**EXERCICES SUR LA THEORIE DE L’INFORMATION & COMPRESSION ENTROPIQUE**

**Exercice 1**

Calculer l’Entropie HX d’une source binaire X = {0,1} telle que

prob (x=0) = p prob (x=1) = 1-p

Etudier et interpréter HX(p)

# Exercice 2

On considère la distribution **X** sur {a, b, c, d, e} définie par

 p (a ) = 0,32

 p (b ) = 0,23

 p (c ) = 0,20

 p (d ) = 0,15

Utiliser l’algorithme de Huffman pour trouver un codage préfixe optimal

Comparer la longueur moyenne de ce codage avec H(X)

Comparer avec un codage binaire naïf

# Exercice 3

On s'intéresse à un nouveau standard de Télévision TBD : Très Basse Définition pour des applications militaires. Une image est composée de Pixels. Chaque pixel est défini par de l’une des 3 couleurs : R, V, B. On a mesuré sur une grande série d'images la répartition en probabilités des 3 couleurs :

 R : 1/4 , V :1/2, B : 1/4

***Question 1***

Quelle est la quantité d'information associée à chacune des 3 couleurs ?

Quelle est l'entropie de la source ?

Donner le codage optimal d’un pixel

La transmission a lieu dans un environnement très bruité.

La matrice de transmission est ainsi définie :

 soit X le symbole émis et Y le symbole reçu, appartenant chacun à l'alphabet {R,V,B}

 Pr (Y/X) = 1/2 si Y = X , 1/4 sinon

***Question 2***

Quelle est la matrice de réception Pr (X/Y) donnant pour chaque symbole reçu la probabilité du symbole émis ?

Calculer l'entropie de la source de chrominance après réception : H(X/Y)

Calculer la quantité d’information transmise I (X; Y)

On donne : log23 = 1,58 log25 = 2,32 log27 = 2,81

**Exercice 4:**

Une source émet des symboles de l'alphabet A = {a,b,c,d,e} avec les probabilités :

P(a) = 0.15, P(b) = 0.04, P(c) = 0.26, P(d) = 0.05 et P(e) = 0.50

1. Calculer l'entropie de cette source

2. Trouver un code de Huffman pour cette source

3. Calculer la longueur moyenne du code de Huffman

# Exercice 5

On considère un ensemble de n éléments : E = {e1, e2, …, en} et une source émettant à intervalles réguliers des sous-ensembles S de E.

***Question 1***

Sachant que la source tire les éléments aléatoirement et indépendamment les uns des autres selon une distribution uniforme

***Quelle est la probabilité d'envoyer un ensemble S de taille quelconque ?***

***Quelle est l'entropie de la source si elle envoie des ensembles S quelconques ?***

***Quelle est la probabilité d'envoyer un ensemble S de taille au plus k ?***

***Si nous savons que la source émet seulement des ensembles de taille au plus k, quelle est son entropie ?***

***Question 2***

Parmi les deux codes possibles :

A. Je code les éléments un à un par la méthode de Huffman, et donc j'envoie pour chaque ensemble la liste des codes de ses éléments.

B. Je code directement un ensemble d'éléments par la méthode de Huffman et donc je n'envoie que les codes des ensembles.

***Laquelle est la plus efficace :***

***- si la source envoie des ensembles de taille quelconque***

***- si la source envoie des ensembles de taille au plus k***

# Exercice 6

Soit une image composée de 10 couleurs. Les pixels de cette image prennent ces couleurs avec les probabilités suivantes :

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Couleur | Noir | Blanc | Bleu | Jaune | Rouge | Vert  | Violet  | Orange | Marron | gris |
| Code | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Probabilité | 0,04 | 0,02 | 0,03 | 0,15 | 0,20 | 0,08 | 0,12 | 0,10 | 0,20 | 0,06 |

On demande de calculer :

1. L’entropie de cette image,
2. le taux de compression en utilisant le codage de Huffman.
3. Trouvez la longueur moyenne du code de Huffman ci-dessus et de la comparer avec l'entropie.
4. mêmes question avec le codage arithmétique

# Exercice 7

On veut optimiser le codage d'images transmises par un système numérique. Les images sont numérisées et codées selon des niveaux de gris.

Une image est constituée de 128x128 pixels. Chaque pixel est ensuite codé par un niveau de gris parmi dix niveaux possibles. Les niveaux de gris sont désignés par la suite par n0, n1, n2, …., n8, n9.

Une étude statistique des images montre que l'on dispose en moyenne de :

* 7410 pixels uniformément répartis sur les niveaux n4 et n5.
* 6900 pixels uniformément répartis sur les niveaux n2, n3, n6 et n7.
* 1620 pixels uniformément répartis sur les niveaux n1 et n8 .
* 454 pixels uniformément répartis sur les niveaux n0 et n9 .

1) Calculer les probabilités pi correspondant aux niveaux ni de la source d'images. En déduire l'entropie de cette image ?

2) On code les pixels à l'aide de bits. On choisit de coder chaque niveau avec un nombre identique de bits. Quel est le nombre minimum de bit par pixel ?

3) Le débit binaire de transmission est de 107 bits/s. Calculer la durée de transmission d'une image.

**Exercice 8**

Considérons une variable aléatoire donnée par :



1. Calculez l’entropie de X
2. Effectuez le codage de Huffman
3. Calculer la longueur moyenne des codes obtenus par Huffman
4. Effectuez le codage arithmétique
5. Calculez la longueur moyenne des codes ainsi obtenus

**Exercice 9**

Soit (X, Y) la distribution conjointe suivante:

****

Calculez H(X), H(Y) et H(X|Y)

**Exercice 10**

Soit (X, Y) la distribution conjointe suivante:

****

Calculez H(X), H(X|Y= 1), H(X|Y= 2) et H(X|Y)

**Exercice 11**

En partant de la définition de l’information mutuelle, démontrer que pour X et Y deux variables aléatoires : I(X;Y)= I(Y;X).

**Exercice 12**

Soit deux sources binaires X et Y tel que

* P(X=0)=p,
* P(X=0|Y=1)= P(X=1|Y=0)=D

Soit aussi une mesure de distorsion d(x,y)=(x+y) mod 2 [sommation modulo 2]

Démontrer que I(X,Y)= Hb(p)- Hb(D)

**Exercice 13**

Considérer le schéma de compression d’une séquence binaire qui comporte les étapes suivantes : On subdivise la séquence binaire en block de taille M.

• Pour chaque bloc on calcule le nombre de 0 : n0.

• Si n0 ≥ M/2, on envoie 0.

• Sinon, on envoie 1.

Considérer maintenant une séquence binaire avec P(0)=0.8.

1. Calculer les débits R1 et R2 et les distorsions D1 et D2 pour M=1 et M=2
2. En utilisant :



 calculer les débits optimaux Ropt1 et Ropt2 correspondant à D1 et D2.

Définition :



1. Un encodeur entropique est utilisé maintenant pour coder la séquence binaire après compression. Déterminer Rentropie1 et R entropie2 correspondant à D1 et D2

**Exercice 14**

Soit le fichier $F= \left\{BCADADBCBCADBCAD\right\}. $Calculez l’entropie d’ordre 1 de ces données. A cet ordre le fichier est-il compressible ?

**Exercice 15 :**

Soit un bloc d’images de taille 4 × 4, représenté ci-dessous. Les niveaux de gris de chacun des pixels de ce bloc sont représentés dans le tableau à côté.



Bloc de taille 4 × 4

1. Calculez la taille en bits de ce bloc en supposant que chaque pixel est codé sur 8 bits.
2. Calculez les probabilités d’apparition de chaque niveau de gris
3. Appliquer la codage de Huffman pour compresser ce bloc
4. Quel est alors le taux de compression obtenu
5. Déterminer l’entropie du bloc
6. En déduire l’efficacité du codage obtenu

**Exercice 16 :**

Soit X une variable aléatoire répartie uniformément sur {1,2,3,4,5,6}.

1. Construisez un code Huffman pour la variable.
2. Quelle est la longueur moyenne des mots de code pour votre code? Comment cela se compare-t-il à l'entropie?
3. Si vous interprétez un mot de code de longueur k comme une probabilité de 2-k, quelle est alors la distribution implicite exprimée par votre code?

**Exercice 17 :**

Quel est le taux d'entropie d'un chevalier marchant sur un échiquier 3×3?

Et un fou?

*Prenez en considération les mouvements possibles des pièces.*



**Exercice 18 :** code Morse

Un alphabet contient un point qui prend une unité de temps pour transmettre et un tiret qui en prend deux.

1. Lorsque les deux symboles ont une probabilité p et 1-p, quel est le taux d’entropie de ce processus?
2. Pour quel choix de p ce taux d'entropie est-il le plus grand?

**Exercice 19 :**

Une source d'informations produit des pixels X1, X2, X3,. . . avec :

**Pr {Xi = BLANC} = 0,995**

**Pr {Xi = NOIR} = 0,005**

On décide d’effectuer un codage par la force brute (Brut-force encoding : toutes les possibilités) des sorties de cette source, 100 pixels à la fois, au moyen d'un tableau de mots de code de longueurs égales. On inclue alors toutes les séquences avec trois pixels noirs ou moins dans le tableau et en acceptant qu'il y aura une erreur dans les autres cas.
1. Calculez le nombre de mots de code dont on aura besoin.
2. De combien de bits avons-nous besoin pour encoder autant de séquences? Comment cela se compare-t-il au minimum théorique?
3. Quelles sont les options pour améliorer ces performances, théoriquement et pratiquement?
4. Découvrez la probabilité que ce schéma de codage rencontre une séquence non tabulée.