

Une urne contient des boules de couleurs $C_1, C_2, \dots, C_k$ ($k \in \mathbb{N}$, $k \geq 2$). Les boules de couleur C_i sont en proportion p_i non nulle avec $\sum_{i=1}^k p_i = 1$. Étant donné n entier naturel non nul, on tire

successivement dans l'urne n boules, avec remise après chaque tirage.

Pour i , entier entre 1 et k , on appelle X_i la variable aléatoire désignant le nombre de boules de couleur C_i obtenues lors du tirage.

1. a. Déterminer l'espérance et la variance de X_i .

b. Pour $i \neq j$, déterminer la loi de $X_i + X_j$ et en déduire que $\text{Cov}(X_i; X_j) = -np_i p_j$.

2. Soit k entier supérieur ou égal à 2. On pose $Y_i = \frac{X_i - np_i}{\sqrt{np_i}}$ et on note M la matrice de covariance des variables aléatoires $Y_1, Y_2, \dots, Y_k$, définie par $M = (\text{Cov}(Y_i; Y_j))_{\substack{1 \leq i \leq k \\ 1 \leq j \leq k}}$.

a. Montrer que pour $i \neq j$, $\text{Cov}(Y_i; Y_j) = -\sqrt{p_i p_j}$ et que $\text{Cov}(Y_i; Y_i) = 1 - p_i$. En déduire l'expression de la matrice M .

b. Dans la suite, $\mathfrak{M}_{n,p}(\mathbb{R})$ désigne l'ensemble des matrices à n lignes et p colonnes à coefficients réels. On notera I_k la matrice unité de $\mathfrak{M}_{k,k}(\mathbb{R})$.

Préciser la matrice P telle que $M = I_k - P$, puis la matrice $C \in \mathfrak{M}_{k,1}(\mathbb{R})$ vérifiant $P = C \cdot {}^t C$, où ${}^t C$ désigne la transposée de la matrice C .

c. Déterminer le rang de P .

d. Calculer P^2 . Préciser les valeurs propres de P et leur multiplicité.

3. Soit $J \in \mathfrak{M}_{k,k}(\mathbb{R})$ dont les coefficients a_{ij} vérifient : $a_{ij} = \begin{cases} 1 & \text{si } i = j \leq k-1 \\ 0 & \text{sinon.} \end{cases}$

Justifier l'existence d'une matrice $S \in \mathfrak{M}_{k,k}(\mathbb{R})$ vérifiant : ${}^t S \cdot S = I_k$ et ${}^t S M S = J$.

4. On définit les variables aléatoires $Z_1, Z_2, \dots, Z_k$ par la relation : $\begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix} = {}^t S \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix}$

et on pose $S = (s_{ij})_{\substack{1 \leq i \leq k \\ 1 \leq j \leq k}}$.

a. Prouver que, pour tout entier i entre 1 et k , Z_i est centrée.

b. Déterminer la matrice de covariance de $Z_1, Z_2, \dots, Z_k$. En déduire que Z_k est la variable certaine égale à zéro.

c. On pose $Q = \sum_{i=1}^k Y_i^2$. Calculer Q en fonction de $Z_1, Z_2, \dots, Z_k$.

d. On admet que n est suffisamment grand pour que la loi de chaque variable X_i puisse être approchée par une loi normale. Dans ces conditions, par quelle loi peut-on approcher la loi de Y_i ? la loi de Z_i ? la loi de Q ?