Exercices 6 à 22 (entropie, codage, décision)

Exercice 6 (TP: construction d'un arbre de décision binaire). Énoncé et données disponibles sur le site web du cours.

Exercice 7. Soit X et Y deux variables aléatoires à valeurs dans $\{1, 2, 3\}$ dont la distribution jointe $p(x, y)$ est définie par le tableau suivant

$p(x,y)$	$y = 1$	$y = 2$	$y = 3$
$x = 1$	0	1/4	0
$x = 2$	1/4	0	1/4
$x = 3$	1/8	1/8	0

a) Calculer $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, $I(X;Y)$.
b) Représenter ces quantités sur un diagramme de Venne.

Exercice 8 (Inégalité de Gibbs). En utilisant la convexité de la fonction $-\log$, montrer que pour toutes distributions de probabilités discrètes $(p_i)_{1\leqslant i\leqslant n}$ et $(q_i)_{1\leqslant i\leqslant n}$, on a

$$\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \geqslant 0.$$

Dans quels cas a-t-on égalité ?

Exercice 9. Soient X et Y deux variables aléatoires discrètes telles que $H(Y|X) = 0$. Montrer qu'il existe une fonction f telle que $Y = f(X)$ avec probabilité 1.

Exercice 10. Soient X et Y deux variables aléatoires discrètes identiquement distribuées mais non nécessairement indépendantes. On suppose $H(X) \neq 0$ et l'on pose

$$\rho = 1 - \frac{H(Y|X)}{H(X)}.$$

a) Montrer que $\rho = I(X;Y)/H(X)$.
b) Montrer que $0 \leqslant \rho \leqslant 1$ et décrire les cas d'égalité.

Exercice 11. a) Trouver un exemple de variables aléatoires X, Y, Z telles que

$$I(X;Y|Z) < I(X;Y).$$

b) Trouver un exemple de variables aléatoires X, Y, Z telles que

$$I(X;Y|Z) > I(X;Y).$$

c) Proposer, à l'aide d'un diagramme de Venne, une définition pour $I(X;Y;Z)$. A-t-on nécessairement $I(X;Y;Z) \geqslant 0$?

Exercice 12. Soient X, Y, Z trois variables aléatoires telles que

$$I(X;Y) = I(X;Z) = 0.$$

a) A-t-on $I(X;Y,Z) = 0$? (démontrer l'égalité ou trouver un contre-exemple)
b) A-t-on $I(Y;Z) = 0$? (démontrer l'égalité ou trouver un contre-exemple)

Exercice 13. Soit X et Y deux variables aléatoires prenant des valeurs respectivement dans $\mathcal{X}$ et $\mathcal{Y}$, deux sous-ensembles finis de $\mathbb{R}$. On pose $Z = X + Y$.

a) Montrer que $H(Z|Y) = H(X|Y)$. En déduire que si X et Y sont indépendantes, alors

$$H(Z) \geqslant \max(H(X), H(Y))$$

(i.e. l'addition de deux variables aléatoires indépendantes augmente l'incertitude).

b) Donner un exemple de deux variables aléatoires X et Y (nécessairement corrélées) telles que $H(Z) < H(X))$ et $H(Z) < H(Y)$.

c) Dans quels cas a-t-on $H(Z) = H(X) + H(Y)$?

Exercice 14 (convexité de la distance de Kullback).
a) Montrer que $\forall x_1, x_2 \in \mathbb{R}_+$, $\forall \alpha \in [0, 1]$,

$$\Big(\alpha x_1 + (1-\alpha)x_2\Big) \log \Big(\alpha x_1 + (1-\alpha)x_2\Big) \ \leqslant\ \alpha x_1 \log x_1 + (1-\alpha)x_2 \log x_2.$$

b) En déduire que $\forall (p_1, p_2) \in \mathbb{R}_+^2$, $\forall (q_1, q_2) \in \mathbb{R}_+^{*\,2}$, $\forall \lambda \in [0, 1]$,

$$\Big(\lambda p_1 + (1-\lambda)p_2\Big) \log \frac{\lambda p_1 + (1-\lambda)p_2}{\lambda q_1 + (1-\lambda)q_2} \leqslant \lambda p_1 \log \frac{p_1}{q_1} + (1-\lambda)p_2 \log \frac{p_2}{q_2}$$

(poser $x_1 = p_1/q_1$ et $x_2 = p_2/q_2$).
c) En déduire que la fonction

$$\left\{ \begin{array}{ccc} S_n \times S_n & \to & [0, +\infty] \\ (p, q) & \mapsto & D(p\|q) \end{array} \right.$$

est convexe, où

$$S_n = \left\{ p \in [0,1]^n,\ \sum_{i=1}^{n} p_i = 1 \right\}.$$

On rappelle que si la distribution q n'est pas absolument continue par rapport à p (i.e., s'il existe i tel que $p_i = 0$ et $q_i \neq 0$), alors $D(p\|q) = +\infty$.

Exercice 15 (principe d'équirépartition asymptotique, avec Scilab). Soient $X_1, X_2, \dots X_n$ des variables aléatoires indépendantes et indentiquement distribuées selon une loi de Bernouilli de paramètre $p \in]0,1[$, i.e

$$P(X_i = 1) = p \quad \text{et} \quad P(X_i = 0) = 1 - p.$$

a) 0n pose $S_n = X_1 + X_2 + \dots X_n$. Montrer que $P(X_1, X_2, \dots X_n) = f(S_n)$, où $f(k) = p^k (1-p)^{n-k}$. Que vaut $P(S_n = k)$?

b) On suppose désormais que $p = 0.1$ et $n = 1000$. Calculer à l'aide de commandes Scilab un tableau H à $n+1$ lignes tels que

$$\forall k, 0 \leqslant k \leqslant n, \quad H(k+1) = -\frac{1}{n} \log_2 f(k).$$

c) Calculer en Scilab un tableau C à $n+1$ lignes tels que

$$\forall k, 0 \leqslant k \leqslant n, \quad C(k+1) = \log_2 \binom{n}{k} = \sum_{i=1}^{k} \log_2(n-k+i) - \log_2(i).$$

d) En utilisant les tableaux H et C calculés précédemment, calculer simplement un tableau P à $n+1$ lignes tel que

$$\forall k, 0 \leqslant k \leqslant n, \quad P(k+1) = P(S_n = k).$$

e) Que tracent les commandes scilab suivantes ?

```
xbasc();plot2d(H,P)
```

Expliquer l'allure du graphe obtenu à l'aide d'un théorème du cours. Autour de quelle valeur h y a-t-il accumulation ? Calculer cette valeur h puis commenter le résultat des commandes scilab

```
J=(abs(H-h)<0.1);sum(P(J))
```

Exercice 16 (loi des grands nombres).

a) Soit X une variable aléatoire positive. Montrer que pour tout $\delta > 0$,

$$P(X \geqslant \delta) \leqslant \frac{E(X)}{\delta}$$

(inégalité de Markov). On pourra se limiter au cas d'une variable aléatoire discrète.

b) Soit Y une variable aléatoire telle que $E(Y) = \mu$ et $E((Y - \mu)^2) = \sigma^2$. Montrer que pour tout $\varepsilon > 0$,

$$P(|Y - \mu| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}$$

(inégalité de Tchebychev).

c) Soient $Z_1, Z_2, \ldots Z_n$ des variables aléatoires i.i.d (indépendantes et identiquement distribuées). On pose

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^{n} Z_i.$$

Calculer $E(\bar{Z}_n)$ et $E((\bar{Z}_n - \mu)^2)$, puis montrer que pour tout $\varepsilon > 0$,

$$P(|\bar{Z}_n - \mu| \geqslant \varepsilon) \underset{n \to +\infty}{\longrightarrow} 0$$

(loi faible des grands nombres).

Exercice 17 (entropie conditionnelle). On considère un jeu de 52 cartes, dont 26 sont noires et 26 rouges. On mélange le jeu, puis on retourne les cartes une à une. Soit $X_k \in \{R, N\}$ la couleur de la k-ième carte (R pour rouge, N pour noire).

1. Comparer et calculer $H(X_1)$, $H(X_2)$, $H(X_2|X_1)$, $H(X_3|X_1)$.

2. Montrer que
$$H(X_3|X_1, X_2) - H(X_2|X_1) = I(X_2; X_3|X_1).$$
En déduire que
$$f(k) = H(X_k|X_1, \ldots X_{k-1})$$
est une fonction décroissante de k.

3. Calculer $f(52)$. En déduire que
$$H := H(X_1, \ldots X_{52}) \leqslant 51.$$

4. Expliquer, en utilisant la notion de message représentatif, pourquoi on doit s'attendre à ce que H soit plutôt proche de 50 que de 0.

5. Montrer que
$$H = \log \binom{52}{26}.$$

6. Donner un développement asymptotique à trois termes quand $n \to +\infty$ de $\log \binom{2n}{n}$. *Indication: utiliser la formule de Stirling,* $n! \sim n^n e^{-n} \sqrt{2\pi n}$.

7. En déduire une approximation de H. Calculer numériquement H sous Scilab et comparer.

8. Calculer f sous Scilab et représenter son graphe.

Exercice 18 (propriétés des codes).

1. Pour chacun des codes c_1, c_2, c_3, c_4 et c_5 définis ci-dessous, répondre aux questions suivantes. Le code est-il régulier ? déchiffrable ? instantanné ? complet ? Peut-il être un code de Huffman ? Justifier les réponses.

x	$c_1(x)$	$c_2(x)$	$c_3(x)$	$c_4(x)$	$c_5(x)$
1	1	000	01	001	00
2	01	001	011	01	01
3	11	100	001	100	100
4	0	101	101	101	101
5	00	111	11	11	11

2. Pour ceux des codes ci-dessus qui sont déchiffrables, donner un code instantanné complet dont la longueur de chaque mot n'augmente pas par rapport au code initial (si le code initial est déjà instantanné et complet, il n'y a rien à faire bien sûr).

Exercice 19 (codage de Shannon-Fano-Elias).

1. Soient a, b deux réels tels que $0 \leqslant a < b \leqslant 1$, et
$$l = l(a,b) := \lceil \log_2 \frac{1}{b-a} \rceil + 1$$
($\lceil t \rceil$ étant le plus petit entier supérieur ou égal à t). Montrer que
$$2^l(b-a) \geqslant 2,$$
et en déduire qu'il existe un entier $k \in \{0, 1, \ldots 2^l - 1\}$ tel que
$$\left[\frac{k}{2^l}, \frac{k+1}{2^l}\right[\subset [a, b[.$$

2. On note $C(a,b)$ l'écriture en base 2 avec $l(a,b)$ chiffres binaires de l'entier k trouvé à la question 1 (on complète éventuellement par des zéros à gauche pour avoir exactement $l(a,b)$ chiffres binaires). Montrer que si $[a,b[$ et $[a',b'[$ sont des intervalles disjoints (avec $0 \leqslant a < b \leqslant 1$ et $0 \leqslant a' < b' \leqslant 1$), alors aucun des deux mots binaires $C(a,b)$ et $C(a',b')$ n'est préfixe l'un de l'autre.

   Indication: on pourra utiliser le fait que si $C(a,b) = b_1 b_2 \ldots b_l$ avec $l = l(a,b)$ et $b_i \in \{0,1\}$ pour tout i, alors l'écriture en base 2 du réel $k/2^l$ est $0, b_1 b_2 \ldots b_l$.

3. Écrire une fonction Scilab `s=code(a,b)`, qui prend en entrée deux réels a et b vérifiant $0 \leqslant a < b \leqslant 1$, et renvoie dans s la chaîne binaire $C(a,b)$.

   Indication: utiliser la fonction Scilab `ceil(t)` *pour calculer* $\lceil t \rceil$.

4. Soit X une variable aléatoire à valeurs dans $\mathcal{X} = \{1, 2, \dots n\}$ et telle que

$$\forall x \in \mathcal{X}, \quad p(x) > 0,$$

   où $p(x) = P(X = x)$. On définit, pour tout $x \in \mathcal{X}$,

$$F(x) = \sum_{y<x} p(y),$$

   puis le code $c : \mathcal{X} \to \{0,1\}^*$ par

$$c(x) = C\Big(F(x), F(x) + p(x)\Big).$$

   Montrer que c est un code instantanné pour X.

5. Montrer que la longueur moyenne L du code c de X vérifie

$$L < H(X) + 2.$$

6. Calculer le code c pour la variable aléatoire X définie par la distribution $p(x)$ suivante :

x	1	2	3	4	5
$p(x)$	0,15	0,2	0,25	0,15	0,25

   Calculer numériquement l'entropie de X, ainsi que la longueur moyenne L de ce code.

7. Représenter l'arbre de codage associé au code obtenu à la question 7. Ce code est-il complet ? (rappel: un code est complet si l'inégalité de Kraft est saturée).

8. Montrer que l'on peut racourcir (en supprimant un ou plusieurs chiffres terminaux) certains mots du code c obtenu à la question 7 en conservant la propriété de code instantanné. Quel est le code instantanné le plus court que l'on peut obtenir ainsi ? Quelle est sa longueur moyenne ?

9. Calculer le code de Huffman associé à la distribution p de la question 7, ainsi que sa longueur moyenne L'. Commenter la différence entre L et L'.

10. Écrire une fonction Scilab `c=build_code(p)`, qui prend en entrée un vecteur colonne p de taille n représentant la distribution de probabilité $p(x)$, et renvoie dans `c` un tableau de même taille dont les éléments sont les chaînes binaires correspondant au codage c associé à la distribution $p(x)$. Cette fonction pourra faire appel à la fonction `code` définie à la question 4.

11. Vérifier le code obtenu à la question 7 au moyen de la fonction `build_code`.

Exercice 20 (arbre de décision).

Parmi 6 bouteilles de vin (numérotées de 1 à 6), exactement une est mauvaise mais on ignore laquelle. L'analyse visuelle des bouteilles permet d'estimer la probabilité a priori pour chaque bouteille d'être mauvaise (probabilité p_i pour la bouteille i) :

i	1	2	3	4	5	6
p_i	$\frac{4}{26}$	$\frac{3}{26}$	$\frac{7}{26}$	$\frac{4}{26}$	$\frac{5}{26}$	$\frac{3}{26}$

1. Si l'on goûte les bouteilles une à une, dans quel ordre faut-il les goûter pour minimiser le nombre d'essais en moyenne ? Que vaut ce nombre moyen d'essais dans ce cas ?

2. On s'autorise maintenant à goûter des mélanges réalisés avec le contenu de plusieurs bouteilles.

   (a) Si l'on ne goûte qu'un seul mélange, lequel peut-on réaliser pour maximiser la probabilité de trouver la mauvaise bouteille ?

   (b) On s'autorise maintenant à goûter plusieurs mélanges successifs. Proposer un arbre de décision décrivant des mélanges successifs à réaliser de façon à minimiser le nombre d'essais moyen. Quels mélanges peut-on tester en premier sans perdre le caractère minimal du nombre d'essais moyen ?

Exercice 21 (exemple de codage de Huffman). On considère une variable aléatoire X distribuée selon la distribution $p(x)$ donnée par le tableau suivant :

x	1	2	3	4	5	6	7
$p(x)$	0.5	0.26	0.11	0.04	0.04	0.03	0.02

1. Construire un code de Huffman pour X.

2. Calculer la longueur moyenne du code obtenu, puis l'entropie de X, et les comparer.

3. Construire un code de Huffman ternaire pour X.

Exercice 22 (stratégie optimale).

1. Un ordinateur tire au hasard un entier X selon une distribution de probabilité connue $p(x), x \in \{1, 2, \ldots, 100\}$. Le joueur pose la question "est-ce que $X = i$?" et se voit répondre "oui", "$X > i$" ou "$X < i$" selon les cas. Le jeu lui donne droit à 6 questions. S'il obtient la réponse "oui" à l'une des 6 questions, il reçoit un prix d'une valeur $v(x)$. Comment le joueur doit-il procéder pour maximiser son espérance de gain ?

2. On suppose maintenant que le joueur a le droit de poser séquentiellement des questions quelconques à réponse binaire jusqu'à ce qu'il puisse déterminer X avec certitude. Chaque question coûte la valeur 1. Comment le joueur doit-il procéder pour maximiser son espérance de gain ? Que vaut-elle au maximum ?

3. Avec les règles de la question 2, quelle est la distribution p la plus défavorable pour le joueur (la fonction v étant fixée) ? Quelle est son espérance de gain dans ce cas ?
