

Théorie de l'information

TP N°1: Compression de l'information sans pertes

But du TP : Les étudiants sont conduits à manipuler les notions relatives à la mesure de l'information et de l'entropie.

Introduction : La quantité d'information véhiculée lors d'une communication numérique est étroitement liée à la nature statistique de la source. La mesure mathématique de l'information doit prendre en considération la fréquence d'apparition (probabilité d'occurrence) de chaque symbole en respectant les conditions suivantes :

1. La quantité d'information contenue dans chaque symbole est inversement proportionnelle à sa certitude.
2. La quantité d'information associée à des données statistiquement indépendantes est additive.

Rappels :

Q0 : On considère un alphabet de M symboles que nous notons x_i , avec $1 \leq i \leq M$.

Chaque x_i symbole apparaît avec une probabilité $p = P(x_i)$.

- Vérifier qualitativement que le contenu informatif de chaque symbole x_i peut alors être mesurée par une grandeur de la forme de $I(x_i)$, définie par :

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

- Nous définissons l'entropie associée à cet alphabet de M symboles par $H(X)$. Comparer cette grandeur au contenu informatif moyen (espérance mathématique des I_i) associé à cet alphabet.

$$H(X) = - \sum_1^M p_i \log_m(p_i)$$

Observation d'une densité de probabilité

Q1 : Saisir le code ci-dessous dans l'éditeur de MATLAB et procéder à son exécution.

- Quelle est la fonction réalisée par **hist(x,h)** à la ligne 10 ?
- Relever la sortie de ce programme pour $L=1, 2, 3$ et 4 en définissant les axes.
- Quelle est la nature de la séquence de symboles générées ?

```

1 -  clc; clf; clear;
2
3 -  L=3; m=2; M=m^L;    % M= nombre de symb. de la source
4 -  N=100000;           % m= nombre de symboles du code
5                        % L= longueur du code
6
7 -  x=randint(N,1,M); h=0:0.01:M-1;
8
9 -  figure(1);
10 - subplot(221); hist(x,h);    % Remplir les vides "... "
11 - xlabel('...'); ylabel('...'); % avec la définition des axes
12 - title('...'); grid;        % et donner un titre à la fig.

```

Frag1 : Génération et observation

Mesure d'une entropie

Q2 : Compléter le premier programme avec celui qui est indiqué ci-dessous.

- Quelle est le contenu de P à l'issue de l'exécution de frag2 ?
- Compléter la ligne 21 pour calculer la quantité d'information associée à chaque symbole ainsi que l'entropie de l'alphabet généré.
- Justifier le choix de la base 2 pour le calcul du logarithme.
- Relever les valeurs de H pour L=1, 2, 3 et 4. Commentaires. Validation Q0 et Q2.

```

14 - frequence=hist(x,h); Lf=length(frequence);
15 - P=zeros(1,M); I=zeros(1,M); j=1;
16 - for i=1:Lf
17 -     if frequence(i)>0
18 -         P(j)=frequence(i); j=j+1;
19 -     end;
20 - end;
21 - P=P/N; % I=-log2(...); H=....

```

Frag2 : Génération et observation

Génération d'une source

Q3 : Nous nous intéressons à la génération de la source z comme celle programmée par la ligne 29 de frag3.

- Compléter et exécuter ce programme. Comparer P et Pz.
- Quelle est la nature statistique de la source générée ?
- Compléter la ligne 43 afin de calculer la quantité d'information associée à chaque symbole ainsi que l'entropie de la source.
- Comparer l'entropie des deux sources x et z. Commentaires.

```

27 - M=8; symbols=[0:M-1];
28 - Pz=[0.0078 0.0625 0.2539 0.4713 0.1250 0.060 0.0156 0.0039];
29 - z=randsrc(N,1,[symbols;Pz]);
30
31 - subplot(222); hist(z,h);
32 - xlabel('...'); ylabel('...');
33 - title('...'); grid;
34
35 - frequence=hist(z,h); Lf=length(frequence);
36 - P=zeros(1,M); Iz=zeros(1,M);
37 - j=1;
38 - for i=1:Lf
39 -     if frequence(i)>0
40 -         P(j)=frequence(i)/N; j=j+1;
41 -     end;
42 - end;
43 - P; %Iz=-log2(...); Hz=...;

```

Frag3 : Programmation d'une source.

Partie 2 :

Considérant la sortie d'une source **texte = On peut tout te prendre; tes biens, tes plus belles années, l'ensemble de tes joies, et l'ensemble de tes mérites, jusqu'à ta dernière chemise. Il te restera toujours tes rêves pour réinventer le monde que l'on t'a confisqué.**

Ecrire un code Matlab permettant de faire les tâches suivantes :

1. Créer un vecteur **symboles** contenant les caractères dans le vecteur **texte**.
2. Calcul du vecteur correspondant aux probabilités d'apparition des caractères de l'alphabet dans le texte.
3. Affichage d'un message d'erreur au cas où la probabilité totale est différente de 1.
4. Calcul de l'information propre (quantité d'information).
5. Calcul de l'entropie de la source.
6. Combien de bits sont nécessaires pour coder les symboles pour le cas où on considère les symboles équiprobables.
7. Est-il possible d'avoir des codes inférieurs au nombre calculé en question 6, justifier votre réponse.
8. Calculer de l'efficacité et la redondance de la source.