Troisième Année Licence M.I.A.S.H.S. 2024 – 2025

Statistique 2

Contrôle Continu 1, Mars 2025

Examen de 1h30. Tout document ou calculatrice est interdit.

Exercice 1 (Sur 14 points)

Soit X une variable aléatoire discrète à valeurs dans $I \subset \mathbf{N}$.

1. Montrer que $\mathbb{P}(X = k) = \mathbb{P}(X > k-1) - \mathbb{P}(X > k)$ pour $k \in \mathbf{N}$ **(1pt)**.
2. Montrer que $\sum_{k=0}^{n} k\,\mathbb{P}(X = k) = \sum_{k=0}^{n-1} \mathbb{P}(X > k) - n\,\mathbb{P}(X > n)$ pour tout $n \in \mathbf{N}^*$ **(1.5pts)**.
3. On suppose que $\mathbb{E}[X] < \infty$. Montrer que $n\,\mathbb{P}(X > n) \underset{n\to\infty}{\longrightarrow} 0$ **(3pts)**. En déduire que
$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \mathbb{P}(X > k) \quad \textbf{(1pt)}.$$
4. Dans une urne, on trouve n boules numérotées de 1 à n. On effectue m tirages indépendants d'une boule avec remise et on note X_n le plus grand numéro tiré parmi les m. Déterminer la fonction de répartition de X_n **(2pts)** et en déduire la loi de X_n **(0.5pts)**.
5. Déterminer la limite quand $n \to \infty$ de $\frac{1}{n}\sum_{k=1}^{n}\left(\frac{k}{n}\right)^m$ **(1pt)**. En déduire un équivalent quand $n \to \infty$ de $\mathbb{E}[X_n]$ **(2pts)**.
6. On suppose maintenant que Y_n est le maximum de m variables aléatoires indépendantes de loi uniforme dans $[1, n]$. Déterminer $\mathbb{E}[Y_n]$ et comparer avec le résultat précédent **(2pts)**.

Proof.

1. La variable est discrète donc $\mathbb{P}(X = k) = F_X(k) - F_X(k-1)$ pour tout $k \in \mathbf{N}$, avec F_X la fonction de répartition de X. Mais $F_X(k) = 1 - \mathbb{P}(X > k)$, d'où $\mathbb{P}(X = k) = 1 - \mathbb{P}(X > k) - \big(1 - \mathbb{P}(X > k)\big) = \mathbb{P}(X > k-1) - \mathbb{P}(X > k)$.
2. Pour $n \in \mathbf{N}^*$, $\sum_{k=0}^{n} k\,\mathbb{P}(X = k) = \sum_{k=0}^{n} k\big(\mathbb{P}(X > k-1) - \mathbb{P}(X > k)\big) = \sum_{k=0}^{n} k\,\mathbb{P}(X > k-1) - \sum_{k=0}^{n} k\,\mathbb{P}(X > k)$ d'après la question précédente. Mais $\sum_{k=0}^{n} k\,\mathbb{P}(X > k-1) = \sum_{k=1}^{n} k\,\mathbb{P}(X > k-1) = \sum_{k=0}^{n-1}(k+1)\,\mathbb{P}(X > k)$ par changement d'indice. D'où $\sum_{k=0}^{n} k\,\mathbb{P}(X = k) = \sum_{k=0}^{n-1}(k+1)\,\mathbb{P}(X > k) - \sum_{k=0}^{n} k\,\mathbb{P}(X > k) = \sum_{k=0}^{n-1} \mathbb{P}(X > k) - n\,\mathbb{P}(X > n)$ pour tout $n \in \mathbf{N}^*$.
3. Si $\mathbb{E}[X] < \infty$ alors $\sum_{k=0}^{\infty} k\,\mathbb{P}(X = k) < \infty$. On en déduit donc que le reste de cette série numérique à termes positifs tend vers 0. soit $\sum_{k=n}^{\infty} k\,\mathbb{P}(X = k) \underset{n\to\infty}{\longrightarrow} 0$. Or $\sum_{k=n}^{\infty} k\,\mathbb{P}(X = k) \geq \sum_{k=n}^{\infty} n\,\mathbb{P}(X = k) \geq n\sum_{k=n+1}^{\infty} \mathbb{P}(X = k) \geq n\,\mathbb{P}(X > n)$. Donc si $\sum_{k=n}^{\infty} k\,\mathbb{P}(X = k) \underset{n\to\infty}{\longrightarrow} 0$ alors $n\,\mathbb{P}(X > n) \underset{n\to\infty}{\longrightarrow} 0$.

   D'après la question 2., $\lim_{n\to\infty} \sum_{k=0}^{n} k\,\mathbb{P}(X = k) = \lim_{n\to\infty}\Big(\sum_{k=0}^{n-1} \mathbb{P}(X > k) - n\,\mathbb{P}(X > n)\Big)$. Comme $\lim_{n\to\infty} \sum_{k=0}^{n} k\,\mathbb{P}(X = k) = \mathbb{E}[X] < \infty$, et comme d'après ce qui précède on en déduit que $\lim_{n\to\infty} n\,\mathbb{P}(X > n) = 0$, on en déduit que $\lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbb{P}(X > k) = \mathbb{E}[X]$, ce qui est aussi vrai pour $\lim_{n\to\infty} \sum_{k=0}^{n} \mathbb{P}(X > k)$.
4. On peut noter U_i la variable aléatoire désignant le ième tirage. Chaque U_i admet une loi uniforme dans $\{1, 2, \ldots, n\}$. Et $X_n = \max_{1\leq i\leq m}(U_i)$, les U_i étant indépendantes. Donc X_n prend ses valeurs dans $\{1, \ldots, n\}$ et:
$$F_{X_n}(x) = 0 \text{ si } x < 1, \text{ et } F_{X_n}(x) = 1 \text{ si } x \geq n.$$
   De plus, pour $1 \leq x \leq n$, $F_{X_n}(x) = \mathbb{P}(\max_{1\leq i\leq m}(U_i) \leq x) = \mathbb{P}(U_1 \leq x)^m$ car les U_i sont indépendantes et ont même distribution. Or $\mathbb{P}(U_1 \leq x) = \sum_{k\leq[x]} \mathbb{P}(U_1 = k) = \sum_{k\leq[x]} 1/n = [x]/n$. D'où $F_{X_n}(x) = \big([x]/n\big)^m$.
   On en déduit que X_n a pour loi discrète:
$$\mathbb{P}(X_n = k) = \big(k/n\big)^m - \big((k-1)/n\big)^m \quad \text{pour tout } k \in \{1, \ldots, n\}.$$
5. C'est une somme de Riemann, avec $f(x) = x^m$ qui est une fonction continue. Donc $\frac{1}{n}\sum_{k=1}^{n}\left(\frac{k}{n}\right)^m \underset{n\to\infty}{\longrightarrow} \int_0^1 x^m\,dx = 1/(m+1)$.
   On a pour tout $n \in \mathbf{N}^*$, $\mathbb{E}[X_n] = \sum_{k=0}^{n} k\,\mathbb{P}(X_n = k) = \sum_{k=0}^{n-1} \mathbb{P}(X_n > k) - n\,\mathbb{P}(X_n > n)$. On a vu que pour $k \in \{0, \ldots, n\}$, $F_{X_n}(k) = \big(k/n\big)^m$, donc $\mathbb{P}(X_n > k) = 1 - \big(k/n\big)^m$ et ainsi $\sum_{k=0}^{n-1} \mathbb{P}(X_n > k) = n\big(1 - \frac{1}{n}\sum_{k=0}^{n-1}\big(k/n\big)^m\big)$, donc grâce à la question précédente $\sum_{k=0}^{n-1} \mathbb{P}(X_n > k) \sim \frac{m}{m+1}\,n$. De plus, $n\,\mathbb{P}(X_n > n) = n\,0 = 0$. Par conséquent $\mathbb{E}[X_n] \sim \frac{m}{m+1}\,n$.

6. On a Y_n qui prend ses valeurs dans $[1,n]$, donc $F_{Y_n}(y) = 0$ si $y < 1$ et $F_{Y_n}(y) = 1$ si $y \geq n$. Et pour $y \in [1,n]$, $F_{Y_n}(y) = \big((y-1)/(n-1)\big)^m$, d'où $f_{Y_n}(y) = m\,(y-1)^{m-1}(n-1)^{-m}\,\mathbb{I}_{y\in[1,n]}$. On en déduit que:

$$\begin{aligned}
\mathbb{E}[Y_n] &= \int_1^n y\,m\,(y-1)^{m-1}(n-1)^{-m}\,dy \\
&= m\,(n-1)^{-m}\int_0^{n-1}(y+1)\,y^{m-1}\,dy \\
&= m\,(n-1)^{-m}\Big(\frac{(n-1)^{m+1}}{m+1}+\frac{(n-1)^m}{m}\Big) \\
&\sim \frac{m}{m+1}\,n \quad \text{pour } n\to\infty\text{: même résultat que précédemment!}
\end{aligned}$$

□

Exercice 2 (Sur 9 points)

Soit le vecteur aléatoire X défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ et on suppose que la loi de X a pour densité f_X par rapport à la mesure de Lebesgue sur $\mathbf{R}^2$:

$$f_X(x_1,x_2) = C\,\mathbb{I}_{x_1^2+x_2^2\leq 1} \qquad \text{pour tout } (x_1,x_2)\in\mathbf{R}^2,$$

avec $C \in \mathbf{R}$.

1. Déterminer C pour que f_X soit bien une densité **(1.5pts)**.
2. En posant $X = (X_1, X_2)$, déterminer les lois de X_1 et de X_2 **(2pts)**.
3. Déterminer $\mathbb{E}[X_1]$ **(1pt)**.
4. Déterminer la loi de $Y = X_1^2 + X_2^2$ **(2pts)** et montrer que $\mathbb{E}[Y] = 1/2$ **(0.5pts)**. En déduire $\mathrm{var}(X_1)$ **(1pt)**.
5. A-t-on X_1 et X_2 indépendantes (justifier!) **(1pt)**?

Proof.

1. On a $C > 0$ pour avoir f_X, qui est mesurable et qui est aussi positive. De plus il faut $\int_{\mathbf{R}^2} f_X(x)\,d\lambda_2(x) = 1$, ce qui implique $C\int\int_{x^2+y^2\leq 1} dxdy = 1$, soit $C\,\pi = 1$ car l'intégrale est la surface du disque unité, donc $C = 1/\pi$.
2. On a d'abord nécessairement X_1 et X_2 dans $[-1,1]$. De plus X_1 est une variable absolument continue du fait que X est absolument continu. On peut donc directement écrire que:

$$f_{X_1}(x_1) = \int_{-1}^1 f_X(x_1,x_2)\,dx_2 = \mathbb{I}_{|x_1|\leq 1}\,\frac{1}{\pi}\int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} dx_2 = \frac{2}{\pi}\,\sqrt{1-x_1^2}\,\mathbb{I}_{|x_1|\leq 1}.$$

   Et X_2 admet la même densité par symétrie.
3. Comme la densité de X_1 est paire, on a $\mathbb{E}[X_1] = 0$.
4. Comme X est dans le disque unité, Y représente OM^2 où M a pour coordonnées (X_1, X_2) donc $Y \in [0,1]$. Donc pour $y < 0$, $F_Y(y) = \mathbb{P}(Y \leq y) = 0$ et $F_Y(y) = 1$ pour $y \geq 1$. Pour $y \in [0,1]$,

$$F_Y(y) = \mathbb{P}(X_1^2+X_2^2\leq y) = \frac{\pi\,y}{\pi} = y,$$

   car la surface définie par le disque $\{(X_1,X_2)\in\mathbf{R}^2,\ X_1^2+X_2^2\leq y\}$ est $\pi\,y$. D'où Y est une v.a. absolument continue de densité:

$$f_Y(y) = \mathbb{I}_{y\in[0,1]} \qquad \Longrightarrow \qquad Y \text{ suit une loi uniforme sur } [0,1].$$

   Comme cela est bien connu pour loi uniforme sur $[a,b]$, $\mathbb{E}[Y] = \frac{1}{2}(a+b) = \frac{1}{2}$ ici.
   On en déduit que $\mathbb{E}[X_1^2+X_2^2] = 2\,\mathbb{E}[X_1^2] = 2\,\mathrm{var}(X_1) = \frac{1}{2}$ puisque $\mathbb{E}[X_1] = 0$. On en déduit que $\mathrm{var}(X_1) = \frac{1}{4}$.
5. Il est clair que $f_X(x_1,x_2) = \frac{1}{\pi}\,\mathbb{I}_{x_1^2+x_2^2\leq 1} \neq f_{X_1}(x_1)\,f_{X_2}(x_2) = \frac{4}{\pi^2}\,\sqrt{(1-x_1^2)(1-x_2^2)}\,\mathbb{I}_{|x_1|\leq 1,|x_2|\leq 1}$ (par exemple pour $x_1 = x_2 = 3/4$, $f_X(x_1,x_2) = 0$ et $f_{X_1}(x_1)\,f_{X_2}(x_2) = \frac{7}{4\,\pi^2}$): les v.a. X_1 et X_2 ne sont pas indépendantes.

□