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 $Z_k = {}^t u\, Y^{(k)}$ pour $k \in \mathbf{N}$. Montrer que si u n'est pas colinéaire à r:
$$\sqrt{n}\,\overline{Z}_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0\,,\,{}^t u\, P_{\{r\}^\perp} u) \quad \text{où} \quad \overline{Z}_n = \frac{1}{n}\sum_{j=1}^n Z_j \quad \textbf{(1pt)}.$$
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
$$n \sum_{i=1}^d \frac{(\widehat{p}_i - p_i)^2}{p_i} \xrightarrow[n\to\infty]{\mathcal{L}} \chi^2(d-1) \quad \textbf{(3pts)}.$$

Proof.

1. X_i prend pour valeurs 0 et 1, elle suit donc une loi de Bernoulli $\mathcal{B}(p_i)$ puisque $\mathbb{P}(X_i = 1) = p_i$.
2. Pour $i \neq j$, $\operatorname{cov}(X_i, X_j) = \mathbb{E}[X_i\, X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j] = 0 - p_i\, p_j = -p_i\, p_j$ car si $X_i = 1$ alors $X_j = 0$ et réciproquement. Donc $\operatorname{cov}(X_i, X_j) \neq 0$: les variables (X_i) ne sont pas indépendantes.
3. Une base orthonormale de la droite vectorielle engendrée par r est $(r/\|r\|)$. Or pour une projection orthogonale sur un sev F de base orthonormale $(f_1, \ldots, fd)$ alors pour tout $x \in \mathbf{R}^d$, $P_F(x) = \sum_{i=1}^p < x, f_i > f_i$, où $< x, f_i >$ est le produit scalaire de x et f_i. Par conséquent, ici on a $P_{[r]}(x) = < x, r > r/\|r\|^2 = \|r\|^{-2} r\, {}^t r\, x$. D'où le résultat.
On sait que $P_{[r]}(x) + P_{\{r\}^\perp}(x) = x$ pour tout $x \in \mathbf{R}^d$, d'où $P_{\{r\}^\perp} = I_d - \|r\|^{-2} r\, {}^t r$.
4. D'après ce qui précède $\mathbb{E}[Y_i] = (\mathbb{E}[X_i] - p_i)/\sqrt{p_i} = 0$ pour tout i, le vecteur Y est bien centré.
De plus, $\operatorname{var}(Y_i) = \frac{1}{p_i}\operatorname{var}(X_i) = 1 - p_i$, et pour $i \neq j$, $\operatorname{cov}(Y_i, Y_j) = \frac{1}{\sqrt{p_i\, p_j}}\operatorname{cov}(X_i, X_j) = -\sqrt{p_i\, p_j}$. Or avec $r = {}^t(\sqrt{p_1}, \ldots, \sqrt{p_d})$, on a $\|r\|^2 = 1$, donc $P_{\{r\}^\perp} = I_d - {}^t(\sqrt{p_1}, \ldots, \sqrt{p_d})\,(\sqrt{p_1}, \ldots, \sqrt{p_d}) = I_d - (\sqrt{p_i\, p_j})_{1 \leq i,j \leq d}$. En conséquence, le terme de cette matrice pour $i = j$ est $1 - \sqrt{p_i\, p_i} = 1 - p_i$ et pour $i \neq j$ il vaut $-\sqrt{p_i\, p_j}$: c'est bien la matrice de covariance de Y.

5. Il est clair que comme $(Y^{(k)})_k$ est une suite de vecteurs aléatoires i.i.d., alors $(Z_k)_k$ est une suite de v.a.i.i.d. De plus $\mathbb{E}[Z_1] = {}^t u\, \mathbb{E}[Y_1] = 0$ et $\mathrm{var}(Z_1) = {}^t u \,\mathrm{cov}(Y_1)\, u = {}^t u\, P_{\{r\}^\perp}\, u < \infty$. On peut donc appliquer le TLC et on obtient le résultat demandé.
D'après le cours, on sait que la fonction caractéristique de $\frac{1}{\sqrt{n}} \sum_{k=1}^n Y^{(k)}$ est donnée pour tout u comme $\mathbb{E}\Big[\exp\Big(i \frac{1}{\sqrt{n}} \sum_{k=1}^n {}^t u\, Y^{(k)}\Big]= \mathbb{E}\big[\exp\big(i\sqrt{n}\,\overline{Z}_n\big)\big]$. Comme cette fonction caractéristique converge vers celle de $\mathcal{N}\big(0\,,\,{}^t u\, P_{\{r\}^\perp} u\big)$ pour presque tout u, on en déduit que $\frac{1}{\sqrt{n}} \sum_{k=1}^n Y^{(k)}$ converge en loi vers $\mathcal{N}_d\big(0\,,\,P_{\{r\}^\perp}\big)$.

6. On utilise le Théorème de Cochran car $S = P_{\{r\}^\perp} V$ avec $V \overset{\mathcal{L}}{\sim} \mathcal{N}_d(0\,,\,I_d)$ puisque $P_{\{r\}^\perp}{}^t P_{\{r\}^\perp} = P_{\{r\}^\perp} P_{\{r\}^\perp} = P_{\{r\}^\perp}$. Or $\dim(P_{\{r\}^\perp}) = \dim(\mathbf{R}^d) - \dim([r]) = d-1$, donc $S \overset{\mathcal{L}}{\sim} \chi^2(\dim(P_{\{r\}^\perp})) = \chi^2(d-1)$.

7. Pour $k \in \mathbf{N}$, $X_i^{(k)} = \mathrm{I\!I}_{W_k = i}$ pour $1 \leq i \leq d$. Alors on a bien $X^{(k)}$ qui a la même loi que X. Donc en considérant $Y^{(k)}$ tel que $Y_i^{(k)} = (\mathrm{I\!I}_{W_k=i} - p_i)/\sqrt{p_i}$, on peut appliquer le résultat du 5. Mais $\frac{1}{\sqrt{n}} \sum_{k=1}^n Y^{(k)} = \Big(\sqrt{n}\Big(\frac{\widehat{p_i} - p_i}{\sqrt{p_i}}\Big)\Big)_{1 \leq i \leq d}$. On a donc:
$$\Big(\sqrt{n}\Big(\frac{\widehat{p_i} - p_i}{\sqrt{p_i}}\Big)\Big)_{1 \leq i \leq d} \xrightarrow[n\to\infty]{\mathcal{L}} S.$$
Enfin, la fonction $x \in \mathbf{R}^d \mapsto \|x\|^2$ est une fonction continue, on peut donc l'appliquer à la convergence en loi précédente, ce qui donne le résultat demandé.

□

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 $(X_1, \ldots, X_n)$ 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 $\widehat{m}_n = \min(X_1, \ldots, X_n)$ **(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 $\mathbb{E}[\sqrt{n}\,|\widehat{m}_n - m|] \xrightarrow[n\to\infty]{} 0$ **(0.5pts)**, et en déduire que pour tout $\lambda > 0$:
$$\sqrt{n}\,(\widehat{\lambda}_n - \lambda) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0\,,\,\lambda^2) \quad \textbf{(3pts)}.$$
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)**.

Proof. 1. Comme F_θ est positive et croissante, comme $F_\theta(m) = 0$, alors pour $x < m$, $F_\theta(x) = 0$.

2. Il est clair que pour $x > m$, F_θ est de classe $\mathcal{C}^\infty$, on vient de voir que pour $x < m$, elle l'est également car nulle. Et en $x = m$, F_θ est clairement continue. Donc F_θ est continue sur $\mathbf{R}$, et $\mathcal{C}^1$ par morceaux, donc c'est la fonction de répartition d'une variable "continue" et sa densité est:
$$f_\theta(x) = \lambda \exp\big(-\lambda(x-m)\big)\, \mathrm{I\!I}_{x \geq m}.$$

3. Soit $Y = X - m$. Alors Y prend ses valeurs dans $\mathbf{R}^+$. De plus pour $y \geq 0$, $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq y+m) = F_\theta(y+m) = 1 - \exp(-\lambda y)$: il s'agit de la fonction de répartition d'une loi exponentielle de paramètre λ.
4. On sait que $\mathbb{E}[Y] = 1/\lambda$ et $\mathrm{var}(Y) = 1/\lambda^2$, donc $\mathbb{E}[X] = m + 1/\lambda$ et $\mathrm{var}(X) = 1/\lambda^2$.
5. Comme on suppose que λ est connu, le modèle statistique est paramètrique et s'écrit $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \mathbb{P}_m^{\otimes n}, m \in \mathbf{R})$, avec $\mathbb{P}_m$ la mesure de probabilité de X.
La vraisemblance va s'écrire en fonction de m. Pour $(x_1, \ldots, x_n) \in \mathbf{R}$, grâce à l'indépendance et l'identique distribution, on a:
$$L_m(x_1, \ldots, x_m) = \prod_{j=1}^n \lambda \exp\big(-\lambda(x_j - m)\big) \, \mathbb{I}_{x_j \geq m} = \lambda^n \exp\Big(-\lambda \sum_{j=1}^n (x_j - m)\Big) \, \mathbb{I}_{\min(x_1, \ldots, x_n) \geq m}.$$
Donc en tant que fonction de m, $L_m(x_1, \ldots, x_m) = 0$ si $m > \min(x_1, \ldots, x_n)$ et $L_m(x_1, \ldots, x_m) = \lambda^n \exp\Big(n\lambda m - \lambda \sum_{j=1}^n x_j\Big)$ si $m \leq \min(x_1, \ldots, x_n)$, qui est clairement une fonction strictement croissante en m. Donc le maximum de $L_m(X_1, \ldots, X_m)$ est atteint en $\widehat{m}_n = \min(X_1, \ldots, X_n)$.
6. La variable $\widehat{m}_n$ prend ses valeurs dans $[m, \infty[$. Donc pour $x < m$, alors $F_{\widehat{m}_n}(x) = 0$. Pour $x \geq m$, on a:
$$F_{\widehat{m}_n}(x) = 1 - \mathbb{P}(\widehat{m}_n > x) = 1 - \mathbb{P}(X_1 > x \cap X_2 > x \cap \ldots \cap X_n > x) = 1 - \big(\exp\big(-\lambda\,(x - m)\big)\big)^n = 1 - \exp\big(-n\,\lambda\,(x-m)\big),$$
en utilisant le fait que les (X_i) sont des v.a.i.i.d. On obtient ainsi une fonction dérivable sur $]m, \infty[$, nulle sur $]-\infty, m[$ et continue en m (on obtient 0 dans les 2 cas). Il s'agit donc de la fonction de répartition d'une variable "continue".
7. Il est clair que la fonction de répartition de $\widehat{m}_n$ est celle de X en lorsque λ est remplacé par $n\,\lambda$. Comme on a obtenu $\mathbb{E}[X]$ précédemment, on peut directement écrire que $\mathbb{E}[\widehat{m}_n] = m + \frac{1}{n\,\lambda}$: l'estimateur est biaisé car $\mathbb{E}[\widehat{m}_n] \neq m$ (mais asymptotiquement non biaisé).
8. Grâce à $\mathrm{var}(X)$ calculée précédemment, on déduit sans calcul que $\mathrm{var}(\widehat{m}_n) = \frac{1}{n^2\,\lambda^2}$. Par conséquent $\mathrm{var}(\widehat{m}_n) \underset{n\to\infty}{\longrightarrow} 0$ et comme $\widehat{m}_n$ est asymptotiquement non biaisé, on en déduit que $\widehat{m}_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} m$: l'estimateur est convergent.
9. Il est clair que $Z_n = n\,(\widehat{m}_n - m)$ prend ses valeurs dans $[0, +\infty[$. Pour $z \geq 0$, $F_{Z_n}(z) = \mathbb{P}(n\,(\widehat{m}_n - m) \leq z) = \mathbb{P}(\widehat{m}_n \leq m + z/n) = 1 - \exp\big(-\lambda\,z\big)$: il s'agit de la fonction de répartition d'une loi exponentielle de paramètre λ.
On sait que $\widehat{m}_n \geq m$ et donc $\mathbb{P}(m \leq \widehat{m}_n \leq m + z/n) = 1 - \exp\big(-\lambda\,z\big)$. Donc si on choisit que cette probabilité vaille 0.95, on obtient l'intervalle de confiance demandé. Pour cela, il suffit de choisir z tel que $1 - \exp\big(-\lambda\,z\big) = 0.95$, soit $-\lambda\,z = \ln(0.05)$, d'où $z = \ln(20)/\lambda$. Par conséquent un intervalle de confiance de niveau 95% sur m est:
$$\Big[\widehat{m}_n - \frac{\ln(20)}{\lambda\,n}\,,\,\widehat{m}_n\Big].$$
10. Les $(X_i - m)_i$ sont des v.a.i.i.d. de même loi $\mathcal{E}(\lambda)$, donc d'espérance $1/\lambda$ et de variance finie $1/\lambda^2$, et ainsi on peut appliquer le TLC:
$$\sqrt{n}\,\frac{\frac{1}{n}\sum_{i=1}^n (X_i - m) - \frac{1}{\lambda}}{\frac{1}{\lambda}} = \sqrt{n}\,\Big(\frac{\lambda}{n}\sum_{i=1}^n (X_i - m) - 1\Big) \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0,1).$$
On sait que $n\,(\widehat{m}_n - m) \overset{\mathcal{L}}{\sim} \mathcal{E}(\lambda)$, donc $\mathbb{E}\big[n\,(\widehat{m}_n - m)\big] = \sqrt{n}\,\mathbb{E}\big[\sqrt{n}\,|\widehat{m}_n - m|\big] = 1/\lambda$. On en déduit donc que $\mathbb{E}\big[\sqrt{n}\,|\widehat{m}_n - m|\big] = 1/(\lambda\sqrt{n}) \underset{n\to\infty}{\longrightarrow} 0$.
On peut réecrire le TLC précédent sous la forme:
$$\sqrt{n}\,\Big(\frac{\lambda}{n}\sum_{i=1}^n (X_i - \widehat{m}_n) - 1\Big) + \lambda\,\sqrt{n}\,(\widehat{m}_n - m) \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0,1).$$
Or d'après ce qui précède, l'inégalité de Markov permet de montrer que $\sqrt{n}\,(\widehat{m}_n - m) \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} 0$. Donc en utilisant le Lemme de Slutsky, on obtient:
$$\sqrt{n}\,\Big(\frac{1}{n}\sum_{i=1}^n (X_i - \widehat{m}_n) - \frac{1}{\lambda}\Big) \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\Big(0\,,\,\frac{1}{\lambda}^2\Big).$$
On peut maintenant appliquer la Delta-méthode avec $g(x) = 1/x$ qui est bien de classe $\mathcal{C}^1$ avec $g'(1/\lambda) = \lambda^2$, d'où le TLC demandé.
11. D'après le TLC précédent, on a $\widehat{\lambda}_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \lambda$. On peut donc utiliser le Lemme de Slutsky et obtenir le TLC 2 suivant:
$$\sqrt{n}\,\frac{\big(\widehat{\lambda}_n - \lambda\big)}{\widehat{\lambda}_n} \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\big(0\,,\,1\big).$$
On en déduit l'intervalle de confiance asymptotique à 95% suivant:
$$\Big[\widehat{\lambda}_n - q\,\frac{\widehat{\lambda}_n}{\sqrt{n}}\,,\,\widehat{\lambda}_n + q\,\frac{\widehat{\lambda}_n}{\sqrt{n}}\Big].$$
Par ailleurs, on peut également utiliser la convergence en loi précédente $n\,(\widehat{m}_n - m) \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{E}(\lambda)$, ce qui revient à $n\,\lambda\,\big(\widehat{m}_n - m\big) \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{E}(1)$. Comme $\widehat{\lambda}_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \lambda$, on peut encore utiliser le Lemme de Slutsky et obtenir:
$$n\,\widehat{\lambda}_n\,\big(\widehat{m}_n - m\big) \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{E}(1).$$
D'où l'intervalle de confiance asymptotique à 95% suivant:
$$\Big[\widehat{m}_n - \frac{\ln(20)}{n\,\widehat{\lambda}_n}\,,\,\widehat{m}_n\Big].$$

$\square$