# TD 2 : Codage de Source
# Compression de l'information

Yann Rotella

2026

**Exercice 1.** *De quel entier on parle ?.*
L'objectif consiste à se familiariser un peu plus avec la notion d'entropie.

1. Quel est le plus petit nombre (on ne cherchera pas à le prouver, restons informel) de questions fermées (à réponses oui ou non) nécessaires pour identifier un entier pris aléatoirement entre 0 et $2^6 - 1$ ?
2. À quoi correspond ce nombre ?
3. Donner un ensemble de questions permettant de déterminer exactement l'entier dont on parle.
4. Montrer que pour un ensemble de taille $K$, il est *nécessaire* de poser au minimum $\log_2(K)$ questions fermées.
5. Montrer que si une seule question n'est pas « équilibrée » et que $K = 2^k$ pour un certain $k$, alors il faudra au moins une question supplémentaire (au moins $k + 1$ questions).

**Exercice 2.** *Huffman et application du théorème de Shannon.*
Soit $X$ une variable aléatoire à valeur dans $\{1, \dots, n\}$. On note $p_i = Pr[X = i]$. On veut déterminer le plus rapidement possible la valeur de $X$ en posant des questions fermées (oui/non). On peut demander n'importe quel type de questions.

1. Donner une borne inférieure maximale sur le nombre moyen de questions à poser.
2. Donner une borne supérieure minimale sur le nombre moyen de questions à poser.
3. Trouver un ensemble optimal de questions à poser pour le cas suivant.

| $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ |
|---|---|---|---|---|---|
| $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{8}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |

**Exercice 3.** *Codes de Huffman.*
On considère une variable aléatoire $X$. On note alors $X^{\otimes i}$ la variable aléatoire $(X_1, X_2, \dots, X_i)$ où les $X_i$ sont des v.a. suivant la distribution de $X$ et indépendantes.

1. Donner alors des codes de Huffman pour $X^{\otimes 2}$ et $X^{\otimes 3}$ avec
$$P[X = 0] = \frac{9}{10}$$
2. Calculer leurs longueurs moyennes $|\phi(X^{\otimes 2})|$ et $|\phi(X^{\otimes 3})|$ et comparez avec $H(X^{\otimes 2})$ et $H(X^{\otimes 3})$. Calculez les ratios correspondant.
3. Ce que vous obtenez est-il normal ? Expliquer

**Exercice 4.** *Codes de Huffman et préfixes.*
Notre but ici est de montrer que le codage de Huffman $\phi_H$ est un code préfixe optimal. Optimal dans le sens où il n'y a pas de code préfixe $\phi$ pour une source $X$ tel que

$$|\phi(X)| < |\phi_H(X)|$$

1. Montrer qu'il existe un code préfixe optimal $\phi$ tel que
   (a) Si $p(x_j) > p(x_k)$, alors $|\phi(x_j)| \leq |\phi(x_k)|$.
   (b) Les deux plus longs mots de codes ont la même longueur et correspondent aux symboles les moins fréquents.
   c) Ces deux mots de code diffèrent seulement au dernier bit.
2. En déduire que le codage de Huffman est optimal.

**Exercice 5.** *Combien d'aléa avons-nous besoin ?.*
L'objectif de cet exercice consiste à chercher combien de lancer de pièces nous avons besoin de faire afin de simuler une distribution qui n'est pas uniforme.

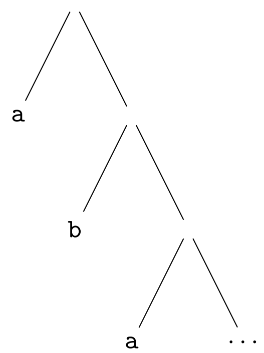
1. On considère la distribution suivante :

| $x_i$ | $p(x_i)$ |
|---|---|
| a | $\frac{1}{2}$ |
| b | $\frac{1}{4}$ |
| c | $\frac{1}{4}$ |

   ainsi que l'accès à une pièce de monnaie, i.e. une v.a. $X$ uniformément distribuée sur $\{0, 1\}$. Donner deux stratégies qui permettent de tirer nos trois lettres selon la distribution donnée, afin qu'en moyenne, on ait besoin de tirer la pièce de monnaie
   (i) deux fois
   (ii) $1,5$ fois
2. Calculer l'entropie de la distribution. Comment interpréter ce résultat ? Pouvons-nous faire mieux ?

Maintenant on va chercher à généraliser. On considère une distribution $(p(x))_{x\in\mathcal{X}}$. On veut tirer $x$ selon cette distribution en lançant des pièces de monnaie non-baisées.

3. Tirer des pièces est équivalent à parcourir un arbre binaire. Dessiner les arbres associés aux stratégies données à la première question.
4. Soit $\mathcal{X} = \{\mathtt{a}, \mathtt{b}\}$ avec $p(a) = 2/3$, montrer que l'arbre infini suivant permet d'obtenir cette distribution.




**Exercice 6.** *Codage de suites d'entiers.*
Le but est de réaliser un codage universel des entiers naturels strictement positifs. C'est-à-dire tel que

$$\phi : \mathbb{N} \to \{0, 1\}^*$$

tel que

$$\Phi : \mathbb{N}^* \to \{0, 1\}^*$$

soit injectif avec $\Phi(x_1, \ldots, x_i) = \phi(x_1)|| \cdots ||\phi(x_i)$ pour tout $i$.

Pour tout entier $a > 0$ se décomposant en base 2 sous la forme $2^{\ell-1} + \sum_{i=0}^{\ell-2} a_i 2^i$, nous noterons $\ell(a) = \ell$ la longueur de sa décomposition en base 2 et $B_2(a) = a_{\ell(a)-2} \ldots a_1 a_0$ c'est-à-dire son écriture en base 2 sans le '1' de tête. Pour $a = 1$, $B_2(1)$ est la séquence vide.

Nous noterons $b^k$ la séquence binaire composée de $k$ fois le symbole $b$ et $s_1 \parallel s_2$ sera la concaténation des séquences $s_1$ et $s_2$.

1. Pourquoi ne peut t'on pas utiliser $\varphi(a) = 1 \parallel B_2(a)$ ?

2. Le code $\varphi_1(a) = 0^{\ell(a)-1} \parallel 1 \parallel B_2(a)$ est il à décodage unique ? Si oui, donner un algorithme de codage et de décodage d'une séquence d'entiers.
3. Montrer que toute séquence binaire est une séquence codée (ou le début d'une séquence codée) par le codage associé à $\varphi_1$.
4. On peut donc, en utilisant le décodeur de $\varphi_1$, construire une source produisant des éléments de $\mathbf{N} \setminus \{0\}$ à partir d'une source binaire. Nous supposerons que la source binaire en question est sans mémoire et uniformement distribuée (*i.e.* '0' et '1' sont émis avec une probabilité 1/2). Quelle est la probabilité d'obtenir un entier donné $n$ ?
5. Quelle est l'entropie de $\mathbf{N} \setminus \{0\}$ muni de cette loi ?
6. Combien de bits de la source faudra t'il utiliser en moyenne pour produire un entier ?