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

Statistique 2

Examen terminal, Mai 2024

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

Exercice 1 (Sur 12 points)

Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire à valeurs dans $\{e_1, \ldots, e_d\}$, où $(e_1, \ldots, e_d)$ est la base canonique classique de $\mathbf{R}^d$ et tel que $\mathbb{P}(X = e_i) = p_i$ pour tout $1 \leq i \leq d$, où les $p_i \in [0, 1]$ sont tels que $p_1 + \cdots + p_d = 1$.

Notation: pour un vecteur colonne $U = {}^t(U_1, \ldots, U_d) \in \mathbf{R}^d$, on note $\|U\|^2 = {}^tU\, U = \sum_{k=1}^d U_d^2$, où ${}^tU$ est le vecteur transposé de U.

1. Déterminer la loi de X_i pour $1 \leq i \leq d$ **(0.5pts)**.
2. Démontrer que pour $i \neq j$, X_i et X_j ne sont pas indépendantes **(1pt)**.
3. Soit $r \in \mathbf{R}^d$ un vecteur colonne tel que $\|r\| > 0$. Montrer que P_r la matrice dans $(e_1, \ldots, e_d)$ de la projection orthogonale sur la droite vectorielle engendrée par r est telle que $P_r = \frac{1}{\|r\|^2}\, r\, {}^t r$ **(1pt)**. En notant $\{r\}^\perp$ le sous-espace vectoriel orthogonal de $\{r\}$, déduire la matrice $P_{\{r\}^\perp}$ dans $(e_1, \ldots, e_d)$ de la projection orthogonale sur $\{r\}^\perp$ **(1pt)**.
4. On considère $Y = (Y_1, \ldots, Y_d)$ telle que $Y_i = \dfrac{X_i - p_i}{\sqrt{p_i}}$ pour $1 \leq i \leq d$. Démontrer que Y est un vecteur centré de matrice de covariance $P_{\{r\}^\perp}$, où $r = {}^t(\sqrt{p_1}, \ldots, \sqrt{p_d})$ **(1.5pts)**.
5. Soit $(Y^{(k)})_{k \in \mathbf{N}}$ une suite de vecteur aléatoire indépendants et identiquement distribués de même loi que Y. Soit $u = {}^t(u_1, \ldots, u_d) \in \mathbf{R}^d$ et soit $Z_k = {}^tu\, Y^{(k)}$ pour $k \in \mathbf{N}$. Montrer que si u n'est pas colinéaire à r:
$$\sqrt{n}\, \overline{Z}_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0\,,\, {}^tu\, P_{\{r\}^\perp} u) \quad \text{où} \quad \overline{Z}_n = \frac{1}{n}\sum_{j=1}^n Z_j \quad \textbf{(1pt)}.$$
En déduire que $\frac{1}{\sqrt{n}} \sum_{k=1}^n Y^{(k)} \xrightarrow[n\to\infty]{\mathcal{L}} S$, où S est un vecteur gaussien à valeurs dans $\mathbf{R}^d$ centré de matrice de covariance $P_{\{r\}^\perp}$ **(1pt)**.
6. Démontrer que $\|S\|^2 \overset{\mathcal{L}}{\sim} \chi^2(d-1)$ **(2pts)**.
7. Soit $(W_k)_{k \in \mathbf{N}}$ une suite de v.a.i.i.d. telle que W_1 prend d valeurs possibles $\{w_1, \ldots, w_d\}$ et vérifie $\mathbb{P}(W_1 = w_i) = p_i$ pour $i \in \{1, \ldots, d\}$. En notant $\widehat{p}_i = \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{W_k = w_i}$ pour $i \in \{1, \ldots, d\}$, en déduire
$$n \sum_{i=1}^d \frac{(\widehat{p}_i - p_i)^2}{p_i} \xrightarrow[n\to\infty]{\mathcal{L}} \chi^2(d-1) \quad \textbf{(3pts)}.$$

Exercice 2 (Sur 20 points)

Soit $\theta = (m, \lambda) \in \mathbf{R} \times]0, \infty[$. Considérons la fonction de répartition définie par:
$$F_\theta(x) = 1 - \exp\big(-\lambda(x - m)\big) \quad \text{pour } x \geq m.$$

1. Que vaut $F_\theta(x)$ pour $x < m$ (justifier) **(0.5pts)**?
2. Montrer que F_θ est la fonction de répartition d'une v.a. "continue" dont on précisera la densité f_θ **(1pt)**.
3. Soit X une v.a. "continue" de densité f_θ. Quelle est la loi de $X - m$ (justifier) **(1pt)**?

4. Calculer $\mathbb{E}[X]$ et $\mathrm{var}(X)$ **(1pt)**.

5. On suppose que m est inconnu et on considère $(X_i)_{i\in\mathbf{N}}$ des v.a.i.i.d. de même loi que X. On suppose $(X_1,\dots,X_n)$ observé. Quand λ est un réel donné connu, préciser le modèle statistique **(0.5pts)** et démontrer que l'estimateur du maximum de vraisemblance de m est $\widehat{m}_n=\min(X_1,\dots,X_n)$ **(2pts)**.

6. Déterminer la fonction de répartition de $\widehat{m}_n$ **(1.5pts)**. Est-ce une v.a. "continue" (justifier) **(0.5pts)**?

7. Prouver que $\widehat{m}_n$ est un estimateur biaisé de m **(1pt)**.

8. Démontrer que $(\widehat{m}_n)_{n\in\mathbf{N}}$ converge en probabilité **(1pt)**.

9. Soit $Z_n=n\,(\widehat{m}_n-m)$. Montrer que Z_n suit une loi de probabilité bien connue **(1pt)**. En déduire un intervalle de confiance de niveau 95% sur m dépendant de λ **(1.5pts)**.

10. On considère maintenant que λ est également inconnu et on définit $\widehat{\lambda}_n$ estimateur de λ tel que:

$$\widehat{\lambda}_n=\frac{n}{\sum_{i=1}^n(X_i-\widehat{m}_n)}.$$

Démontrer un théorème de la limite centrale satisfait par les $(X_i-m)_i$ **(1pt)**. Montrer que $\mathbb{E}[\sqrt{n}\,|\widehat{m}_n-m|]\underset{n\to\infty}{\longrightarrow}0$ **(0.5pts)**, et en déduire que pour tout $\lambda>0$:

$$\sqrt{n}\,(\widehat{\lambda}_n-\lambda)\ \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}\ \mathcal{N}(0\,,\,\lambda^2)\quad \textbf{(3pts)}.$$

11. En utilisant le quantile q d'ordre 97.5% d'une loi $\mathcal{N}(0,1)$, déterminer un intervalle de confiance asymptotique de niveau 95% pour λ **(1.5pts)**. En déduire également un intervalle de confiance asymptotique de niveau 95% pour m s'écrivant sans que λ soit connu **(1.5pts)**.