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

Statistique 2

Contrôle Continu 2, Avril 2025

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

Exercice 1 (Sur 11 points)

On suppose que (X,Y) est un vecteur aléatoire de loi absolument continue par rapport à la mesure de Lebesgue sur $\mathbf{R}^2$ avec pour densité:

$$f(x,y) = K\, e^{-\frac{1}{8}\,(x^2+4y^2)}\, \mathbb{I}_{xy\geq 0} \qquad \text{pour } (x,y)\in\mathbf{R}^2,$$

où $K\in\mathbf{R}$.

1. Tracer dans le plan l'ensemble $\{(x,y)\in\mathbf{R}^2,\ xy\geq 0\}$ **(1pt)**.
2. Démontrer que $K=(2\pi)^{-1}$ **(1.5pts)**.
3. Déterminer les lois marginales de X et de Y **(2pts)**. Que peut-on en conclure quant à l'indépendance entre X et Y **(1pt)**?
4. Démontrer que la corrélation entre X et Y est telle que $\mathrm{cor}(X,Y)=2/\pi$ **(2.5pts)**. En déduire la densité d'un vecteur gaussien $Z=(Z_1,Z_2)$ de mêmes lois marginales que celles de X et de Y et de corrélation $2/\pi$ **(2.5pts)**. En déduire que (X,Y) n'est pas un vecteur gaussien **(0.5pts)**.

Proof.

1. Il s'agit des deux quarts de plan, $\{(x,y)\in[0,+\infty[^2\}$ et $\{(x,y)\in]-\infty,0[^2\}$.
2. On a bien une fonction continue (donc mesurable) et positive si $K\geq 0$. Il reste à trouver K tel que $\int\int_{\mathbf{R}^2} f(x,y)dxdy=1$. Cela revient à l'équation:
$$2\,K\int_0^\infty\int_0^\infty e^{-\frac{1}{8}\,(x^2+4y^2)}dxdy = 2\,K\Big(\int_0^\infty e^{-\frac{1}{8}\,x^2}dx\Big)\Big(\int_0^\infty e^{-\frac{1}{2}\,y^2}dy\Big)=1,$$
grâce à Fubini. Mais pour $\sigma>0$, $\int_0^\infty e^{-\frac{1}{2\sigma^2}\,z^2}dz=\frac{1}{2}\int_{-\infty}^\infty e^{-\frac{1}{2\sigma^2}\,z^2}dz=\frac{1}{2}\sqrt{2\pi}\,\sigma$ (densité gaussienne centrée). D'où
$$2\,K\Big(\frac{1}{2}\sqrt{2\pi}\,2\Big)\Big(\frac{1}{2}\sqrt{2\pi}\Big)=1 \quad\Longrightarrow\quad K\,2\pi=1.$$
3. On sait que X et Y sont des variables absolument continues puisque (X,Y) est absolument continu. La densité de X, f_X est telle que si $x\geq 0$:
$$f_X(x)=(2\pi)^{-1}\int_0^{+\infty} e^{-\frac{1}{8}\,(x^2+4y^2)}dy=(2\pi)^{-1}e^{-\frac{1}{8}\,x^2}\int_0^{+\infty}e^{-\frac{1}{2}\,y^2}dy=(2\pi)^{-1}e^{-\frac{1}{8}\,x^2}\,\frac{1}{2}\sqrt{2\pi}=\frac{1}{2\sqrt{2\pi}}\,e^{-\frac{1}{8}\,x^2}.$$
Par symétrie on obtient la même chose pour $x\leq 0$. Au final $X \overset{\mathcal{L}}{\sim} \mathcal{N}(0,2^2)$.

   De la même manière, on obtient que $Y \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$.

   Pour $x>0$ et $y<0$, on obtient $f(x,y)=0$ alors que $f_X(x)f_Y(y)>0$: X et Y ne sont pas indépendantes.
4. On commence par calculer $\mathrm{cov}(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]=\mathbb{E}[XY]$. Mais:
$$\mathbb{E}[XY]=\frac{2}{2\pi}\int_0^\infty\int_0^\infty xy\,e^{-\frac{1}{8}\,(x^2+4y^2)}dxdy=\frac{1}{\pi}\Big(\int_0^\infty x\,e^{-\frac{1}{8}\,x^2}dx\Big)\Big(\int_0^\infty y\,e^{-\frac{1}{2}\,y^2}dy\Big)=\frac{1}{\pi}\Big[-4\,e^{-\frac{1}{8}\,x^2}\Big]_0^\infty\Big[-e^{-\frac{1}{2}\,y^2}\Big]_0^\infty=\frac{4}{\pi}.$$
Comme $\mathrm{var}(X)=4$ et $\mathrm{var}(Y)=1$, on en déduit que $\mathrm{cor}(X,Y)=2/\pi$.

   Si (X,Y) était gaussien, il serait centré et sa matrice de variance-covariance serait: $\begin{pmatrix} 4 & \frac{4}{\pi} \\ \frac{4}{\pi} & 1\end{pmatrix}$, d'où la densité:
$$\frac{1}{2\pi 2\sqrt{1-\frac{4}{\pi^2}}}\,\exp\Big(-\frac{1}{8(1-4/\pi^2)}\,\big(x^2-\frac{8}{\pi}xy+4y^2\big)\Big) \quad \text{for all } (x,y)\in\mathbf{R}^2.$$
Clairement la densité de (X,Y) n'est pas celle ci-dessus: le vecteur n'est pas gaussien.

$\square$

Préambule de l'exercice 2 (Sur 5 points)

1. Soit U et V deux variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilités. Démontrer que pour tout $\varepsilon \in \mathbf{R}$, $\mathbb{P}(U+V \geq 2\,\varepsilon) \leq \mathbb{P}(U \geq \varepsilon) + \mathbb{P}(V \geq \varepsilon)$ (vous pouvez vous aider d'un dessin) **(2pts)**.
2. Soit (Z_n) une suite de variables aléatoires convergeant en probabilités. En vous aidant de la question précédente, démontrer que la suite $(Z_{n+1} - Z_n)_n$ converge en probabilité vers 0 **(3pts)**.

Proof.

1. Un graphe permet de voir que $\{(x,y) \in \mathbf{R}^2,\ x+y \geq 2\varepsilon\} \subset \{(x,y) \in \mathbf{R}^2,\ x \geq \varepsilon\} \bigcup \{(x,y) \in \mathbf{R}^2,\ x \geq \varepsilon\}$ d'où le résultat.
2. Notons Z_∞ la v.a. telle que $Z_n \xrightarrow[n\to+\infty]{\mathcal{P}} Z_\infty$. Par l'inégalité triangulaire, on a $|Z_{n+1} - Z_n| \leq |Z_{n+1} - Z_\infty| + |Z_n - Z_\infty|$, d'où "$|Z_{n+1} - Z_n| \geq \varepsilon$" $\subset$ "$|Z_{n+1} - Z_\infty| + |Z_n - Z_\infty| \geq \varepsilon$" et ainsi pour tout $\varepsilon > 0$,
$$\mathbb{P}\big(|Z_{n+1} - Z_n| \geq \varepsilon\big) \leq \mathbb{P}\big(|Z_{n+1} - Z_\infty| + |Z_n - Z_\infty| \geq \varepsilon\big).$$
Grâce à la question précédente, $\mathbb{P}\big(|Z_{n+1} - Z_\infty| + |Z_n - Z_\infty| \geq \varepsilon\big) \leq \mathbb{P}\big(|Z_{n+1} - Z_\infty| \geq \varepsilon/2\big) + \mathbb{P}\big(|Z_n - Z_\infty| \geq \varepsilon/2\big)$. Mais comme $Z_n \xrightarrow[n\to+\infty]{\mathcal{P}} Z_\infty$, $\mathbb{P}\big(|Z_{n+1} - Z_\infty| \geq \varepsilon/2\big) \xrightarrow[n\to\infty]{} 0$ and $\mathbb{P}\big(|Z_n - Z_\infty| \geq \varepsilon/2\big) \xrightarrow[n\to\infty]{} 0$, d'où $\mathbb{P}\big(|Z_{n+1} - Z_n| \geq \varepsilon\big) \xrightarrow[n\to\infty]{} 0$.

□

Exercice 2 (Sur 15 points)

On considère $(\varepsilon_k)_{k\in\mathbf{N}^*}$ une suite de variables aléatoires indépendantes de même loi $\mathcal{N}(0, \sigma_\varepsilon^2)$ avec $\sigma_\varepsilon^2 > 0$. Soit X_0 une variable gaussienne de loi $\mathcal{N}(0, \sigma^2)$, avec $\sigma^2 > 0$ et X_0 indépendante de $(\varepsilon_k)_{k\in\mathbf{N}^*}$. Pour $\alpha \in]-1, 1[$, on définit alors par récurrence:
$$X_n = \alpha\, X_{n-1} + \varepsilon_n \quad \text{pour } n \in \mathbf{N}^*.$$

1. Montrer que (X_0, X_1) est un vecteur gaussien **(1pt)**, donner la loi de X_1 **(1pt)**. Suivant les valeurs de α, les variables X_0 et X_1 sont-elles indépendantes **(1pt)**?
2. Démontrer que pour tout $n \geq 1$, $X_n = \alpha^n X_0 + \sum_{k=0}^{n-1} \alpha^k \varepsilon_{n-k}$ **(1.5pts)**. En déduire que pour tout $n \geq 1$, $(X_0, X_1, \ldots, X_n)$ est un vecteur gaussien **(1.5pts)**. Déterminer la loi de X_k pour $k \geq 1$ **(1pt)**.
3. Démontrer que les X_i ont toutes la même loi si et seulement si $(1-\alpha^2)\,\sigma^2 = \sigma_\varepsilon^2$ **(1pt)**. Dans un tel cas, démontrer que $\mathrm{cov}(X_k, X_\ell) = \frac{\sigma_\varepsilon^2}{1-\alpha^2}\, \alpha^{|\ell-k|}$ pour $(k,\ell) \in \mathbf{N}^2$ **(2pts)**. En déduire la loi du vecteur $(X_0, X_1, \ldots, X_n)$ **(1pt)**.
4. On ne suppose plus nécessairement que $(1-\alpha^2)\,\sigma^2 = \sigma_\varepsilon^2$. Démontrer que (X_n) converge en loi et préciser sa loi limite **(2pts)**.
5. Que peut-on dire quant à la convergence en probabilité de (X_n) (utiliser le préambule) **(2pts)**?

Proof.

1. On a $X_1 = \alpha X_0 + \varepsilon_1$, d'où $\begin{pmatrix} X_1 \\ X_0 \end{pmatrix} = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ X_0 \end{pmatrix}$. Or X_0 et ε_0 sont deux v.a. gaussiennes indépendantes donc $\begin{pmatrix} \varepsilon_1 \\ X_0 \end{pmatrix}$ est un vecteur gaussien. Puisque $\begin{pmatrix} X_1 \\ X_0 \end{pmatrix}$ s'écrit comme une matrice de réels multipliée par $\begin{pmatrix} \varepsilon_1 \\ X_0 \end{pmatrix}$, c'est donc aussi un vecteur gaussien.
De ce qui précède X_1 est une v.a. gaussienne (marginale d'un vecteur gaussien). On a $\mathbb{E}[X_1] = \alpha\mathbb{E}[X_0] + \mathbb{E}[\varepsilon_1] = 0$ et $\mathrm{var}(X_1) = \alpha^2 \mathrm{var}(X_0) + \mathrm{var}(\varepsilon_1) = \alpha^2\sigma^2 + \sigma_\varepsilon^2$ puisque X_0 et ε_1 sont indépendantes. Donc $X_1 \overset{\mathcal{L}}{\sim} \mathcal{N}\big(0,\, \alpha^2\sigma^2 + \sigma_\varepsilon^2\big)$.
Puisque $\begin{pmatrix} X_1 \\ X_0 \end{pmatrix}$ est un vecteur gaussien, il suffit de déterminer $\mathrm{cov}(X_0, X_1)$ pour savoir si X_0 et X_1 sont indépendantes. Mais $\mathrm{cov}(X_0, X_1) = \mathrm{cov}(X_0, \alpha X_0 + \varepsilon_1) = \alpha \mathrm{var}(X_0) + 0$ puisque X_0 et ε_1 sont indépendantes. Donc si $\alpha = 0$, alors il y a indépendance, et si $\alpha \neq 0$, il n'y a pas indépendance.
2. On le montre par récurrence. Vrai pour $n = 0$ ou $n = 1$ directement. Si vrai au rang n, alors comme $X_{n+1} = \alpha\, X_n + \varepsilon_{n+1}$, on obtient:
$$X_{n+1} = \alpha\,\Big(\alpha^n X_0 + \sum_{k=0}^{n-1} \alpha^k \varepsilon_{n-k}\Big) + \varepsilon_{n+1} = \alpha^{n+1} X_0 + \sum_{k=0}^{n-1} \alpha^{k+1} \varepsilon_{n-k} + \varepsilon_{n+1} = \alpha^{n+1} X_0 + \sum_{k'=1}^{n} \alpha^{k'} \varepsilon_{n+1-k'}) + \varepsilon_{n+1} = \alpha^{n+1} X_0 + \sum_{k'=0}^{n} \alpha^{k'} \varepsilon_{n+1-k'}.$$
Donc vrai au rang $n+1$.
Avec cette formule, on peut écrire le vecteur $\begin{pmatrix} X_n \\ \vdots \\ X_0 \end{pmatrix}$ sous la forme $A \begin{pmatrix} \varepsilon_n \\ \vdots \\ \varepsilon_1 \\ X_0 \end{pmatrix}$ avec A une matrice de réels. Comme $\begin{pmatrix} \varepsilon_n \\ \vdots \\ \varepsilon_1 \\ X_0 \end{pmatrix}$ est un vecteur gaussien (vecteur composé de v.a. gaussiennes indépendantes), on en déduit que $\begin{pmatrix} X_n \\ \vdots \\ X_0 \end{pmatrix}$ est aussi un vecteur gaussien.
De ce qui précède, X_n est une v.a. gaussienne, son espérance est nulle et comme les variables X_0 et ε_k sont indépendantes, $\mathrm{var}(X_n) = \mathrm{var}\big(\alpha^n X_0 + \sum_{k=0}^{n-1} \alpha^k \varepsilon_{n-k}\big) = \alpha^{2n}\sigma^2 + \sigma_\varepsilon^2 \sum_{k=0}^{n-1} \alpha^{2k} = \alpha^{2n}\sigma^2 + \sigma_\varepsilon^2\, \frac{1-\alpha^{2n}}{1-\alpha^2}$. Donc $X_n \overset{\mathcal{L}}{\sim} \mathcal{N}\big(0,\, \alpha^{2n}\sigma^2 + \sigma_\varepsilon^2\, \frac{1-\alpha^{2n}}{1-\alpha^2}\big)$.

3. X_k a la même loi que X_ℓ si et seulement si leurs variances sont identiques, soit $\alpha^{2k}\sigma^2+\sigma_\varepsilon^2\,\frac{1-\alpha^{2k}}{1-\alpha^2}=\alpha^{2\ell}\sigma^2+\sigma_\varepsilon^2\,\frac{1-\alpha^{2\ell}}{1-\alpha^2}$ ce qui revient à écrire que $(\alpha^{2k}-\alpha^{2\ell})\sigma^2=\sigma_\varepsilon^2\,\frac{\alpha^{2k}-\alpha^{2\ell}}{1-\alpha^2}$, d'où $(1-\alpha^2)\sigma^2=\sigma_\varepsilon^2$.
On a alors pour $0\leq k\leq \ell$,

$$\begin{aligned}
\mathrm{cov}(X_k,X_\ell) &= \mathrm{cov}\Big(\alpha^k X_0+\sum_{j=0}^{k-1}\alpha^j\varepsilon_{k-j},\alpha^\ell X_0+\sum_{i=0}^{\ell-1}\alpha^i\varepsilon_{\ell-i}\Big)\\
&= \mathrm{cov}\Big(\alpha^k X_0+\sum_{j=1}^{k}\alpha^{k-j}\varepsilon_{j},\alpha^\ell X_0+\sum_{i=1}^{\ell}\alpha^{\ell-i}\varepsilon_{i}\Big)\\
&= \alpha^{k+\ell}\,\frac{\sigma_\varepsilon^2}{1-\alpha^2}+\sigma_\varepsilon^2\sum_{j=1}^{k}\alpha^{k-j}\alpha^{\ell-j}\\
&= \alpha^{k+\ell}\,\frac{\sigma_\varepsilon^2}{1-\alpha^2}\big(1-(1-\alpha^{-2k})\big)=\alpha^{\ell-k}\,\frac{\sigma_\varepsilon^2}{1-\alpha^2}.
\end{aligned}$$

On en déduit que $(X_1,\dots,X_n)$ suit une loi gaussienne centrée, et de matrice de covariance $\frac{\sigma_\varepsilon^2}{1-\alpha^2}\,\big(\alpha^{|j-i|}\big)_{1\leq i,j\leq n}$.

4. La loi de X_n est $\mathcal{N}\big(0\,,\,\alpha^{2n}\sigma^2+\sigma_\varepsilon^2\,\frac{1-\alpha^{2n}}{1-\alpha^2}\big)$. Mais lorsque $n\to\infty$, $\alpha^{2n}\sigma^2+\sigma_\varepsilon^2\,\frac{1-\alpha^{2n}}{1-\alpha^2}\underset{n\to\infty}{\longrightarrow}\sigma_\varepsilon^2\,\frac{1}{1-\alpha^2}$ car $\alpha\in]-1,1[$. Donc $X_n\underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}\mathcal{N}\big(0\,,\,\sigma_\varepsilon^2\,\frac{1}{1-\alpha^2}\big)$.

5. On va utiliser le préambule. Si on détermine la loi de X_n+1-X_n, il s'agit encore d'une loi gaussienne centrée et sa variance est celle de $(\alpha-1)X_n+\varepsilon_{n+1}$ donc $(\alpha-1)^2\mathrm{var}(X_n)+\sigma_\varepsilon^2$ du fait que X_n et ε_{n+1} sont indépendantes. Comme on a obtenu que $\mathrm{var}(X_n)\underset{n\to\infty}{\longrightarrow}\sigma_\varepsilon^2\,\frac{1}{1-\alpha^2}$, on a donc $\mathrm{var}(X_{n+1}-X_n)\underset{n\to\infty}{\longrightarrow}\sigma_\varepsilon^2\,\frac{2-\alpha^2}{1-\alpha^2}>0$: donc $(X_{n+1}-X_n)$ ne peut pas converger vers 0 en probabilité (sa loi limite est une loi gaussienne à variance positive), donc (X_n) ne converge pas en probabilité.

$\square$