**TD1 : CODAGES ENTROPIQUE**

**CODAGE ET COMPRESSION MASTER 1 RT ET ST**

**Exercice 1**

Soit la variable aléatoire discrète (va) X dont les valeurs x sont respectivement les codes ASCII (American Standard Code for Information Interchange) écrits en décimal des lettres majuscules {M, N, O, P, Q}. Nous supposons que ces caractères obéissent à deux lois de probabilités pour deux cas de figures différents. Nous proposons aussi de les coder différemment d’abord avec le code ASCII de base et ensuite avec un code C donné

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 81 | 80 | 79 | 78 | 77 | x |
| 0.10 | 0.40 | 0.20 | 0.15 | 0.15 | P1(x) |
| 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | P2(x) |

1. Tracez la fonction de densité de probabilité de X pour les deux lois de probabilités,

2. Calculez l’espérance mathématique E(X), la variance VAR(X) et l’écart type σ, pour les deux lois de probabilités

3. Calculez l’entropie de la source (de la variable aléatoire X) pour les deux lois de probabilités

4. Calculez la redondance de la source (ou variable aléatoire X) par rapport à la première loi de probabilité. On rappelle que la redondance d’une source c’est la différence entre l’entropie maximale obtenue pour une distribution uniforme et l’entropie réelle. Que peut-on conclure à partir de cette redondance calculée?

5. Comme nous prenons en considération dans cet exercice le code ASCII de base, où chaque caractère est sur 7 bits, comparez l’entropie calculée en (3) avec la longueur moyenne (lmoy) d’un caractère (comme il s’agit d’un codage ASCII à taille fixe, donc lmoy=7 bits). Que pouvez-vous conclure?

6. Calculez l’efficacité de codage (ASCII) et la redondance du code (différente de la redondance source) pour chacune des lois de probabilité. Que pouvez-vous conclure?

On rappelle que :

- l’efficacité de codage **η = H(X)/lmoy**

- le **rendement du code= lmoy - H(X)** est donc : **rendement/H(X)** représente en pourcentage de bits supplémentaires par rapport à un code optimal

7. Si nous utilisons maintenant le codage C formé des mots de code suivants : {0, 10, 110, 1110, 1111}.

- Que peut dire par rapport à ce codage, est il déchiffrable ? Expliquez

- A votre avis quel mot de code doit on affecter aux différentes valeurs de x, en se limitant à la première loi de probabilité ? Justifiez

- Qu’en est il de l’efficacité de codage et du rendement de ce nouveau code et pour la première loi probabilité ? Que peut-on conclure ?

**Exercice 2.**

Un codage Huffman est utilisé pour encoder de manière compacte les espèces d’oiseaux menacés dans la région humide d’Annaba/El Tarf (les lacs). Si 50% de ces oiseaux sont des chardonnerets et que le reste est réparti uniformément (même probabilité) entre 5 autres espèces,

1. Représentez Graphiquement la fonction de densité de probabilité de ces espèces d’oiseaux

2. Calculez l’entropie de cette variable aléatoire. En déduire la redondance de cette source

3. Combien de bits seraient utilisés pour coder selon l’algorithme de Huffman l'espèce chardonneret par rapport aux autres oiseaux ?

**Exercice3**

Une variable aléatoire X est distribuée selon les probabilités ponctuelles du tableau suivant :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| e | d | c | b | a | x |
| 1/12 | 1/6 | 1/6 | 1/4 | 1/3 | P(X) |

(a) Calculez H(X). Que pouvez-vous dire sur la redondance de cette source ?

(b) Construisez un code Huffman pour la variable et calculez sa longueur moyenne lmoy . En déduire l’efficacité du codage

(c) Décodez le message 00101100001 selon votre code

**Exercice4**

Soit une variable aléatoire distribuée uniformément sur l'ensemble {1, 2, 3, 4, 5, 6}.

(a) Construisez un code Huffman pour la variable.

(b) Quelle est la longueur moyenne du mot de code pour votre code ? Comment cela se compare-t-il à l’entropie ?

(c) 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 5**

1. Une source émet des symboles Xi, 1≤i≤6, au format DCB (Décimal Codé Binaire ou BCD : Binary Coded Decimal) avec des probabilités P(Xi) selon le tableau suivant, à un débit Rs = 9,6 kbaud (baud = symbole / seconde).

- Trouvez le débit d’information (ou taux d'information) R = H×Rs

- Trouvez le débit de données de la source, c'est-à-dire le nombre de bits/symboles × Rs

|  |  |  |
| --- | --- | --- |
| Mot DCB | P(X) | X |
| 000 | 0.30 | A |
| 001 | 0.10 | B |
| 010 | 0.02 | C |
| 011 | 0.15 | D |
| 100 | 0.40 | E |
| 101 | 0.03 | F |

2. Appliquer le codage Shannon-Fano au signal source décrit dans le tableau. Y a-t-il des inconvénients dans les mots de code résultants?

3. Quelle est la séquence de symboles originale du signal codé Shannon-Fano 110011110000110101100?

4. Quel est le débit de données du signal après le codage Shannon-Fano? Quel facteur de compression a été atteint?

5. Calculez l'efficacité de codage du signal DCB non codé ainsi que du signal codé Shannon-Fano.

6. Répétez les parties 2 à 5 mais cette fois avec le codage Huffman

**Exercice 6 :**

Considérez les deux arbres de décodage de Huffman suivants pour un code de longueur variable impliquant 5 symboles : A, B, C, D et E.

**0**

**0**

**0**

**A**

**1**

**B**

**0**

**1**

**1**

**C**

**D**

**1**

**E**

**Arbre 1**

**0**

**0**

**1**

**A**

**0**

**1**

**B**

**0**

**C**

**1**

**D**

**1**

**E**

**Arbre 2**

1. En déduire le code de chaque symbole pour les deux arbres

2. À l'aide de l'arbre n ° 1, décodez le message codé suivant : "01000111101".

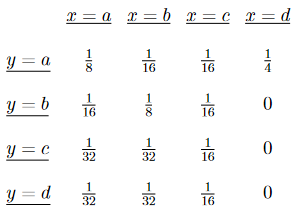
3. Supposons que nous encodions des messages avec les probabilités suivantes pour chacun des 5 symboles: P(A) = 0,5 et P(B) = P(C) = P(D) = P(E) = 0,125. Lequel des deux encodages ci-dessus (arbre n ° 1 ou arbre n ° 2) produirait les messages encodés les plus courts en moyenne sur de nombreux messages?

4. En utilisant les probabilités de la partie (3), si vous apprenez que le premier symbole d'un message est "B", combien de bits d'information avez-vous reçu?

5. En utilisant les probabilités de la partie (B), si l'arbre n ° 2 est utilisé pour coder des messages, quelle est la longueur moyenne des messages de 100 symboles, en moyenne sur de nombreux messages?

**Exercice 7**

La source d'entrée d'un canal de communication bruyant est une variable aléatoire X sur les quatre symboles a, b, c, d. La sortie de ce canal est une variable aléatoire sur ces mêmes quatre symboles. La densité de probabilité conjointe f(x,y) de ces deux variables aléatoires est la suivante:



(a) Calculez la distribution marginale pour X et calculez l'entropie marginale H(X) en bits.

***Rappel : La densité de probabilité ou distribution marginale de X notée fX(x) (respectivement de Y) est la distribution de X (respectivement Y) sur l’échantillon, calculée à partir de la distribution conjointe f(x,y) :***

******

(b) Calculez la distribution marginale pour Y et calculez l'entropie marginale H(Y) en bits.

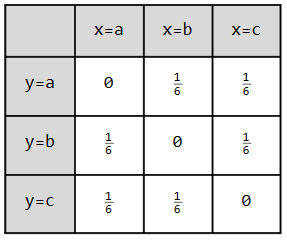
(c) Quelle est l'entropie conjointe H(X, Y) des deux variables aléatoires en bits?

(d) Quelle est l'entropie conditionnelle H(Y | X) en bits?

(e) Quelle est l'information mutuelle I(X;Y) entre les deux variables aléatoires en bits?

**Exercice 8**

La source d'entrée d'un canal de communication bruyant (avec bruit) est une variable aléatoire X sur les trois symboles a, b et c. La sortie de ce canal est une variable aléatoire Y sur ces trois mêmes symboles. La distribution conjointe de ces deux variables est la suivante.



**Mêmes questions que pour l’exercice 7**

**Exercice 9**

1. Que peut-on dire des codes suivants?
   1. C1={0; 11; 101}
   2. C2={00; 01; 001}
   3. C3={0; 01; 10}
   4. C4={000; 001; 01; 1}
   5. C5={000100; 100101; 010101; 111000}
   6. C6={0; 01; 11}.
   7. C7={0; 10; 110; 1110}.
   8. C8={0; 10; 110; 1111}.
   9. C9={0; 01; 011; 0111}.
   10. C10={0,10,101,0101}.
   11. C11={0,10,110,111}.
2. A partir de l’inégalité de Kraft, que pouvez-vous dire à propos des codes suivants:

* C1 = {00, 01, 10, 11}.
* C2 = {0, 100, 110, 111}.
* C3 = {0, 10, 110, 11}.