

TD1 CODAGE ET COMPRESSION
S2 MASTER TELECOM

EXERCICE 1

On considère un dé cubique truqué dont les faces sont numérotées de 1 à 6 et on note X la variable aléatoire donnée par le numéro de la face du dessus. On suppose que le dé est truqué de sorte que la probabilité d'obtenir une face est proportionnelle au numéro inscrit sur cette face.

1. Déterminer la loi de X , , calculer son espérance.
2. On pose $Y=1/X$. Déterminer la loi de Y , et son espérance

EXERCICE 2

Soit un alphabet $A = \{a_1, a_2, a_3, a_4\}$. Calculer l'entropie en bits pour les situations suivantes :

(a) $P(a_1)=0, P(a_2)=0, P(a_3)=0, P(a_4)=1$

(b) $P(a_1)=0.25, P(a_2)=0.25, P(a_3)=0.25, P(a_4)=0.25$

(c) $P(a_1)=0.505, P(a_2)=0.25, P(a_3)=0.125, P(a_4)=0.12$

EXERCICE 3

Un joueur tire sur une cible de 10cm de rayon, constituée de couronnes concentriques, délimitées par des cercles de rayons 1,2, ..., 10 cm, et numérotées respectivement de 10 à 1. La probabilité d'atteindre la couronne k est proportionnelle à l'aire de cette couronne, et on suppose que le joueur atteint sa cible à chaque lancer. Soit X la variable aléatoire qui à chaque lancer associe le numéro de la cible.

1. Quelle est la loi de probabilité de X ?
2. Le joueur gagne k DA (*Dinars Algériens*) s'il atteint la couronne numérotée k pour k compris entre 6 et 10, tandis qu'il perd 2 DA s'il atteint l'une des couronnes périphériques numérotées de 1 à 5. Le jeu est-il favorable au joueur ?

Indication ▼

1. Calculer l'aire de chaque couronne.
2. Introduire la variable aléatoire égale au gain algébrique du joueur, puis calculer l'espérance de cette variable aléatoire

EXERCICE 4

Une source sans mémoire S génère des symboles A, B, C, D, E, F et G avec les probabilités respectives $1/16, 1/16, 1/16, 1/16, 1/4, 1/4, 1/4$.

1. Réaliser un codage de Huffman des 7 valeurs de S . Comparer le nombre moyen d'éléments binaires utilisés par ce code pour représenter une valeur de S avec l'entropie de S .

On désigne par U la source binaire obtenue à partir de S par le codage de Huffman précédent. (prière de répondre aux questions dans l'ordre dans lequel elles sont posées)

2. Calculer l'entropie de U .
3. En utilisant la loi des grands nombres, déterminer la loi de U .
4. Que conclure des résultats précédents quant à la mémoire de U ?
5. Généraliser les résultats précédents pour donner une signification à la notion de codage de source optimum.

EXERCICE 5

On cherche à dépister une maladie détectable à l'aide d'un examen sanguin. On suppose que dans notre population, il y a une proportion p de personnes qui n'ont pas cette maladie.

1. On analyse le sang de r personnes de la population, avec r entier au moins égal à 2. On suppose que l'effectif de la population est suffisamment grand pour que le choix de ces r personnes s'apparente à un tirage avec remise. Quelle est la probabilité qu'aucune de ces personnes ne soit atteinte de la maladie?
2. On regroupe le sang de ces r personnes, puis on procède à l'analyse de sang. Si l'analyse est négative, aucune de ces personnes n'est malade et on arrête. Si l'analyse est positive, on fait toutes les analyses individuelles (on avait pris soin de conserver une partie du sang recueilli avant l'analyse groupée). On note Y la variable aléatoire qui donne le nombre d'analyses de sang effectuées. Donner la loi de probabilité de Y et calculer son espérance en fonction de r et de p .
3. On s'intéresse à une population de n personnes, et on effectue des analyses collectives après avoir mélangé les prélèvements par groupe de r personnes, où r est un diviseur de n . Montrer que le nombre d'analyses que l'on peut espérer économiser, par rapport à la démarche consistant à tester immédiatement toutes les personnes, est égal à $np^r - n/r$.
4. Dans cette question, on suppose que $p=0,9$ et on admet qu'il existe un réel $a>1$ de sorte que la fonction $x \mapsto p^x - 1/x$ est croissante sur $[1, a]$ et décroissante sur $[a, +\infty[$. Écrire un algorithme permettant de déterminer pour quelle valeur de l'entier r le nombre $p^r - 1/r$ est maximal.

Indication ▼

- 1.
2. Quelles sont les valeurs prises par Y ?
3. Utiliser la valeur de l'espérance calculée à la question précédente.
- 4.

EXERCICE 6

Les vaches laitières sont atteintes par une maladie M avec la probabilité $p=0,15$. Pour dépister la maladie M dans une étable de n vaches, on fait procéder à une analyse de lait. Deux méthodes sont possibles :

- Première méthode : On fait une analyse sur un échantillon de lait de chaque vache.
- Deuxième méthode : On effectue d'abord une analyse sur un échantillon de lait provenant du mélange des n vaches. Si le résultat est positif, on effectue une nouvelle analyse, cette fois pour chaque vache.

On voudrait connaître la méthode la plus économique (=celle qui nécessite en moyenne le moins d'analyse). Pour cela, on note X_n la variable aléatoire du nombre d'analyses réalisées dans la deuxième étape. On pose $Y_n = X_n/n$.

1. Déterminer la loi de Y_n , et montrer que son espérance vaut : $1 + 1/n - (0.85)^n$
2. Etudier la fonction $f(x) = ax + \ln x$, pour $a = \ln(0.85)$. Donner la liste des entiers n tels que $f(n) > 0$
3. Montrer que $f(n) > 0$ équivaut à $E(Y_n) < 1$. En déduire la réponse (en fonction de n) à la question posée).

Indication ▼

1. Y_n ne prend que deux valeurs. Calculer la probabilité de chacune.
2. Montrer que f admet un maximum en $-1/a$
3. Pour la première méthode, on fait constamment n prélèvements. Donc, il faut choisir la deuxième méthode si, et seulement si, $E(Y_n) < 1$.

EXERCICE 7

Soit X, Y deux variables aléatoires indépendantes suivant la loi uniforme sur $\{1, \dots, n\}$.

1. Déterminer $P(X=Y)$.
2. Déterminer $P(X \geq Y)$.
3. Déterminer la loi de $X+Y$.

Indication ▼

1. $X=Y \Leftrightarrow \exists k \in \{1, \dots, n\}, X=k \text{ et } Y=k$.
2. Utiliser $P(X > Y)$.
3. $X+Y=k \Leftrightarrow \exists i \in \{1, \dots, k-1\}, X=i \text{ et } Y=k-i$.

EXERCICE 8

Soit X une variable aléatoire discrète finie prenant la valeur x_i avec probabilité p_i , pour $i=1,\dots,n$. On définit l'entropie de X par :

$$H(X) = -\sum_{i=1}^n p_i \ln(p_i)$$

avec la convention $x \ln x = 0$ si $x=0$ (ce qui correspond au prolongement par continuité en 0 de la fonction $x \mapsto x \ln x$).

1. Démontrer que $H(X) \geq 0$.
2. Démontrer que $H(X)=0$ si et seulement si X est presque sûrement constante, c'est-à-dire s'il existe $i \in \{1,\dots,n\}$ tel que $p_i=1$.
3. Vérifier que, pour tout $k=1,\dots,n$, on a

$$(-np_k) \ln(np_k) \leq 1 - np_k$$

avec égalité si et seulement si $np_k=1$.

4. En déduire que $H(X) \leq \ln n$.
5. Démontrer que $H(X)=\ln n$ si et seulement si X est équirépartie, ie si $p_i=1/n$ pour tout $i=1,\dots,n$.

EXERCICE 9

On considère un couple de variables aléatoires I et J dépendantes l'une de l'autre. Ces variables peuvent respectivement prendre les valeurs $i \in \{A,B,C\}$ et $j \in \{D,E,F,G\}$.

1. Démontrer la formule suivante :

$$H(i,j) \leq H(i) + H(j)$$

2. Dans quel cas a-t-on égalité ?

NB: on peut utiliser la relation $\log_e(x) = \ln(x) \leq x-1$.

EXERCICE 10

On considère 8 symboles équiprobables codés en huit mots binaires:

$$x_1 = 0000 \quad x_5 = 1001$$

$$x_2 = 0011 \quad x_6 = 1010$$

$$x_3 = 0101 \quad x_7 = 1100$$

$$x_4 = 0110 \quad x_8 = 1111$$

Ces mots code transitent sur un canal binaire symétrique de probabilité d'erreur p . On suppose que le mot reçu est $y = 0000$.

1. Calculer l'information apportée par la connaissance du premier bit reçu ("0") sur l'événement $\{x_1 \text{ a été émis}\}$.
2. Calculer l'information apportée par la connaissance du second bit reçu ("0") sur l'événement $\{x_1 \text{ a été émis}\}$ conditionnellement à la connaissance du premier bit reçu ("0"). Que constate-t-on?
3. Que peut-on induire quant à l'information apportée par la connaissance du troisième (resp. quatrième) bit reçu ("0") sur l'événement $\{x_1 \text{ a été émis}\}$ conditionnellement à la connaissance des deux premiers (resp. des trois premiers) bits reçus ("0")?

EXERCICE 11

Soit un ensemble de n pièces d'or dont une seule a un poids inférieur au poids nominal. Pour identifier cette pièce, on dispose d'une balance à deux plateaux qui permet de comparer les poids a et b de deux ensembles A et B de pièces. Le résultat d'une pesée est donc l'un des trois événements : $a < b$, $a = b$, $a > b$.

1. Calculer l'incertitude moyenne liée à la détermination de la pièce fausse.
2. Exprimer l'information moyenne apportée par une pesée sur la détermination de la pièce fausse en fonction de l'incertitude moyenne liée au résultat de la pesée. Généraliser ce résultat au cas de m pesées.
3. Quelle est l'information moyenne maximum apportée par une pesée sur l'identification de la pièce fausse? Quand rencontre-t-on cette situation?
4. On suppose que l'on a mis au point une stratégie permettant, assurément, de découvrir la pièce fausse en m pesées au plus.
 - Donner, en fonction de n , une borne inférieure pour m .
 - Dans le cas particulier où n est une puissance de 3, élaborer la stratégie permettant, à chaque pesée, d'obtenir une quantité d'information moyenne maximale sur la détermination de la pièce fausse. En quoi cette stratégie est-elle optimale?
5. On considère le problème précédent avec des hypothèses différentes : On dispose d'une balance à un seul plateau permettant de déterminer le poids d'un ensemble de pièces et le nombre de pièces fausses est inconnu mais chacune pèse un poids f inférieur au poids nominal v .
 - Montrer qu'une pesée permet de déterminer le nombre de pièces fausses parmi celles pesées.
 - Soit une stratégie permettant assurément de connaître le nombre de pièces fausses et de les identifier en m pesées au plus. Donner, en fonction de n , une borne inférieure pour m .

EXERCICE 12

Pour une certaine région, les prévisions d'un météorologiste M se distribuent suivant les fréquences relatives données par le tableau :

Les colonnes correspondent au temps effectif et les lignes aux prévisions.

	pas de pluie	pluie
pas de pluie	$5/8$	$1/16$
pluie	$3/16$	$1/8$

1. Un étudiant astucieux montre que M se trompe une fois sur quatre et qu'il peut faire mieux en ne prévoyant que des jours sans pluie. Vérifier cette assertion. Il postule donc pour le poste de M. Le directeur de M refuse l'offre de l'étudiant. Commenter cette décision à l'aide d'un raisonnement faisant intervenir des notions de théorie de l'information.

2. L'étudiant revient quelque temps plus tard en garantissant les résultats suivants :

Les colonnes correspondent au temps effectif et les lignes aux prévisions de l'étudiant.

	pas de pluie	pluie
pas de pluie	$403/512$	$93/512$
pluie	$13/512$	$3/512$

Comment le directeur doit-il réagir à cette nouvelle offre?

3. Le directeur souhaite stocker dans ses archives le temps T (pluie ou absence de pluie) et les prévisions de M. Quelle est la taille minimale du fichier qu'il doit prévoir (en nombre de bits par réalisation de T ou de M)?

Si on suppose que prévision et temps effectif pour un jour donné sont indépendants du passé, quel résultat obtient-on par codage de Huffman des couples (T, M) ?

EXERCICE 13

Soit X une variable aléatoire à n valeurs possibles $x_1, x_2, \dots, x_{n-1}, x_n$ de probabilités

$$\frac{1}{2}, \frac{1}{2^2}, \dots, \frac{1}{2^{n-1}}, \frac{1}{2^{n-1}}.$$

1. Effectuer un codage de Huffman des n valeurs possibles de X .
2. Comparer la longueur moyenne des mots code $\bar{n}$ à l'entropie $H(X)$ de la source (on se contentera d'exprimer $\bar{n}$ et $H(X)$ sous la forme d'une somme finie de termes). Que constate-t-on? En quoi ce résultat est-il remarquable? Comment expliquer ce résultat remarquable?

EXERCICE 14

On considère une source binaire sans mémoire U de loi de probabilité: $P\{U=0\}=0,9$ et $P\{U=1\}=0,1$. Les zéros étant beaucoup plus fréquents que les uns, on se propose de coder les séquences issues de la source en tenant compte du nombre de zéros entre deux uns consécutifs. L'opération consiste en deux étapes :

- Première étape

On compte le nombre de zéros entre deux uns consécutifs. On obtient ainsi un entier que l'on appelle entier intermédiaire.

- Deuxième étape

On code l'entier intermédiaire en un mot binaire constitué de quatre éléments binaires si l'entier est inférieur ou égal à 7 et on choisit un mot de un élément binaire si l'entier est 8. Si l'entier intermédiaire dépasse 8, on codera les suites de 8 zéros consécutifs par un bit correspondant à l'entier intermédiaire autant de fois que nécessaire. On obtient ainsi la table de correspondance entre les séquences source et les entiers intermédiaires :

séquences source	entiers intermédiaires
1	0
01	1
001	2
0001	3
...	...
...	...
00000001	7
00000000	8

Exemple concernant la première étape d'attribution des entiers intermédiaires:

1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 1
0 2 8 2 0 4

1. Les contraintes imposées permettent-elles de choisir un code uniquement déchiffrable (on ne demande pas d'explicitier un code particulier)?
2. Calculer le nombre moyen $\bar{n}_1$ de bits source par entier intermédiaire.
3. Calculer le nombre moyen $\bar{n}_2$ de bits encodés par entier intermédiaire.
4. On considère une séquence de bits source de longueur n avec n très grand. En appliquant la loi faible des grands nombres, exprimer le rapport du nombre de bits utilisés pour coder cette séquence au nombre de bits source (n) en fonction de $\bar{n}_1$ et $\bar{n}_2$. Calculer numériquement la valeur de ce rapport. Comparer avec la limite en dessous de laquelle il n'est pas possible de descendre.
5. Effectuer un codage de Huffman de l'extension d'ordre 4 de la source U . Calculer (numériquement) le nombre moyen de bits utilisés pour coder un élément binaire issu de la source U . Comparer avec les résultats du 4.

EXERCICE 15

Soit une source qui génère des lettres de l'alphabet $A = \{a_1, a_2, a_3, a_4, a_5\}$ avec les probabilités suivantes : $P(a_1)=0.15$, $P(a_2)=0.04$, $P(a_3)=0.26$, $P(a_4)=0.05$, $P(a_5)=0.5$.

- (a) Calculer l'entropie de la source.
- (b) Trouver le code de Huffman de la source.
- (c) Calculer la longueur moyenne de ce code.
- (d) Calculer la redondance de ce code.
- (e) Calculer l'efficacité de ce code.

EXERCICE 16

Soit une source qui génère des lettres de l'alphabet $A = \{a_1, a_2, a_3, a_4\}$ avec les probabilités suivantes : $P(a_1)=0.1$, $P(a_2)=0.3$, $P(a_3)=0.25$, $P(a_4)=0.35$.

- (a) Trouver un code de Huffman selon la procédure habituelle.
- (b) Trouver un code de Huffman avec la procédure de variance minimale.
- (c) Calculer la variance de ces deux codes.
- (d) Discuter des performances de ces deux codes.