Thème : variables aléatoires Série 3

Exercice 1

Une boîte contient deux billes noires et trois billes blanches. On tire une bille à la fois, sans remise, jusqu'à ce que la première bille noire apparaisse. Dès que la bille noire est tirée, aucune autre bille est extraite de la boîte. Désignons par X le nombre de tirages nécessaires. En utilisant un arbre binaire,

a) déterminer l'ensemble fondamental Ω de l'expérience aléatoire à l'aide des événements :

N : "Une bille noire est tirée" et B : "Une bille blanche est tirée";

b) calculer la loi de probabilité de X.

Exercice 2

Des signaux binaires (0 ou 1) sont transmis d'un point A à un point B par un canal de communication. Cependant, une erreur de transmission peut toujours résulter de perturbations aléatoires agissant sur le canal. Supposons qu'un signal émis en A soit correctement enregistré en B avec probabilité 0.8 et ce indépendamment d'un signal à l'autre. Pour essayer d'améliorer la qualité de transmission, on se propose d'émettre chaque signal trois fois de suite, i.e. 000 au lieu de 0 et 111 au lieu de 1.

a) Déterminer les réceptions possibles par émission par blocs de 3 signaux identiques.

b) Supposons que le signal 0 a été émis en A par le bloc 000 et définissons par X la variable aléatoire qui indique le nombre d'erreurs possibles à la réception du bloc. Déterminer les valeurs que peut prendre la variable aléatoire X.

c) Calculer la loi de probabilité de X.

d) Le décodage à la réception se réalise suivant le principe majoritaire. Par exemple, 010 est décodé en 0. Calculer la probabilité d'erreur pour le signal 0 émis par le bloc 000. Est-elle inférieure à 0.2, probabilité sans répétition du signal ? La qualité de transmission est-elle améliorée ?

Exercice 3

La fonction de répartition d'une variable aléatoire X est

$$F_X(x) = \begin{cases} 0 & \text{si} \quad x < 5; \\ 0.2 & \text{si} \quad 5 \leq x < 7; \\ 0.3 & \text{si} \quad 7 \leq x < 8; \\ 0.7 & \text{si} \quad 8 \leq x < 10; \\ 0.8 & \text{si} \quad 10 \leq x < 11; \\ 1 & \text{si} \quad x \geq 11. \end{cases}$$

a) Représenter graphiquement la fonction de répartition de X.

b) Déterminer les réalisations de la variable aléatoire X ainsi que sa loi de probabilité.

c) Calculer les probabilités $P(X \leq 6.5)$, $P(8 < X \leq 10)$, $P(8 \leq X \leq 10)$ et $P(X \geq 10)$.

d) Sans calcul, déterminer l'espérance de X.

Exercice 4

Imaginons qu'on souhaite transmettre les réalisations $x_1, x_2, \ldots, x_n$ d'une variable aléatoire discrète X d'un point d'observation A à un point de réception B à l'aide d'un canal de communication ne pouvant transférer que des 0 ou des 1. Ainsi, les valeurs prises par X devront être codées en chaînes formées uniquement de 0 et de 1 avant d'être transmises. Pour éviter toute ambiguïté, on exige qu'un code ne puisse pas être une extension d'un autre.

Comme exemple, supposons que les réalisations de X sont x_1, x_2, x_3 et x_4. Comme code possible, on peut envisager

$$x_1 \leftrightarrow 00, \quad x_2 \leftrightarrow 01, \quad x_3 \leftrightarrow 10, \quad x_4 \leftrightarrow 11. \tag{1}$$

Ainsi, si X prend la valeur x_1, le message envoyé en B sera 00, il vaudra 01 si $X = x_2$ et ainsi de suite. Un autre code possible est

$$x_1 \leftrightarrow 0, \quad x_2 \leftrightarrow 10, \quad x_3 \leftrightarrow 110, \quad x_4 \leftrightarrow 111. \tag{2}$$

En revanche, le codage

$$x_1 \leftrightarrow 0, \quad x_2 \leftrightarrow 1, \quad x_3 \leftrightarrow 00, \quad x_4 \leftrightarrow 01$$

n'est pas admis étant donné que x_3 et x_4 sont des extensions de x_1.

Un objectif du codage consiste tout naturellement à minimiser le nombre espéré de bits nécessaires pour transmettre l'information. Ainsi, un code est dit *plus efficace* qu'un autre si son nombre espéré de bits est plus petit que celui nécessaire à l'autre code.

a) Supposons que la loi de probabilité de la variable aléatoire X est

$X = x$	x_1	x_2	x_3	x_4
$P(X = x)$	$1/2$	$1/4$	$1/8$	$1/8$

et considérons les codes donnés ci-dessus par (1) et (2).

Pour cette distribution de X, quel est le code le plus efficace ?

b) Considérons en toute généralité une variable aléatoire discrète X prenant ses valeurs dans l'ensemble $\{x_1, x_2, \ldots, x_n\}$ avec probabilités correspondantes $p_1, p_2, \ldots, p_n$. La grandeur

$$H(X) = -\sum_{i=1}^{n} p_i \cdot \log_2(p_i)$$

est appelée, en théorie de l'information, *entropie* de la variable aléatoire X. Par convention, si l'une des probabilités p_i est nulle, on pose $0 \cdot \log_2(0) = 0$. L'entropie représente en quelque sorte la quantité d'incertitude relative à X. Autrement dit, on considère $H(X)$ comme l'*information* liée à l'observation de X.

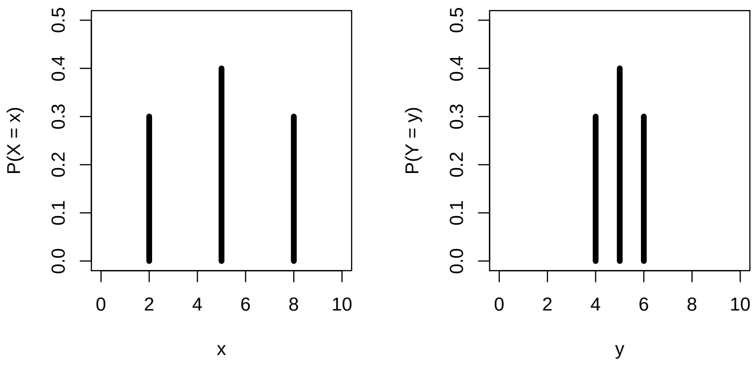
Calculer l'entropie de X selon la distribution donnée en a).

c) Selon le théorème du codage sans bruit, tout codage nécessite un nombre espéré de bits au minimum égal à l'entropie de X.

L'un des deux codes présentés ci-dessus est-il le code le plus efficace qu'il puisse exister pour X ? Est-ce surprenant ?

Exercice 5

Considérons deux variables aléatoires X et Y dont les histogrammes se trouvent dans la figure ci-dessous.



En n'effectuant aucun calcul, l'écart-type de X est-il plus grand que celui de Y ?

Exercice 6

La distribution de Benford prétend que pour des mesures d'une certaine quantité physique, la probabilité de tomber sur un 1 est de l'ordre de 30 % alors que celle de tomber sur un 9 est de l'ordre de 4.6 %. Cette distribution n'est applicable que pour un grand nombre d'objets mesurables. Elle s'applique par exemple aux longueurs de tous les fleuves du monde, aux prix des articles d'un supermarché ou encore aux volumes des tables de logarithmes d'une encyclopédie. En effet, les premières pages de chaque volume sont toujours plus usées que les suivantes, celles du premier volume étant plus utilisées que celles du second et ainsi de suite. Il en sera de même pour les nombres d'une comptabilité. Ainsi, pour détecter d'éventuelles fraudes dans les déclarations d'impôts de grandes entreprises, le fisc américain a utilisé la distribution de Benford. En effet, dans une comptabilité de grande envergure, qui bien souvent est formée de nombreuses pages de chiffres, les comptables essayaient pour déjouer le fisc de créer des données truquées en tirant des nombres au hasard. L'idée n'était pas très bonne. En se livrant à une petite analyse statistique à l'aide de la distribution de Benford, il était possible d'identifier des fraudes dans les déclarations d'impôts de grandes sociétés.

La distribution de probabilités de Benford se trouve dans la table ci-dessous.

valeur	1	2	3	4	5	6	7	8	9
probabilité	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

a) Calculer l'espérance mathématique d'une variable aléatoire X issue d'une distribution de Benford.

b) Déterminer la probabilité $P(X < 7 \,|X \geq 4)$.