

TP 3 : Compression de l'information par codage de Huffman

Question 0

On sait que :

- La quantité d'information est positive ou nulle.
- La quantité d'information contenue dans chaque symbole est inversement proportionnelle à sa certitude.
- La quantité d'information associée à des données statistiquement indépendantes est additive.

On a donc :

$$I(x_i) = \log_m \left(\frac{1}{p_i} \right)$$

$I(x_i)$ représente la quantité d'information. $\left(\frac{1}{p_i}\right)$ représente l'inverse de la certitude. Le logarithme est une fonction croissante.

Additivité :

Soit 2 symboles tels que :

$$1 > p_b > p_a$$

$$1 < \frac{1}{p_b} < \frac{1}{p_a}$$

$$0 < \log_m \left(\frac{1}{p_b} \right) < \log_m \left(\frac{1}{p_a} \right)$$

$$0 < I(x_b) < I(x_a)$$

Donc l'information portée par un symbole est d'autant plus grand que sa probabilité d'apparition est faible.

$$I(x_a) + I(x_b) = \log_m \left(\frac{1}{p_a} \right) + \log_m \left(\frac{1}{p_b} \right)$$

$$I(x_a) + I(x_b) = \log_m \left(\frac{1}{p_a} \times \frac{1}{p_b} \right)$$

$$I(x_a) + I(x_b) = I(x_a \text{ ET } x_b)$$

Car x_a et x_b sont indépendants

I est donc indépendante et additive (3^{ème} propriété).

$$E(I) = \sum_{i=1}^M p_i x_i$$

Entropie et contenu informatif :

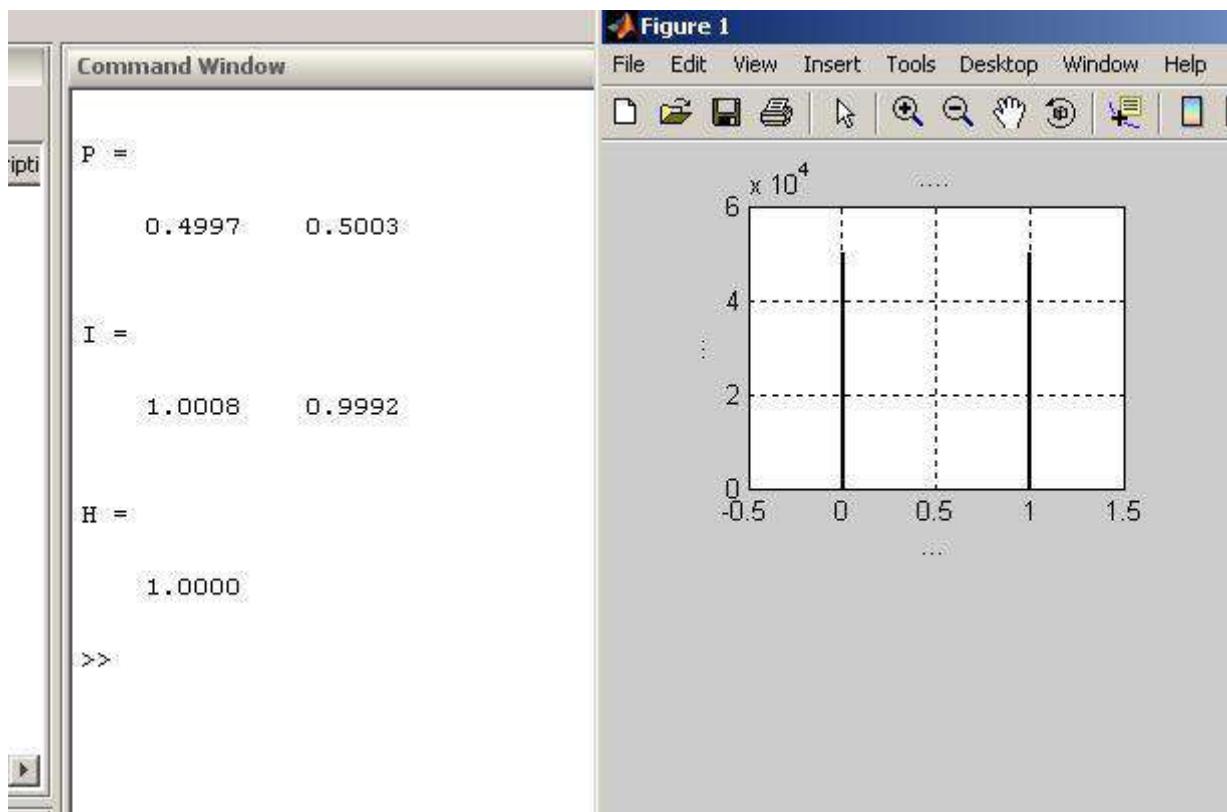
$$H(X) = - \sum_{i=1}^M p_i \times \log_m(p_i)$$

$$H(X) = \sum_{i=1}^M p_i \times I(x_i)$$

Si la source X génère les symboles de manière équiprobable alors l'entropie de la source sera égale au contenu informatif.

Question 2

P est un tableau contenant les probabilités p_i :



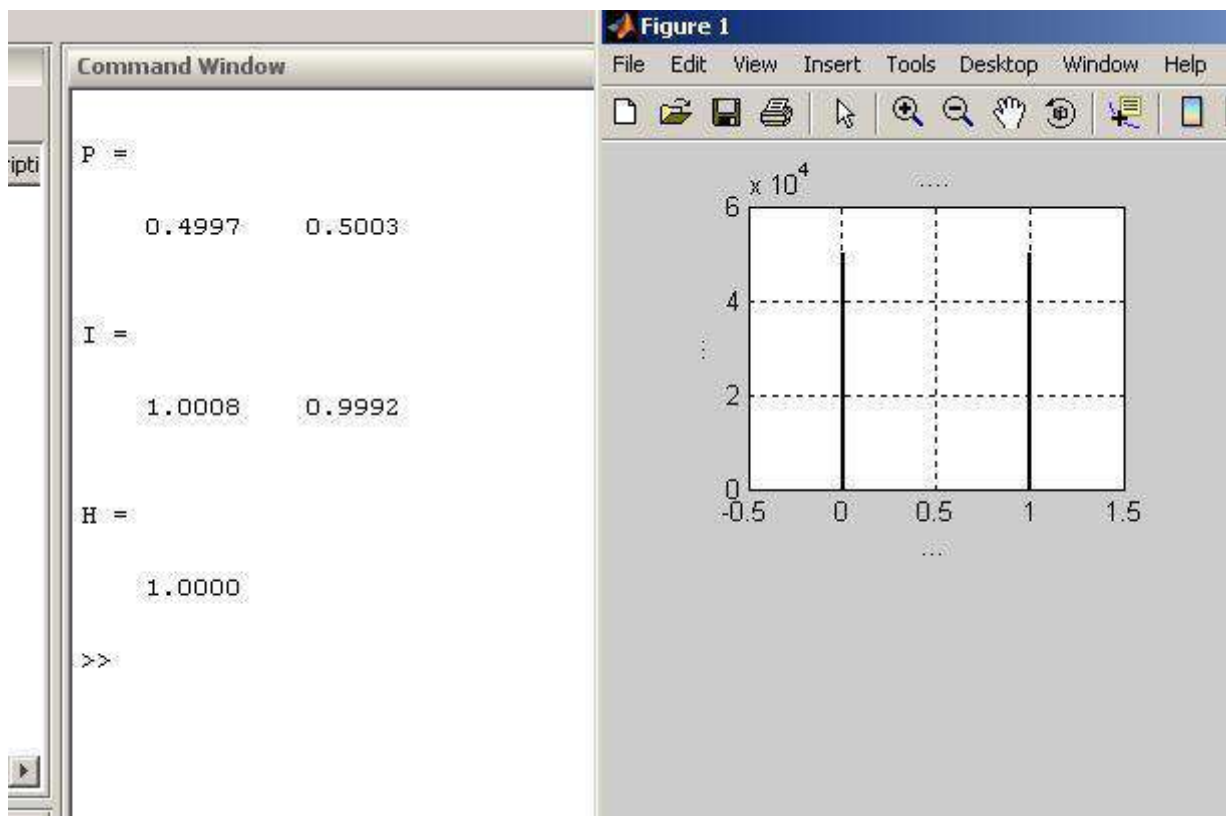
Voici la ligne 21 complétée

```

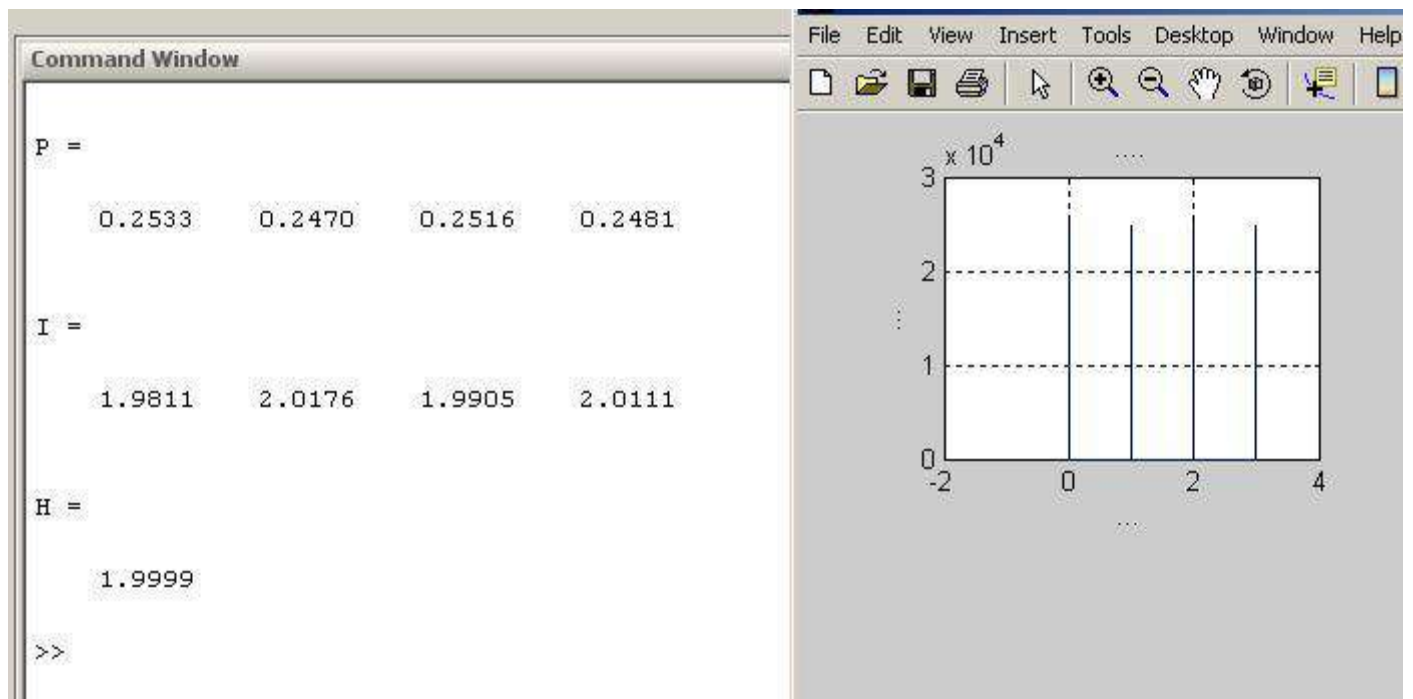
P=P/N
I=-log2 (P)

H=0;
for i=1:M
    H=H-P (i) *log2 (P (i)) ;
end;
H
  
```

Pour $L=1$:



Pour $L=2$:



Pour $L=3$:

P =

0.1259 0.1259 0.1245 0.1267 0.1251 0.1236 0.1235 0.1248

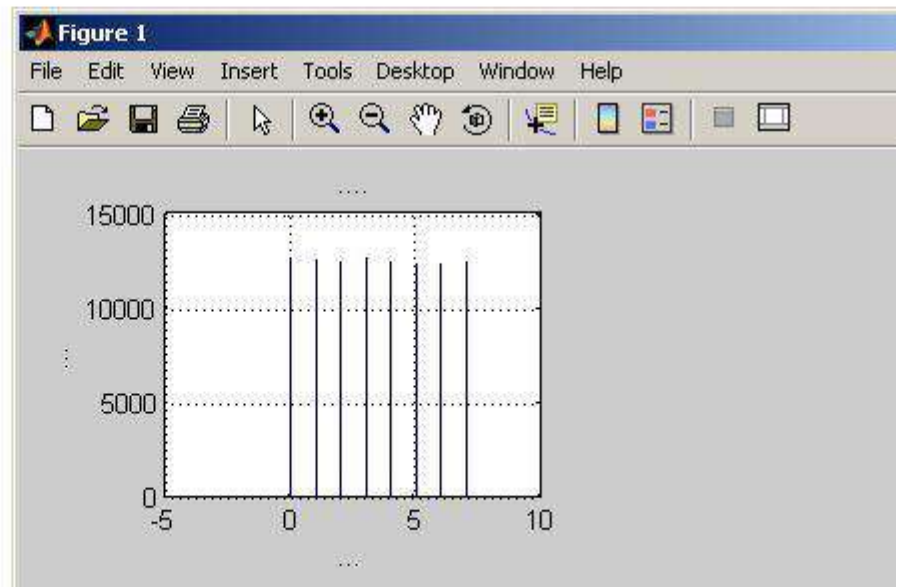
I =

2.9901 2.9900 3.0053 2.9806 2.9992 3.0162 3.0171 3.0018

H =

2.9999

>>



Pour L=4 :

P =

Columns 1 through 11

0.0612 0.0632 0.0630 0.0620 0.0623 0.0624 0.0632 0.0612 0.0635 0.0624 0.0619

Columns 12 through 16

0.0617 0.0635 0.0621 0.0628 0.0636

I =

Columns 1 through 11

4.0306 3.9839 3.9896 4.0121 4.0049 4.0014 3.9832 4.0294 3.9773 4.0032 4.0137

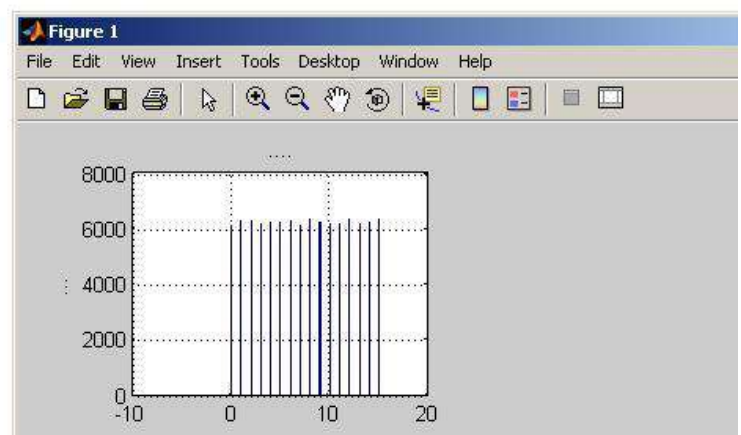
Columns 12 through 16

4.0174 3.9778 4.0097 3.9936 3.9739

H =

3.9999

>>



Pour chaque L , $H=L$. L'entropie est donc égale à la longueur du code. De plus, plus la longueur du code est petite et plus l'entropie est faible (et inversement).

L'entropie est également égale à la moyenne des quantités d'information.

De plus si un code est déchiffrable, alors sa longueur L est toujours supérieure ou égale à l'entropie de la source :

$$L \geq H(S)$$

On notera qu'un code qui vérifie $H = L$ est un code « absolument optimal ».

Question 3

Comparaison de P et Pz :

P =

0.1240 0.1225 0.1269 0.1250 0.1247 0.1266 0.1257 0.1247

I =

3.0119 3.0290 2.9783 3.0001 3.0031 2.9818 2.9924 3.0039

H =

2.9999

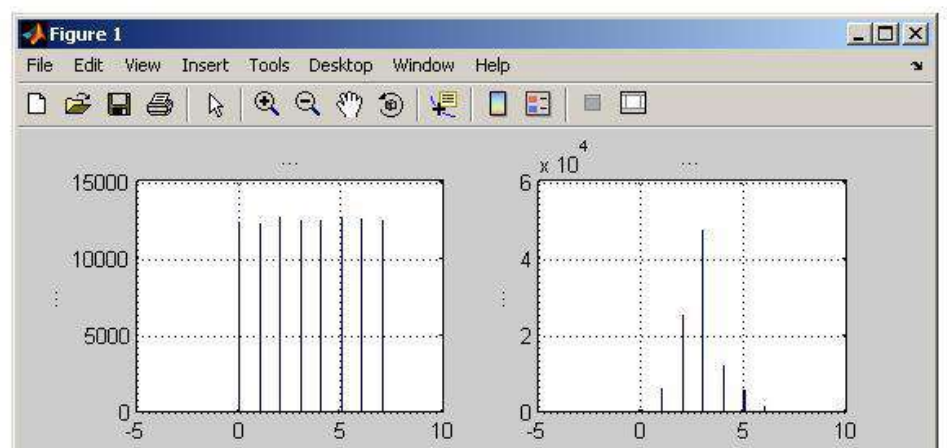
Pz =

0.0078 0.0625 0.2539 0.4713 0.1250 0.0600 0.0156 0.0039

H_z =

2.0616

>>



Pour P, la source génère des probabilités équiprobables.

Pour Pz, on rentre les valeurs manuellement.

La source générée est une source aléatoire. Cette source génère des symboles selon leur probabilité. Ces probabilités sont entrées au préalable.

```

Pz

Hz=0;
for i=1:M
    Hz=Hz-Pz(i)*log2(Pz(i));
end;
Hz

```

Comme précédemment, on applique la formule de l'entropie à travers une boucle for.

L'entropie de la source x est égale à la longueur de son code car les symboles sont équiprobables. L'entropie de la source z est inférieure à la longueur de son code car les symboles ne sont pas équiprobables (entrés manuellement).

Question 4

Dicoz =

```

[0]    [1x7 double]
[1]    [1x4 double]
[2]    [1x2 double]
[3]    [          1]
[4]    [1x3 double]
[5]    [1x5 double]
[6]    [1x6 double]
[7]    [1x7 double]

```

LMoyz =

```

2.0796

```

Dicoz contient le symbole associé à son codage et sa longueur de code.

LMoyz contient la moyenne des longueurs de code par rapport à la probabilité d'apparitions des symboles.

Degré de compression :

Le degré de compression est le rapport entre la longueur du code et la longueur moyenne :

$$Compression = \frac{L}{L_{moyz}}$$

Pour notre codage de Huffman, on obtient un taux de compression de 69%.

Question 5

Lorsque l'on recommence l'expérience pour une source équiprobable.

La longueur moyenne est égale à LMoyz=3.

Pour une source équiprobable il n'y a aucun intérêt à utiliser un tel codage car ce code est absolument optimal. C'est-à-dire que La longueur du code est égale à l'entropie et que la compression est inutile.