Faculté des Sciences
Aix-Marseille Université

Année universitaire 2022-2023

Site : ☐ Luminy ☒ St-Charles ☐ St-Jérôme ☐ Cht-Gombert ☐ Aix-Montperrin ☐ Aubagne-SATIS
Sujet Session de ☒ 1er semestre - ☐ 2ème semestre - ☒ Session unique Durée de l'épreuve : 3h
Examen de ☐ L1/☐ L2/☐ L3 - ☒ M1/☐ M2 - ☐ LP - ☐ DU Nom Diplôme : MAS
Code Apogée du module : SMSAU02C Libellé du module : **Statistique.**
Document autorisé : ☒ OUI-☐ NON Calculette autorisée ☐ OUI-☒ NON

Correction de l'examen à mi-parcours du Lundi 31 Octobre 2022.

Les documents, notes de cours et calculatrices sont interdits, à l'exception d'une feuille A4 recto-verso. Le soin apporté à la rédaction sera un élément d'appréciation. Le barême est donné à titre purement indicatif.

Exercice 1 (5 points). Soient X et Y des variables aléatoires indépendantes de loi gaussienne centrée réduite, $X \sim Y \sim \mathcal{N}(0,1)$.
Le terme "indépendantes" manquait dans l'énoncé.

1. Montrez que la loi du couple $(X, Z) = (X, X^2 + Y^2)$ a pour densité

$$f_{(X,Z)}(x, z) = \frac{1}{2\pi} e^{-\frac{z}{2}} \frac{1}{\sqrt{z - x^2}} 1_{z \geq x^2}.$$

Soit h une fonction mesurable bornée de $\mathbb{R}^2$ dans $\mathbb{R}$

$$\begin{aligned}\mathbb{E}(h(X,Z)) &= \iint h(x, x^2+y^2) \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) dx\, dy \\ &= 2\int_0^{+\infty} dy \int_{-\infty}^{+\infty} dx\ h(x, x^2+y^2)\frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)\end{aligned}$$

On fait le changement de variable $(x, y) \in \mathbb{R} \times \mathbb{R}^+ \mapsto (u, v) = (x, x^2 + y^2) \in D$. On a $x = u, y = \sqrt{v - u^2}$. Le déterminant de la matrice Jacobienne vaut en valeur absolue $1/2\sqrt{v - u^2}$, et le domaine $D = \{(u, v) \in \mathbb{R} \times \mathbb{R}^+; u^2 \leq v\}$. On obtient donc :

$$\mathbb{E}(h(X,Z)) = \iint_D h(u,v) \frac{1}{2\pi} \exp\left(-\frac{v}{2}\right) \frac{1}{\sqrt{v-u^2}} du\, dv$$

et la densité du couple (X, Z) est la fonction $\frac{e^{-z/2}}{2\pi\sqrt{z-x^2}}\ \mathbb{1}_D(x, z)$

2. Quelle est la loi de $X^2 + Y^2$,
Par définition de la loi du χ^2_d, $X^2 + Y^2 \sim \chi^2_2$.

3. Montrer que la loi conditionnelle de $X \,|\, X^2+Y^2=z$ a pour densité
$$f_{X\,|\,X^2+Y^2=z}(x) = \frac{1}{\pi}\frac{1}{\sqrt{z-x^2}}1_{x^2\leq z}.$$

La densité conditionnelle de $X \,|\, X^2+Y^2=z$ est la fonction $x \in [-\sqrt{z}, +\sqrt{z}] \mapsto \frac{1}{C(z)\sqrt{z-x^2}}$, où la constante de normalisation $C(z)$ vaut
$$C(z) = \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{z-x^2}}\,dx = \int_{-1}^{1} \frac{1}{\sqrt{1-u^2}}\,du = \int_{-\pi/2}^{\pi/2} d\theta = \pi\,,$$
où le deuxième égalité vient du changement de variable $u = x/\sqrt{z}$, et la troisième du changement de variable $u = \sin(\theta)$.

4. En déduire que
$$\mathbb{E}[\,|X|\,|\,X^2+Y^2] = 2\frac{\sqrt{X^2+Y^2}}{\pi}$$

On a donc
$$\mathbb{E}[\,|X|\,|\,X^2+Y^2=z] = \int_{-\sqrt{z}}^{\sqrt{z}} \frac{|x|}{\pi\sqrt{z-x^2}}dx = 2\sqrt{z}\int_0^1 \frac{u}{\pi\sqrt{1-u^2}}du = 2\sqrt{z}\int_0^{\pi/2}\sin(\theta)\frac{d\theta}{\pi} = \frac{2}{\pi}z\,,$$
en effectuant les mêmes changements de variable qu'en question 3. On a donc
$$\mathbb{E}[\,|X|\,|\,X^2+Y^2] = \frac{2(X^2+Y^2)}{\pi}\,.$$

Exercice 2 (8 points). On considère un échantillon indépendant et identiquement distribué (i.i.d.) $(X_i)_{1\leq i\leq n}$, avec X_1 de densité
$$f_\theta(x) = \frac{2x}{\theta^2}\,1_{[0,\theta]}(x), \quad \text{avec } \theta>0 \text{ un paramètre inconnu.}$$

1. On note $X_{(n)} = \max_{1\leq i\leq n} X_i$.
   (a) Calculer la fonction de répartition de $X_{(n)}$. Quelle est la densité de $X_{(n)}$?
   Les $(X_i)_{1\leq i\leq n}$ étant i.i.d., pour tout $t\in\mathbb{R}$, $\mathbb{P}_\theta(X_{(n)}\leq t) = \mathbb{P}_\theta(X_1\leq t)^n$, avec
$$\mathbb{P}_\theta(X_1\leq t) = \begin{cases} 0 & \text{si } t\leq 0\,,\\ 1 & \text{si } t\geq\theta\,,\\ \int_0^t \frac{2x}{\theta^2}dx = \frac{t^2}{\theta^2} & \text{si } 0\leq t\leq\theta. \end{cases}$$
   On a donc
$$\mathbb{P}_\theta(X_{(n)}\leq t) = \begin{cases} 0 & \text{si } t\leq 0\,,\\ 1 & \text{si } t\geq\theta\,,\\ \frac{t^{2n}}{\theta^{2n}} & \text{si } 0\leq t\leq\theta. \end{cases}$$
   et la densité de $X_{(n)}$ vaut donc $2n\frac{t^{2n-1}}{\theta^{2n}}\,\mathbb{1}_{[0;\theta]}(t)$.

(b) Donner les deux premiers moments de $X_{(n)}$. En déduire sa variance.

On a

$$\mathbb{E}_\theta(X_{(n)}) = \int_0^\theta 2n\frac{t^{2n}}{\theta^{2n}}dt = \frac{2n}{2n+1}\theta.$$

$$\mathbb{E}_\theta(X_{(n)}^2) = \int_0^\theta 2n\frac{t^{2n+1}}{\theta^{2n}}dt = \frac{n}{n+1}\theta^2.$$

$$\mathrm{var}_\theta(X_{(n)}) = \left[\frac{n}{n+1} - \left(\frac{2n}{2n+1}\right)^2\right]\theta^2 = \frac{n}{n+1}\frac{1}{(2n+1)^2}\theta^2.$$

(c) Quel est le risque quadratique de $X_{(n)}$ comme estimateur de θ ?

$$R_{X_{(n)}}(\theta) = \mathbb{E}_\theta\left[(X_{(n)}-\theta)^2\right] = \text{biais}^2 + \text{ variance}$$
$$= \left(\frac{2n}{2n+1}-1\right)^2\theta^2 + \frac{n}{n+1}\frac{1}{(2n+1)^2}\theta^2 = \frac{\theta^2}{(2n+1)^2}.$$

(d) Montrer que $X_{(n)}$ converge en probabilité vers θ.

Pour tout $\epsilon > 0$, l'inégalité de Markov donne

$$\mathbb{P}_\theta(\left|X_{(n)}-\theta\right| \geq \epsilon) \leq \frac{1}{\epsilon^2}\mathbb{E}_\theta\left[(X_{(n)}-\theta)^2\right] = \frac{1}{\epsilon^2}\frac{\theta^2}{(2n+1)^2}.$$

On en déduit que $X_{(n)}$ converge en probabilité vers θ.

2. On pose maintenant $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

(a) Montrer que $\frac{3}{2}\bar{X}_n$ est un estimateur sans biais et consistant de θ.

$\mathbb{E}_\theta(X_1) = \mathbb{E}_\theta(X_{(1)}) = 2\theta/3$. Par linéarité de l'espérance, $\mathbb{E}_\theta(\bar{X}_n) = \mathbb{E}_\theta(X_1) = 2\theta/3$. $3\bar{X}_n/2$ est donc un estimateur sans biais de θ.

Les $(X_i)_{1\leq i\leq n}$ étant i.i.d. et intégrable, $\bar{X}_n$ converge p.s. vers $\mathbb{E}_\theta(X_1)$ par la loi forte des grands nombres, et $3\bar{X}_n/2$ est donc un estimateur consistant de θ.

(b) Quel est le risque quadratique de $\frac{3}{2}\bar{X}_n$ comme estimateur de θ ?

$3\bar{X}_n/2$ étant sans biais, son risque est égal à sa variance. Les $(X_i)_{1\leq i\leq n}$ étant i.i.d. , on a alors

$$R_{3\bar{X}_n/2}(\theta) = \frac{9}{4n}\mathrm{var}_\theta(X_1) = \frac{9}{4n}\mathrm{var}_\theta(X_{(1)}) = \frac{9}{4n}\frac{\theta^2}{2(9)} = \frac{\theta^2}{8n}.$$

(c) Qui de $X_{(n)}$ ou $\frac{3}{2}\bar{X}_n$ choisiriez-vous pour estimer θ ?

On vérifie facilement que pour tout $n \geq 1$, $3n \leq (2n+1)^2$. On a donc pour tout $n \geq 1$ et tout $\theta \in \mathbb{R}$, $R_{3\bar{X}_n/2}(\theta) \geq R_{X_{(n}}(\theta)$. $X_{(n)}$ est un meilleur estimateur de θ que $3\bar{X}_n/2$.

3. On s'intéresse à une estimation ensembliste de θ en utilisant l'estimateur $X_{(n)}$.

(a) Montrer que $\frac{X_{(n)}}{\theta}$ est un pivot pour l'estimation de θ.

Pour tout $t \in \mathbb{R}$,

$$\mathbb{P}_\theta\left(\frac{X_{(n)}}{\theta} \leq t\right) = \mathbb{P}_\theta(X_{(n)} \leq \theta t) = \begin{cases} 0 & \text{si } t \leq 0\,, \\ 1 & \text{si } t \geq 1\,, \\ t^{2n} & \text{si } 0 \leq t \leq 1. \end{cases}$$

La loi de $X_{(n)}/\theta$ ne dépend donc pas de θ, i.e. $X_{(n)}/\theta$ est un pivot pour l'estimation de θ.

(b) Construire un intervalle de confiance pour θ de niveau $(1-\alpha)$, basé sur $X_{(n)}$.

Soit t tel que $\mathbb{P}_\theta\left(\frac{X_{(n)}}{\theta} \geq t\right) = 1-\alpha$. D'après ce qui précède, on a $t = \alpha^{1/2n}$, et pour tout θ,

$$\mathbb{P}_\theta(0 \leq \theta \leq \alpha^{-1/2n} X_{(n)}) = 1-\alpha.$$

Remarque : d'autres choix sont possibles, mais l'intervalle choisi ici est celui dont la longueur est minimale.

4. Nous cherchons maintenant à déterminer une estimation ensembliste de θ à partir de $\bar{X}_n$.

(a) Montrer que $\frac{\sqrt{8n}}{\sqrt{3}} \frac{\frac{3}{2}\bar{X}_n - \theta}{\theta}$ est un pivot asymptotique pour l'estimation de θ.

Une autre erreur dans l'énoncé... Le facteur $\sqrt{3}$ doit être supprimé.

Les $(X_i)_{1\leq i\leq n}$ étant i.i.d. et de carré intégrable, le théorème de la limite centrale affirme que $\frac{\frac{3}{2}\bar{X}_n - \theta}{\sqrt{\mathrm{var}(\frac{3}{2}\bar{X}_n)}}$ converge en loi vers une loi $\mathcal{N}(0,1)$. De plus, on a vu que $\mathrm{var}(\frac{3}{2}\bar{X}_n) = \theta^2/8n$. Par conséquent, $\sqrt{8n}\frac{\frac{3}{2}\bar{X}_n - \theta}{\theta}$ converge en loi vers une loi $\mathcal{N}(0,1)$ indépendante de θ, et est donc un pivot asymptotique.

Remarque : En fait $\frac{\frac{3}{2}\bar{X}_n - \theta}{\theta} = \sum_{i=1}^n Y_i - 1$ où les variables $Y_i = 3X_i/2\theta$ sont i.i.d. de loi indépendante de θ. Par conséquent $\frac{\frac{3}{2}\bar{X}_n - \theta}{\theta}$ est un pivot de θ quelle que soit la valeur de n. Mais on ne connaît pas sa loi à n fixé. On peut toutefois approcher cette loi par la méthode de Monte-Carlo.

(b) En déduire un intervalle de confiance pour θ de niveau asymptotique $(1-\alpha)$, basé sur $\bar{X}_n$.

Soit $t_{1-\frac{\alpha}{2}}$ le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$. Quand $n \to \infty$, on a

$$\mathbb{P}\left[\left|\sqrt{8n}\frac{\frac{3}{2}\bar{X}_n - \theta}{\theta}\right| \leq t_{1-\frac{\alpha}{2}}\right] \simeq 1-\alpha$$
$$\iff \mathbb{P}\left[\frac{\frac{3}{2}\bar{X}_n}{1+\frac{t_{1-\frac{\alpha}{2}}}{\sqrt{8n}}} \leq \theta \leq \frac{\frac{3}{2}\bar{X}_n}{1-\frac{t_{1-\frac{\alpha}{2}}}{\sqrt{8n}}}\right] \simeq 1-\alpha$$

Exercice 3 (7 points). Soit $\theta \in \mathbb{R}$ un paramètre inconnu. On considère X_1, ..., X_n un échantillon i.i.d. de loi gaussienne centrée en θ et de variance 1, donc $X_i \sim \mathcal{N}(\theta, 1)$.

1. Dans ce modèle,

(a) Quelle est la vraisemblance $L(\theta)$?

$$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(X_i-\theta)^2\right) = \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2}\sum_{i=1}^n (X_i-\theta)^2\right).$$

(b) Montrer que l'information de Fisher vaut n.

$$\ell(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^n (X_i-\theta)^2\,,\ \ell'(\theta) = \sum_{i=1}^n (X_i-\theta)\,.$$

$$\ell''(\theta) = -n\,,\ \ I_n(\theta) = \mathbb{E}_\theta\left[-\ell''(\theta)\right] = n\,.$$

(c) Montrer que $\hat{\theta}_n = \bar{X}_n$ est l'estimateur du maximum de vraisemblance (EMV) de θ, et qu'il est efficace.

$$\ell'(\theta) = 0 \Longleftrightarrow \sum_{i=1}^{n} X_i = n\theta.$$

$\bar{X}_n$ est donc l'estimateur du maximum de vraisemblance de θ.

$$\mathrm{var}_\theta(\bar{X}_n) = \frac{\mathrm{var}_\theta(X_1)}{n} = \frac{1}{n} = I_n(\theta)^{-1}\,.$$

$\bar{X}_n$ est efficace.

(d) Quelle est la loi de $\sqrt{n}(\bar{X}_n - \theta)$?

Les X_i étant i.i.d. de loi $\mathcal{N}(\theta, 1)$, leur somme est une variable normale de moyenne $n\mathbb{E}_\theta(X_1) = n\theta$, et de variance $n\mathrm{var}_\theta(X_1) = n$. Par conséquent, $\sqrt{n}(\bar{X}_n - \theta) \sim \mathcal{N}(0, 1)$.

2. On s'intéresse au paramètre inconnu η défini par $\eta^3 = \theta$.

(a) Montrer que la vraisemblance $\tilde{L}$ pour ce paramètre est donnée par $\tilde{L}(\eta) = L(\eta^3)$.

En fonction de η, les X_i sont i.i.d. de loi $\mathcal{N}(\eta^3, 1)$, on a évidemment $\tilde{L}(\eta) = L(\eta^3)$.

(b) Quelle est l'information de Fisher pour ce nouveau paramètre ? Quelle est sa valeur pour $\eta = 0$?

On a $\tilde{\ell}(\eta) = \ell(\eta^3)$, $\tilde{\ell}'(\eta) = 3\eta^2\ell'(\eta^3)$,

$$\tilde{\ell}''(\eta) = 6\eta\ell'(\eta^3) + 9\eta^4\ell''(\eta^3) = 6\eta\sum_{i=1}^{n}(X_i - \eta^3) - 9n\eta^4.$$

$$\tilde{I}_n(\eta) = 6\eta\mathbb{E}_\eta\left[\sum_{i=1}^{n}(\eta^3 - X_i)\right] + 9n\eta^4 = 9n\eta^4,\ \text{ et } \tilde{I}_n(0) = 0.$$

(c) Montrer que l'EMV $\hat{\eta}_n$ de η est tel que $\hat{\eta}_n^3 = \hat{\theta}_n$.

Puisque $\tilde{\ell}(\eta) = \ell(\eta^3)$, $\hat{\eta}_n$ maximise $\eta \mapsto \ell(\eta^3)$ et donc $\hat{\eta}_n^3 = \hat{\theta}_n$.

3. On suppose que $\eta \neq 0$. Montrer alors qu'asymptotiquement

$$\sqrt{n}(\hat{\eta}_n - \eta) \ \xrightarrow{\text{loi}}\ \mathcal{N}\left(0, \left(\frac{1}{3\eta^2}\right)^2\right).$$

Si $\eta \neq 0$, $\tilde{I}_n(\eta) > 0$, et on peut appliquer les résultats asymptotique sur les EMV pour en déduire que $\sqrt{\tilde{I}_n(\eta)}(\hat{\eta}_n - \eta) = 3\eta^2\sqrt{n}(\hat{\eta}_n - \eta)$ converge en loi vers une loi $\mathcal{N}(0, 1)$.

On peut aussi utiliser la méthode Delta, en utilisant le fait que $\hat{\eta}_n = \varphi(\hat{\theta}_n)$, où $\varphi(x) = x^{1/3}$ est dérivable pour $x \neq 0$. On en déduit que $\sqrt{n}(\hat{\eta}_n - \eta)$ converge en loi vers une loi normale de variance $\frac{\varphi'(\theta)^2}{I_1(\theta)} = \left(\frac{1}{3\theta^{4/3}}\right)^2 = \frac{1}{9\eta^4}$.

4. On suppose maintenant que $\eta = 0$.

(a) Quelle est la loi de $\sqrt{n}\,\hat{\theta}_n$? En déduire que $n^{(1/6)}\hat{\eta}_n$ a une loi qui ne dépend pas de n, et comparer au résultat de la question 3.

$\sqrt{n}\,\hat{\theta}_n \sim \mathcal{N}(0, 1)$. Comme $\sqrt{n}\,\hat{\theta}_n = (n^{1/6}\hat{\eta}_n)^3$, on en déduit que $n^{1/6}\hat{\eta}_n$ a même loi que $Z^{1/3}$, où $Z \sim \mathcal{N}(0, 1)$.

(b) Montrer que $n^{(1/6)}\hat{\eta}_n$ a pour densité

$$e^{-(t^6/2)}\frac{3t^2}{\sqrt{2\pi}}.$$

Soit h une fonction mesurable bornée.

$$\mathbb{E}(h(n^{1/6}\hat{\eta}_n)) = \mathbb{E}(h(Z^{1/3})) = \int h(z^{1/3})\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz.$$

On pose $t = z^{1/3} \iff z = t^3$.

$$\mathbb{E}(h(n^{1/6}\hat{\eta}_n)) = \int h(t)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^6}{2}\right)3t^2 dt,$$

ce qui donne le résultat.