Université Paris I, Panthéon - Sorbonne

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

Statistique 2

Examen terminal, Mai 2025

Examen de 2h00. Tout document ou calculatrice est interdit.

Exercice 1 (Sur 10 points)

Soit X une variable aléatoire suivant une distribution uniforme sur $[0,1]$. Pour $m \in [0,1/2]$, on définit $Y = |X - m|$.

1. Prouver que pour tout $m \in [0,1/2]$, $\mathbb{E}[Y] = m^2 - m + 1/2$ **(1pt)**.
2. Déterminer la fonction de répartition de Y **(1 pt)** et en déduire que Y est une variable aléatoire absolument continue de densité $f_Y(y) = 2$ pour $y \in [0,m]$, $f_Y(y) = 1$ pour $y \in]m, 1-m]$ et $f_Y(y) = 0$ ailleurs **(0.5pts)**.
3. Supposons que m est inconnu et que $(Y_1,\dots,Y_n)$ est une famille de variables aléatoires i.i.d. observées suivant la même distribution que Y. Avec $\overline{Y}_n$ la moyenne empirique de $(Y_1,\dots,Y_n)$, montrer que $\overline{m}_n = \frac{1}{2}(1 - \sqrt{|4\,\overline{Y}_n - 1|})$ est un estimateur convergent de m **(1pt)**. Montrer que $\mathrm{var}(Y) = \frac{1}{12} - m^2(1-m)^2$ **(1pt)**. En déduire que pour $m \in [0,1/2[$, $\overline{m}_n$ vérifie un théorème central limite que l'on précisera **(2.5pts)**.
4. On définit un second estimateur $\widehat{m}_n = 1 - \max(Y_1,\dots,Y_n)$ de m. Montrer que $\mathbb{P}(m \le \widehat{m}_n \le m+\varepsilon) = 1-(1-\varepsilon)^n$ pour $0 \le \varepsilon \le 1-2m$ **(1pt)**. Déduisez un intervalle de confiance de niveau 95% pour m **(0.5pts)**.
5. Démontrer que $n\,(\widehat{m}_n - m) \underset{n\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{E}(1)$, loi exponentielle de paramètre 1 **(1.5pts)**.

Exercice 2 (Sur 16 points)

Soit $(X_i)_{i\in\mathbf{N}}$ une suite de variables aléatoires identiquement distribuées, d'espérance m et de variance σ^2, telle que pour tout $n \in \mathbf{N}^*$, le vecteur aléatoire $(X_1,\dots,X_n)$ a une matrice de covariance $\Gamma_n = (\gamma_{ij}^{(n)})_{1\le i,j\le n}$ définie positive.

1. Que vaut $\gamma_{11}^{(n)}$ **(0.5pts)**? Montrer que $|\gamma_{ij}^{(n)}| \le \sigma^2$ pour tout $1 \le i,j \le n$ **(0.5pts)**.
2. On note $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^n X_k$. Déterminer $\mathbb{E}[\overline{X}_n]$ **(0.5pts)** et montrer que $\mathrm{var}(\overline{X}_n) = \dfrac{\sigma^2}{n} + \dfrac{2}{n^2}\displaystyle\sum_{1\le i<j\le n} \gamma_{ij}^{(n)}$ **(0.5pts)**.
3. On suppose qu'il existe une fonction $r : \mathbf{N} \to \mathbf{R}$ telle que $\gamma_{ij}^{(n)} = r(|j-i|)$ pour tous $1 \le i,j \le n$ et $n \in \mathbf{N}^*$ telle que $\sum_{k=0}^\infty |r(k)| < \infty$. Montrer que $\mathrm{var}(\overline{X}_n) = \dfrac{\sigma^2}{n} + \dfrac{2}{n}\displaystyle\sum_{k=1}^n (1 - \frac{k}{n})\, r(k)$ **(1.5pts)**. En écrivant que $\sum_{k=1}^n = \sum_{k=1}^{[\sqrt{n}]} + \sum_{k=[\sqrt{n}]+1}^n$, montrer que $n\,\mathrm{var}(\overline{X}_n) \underset{n\to\infty}{\longrightarrow} \sigma^2 + 2\displaystyle\sum_{k=1}^\infty r(k)$ **(2pts)**. En déduire que $\overline{X}_n \underset{n\to+\infty}{\xrightarrow{\mathcal{P}}} m$ **(0.5pts)**.
4. Soit $(Z_n)_{n\in\mathbf{N}}$ une suite de variables aléatoires gaussiennes. Montrer que si les suites $(\mathbb{E}[Z_n])_{n\in\mathbf{N}}$ et $(\mathrm{var}(Z_n))_{n\in\mathbf{N}}$ convergent, alors $(Z_n)_{n\in\mathbf{N}}$ converge en loi vers une loi que l'on précisera **(1pt)**.
5. On suppose désormais que la suite $(X_i)_{i\in\mathbf{N}}$ est telle que le vecteur $(X_1,\dots,X_n)$ est gaussien pour tout $n \in \mathbf{N}^*$. Montrer que $\sqrt{n}\,(\overline{X}_n - m)$ converge en loi vers une limite que l'on précisera **(1pt)**.
6. On suppose que $(X_1,\dots,X_n)$ est observé avec Γ_n connue, mais m inconnue et on veut estimer m. Montrer que maximiser la vraisemblance du modèle en $x = {}^t(x_1,\dots,x_n) \in \mathbf{R}^n$ revient à minimiser ${}^t(x - m\,\mathbb{I}_n)\,\Gamma_n^{-1}(x - m\,\mathbb{I}_n)$, où $\mathbb{I}_n = {}^t(1,1,\dots,1)$ **(1pt)**. En déduire que l'estimateur du maximum de vraisemblance $\widehat{m}_n$ de m est unique et:
$$\widehat{m}_n = ({}^t\mathbb{I}_n\,\Gamma_n^{-1}\mathbb{I}_n)^{-1}\,{}^t\mathbb{I}_n\,\Gamma_n^{-1}\,{}^t(X_1,\dots,X_n) \quad \textbf{(2pts)}.$$
7. Montrer que $\widehat{m}_n$ est un estimateur sans biais **(0.5pts)** et écrire sous forme matricielle la variance de $\widehat{m}_n$ **(1pt)**.
8. Ecrire également $\mathrm{var}(\overline{X}_n)$ sous forme matricielle en utilisant Γ_n et $\mathbb{I}_n$ **(1pt)**. Montrer que pour tout vecteur $U \in \mathbf{R}^n$, $({}^tU\,U)^2 \le ({}^tU\,\Gamma_n\,U)\,({}^tU\,\Gamma_n^{-1}\,U)$ **(1.5pts)**. En déduire que $\mathrm{var}(\widehat{m}_n) \le \mathrm{var}(\overline{X}_n)$ **(1pt)**.