

En générale, les mots de code de longueur n sont ceux qui ne termine pas avec un nombre impair de 2.

b. Pour chaque $x_1 \dots x_{n-1}$ dans $\{0, 1, 2\}^{n-1}$ et chaque x_n dans $\{0, 1\}$, le mot $x = x_1 \dots x_{n-1} x_n$ est dans $C(\mathcal{X}^*)$. Alors $|C(\mathcal{X}^*) \cap \{0, 1, 2\}^n| \geq 2 \cdot 3^{n-1}$. Plus precisement,

$$|C(\mathcal{X}^*) \cap \{0, 1, 2\}^n| = 2(3^{n-1} + 3^{n-3} + \dots + 3^e),$$

où $e = 0$ si n est impair et $e = 1$ si n est pair.

c. Par définition,

$$\begin{aligned} r &= \limsup_{n \rightarrow \infty} \frac{\log_2 |T(\mathcal{X}^*) \cap \mathcal{A}^n|}{\log_2 (|\mathcal{A}^n|)} \geq \limsup_{n \rightarrow \infty} \frac{\log_2 (\varepsilon |\mathcal{A}^n|)}{n \log_2 (|\mathcal{A}|)} \\ &= \limsup_{n \rightarrow \infty} \frac{n \log_2 (|\mathcal{A}|) + \log_2 (\varepsilon)}{n \log_2 (|\mathcal{A}|)} = 1. \end{aligned}$$

Le taux de transmission est alors 1 pour C , en prennant $\varepsilon = 2/3$.

4. Supposez qu'un canal binaire symétrique a probabilité d'erreur 0.1, et, pour $i = 1, 2$ et 3 , que $C_i : \{0, 1\}^i \rightarrow \{0, 1\}^{2^i}$ sont les codages suivants

$$C_1(x_1) = x_1x_1,$$

$$C_2(x_1x_2) = x_1x_2x_1x_3 \text{ où } x_3 = x_1 + x_2,$$

$$C_3(x_1x_2x_3) = x_1x_2x_3x_4x_5x_6 \text{ où } x_4 = x_1 + x_2, x_5 = x_1 + x_3, \text{ et } x_6 = x_2 + x_3,$$

pour l'addition binaire modulo 2.

- a. Est-ce qu'il est possible d'avoir un codage correcteur d'erreurs avec taux de transmission $1/2$, tel que la probabilité qu'un mot reçu ne puisse pas être corrigé est moins que 10^{-16} ?
- b. Quels sont les distances minimums de ces codages ?
- c. Pour $\mathcal{C}_3 = C_3(\{0, 1\}^3)$, combien de mots est-ce qu'il y a dans $\bigcup_{x \in \mathcal{C}_3} B(x, 1)$?
- d. Pour une stratégie de correction d'erreurs, qui, pour chaque mot reçu y , retourne le mot de code $x \in \mathcal{C}_3$ si y est dans $B(x, 1)$, ou y si aucun tel mot x existe : (i) quelle est la probabilité qu'un mot reçu n'est pas corrigé en un mot de code, et (ii) quelle est la probabilité qu'un mot reçu est corrigé en un mot de code autre que le mot transmis ?

Remarque : Vous pouvez utiliser l'estimation

$$H(0.1) = -0.1 \log_2(0.1) - 0.9 \log_2(0.9) \doteq 0.469.$$

Pour **4.d.** vous pouvez utiliser l'observation que les probabilités sont indépendantes du mot transmis, et supposer que 000000 est le mot transmis.

Solution.

- a. Oui. Selon le théorème de Shannon (canal bruyant), il existe un tel codage car

$$r = 1/2 < C = 1 - 0.469 = 0.531.$$

- b. Les distances minimums sont $d(C_1) = d(C_2) = 2$ et $d(C_3) = 3$.
- c. On a $|B(x, 1)| = 7$, et alors l'union de huit boules a 56 éléments (comme $d(C_3) = 3$, les boules sont disjointes).
- d. (i) Pour un canal bruyant avec probabilité d'erreur $\varepsilon = 0.1$, la probabilité de correction de zéro ou d'un erreur est :

$$\sum_{i=0}^1 \binom{6}{i} \varepsilon^i (1 - \varepsilon)^{6-i} = 0.9^6 + 6(0.1)(0.9)^5 = 0.885735$$

(ii) Le complement de cette probabilité sont les mots qui ne sont pas corrigés ou qui sont corrigés vers le mauvais mot :

$$\sum_{i=2}^6 \binom{6}{i} \varepsilon^i (1 - \varepsilon)^{6-i} = 1 - 0.885735 = 0.114265.$$

Mais il y a huit mots exceptionnels y (les 64 mots moins les 56 dans les boules $B(x, 1)$) qui sont a distance au moins 2 de tous les mots de codes :

$C(X)$	y	$d(000000, y)$
000000	100001	2
001011	101010	3
010101	110100	3
011110	111111	6
100110	000111	3
101101	001100	2
110011	010010	2
111000	011001	3

Ce sont les mots qui ne seront pas corrigés. Si on suppose que 000000 est le mot transmis, il y a trois mots exceptionnels à distance 2, quatre mots exceptionnels à distance 3 et un mot exceptionnel à distance 6. Il faut enlever la probabilité de ces mots non-corrigés pour trouver la probabilité d'une mauvaise correction :

$$\sum_{i=2}^6 \binom{6}{i} \varepsilon^i (1 - \varepsilon)^{6-i} - (3 \varepsilon^2 (1 - \varepsilon)^4 + 4 \varepsilon^3 (1 - \varepsilon)^3 + \varepsilon^6) = 0.091665$$