# Int. M1 - Codage et Compression

2018

## * Exo 1 :

1) **H (X) :** $p(1) = p \ll 1$ et $p(0) = 1 - p$.

$$\boxed{H(X) = -\sum_{i=0}^{1} p_i \log_2(p_i)}$$ 0,25

$$= -p_0 \log_2(p_0) - p_1 \log_2(p_1)$$

$$= -(1-p)\log_2(1-p) - p\log_2(p)$$ avec : $(1-p)\log_2(1-p) \simeq \frac{p}{\ln(2)}$

$$= -p\,(\underbrace{\log_2 p - \ln(2)}_{\simeq \log_2(p) \text{ pour } p \ll 1})$$

Donc :

$$\boxed{H(X) \simeq -p\log_2(p)}$$ 0,5

2) **$X^2$ :** a) Probabilités

$P(ij) = P(i) \cdot P(j)$.

| symbole | probabilité |
|---|---|
| 0 0 | $(1-p)^2$ 0,25 |
| 0 1 | $p(1-p)$ 0,25 |
| 1 0 | $p(1-p)$ 0,25 |
| 1 1 | $p^2$ 0,25 |

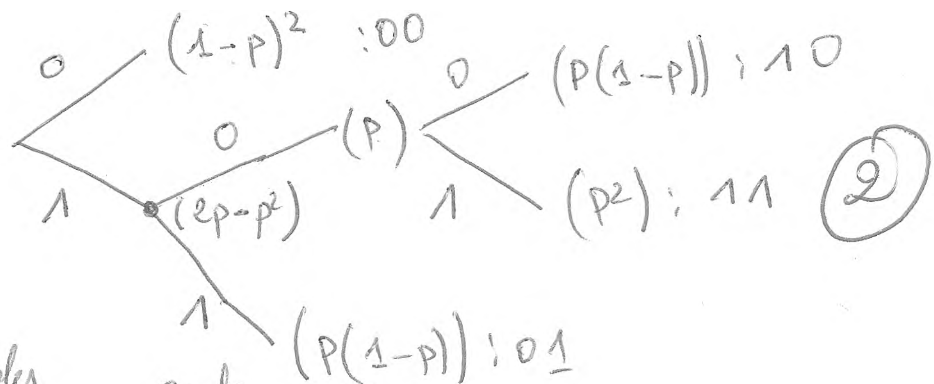
b) **$H(X^2)$ :**

$$H(X^2) = -\sum p_{ij} \log_2(p_{ij}) = -p_{00}\log_2(p_{00}) - p_{01}\log_2(p_{01}) - \cdots$$

$$H(X^2) = 2H(X) \simeq -2p\log_2(p)$$ 1

## c) Codage de Huffman de $X^2$ :

Comme $p \ll 1$ donc : $p^2 < p(1-p) < (1-p)^2$




| symboles | code | |
|---|---|---|
| 00 | 0 | $n=1$ |
| 01 | 11 | $n=2$ |
| 10 | 100 | $n=3$ |
| 11 | 101 | $n=3$ |

0,5

## d) Longueur moyenne $\overline{L}$ :

$$\boxed{\overline{L} = \sum_i p(x_i)\, n_i}$$ 0,25

$x_i$ : symbole
$n_i$ : longueur

$$\overline{L} = (1-p)^2 + 2p(1-p) + 3p(1-p) + 3p^2$$

$$\boxed{\overline{L} = 1 + 3p - p^2}$$ 0,5

## e) Efficacité du code :

$$\boxed{\eta_2 = \frac{H(X^2)}{\overline{L}}} = \frac{-2p\log_2(p)}{1+3p-p^2}$$ 0,5

0,25

De (1) : $\boxed{\eta_1 = -p\log_2(p)}$ et comme : $p \ll 1$ : $p \to 0$

Donc : $\boxed{\eta_2 \simeq 2\eta_1}$ 0,75 exo 2

1)°

$$H(S) = -\sum_{i=1}^{6} P_i \log_2(P_i)$$

$$\boxed{H(S) = 2,14 \text{ bit/symbole}}$$ 0,5

Code a 6 symboles $\Rightarrow L_{min} = 3$ 0,25

$$\eta = \frac{2,14}{3} = 0,714$$ 0,25

2)° $\bar{L}_x = 2,1 < H(S)$ 0,25 $\Rightarrow$ son code ne peut pas être uniquement déchiffrable 0,25 $\Rightarrow$ pas être préfixe.

Code ternaire :

$$\log_3(N) = \frac{\ln(N)}{\ln(3)}$$ on utilise $\log_3$ à la place de $\log_2$

$$H'(S) = -\sum_{i=1}^{6} P_i \log_3(P_i) \Rightarrow \boxed{H'(S) = 1,352 \text{ bit/sym}}$$ 0,5

c'est possible.

3)°

| symb | code |
|---|---|
| a | 0 |
| b | 100 |
| c | 1100 |
| d | 101 |
| e | 111 |
| f | 1101 |

$\Rightarrow$ code préfixe 0,25 $\Rightarrow$ 0,25 uniquement déchiffrable.

1101 | 0 | 1100 | 0 | 101 | 111 $\Rightarrow$ facade
f a c a d e 0,25

$$L'_{moy} = \sum_{i=1}^{6} P_i(x_i)\, n_i \Rightarrow \boxed{L'_{moy} = 2,3} < 3 = L_{min}$$ 0,25

Donc il est plus efficace 0,25

4)° a)° longueur 1 : $\sum 2^{-l_i} \le 1 \Rightarrow \frac{1}{2} + \frac{5}{8} > 1 \Rightarrow$ pas uniquement déchiffrable 0,25

$\cancel{\log_3 N = \frac{\ln(N)}{\ln(3)}}$

0,25

b) puisque on peut pas le faire avec des longueurs de 1
=> il faut le faire avec 2 et 3.

$$\begin{array}{cccccc} l_1 & l_2 & l_3 & l_4 & l_5 & l_6 \\ (2, & 2, & 3, & 3, & 3, & 3) \\ a & b & c & d & e & f \end{array} \Rightarrow \sum_{i=1}^{6} 2^{-l_i}$$ 0,25

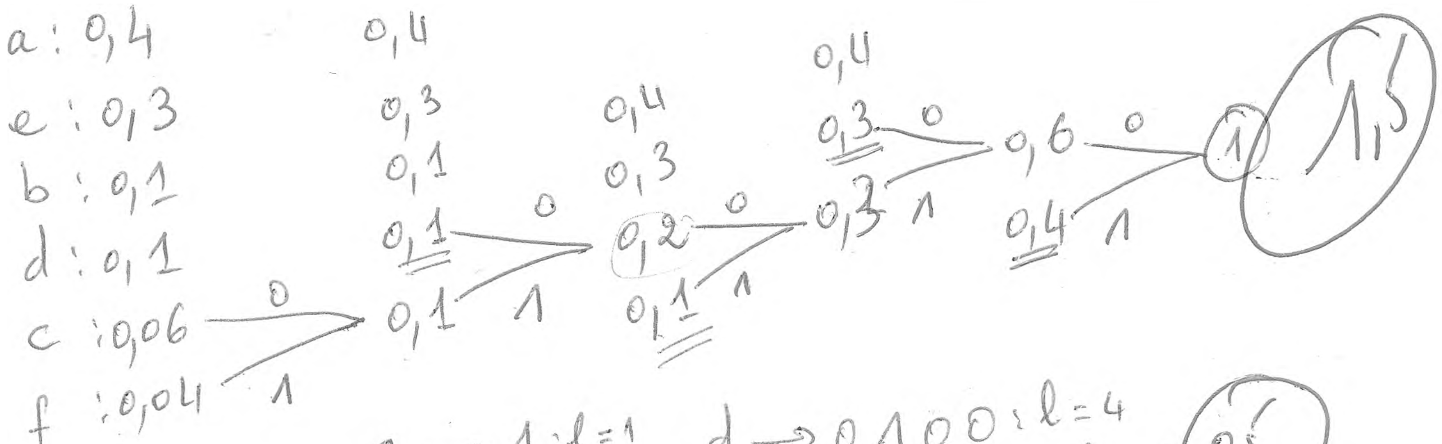
plus efficace => coder les 2 symboles les plus probables (a et e). 0,5

$$\boxed{\begin{array}{l} a \to 00 \\ e \to 01 \end{array}} \quad \begin{array}{l} b \to 100 \\ c \to 101 \end{array} \quad \begin{array}{l} d \to 110. \\ f \to 111. \end{array}$$

$$\boxed{L_{moy} = 2,3}$$ 0,25

Donc $\eta = \frac{H(S)}{L_{moy}}$ la même que Monsieur Y. 0,5

5) <u>Codage de Huffman :</u>




$a \to 1 : l=1$ $\quad d \to 0100 : l=4$
$e \to 00 : l=2$ $\quad f \to 01011 : l=5$ 0,5
$b \to 011 : l=3$ $\quad c \to 01010 : l=5$

$$L_{moy} = \sum_{i=1}^{6} P(x_i)\, l_i \Rightarrow \boxed{L_{moy} = 2,2} \Rightarrow \eta = \frac{2,14}{2,2} = 0,9727$$ 0,25

0,5 $\boxed{\eta = 97,27\%}$