Troisième Année Licence M.I.A.S.H.S. 2023 – 2024

Statistique 2

Contrôle Continu 2, Avril 2024

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

Exercice 1 (Sur 12 points)

Dans la suite, pour un vecteur $U = (U_1, \dots, U_n) \in \mathbf{R}^n$ on note $\|U\|^2 = \sum_{k=1}^n U_i^2$.

1. Soit $Z_1 \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$ et soit g une fonction de classe $\mathcal{C}^1$ sur $\mathbf{R}$ et $C > 0$ tel que $|x\,g(x)| + |g'(x)| \leq C\,(1+x^2)^{-1}\,e^{x^2/2}$ pour tout $x \in \mathbf{R}$. Montrer que $\mathbb{E}[|g'(Z_1)| + |Z_1\,g(Z_1)|] < \infty$ **(1pt)**, puis que $\mathbb{E}[g'(Z_1)] = \mathbb{E}[Z_1\,g(Z_1)]$ **(1.5pts)**. En déduire que $\mathbb{E}[Z_1^4] = 3$ **(1pt)**.

2. Soit $X_1 \overset{\mathcal{L}}{\sim} \mathcal{N}(\mu_1, \sigma^2)$ avec $\mu_1 \in \mathbf{R}$ et $\sigma^2 > 0$. Démontrer que $\mathbb{E}[g'(X_1)] = \dfrac{1}{\sigma^2}\,\mathbb{E}[(X_1 - \mu_1)\,g(X_1)]$ **(1.5pts)**.

3. Soit $(Z_k)_{k\in\mathbf{N}^*}$ une suite de v.a.i.i.d. de même loi $\mathcal{N}(0,1)$. Déterminer la loi de $Z = (Z_1, \dots, Z_n)$ où $n \geq 1$ **(0.5pts)**. Si $h : \mathbf{R}^n \to \mathbf{R}$ est une fonction de classe $\mathcal{C}^1$ sur $\mathbf{R}^n$ telle que $\mathbb{E}[\|\nabla h(Z)\| + \|Z\,h(Z)\|] < \infty$ avec $\nabla h(x) = \Big(\dfrac{\partial}{\partial x_i} h(x)\Big)_{1\leq i\leq n}$, démontrer que $\mathbb{E}[\nabla h(Z)] = \mathbb{E}[Z\,h(Z)]$ **(3pts)**.

4. Soit $(X_k)_{k\in\mathbf{N}^*}$ une suite de v.a. indépendantes, telle que $X_k \overset{\mathcal{L}}{\sim} \mathcal{N}(\mu_k, \sigma^2)$ avec $\mu_k \in \mathbf{R}$ pour tout $k \in \mathbf{N}^*$. Déterminer la loi de $X = (X_1, \dots, X_n)$ où $n \geq 1$ **(0.5pts)**. Soit $h : \mathbf{R}^n \to \mathbf{R}$ une fonction de classe $\mathcal{C}^1$ sur $\mathbf{R}^n$ telle que $\mathbb{E}[\|\nabla h(X)\| + \|X\,h(X)\|] < \infty$. Démontrer que $\sigma^2\,\mathbb{E}[\nabla h(X)] = \mathbb{E}[(X-\mu)\,h(X)]$ en spécifiant μ **(1.5pts)**. En déduire que si $h : (x_1, \dots, x_n) \in \mathbf{R}^n \mapsto (h_1(x_1,\dots,x_n), \dots, h_n(x_1,\dots,x_n))$ une fonction de classe $\mathcal{C}^1$ sur $\mathbf{R}^n$ telle que $\mathbb{E}[\|\nabla h_i(X)\|] < \infty$ pour tout i, alors
$$\frac{1}{\sigma^2}\sum_{i=1}^n \mathrm{cov}(X_i, h_i(X)) = \sum_{i=1}^n \mathbb{E}\Big[\frac{\partial}{\partial x_i} h_i(X)\Big] \quad \textbf{(1.5pts)}. \tag{1}$$

Exercice 2 (Sur 13 points)

On considère $(X_k)_{k\in\mathbf{N}^*}$ une suite de v.a. indépendantes, telle que $X_k \overset{\mathcal{L}}{\sim} \mathcal{N}(m_k, 1)$ avec $m_k \in \mathbf{R}$ pour tout $k \in \mathbf{N}^*$. On suppose que $(X_1, \dots, X_n)$ a été observé et $m = (m_1, \dots, m_n)$ est un vecteur inconnu (avec $n \geq 2$).

1. Déterminer la vraisemblance du modèle statistique, après avoir précisé ce dernier **(1pt)**.

2. Prouver qu'il existe un unique estimateur $\widehat{m}$ de m par maximum de vraisemblance et $\widehat{m} = (X_1, \dots, X_n)$ **(1pt)**.

3. Déterminer le biais de $\widehat{m}$ **(0.5pts)** et montrer que son risque quadratique est $R(\widehat{m}) = n$ **(1pt)**.

4. Soit maintenant l'estimateur $\widetilde{m} = (\widetilde{m}_1, \dots, \widetilde{m}_n)$ tel que
$$\widetilde{m}_k = \widetilde{m}_k(X_1, \dots, X_n) = X_k\,\Big(1 - \frac{n-2}{\sum_{i=1}^n X_i^2}\Big) \quad \text{pour tout } k \in \{1, \dots, n\}.$$
Déterminer la limite en loi de $\widetilde{m}_k$ quand $n \to \infty$ **(2pts)**.

5. Démontrer que $R(\widetilde{m}) = -n + \mathbb{E}[\|\widetilde{m} - X\|^2] + 2\sum_{i=1}^n \mathrm{cov}(X_i, \widetilde{m}_i)$ **(2.5pts)**. En utilisant (1), en déduire que
$$R(\widetilde{m}) = -n + \mathbb{E}[\|\widetilde{m} - X\|^2] + 2\sum_{i=1}^n \mathbb{E}\Big[\frac{\partial}{\partial x_i}\widetilde{m}_i(X)\Big] \quad \textbf{(1pt)}.$$

6. Montrer que $\mathbb{E}[\|\widetilde{m} - X\|^2] = \mathbb{E}\Big[\dfrac{(n-2)^2}{\sum_{i=1}^n X_i^2}\Big]$ **(1pt)** et $\displaystyle\sum_{i=1}^n \frac{\partial}{\partial x_i}\widetilde{m}_i(X) = n - \frac{(n-2)^2}{\sum_{i=1}^n X_i^2}$ **(1.5pts)**. En déduire que
$$R(\widetilde{m}) = n - (n-2)^2\,\mathbb{E}\Big[\frac{1}{\sum_{i=1}^n X_i^2}\Big] \quad \textbf{(1pt)}.$$
Comment expliquer que $R(\widetilde{m}) < R(\widehat{m})$ pour $n \geq 3$ **(0.5pts)**?