

MATHÉMATIQUES
ÉPREUVE B
Durée : 3 heures 30 minutes

L'usage d'une calculatrice est autorisé 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.

Les parties I et II sont indépendantes, la partie III très largement indépendante des parties I et II.

Notations

Dans tout le problème, \mathbb{R} désigne l'ensemble des nombres réels. Pour tous entiers naturels n et k , $n \geq k$, $\binom{n}{k}$ désigne le nombre de combinaisons de k éléments choisis parmi n . Par convention, on prend $\binom{n}{0} = 1$. Pour une variable aléatoire réelle X , $E(X)$ et $V(X)$ désignent respectivement l'espérance et la variance de X (lorsque celles-ci existent).

Préliminaires

Lorsqu'on dispose des résultats chiffrés d'une série d'expériences aléatoires, il est légitime de s'intéresser à la suite de ces résultats rangés dans l'ordre croissant. Cette suite ordonnée est appelée suite des statistiques d'ordre. L'objectif de ce problème est d'étudier par deux exemples la loi des statistiques d'ordre d'un échantillon de variables aléatoires indépendantes.

Dans tout ce problème, n désigne un entier naturel supérieur ou égal à 2 fixé.

I. Première partie

Au cours d'un scrutin, des enquêteurs organisent un sondage à la sortie des bureaux de vote. On considère que le scrutin débute à l'instant 0 et s'achève à l'instant 1. La liste électorale comprend n noms, numérotés de 1 à n . Il ne peut pas y avoir d'abstentions.

On modélise l'instant d'arrivée de l'électeur i , $1 \leq i \leq n$ par une variable aléatoire X_i de loi uniforme sur le segment $[0, 1]$. Les variables X_i sont supposées mutuellement indépendantes.

On note (Y_1, Y_2, \dots, Y_n) les n variables aléatoires ayant pour valeur les valeurs des variables (X_1, X_2, \dots, X_n) ordonnées dans l'ordre croissant. Par exemple, pour $n = 4$ si on obtient $X_1 = 0,3$; $X_2 = 0,1$; $X_3 = 0,7$; $X_4 = 0,2$, on aura $Y_1 = 0,1$; $Y_2 = 0,2$; $Y_3 = 0,3$; $Y_4 = 0,7$.

I.1. On s'intéresse tout d'abord aux instants d'arrivée du premier et du dernier votant, respectivement $Y_1 = \min\{X_1, \dots, X_n\}$ et $Y_n = \max\{X_1, \dots, X_n\}$.

I.1.a. Montrer que pour tout x dans $[0, 1]$, $P(Y_n \leq x) = x^n$.

I.1.b. Donner la fonction de répartition F_n de Y_n . En déduire que Y_n est une variable aléatoire à densité. Montrer que la fonction f_n définie par :

$$f_n : \mathbb{R} \rightarrow \mathbb{R} \\ x \mapsto \begin{cases} nx^{n-1} & \text{si } 0 < x < 1 \\ 0 & \text{sinon} \end{cases}$$

est une densité de Y_n .

I.1.c. Montrer que pour tout x dans $[0, 1]$, $P(Y_1 > x) = (1 - x)^n$.

I.1.d. Donner la fonction de répartition F_1 de Y_1 . En déduire que Y_1 est une variable aléatoire à densité. Montrer que la fonction f_1 définie par :

$$f_1 : \mathbb{R} \rightarrow \mathbb{R} \\ x \mapsto \begin{cases} n(1-x)^{n-1} & \text{si } 0 < x < 1 \\ 0 & \text{sinon} \end{cases}$$

est une densité de Y_1 .

I.2.a. Montrer que Y_n et $1 - Y_1$ ont la même loi. En déduire des relations entre $E(Y_1)$ et $E(Y_n)$ d'une part, $V(Y_1)$ et $V(Y_n)$ d'autre part.

I.2.b. Montrer que $E(Y_n) = \frac{n}{n+1}$. En déduire $E(Y_1)$.

I.2.c. Montrer que $V(Y_n) = \frac{n}{(n+1)^2(n+2)}$. En déduire $V(Y_1)$.

I.3. Pour obtenir une estimation statistique des résultats du scrutin, on décide de s'intéresser au k premiers votes où k est un entier inférieur à n suffisamment grand. On cherche donc maintenant la loi et l'espérance de Y_k , $1 \leq k \leq n$, instant d'arrivée du k -ième votant.

Pour un réel x fixé dans $[0, 1]$, on introduit pour tout entier naturel i , $1 \leq i \leq n$ la variable aléatoire U_i valant 1 si l'événement $\{X_i \leq x\}$ est réalisé, 0 sinon. On introduit également la variable

$$\text{aléatoire } U = \sum_{i=1}^n U_i.$$

I.3.a. Donner la loi de U .

I.3.b. Pour un entier naturel k , $1 \leq k \leq n$, fixé, montrer que les événements $\{Y_k \leq x\}$ et $\{U \geq k\}$ sont égaux. En déduire que $P(Y_k \leq x) = \sum_{j=k}^n \binom{n}{j} x^j (1-x)^{n-j}$.

I.4. Dans la suite de cette partie, on notera F_k la fonction de répartition de Y_k et f_k sa densité continue sur \mathbb{R} pour tout entier naturel k , $2 \leq k \leq n-1$ (on ne cherchera pas à calculer f_k).

I.4.a. Pour tout x dans $[0, 1]$ et tout entier k , $2 \leq k \leq n-1$ montrer que $F_k(x) = 1 - F_{n-k+1}(1-x)$. En déduire $f_k(x) = f_{n-k+1}(1-x)$, puis des relations entre $E(Y_k)$ et $E(Y_{n-k+1})$ d'une part, $V(Y_k)$ et $V(Y_{n-k+1})$ d'autre part.

I.4.b. Montrer que pour tout entier naturel k , $1 \leq k \leq n$, $E(Y_k) = \int_0^1 [1 - F_k(x)] dx$. En déduire que $E(Y_{k+1} - Y_k) = \binom{n}{k} \int_0^1 x^k (1-x)^{n-k} dx$.

I.4.c. On introduit la suite $I_k(n) = \binom{n}{k} \int_0^1 x^k (1-x)^{n-k} dx$ pour k entier naturel, $0 \leq k \leq n$.

Montrer que pour tout k entier naturel, $0 \leq k \leq n-1$, $I_k(n) = I_{k+1}(n)$. En déduire la valeur de $I_k(n)$ pour tout entier naturel k , $0 \leq k \leq n$.

I.4.d. En déduire que pour tout entier naturel k , $1 \leq k \leq n$, $E(Y_k) = \frac{k}{n+1}$.

II. Deuxième partie

Une machine possède n composants qui peuvent tomber en panne de façon indépendante. On suppose pour simplifier que tous les composants ont la même durée de vie moyenne. On modélise l'instant où le composant i tombe en panne par une variable aléatoire X_i , $1 \leq i \leq n$ de loi exponentielle de paramètre 1 sur $[0, +\infty[$. Comme dans la partie précédente, on notera (Y_1, Y_2, \dots, Y_n) les n variables aléatoires de mêmes valeurs ordonnées dans l'ordre croissant.

On s'intéresse au cas où la machine ne peut plus fonctionner dès que l'un des composants est en panne (respectivement dès que tous les composants sont en panne). La durée de fonctionnement de la machine est dans ce cas la variable Y_1 (respectivement Y_n).

II.1.a. Montrer que pour tout x dans $[0, +\infty[$, $P(Y_n \leq x) = (1 - e^{-x})^n$.

II.1.b. Donner la fonction de répartition F_n . En déduire que Y_n est une variable aléatoire à densité et donner une densité f_n de Y_n .

II.1.c. Donner de même la fonction de répartition F_1 de Y_1 . En déduire que Y_1 est une variable aléatoire à densité et donner une densité f_1 de Y_1 . En déduire que Y_1 suit une loi exponentielle dont on précisera le paramètre, puis donner $E(Y_1)$ et $V(Y_1)$.

II.2.a. Montrer que l'intégrale $\int_0^{+\infty} x e^{-x} (1 - e^{-x})^{n-1} dx$ converge. En déduire que $E(Y_n)$ existe.

II.2.b. En utilisant la formule du binôme de Newton, montrer que : $E(Y_n) = \sum_{k=0}^{n-1} \frac{(-1)^k \binom{n}{k+1}}{k+1}$.

II.2.c. Montrer que pour tout couple d'entiers naturels (j, k) compris entre 1 et $n-1$, vérifiant $j \leq k-1$, on a $k \binom{k}{j} = j \binom{k}{j} + k \binom{k-1}{j}$.

II.2.d. On introduit la suite u de terme général $u_k = \sum_{j=0}^{k-1} \frac{(-1)^j \binom{k}{j+1}}{j+1}$ pour tout entier $k \geq 1$.

Montrer, en utilisant II.2.c., que $u_k = u_{k-1} + \frac{1}{k}$ pour tout $k \geq 2$. En déduire que $E(Y_n) = \sum_{k=1}^n \frac{1}{k}$.

II.3.a. Soit $(a_i)_{1 \leq i \leq n} \in \mathbb{R}^n$. Montrer que $\left(\sum_{i=1}^n a_i \right)^2 = 2 \sum_{1 \leq i < j \leq n} a_i a_j - \sum_{i=1}^n a_i^2$.

II.3.b. On introduit la suite v de terme général $v_k = \sum_{j=0}^{k-1} (-1)^j \frac{\binom{k}{j+1}}{(j+1)^2}$ pour tout entier $k \geq 1$.

Montrer que $\forall k \geq 2$, $v_k - v_{k-1} = \frac{1}{k} u_k$.

II.3.c. Montrer que $E(Y_n^2) = 2v_n$ puis, en utilisant II.3.a, montrer que $V(Y_n)$ existe et vaut $\sum_{k=1}^n \frac{1}{k^2}$

III. Troisième partie

On considère maintenant le cas général de n variables aléatoires réelles (X_1, \dots, X_n) mutuellement indépendantes et de même loi de fonction de répartition F continue sur \mathbb{R} , et de classe \mathcal{C}^2 sur \mathbb{R} sauf peut-être en un nombre fini de points.

On notera f une densité de X_i sur \mathbb{R} telle que $f(x) = F'(x)$ en tout point x de \mathbb{R} où F est dérivable. On suppose de plus que les intégrales $\int_{\mathbb{R}} xf(x)dx$ et $\int_{\mathbb{R}} x^2 f(x)dx$ sont convergentes.

Comme dans les parties précédentes, on appellera (Y_1, \dots, Y_n) la suite des X_i ordonnées de façon croissante. Dans toute cette partie, on notera F_k la fonction de répartition de Y_k pour tout entier naturel k , $1 \leq k \leq n$.

III.1. On cherche dans un premier temps à déterminer la loi de Y_k pour tout k , $1 \leq k \leq n$.

III.1.a. Montrer que $P(Y_k \leq x) = \sum_{j=k}^n \binom{n}{j} F(x)^j (1 - F(x))^{n-j}$.

III.1.b. En déduire la fonction de répartition F_k de Y_k , puis qu'une densité f_k de Y_k est donnée pour tout $x \in \mathbb{R}$ par $f_k(x) = n \binom{n-1}{k-1} f(x) F(x)^{k-1} (1 - F(x))^{n-k}$.

III.2. On s'intéresse maintenant à la loi du couple (Y_n, Y_1) .

NOTATION 1 : On appelle fonction de répartition du couple de variable aléatoire (X, Y) l'application F définie pour tout couple (x, y) dans \mathbb{R}^2 par $F(x, y) = P([X \leq x] \cap [Y \leq y])$.

NOTATION 2 : On admettra que la densité de la fonction de répartition F du couple (X, Y) , supposée \mathcal{C}^1 sur \mathbb{R}^2 , s'obtient en dérivant successivement F par rapport aux deux variables.

III.2.a. Pour tout couple (x, y) dans \mathbb{R}^2 , calculer $P([Y_1 > y] \cap [Y_n \leq x])$. En déduire la fonction de répartition Φ du couple (Y_n, Y_1) .

III.2.b. Montrer qu'une densité du couple (Y_n, Y_1) peut s'écrire

$$\phi : \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x, y) \mapsto \begin{cases} n(n-1)f(x)f(y)[F(x) - F(y)]^{n-2} & \text{si } x \geq y \\ 0 & \text{si } x < y \end{cases}$$

III.3. On se place désormais et jusqu'à la fin du problème dans le cas de variables $(X_i)_{1 \leq i \leq n}$ uniformes sur $[0, 1]$. On pourra utiliser tous les résultats de la partie I.

III.3.a. Montrer que $E[(Y_n - Y_1)^2] = \frac{n(n-1)}{(n+1)(n+2)}$.

III.3.b. En déduire que $E(Y_1 Y_n) = \frac{1}{n+2}$.

III.3.c. Calculer la covariance de Y_1 et Y_n ainsi que leur coefficient de corrélation.

FIN