Série 11

1 Calcul d'entropie

Au moyen de la formule

$$H(X) = \sum_{j=1}^{n} p_j \, \log_2\left(\frac{1}{p_j}\right) \tag{1}$$

calculez l'entropie de la séquence de 20 lettres (espaces et point d'exclamation inclus) :

HASTA LA VISTA BABY !

en écrivant tout d'abord le résultat sous la forme :

$$H = a + b \, \log_2(3) + c \, \log_2(5)$$

où a, b, c sont des fractions (positives ou négatives). Estimez ensuite le résultat en utilisant les approximations suivantes : $\log_2(3) \simeq 1.58$ et $\log_2(5) \simeq 2.32$.

2 Un peu de magie noire

a) Calculez l'entropie de la séquence de lettres suivante (sans l'espace) :

AVADA KEDAVRA

• en écrivant d'abord le résultat sous la forme $a + b \, \log_2(3) + c \, \log_2(5)$, où a, b, c sont des fractions de nombres entiers ;

• puis en calculant le résultat numérique ou en l'approximant avec $\log_2(3) \simeq 1.58$ et $\log_2(5) \simeq 2.32$.

b) Créez un dictionnaire pour cette même séquence de lettres à l'aide de l'algorithme de Shannon-Fano. Combien de bits utilisez-vous pour représenter la séquence ? Essayez différentes versions de l'algorithme et comparez.

c) Créez ensuite un dictionnaire à l'aide de l'algorithme de Huffman. Combien de bits utilisez-vous pour représenter la séquence ? À nouveau, essayez différentes versions de l'algorithme et comparez.

d) Comparez les résultats obtenus aux points a), b) et c). Ceci est-il cohérent avec ce que vous avez appris au cours ?

3 De bien mauvais dicos

Pour encoder le mot "BANANA" sous forme de 0 et de 1, six minions ont proposé les dictionnaires suivants :

a)

A	N	B
1	11	111

b)

A	N	B
0	1	11

c)

A	N	B
0	10	110

d)

A	N	B
00	01	10

e)

A	N	B
00	01	1

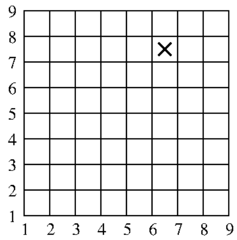
f)

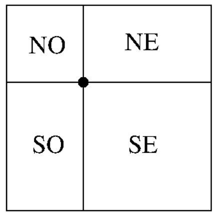
A	N	B
0	01	001

Mais Grou n'est satisfait d'aucun d'entre eux. Pouvez-vous expliquer le(s) défaut(s) de chacun ? Et aussi proposer un meilleur dictionnaire ?

4 À la recherche d'un trésor

Vous êtes à la recherche d'un trésor, dont la position est une case choisie au hasard et indiquée par une croix sur le plan à gauche ci-dessous. Pour trouver la position du trésor, vous devez choisir un point sur la grille et poser la question : dans quelle direction se trouve le trésor ? Un oracle vous indiquera alors s'il est au nord-est (NE), au nord-ouest (NO), au sud-est (SE) ou au sud-ouest (SO), comme indiqué par exemple à droite ci-dessous. Bien sûr, il vous faudra interroger l'oracle en plusieurs points de la grille pour localiser le trésor.





a) Si vous utilisez une stratégie optimale, combien de questions faudra-t-il pour localiser le trésor ?

b) Dans l'exemple ci-dessus à gauche, quelle est la suite des points où vous interrogerez l'oracle avec votre stratégie optimale ? (utilisez la notation (a, b) pour un point, avec a, b allant de 1 à 9, a dénotant l'abscisse et b l'ordonnée du point).

c) Proposez un système *efficace* d'encodage avec des bits pour transmettre l'information de l'emplacement du trésor à un ami. En se basant sur votre système, quelle sera la séquence de bits que vous transmettrez à votre ami dans l'exemple ci-dessus à gauche ?

5 Questions d'examens passés

a)

Une source X émet les symboles suivants :

$$\{A, B, C, D\}$$

avec les probabilités respectives :

$$\left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right\}$$

On note : $H(X)$ l'entropie de la source, $L_H(X)$ la longueur moyenne du code de Huffman et $L_{\mathrm{SF}}(X)$ la longueur moyenne du code Shannon–Fano.

Quelle est la valeur de $L_H(X)$?
- ☐ 1.5
- ☐ 1.625
- ☐ 1.75
- ☐ 2.0

Quelle relation est correcte ?
- ☐ $H(X) < L_H(X) = L_{\mathrm{SF}}(X)$
- ☐ $H(X) = L_H(X) = L_{\mathrm{SF}}(X)$
- ☐ $H(X) = L_H(X) < L_{\mathrm{SF}}(X)$
- ☐ $H(X) < L_H(X) < L_{\mathrm{SF}}(X)$

b)

Pour une source utilisant un alphabet de taille n, l'entropie $H(X)$ vérifie $0 \leq H(X) \leq \log_2 n$.
- ☐ VRAI
- ☐ FAUX

c)

La longueur moyenne d'un code de Huffman est strictement inférieure à l'entropie de la source.
- ☐ VRAI
- ☐ FAUX

d)

On souhaite transmettre la phrase suivante, espaces inclus :

`LA BASE SALSA EST LA`

A. Calculez la fréquence de chaque caractère (lettres et espaces).

B. Construisez un arbre de Huffman et donnez le code binaire de chaque caractère.

C. Calculez le nombre total de bits nécessaires pour encoder le message (i) avec le code de Huffman (B_{Huffman}) et (ii) avec un encodage à taille fixe minimale ($B_{\text{fixe-min}}$). Donnez ensuite le gain de compression en utilisant :

$$\text{gain de compression} = 1 - \frac{B_{\text{Huffman}}}{B_{\text{fixe-min}}}.$$

Les résultats peuvent être donnés sous forme de fraction ou de valeur décimale.

D. Comparez la longueur moyenne théorique minimale à la longueur moyenne obtenue par votre code de Huffman.