MATHÉMATIQUES
Devoir surveillé n°6
Durée : 3 heures

L'usage d'une calculatrice est interdit pour cette épreuve. Si, au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il le signale sur sa copie et poursuit sa composition en expliquant les raisons des initiatives qu'il a été amené à prendre.

Problème

Notation Soient X et Y deux variables aléatoires discrètes, on note $\text{cov}(X,Y)$ leur covariance lorsque celle-ci existe. On rappelle que la covariance est bilinéaire, c'est à dire :
$\forall (X_1, X_2, Y_1, Y_2)$ variables aléatoires, $\forall \lambda \in \mathbb{R}$ tels que les covariances existent, on a :
$\text{cov}(X_1, \lambda Y_1 + Y_2) = \lambda \text{cov}(X_1, Y_1) + \text{cov}(X_1, Y_2)$ et $\text{cov}(\lambda X_1 + X_2, Y_1) = \lambda \text{cov}(X_1, Y_1) + \text{cov}(X_2, Y_1)$

Partie I

Soient n, s deux entiers supérieurs ou égaux à 2 ; on considère une urne contenant des boules de couleurs $C_1, \ldots, C_s$ en proportions $p_1, \ldots, p_s$ respectivement.
On a donc $\sum_{i=1}^{s} p_i = 1$ et on suppose que pour tout $i \in [\![1, s]\!]$, $p_i > 0$.
On effectue n tirages successifs d'une boule avec remise. Pour tout $i \in [\![1, s]\!]$ on note X_i la variable aléatoire égale au nombre de boules de couleur C_i obtenues au cours des n tirages. On remarque que la variable X_i dépend de n et on définit la variable aléatoire U_n par $U_n = \sum_{i=1}^{s} \frac{(X_i - n p_i)^2}{n p_i}$

A. Étude des variables X_i

1. Déterminer la loi de X_i, son espérance et sa variance.
2. Soient $(i, j) \in [\![1, s]\!]^2$ distincts ; déterminer la loi de $X_i + X_j$ et sa variance.
   En déduire $\text{cov}(X_i, X_j) = -n p_i p_j$.

B. On suppose dans cette partie que $s = 2$

1. Montrer que $U_n = Z_1^2$ où $Z_1 = \frac{X_1 - n p_1}{\sqrt{n p_1 p_2}}$
2. Par quelle loi peut-on approcher la loi de Z_1 lorsque n est grand ?

C. On suppose dans cette partie que $s = 3$, $p_1 = p_2 = \frac{1}{4}$ et $p_3 = \frac{1}{2}$

On pose $Z_1 = \frac{2}{\sqrt{n}} \left(X_3 - \frac{n}{2} \right)$ et $Z_2 = \sqrt{\frac{2}{n}} (X_1 - X_2)$

1. Montrer que $U_n = Z_1^2 + Z_2^2$ *(On utilisera la relation $X_1 + X_2 + X_3 = n$).*
2. Déterminer l'espérance de Z_1, de Z_2 ainsi que $\text{cov}(Z_1, Z_2)$.
3. Par quelle loi peut-on approcher la loi de Z_1 lorsque n est grand ?
4. Pour tout $i \in [\![1, n]\!]$, on définit la variable T_i par :
   $T_i = 1$ si on obtient une boule de couleur C_1 au $i$ième tirage ;
   $T_i = -1$ si on obtient une boule de couleur C_2 au $i$ième tirage ;
   $T_i = 0$ si on obtient une boule de couleur C_3 au $i$ième tirage.
   (a) Exprimer $X_1 - X_2$ à l'aide des variables T_i.
   (b) En déduire que l'on peut approcher la loi de Z_2 par une loi normale centrée réduite lorsque n est suffisamment grand.

5. Programme Python

D. On suppose désormais que $s = 4$ **et** $p_1 = p_2 = p_3 = p_4 = \dfrac{1}{4}$

Soient $A \in \mathcal{M}_{p,n}(\mathbb{R})$, $A = (a_{i,j})_{1\leqslant i\leqslant p, 1\leqslant j\leqslant n}$ et ${}^tA$ sa matrice transposée.
On rappelle que pour toutes matrices A, B telles que le produit AB existe, on a ${}^t(AB) = {}^tB\,{}^tA$

1. Pour tout $i \in [\![1,4]\!]$, on pose $Y_i = \dfrac{X_i - np_i}{\sqrt{np_i}}$

   On note M la matrice carrée d'ordre 4 dont le coefficient en ligne i et colonne j vaut $\mathrm{cov}(Y_i, Y_j)$ et N la matrice carrée d'ordre 4 dont tous les coefficients valent $\frac{1}{4}$.
   Exprimer M en fonction de N et I_4.

2. (a) Soient $\vec{e}_1 = \dfrac{1}{\sqrt{2}}(1,-1,0,0)$, $\vec{e}_2 = \dfrac{1}{\sqrt{6}}(1,-1,-2,0)$, $\vec{e}_3 = \dfrac{1}{2\sqrt{3}}(1,1,1,-3)$, $\vec{e}_4 = \dfrac{1}{2}(1,1,1,1)$ 4 vecteurs de $\mathbb{R}^4$.
   Montrer que ces 4 vecteurs sont des vecteurs propres de N et qu'ils forment une base de $\mathbb{R}^4$.

   (b) Soit Q la matrice de passage de la base canonique de $\mathbb{R}^4$ à la base $(\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4)$.
   Montrer que ${}^tQ\,Q = I_4$.
   Expliciter la matrice ${}^tQ\,M\,Q$.

3. On définit les variables aléatoires Z_1, Z_2, Z_3, Z_4 par $\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} = {}^tQ \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$

   (a) Exprimer chaque Z_i en fonction de Y_1, Y_2, Y_3, Y_4 et montrer que $Z_4 = 0$.

   (b) Montrer que les variables Z_1, Z_2, Z_3 sont centrées.

   (c) Montrer que $U_n = Z_1^2 + Z_2^2 + Z_3^2$ *on pourra calculer* $\begin{pmatrix} Y_1 & Y_2 & Y_3 & Y_4 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$.

Partie II

1. Soit r un entier naturel supérieur ou égal à 2.
   Établir la convergence de l'intégrale $\displaystyle\int_0^{+\infty} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} \mathrm{d}x$; on note J_r sa valeur.

2. (a) Montrer que pour tout réel $t > 0$, l'intégrale $\displaystyle\int_t^{+\infty} x^{-\frac{1}{2}} e^{-\frac{x}{2}} \mathrm{d}x$ converge ; on note $G(t)$ sa valeur.

   (b) Montrer à l'aide d'un changement de variable, que pour tout réel $t > 0$, $G(t) = 2\sqrt{2\pi}(1 - \Phi(t))$ où Φ désigne la fonction de répartition de la loi normale centrée réduite.

   (c) En déduire la convergence et la valeur de $J_1 = \displaystyle\int_0^{+\infty} x^{-\frac{1}{2}} e^{-\frac{x}{2}} \mathrm{d}x$

3. Soit r un entier naturel non nul ; on définit la fonction f_r par :

$$\forall x > 0,\ f_r(x) = \frac{1}{J_r} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} \text{ et } \forall x \leqslant 0,\ f_r(x) = 0$$

   (a) Montrer que f_r est une densité de probabilité.
   On dit qu'une variable aléatoire suit une loi du χ^2 à r degrés de liberté si X admet f_r comme densité.

   (b) Quelle loi reconnaît-on pour $r = 2$?
   Graphique ?

Partie III

On reprend les notations de la partie I avec s un entier quelconque supérieur ou égal à 2.

On admet que pour n assez grand, U_n suit la loi du χ^2 à $s-1$ degrés de liberté.
On considère un algorithme générateur de nombres aléatoires compris entre 1 et 4 ; on utilise cet algorithme pour créer un échantillon de 10000 nombres compris entre 1 et 4. On obtient :
2602 fois le nombre 1
2534 fois le nombre 2
2422 fois le nombre 3
2442 fois le nombre 4

On se propose de tester la fiabilité de cet algorithme par l'examen de cet échantillon, avec comme hypothèse que le nombre aléatoire fourni par le générateur suit une loi uniforme sur $[\![1,4]\!]$.

On donne les valeurs suivantes :
$F_3(7,81)=0,95$ où F_3 désigne la fonction de répartition de la loi du χ^2 à 3 degrés de liberté.
$\dfrac{1}{2100}\left(102^2+34^2+78^2+58^2\right)=8,4032$

En introduisant des variables convenables X_1, X_2, X_3, X_4 et la variable U_n associée, justifier le rejet de l'hypothèse de répartition uniforme des nombres fournis par le générateur.