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 06 Novembre 2023.

*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 (7 points).** Soient $U$ et $V$ des variables aléatoires indépendantes de même loi $\mathcal{U}([0,1])$. On pose $X = UV$ et $Y = U(1-V)$.

1. (0,5 points) Justifiez sans calculs que $X$ et $Y$ sont de même loi.

   $V$ et $1-V$ sont deux variables de loi $\mathcal{U}([0,1])$ qui sont toutes deux indépendantes de $U$. Donc $X$ et $Y$ sont de même loi.

2. (1 point) Calculez $\mathbb{E}(U)$, $\mathbb{E}(U^2)$, $\mathbb{E}(X)$, $\mathbb{E}(X^2)$.

$$\mathbb{E}(U) = \int_0^1 u\,du = 1/2\,;\ \mathbb{E}(U^2) = \int_0^1 u^2\,du = 1/3\,.$$

$$\mathbb{E}(X) = \mathbb{E}(UV) = \mathbb{E}(U)\mathbb{E}(V) = 1/4\,.\ \mathbb{E}(X^2) = \mathbb{E}(U^2V^2) = \mathbb{E}(U^2)\mathbb{E}(V^2) = 1/9\,.$$

3. (1 point) Calculez $\mathbb{E}(XY)$. Les variables $X$ et $Y$ sont-elles indépendantes ?

$$\mathbb{E}(XY) = \mathbb{E}(U^2V(1-V)) = \mathbb{E}(U^2)\mathbb{E}(V(1-V)) = \frac{1}{3}(\mathbb{E}(V) - \mathbb{E}(V^2)) = \frac{1}{3}\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{1}{18}.$$

   Comme $\mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)^2 = \frac{1}{16}$, $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$. $X$ et $Y$ ne sont donc pas indépendantes.

4. ( 2 points) Montrez que le couple $(X,Y)$ a pour densité

$$f_{(X,Y)}(x,y) = \frac{1}{x+y}\ \mathbb{1}_{x\geq 0, y\geq 0, x+y\leq 1}\,.$$

   Soit $h$ une fonction mesurable bornée.

$$\mathbb{E}(h(X,Y)) = \mathbb{E}(h(UV, U(1-V))) = \iint_{[0;1]^2} h(uv, u(1-v))\,du\,dv.$$

On pose $\left\{\begin{array}{lcl} x & = & uv \\ y & = & u(1-v) \end{array}\right. \iff \left\{\begin{array}{lcl} u & = & y+x \\ v & = & \frac{x}{y+x} \end{array}\right.$. La matrice jacobienne de ce changement de variables est $J = \begin{pmatrix} 1 & \frac{y}{(y+x)^2} \\ 1 & -\frac{x}{(y+x)^2} \end{pmatrix}$, et $|\det(J)| = \frac{1}{x+y}$.

De plus, $\left\{\begin{array}{l} 0 \leq u \leq 1 \\ 0 \leq v \leq 1 \end{array}\right. \iff \left\{\begin{array}{l} x \geq 0 \\ y \geq 0 \\ x+y \leq 1 \end{array}\right.$. On obtient donc

$$\mathbb{E}(h(X,Y)) = \iint_{x\geq 0, y\geq 0, x+y\leq 1} h(x,y)\,\frac{1}{x+y}\,dx\,dy,$$

ce qui donne le résultat.

5. (2 points) Montrez que la densité conditionnelle de $X$ sachant $Y=y$ (où $y \in ]0,1]$) est

$$f_{X|Y=y}(x) = \frac{-\log(y)}{x+y}\ \mathbb{1}_{[0;1-y]}(x)\,.$$

En déduire que $\mathbb{E}(X|Y) = -\log(Y)(1-Y+Y\log(Y))$.

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x,y)}{\int f_{(X,Y)}(x,y)dx}\,.$$

Or, pour tout $y \geq 0$ fixé,

$$\int f_{(X,Y)}(x,y)dx = \int_{x\geq 0, x\leq 1-y} \frac{dx}{x+y}dx = \ \mathbb{1}_{y\leq 1}\left[\log(x+y)\right]_{x=0}^{x=1-y} \ = \ \mathbb{1}_{y\leq 1}(-\log(y))\,.$$

D'où, pour $y \in [0,1[$ $f_{X|Y=y}(x) = -\frac{1}{\log(y)(x+y)}\ \mathbb{1}_{[0;1-y]}(x)$.

Désolée pour l'erreur dans l'énoncé.

On a alors

$$\begin{aligned}
\mathbb{E}(X|Y) &= \int x\, f_{X|Y}(x)\, dx = \int_0^{1-Y} -\frac{1}{\log(Y)}\frac{x}{x+Y}\,dx \\
&= -\frac{1}{\log(Y)}\int_0^{1-Y} 1 - \frac{Y}{x+Y}\,dx \\
&= -\frac{1}{\log(Y)}\left(1-Y-Y\left[\log(x+Y)\right]_{x=0}^{x=1-Y}\right) \\
&= -\frac{1}{\log(Y)}\left(1-Y+Y\log(Y)\right)
\end{aligned}$$

6. (0,5 points) Sans calcul supplémentaire, quelle est la valeur de $\mathbb{E}(Y|X)$ ? Justifiez votre réponse.

Comme $f_{(X,Y)}(x,y) = f_{(X,Y)}(y,x)$, la loi du couple $(X,Y)$ est la même que la loi du couple $(Y,X)$. Par conséquent, $\mathbb{E}(Y|X) = -\frac{1-X+X\log(X)}{\log(X)}$ .

**Exercice 2 (4 points).** Soit $(X,Y)$ un couple gaussien centré de matrice de covariance $\begin{pmatrix} 9 & -3 \\ -3 & 1 \end{pmatrix}$.

1. (0,5 points) Les variables $X$ et $Y$ sont-elles indépendantes ?

Cov$(X,Y) = -3 \neq 0$. Donc $X$ et $Y$ ne sont pas indépendantes.

2. (0,5 points) Que vaut $\mathbb{P}(X \geq 0)$ ?
   $X \sim \mathcal{N}(0;9)$. La loi de $X$ est donc symétrique en 0 et $\mathbb{P}(X \geq 0) = 1/2$.
3. (1 point) Quelle est la loi de la variable $X+Y$ ?
   $X+Y$ est une combinaison linéaire des coordonnées du vecteur gaussien $(X,Y)$. Par conséquent, $X+Y$ suit une loi normale. On a $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 0$, et $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y) = 9+1-6 = 4$. D'où $X+Y \sim \mathcal{N}(0,4)$.
4. (2 points) Calculez $\mathrm{Var}(X+3Y)$. Qu'en concluez-vous ? Quelle est la valeur de $\mathbb{E}(X|Y)$ ? Le couple $(X,Y)$ admet-il une densité ?
   $\mathrm{Var}(X+3Y) = \mathrm{Var}(X) + 9\mathrm{Var}(Y) + 6\mathrm{Cov}(X,Y) = 9+9-18 = 0$. Par conséquent, $X+3Y = \mathrm{cte} = \mathbb{E}(X+3Y) = 0$. On a donc $X = -3Y$, et $\mathbb{E}(X|Y) = \mathbb{E}(-3Y|Y) = -3Y$. Si le couple $(X,Y)$ admet une densité $f_{(X,Y)}$, on a

$$1 = \mathbb{P}(X+3Y=0) = \iint_{(x,y);x+3y=0} f_{(X,Y)}(x,y)\,dxdy = 0\,.$$

   On obtient alors une contradiction. Donc, le couple $(X,Y)$ n'a pas de densité.

**Exercice 3 ( 9 points).** Soit $\theta > 0$ un paramètre inconnu. On considère $X_1$, ..., $X_n$ un échantillon i.i.d. de loi de densité

$$f_\theta(x) = \frac{1}{\theta} x^{-\left(1+\frac{1}{\theta}\right)} \mathbb{1}_{[1;+\infty[}(x)\,.$$

1. (0,5 points) Vérifiez que $f_\theta$ est bien une densité de probabilité.
   Pour tout $x \in \mathbb{R}$, $f_\theta(x) \geq 0$. De plus, pour $\theta > 0$,

$$\int f_\theta(x)\,dx = \frac{1}{\theta}\int_1^{+\infty} \frac{dx}{x^{1+\frac{1}{\theta}}} = \left[-x^{-\frac{1}{\theta}}\right]_{x=1}^{x=+\infty} = 1\,.$$

2. (1 point) Montrez que

$$\frac{\partial}{\partial\theta}\log f_\theta(x) = \frac{1}{\theta^2}(\log(x) - \theta).$$

$$\frac{\partial^2}{\partial\theta^2}\log f_\theta(x) = -\frac{2}{\theta}\frac{\partial}{\partial\theta}\log f_\theta(x) - \frac{1}{\theta^2}.$$

   Pour tout $x > 1$,

$$\frac{\partial}{\partial\theta}\log f_\theta(x) = \frac{\partial}{\partial\theta}\left[-\log(\theta) - \left(1+\frac{1}{\theta}\right)\log(x)\right] = -\frac{1}{\theta} + \frac{1}{\theta^2}\log(x) = \frac{1}{\theta^2}\left(\log(x) - \theta\right).$$

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}\log f_\theta(x) &= \frac{\partial}{\partial\theta}\left(-\frac{1}{\theta} + \frac{1}{\theta^2}\log(x)\right) = \frac{1}{\theta^2} - \frac{2}{\theta^3}\log(x) = -\frac{2}{\theta}\left[\frac{1}{\theta^2}\log(x) - \frac{1}{\theta}\right] - \frac{1}{\theta^2}\\ &= -\frac{2}{\theta}\frac{\partial}{\partial\theta}\log f_\theta(x) - \frac{1}{\theta^2}.\end{aligned}$$

3. (2 points)Pour cette question, on rappelle les résultats de cours suivants : Si $X$ est de densité $f_\theta$ satisfaisant certaines conditions de régularité, on a

$$\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] = 0\,;\;\; \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right].$$

Déduisez de ce qui précède que $\mathbb{E}_\theta[\log(X)] = \theta$ ; $\mathrm{Var}_\theta[\log(X)] = \theta^2$.

$$\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] = \frac{1}{\theta^2}\mathbb{E}_\theta[\log(X)-\theta] = 0\,.$$

On a donc $\mathbb{E}_\theta[\log(X)] = \theta$ . Par ailleurs, d'après la question précédente et le rappel de cours,

$$\begin{aligned}\mathrm{Var}_\theta[\log(X)] &= \mathbb{E}_\theta\left[(\log(X)-\theta)^2\right] = \theta^4\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right] = -\theta^4\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right]\\ &= -\theta^4\mathbb{E}_\theta\left[-\frac{2}{\theta}\frac{\partial}{\partial\theta}\log f_\theta(x) - \frac{1}{\theta^2}\right] = \frac{\theta^4}{\theta^2} = \theta^2\,.\end{aligned}$$

4. (1 point) Montrez que l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de $\theta$ est donné par

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n \log(X_i)\,.$$

Est-il sans biais ?

La vraisemblance du modèle est donnée par

$$L(\theta;X) = \prod_{i=1}^n f_\theta(X_i) = \frac{1}{\theta^n}\prod_{i=1}^n X_i^{-(1+\frac{1}{\theta})}\ \mathbb{1}_{[1;+\infty[}(\min_{i=1}^n X_i).$$

Si $\min_{i=1}^n X_i \geq 1$, on a donc

$$\frac{\partial}{\partial\theta}\log L(\theta;X) = \frac{\partial}{\partial\theta}\left(-n\log(\theta) - (1+\frac{1}{\theta})\sum_{i=1}^n\log(X_i)\right) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n\log(X_i).$$

L'estimateur du maximum de vraisemblance satisfait donc l'équation :

$$\frac{n}{\hat{\theta}_n} = \frac{1}{\hat{\theta}_n^2}\sum_{i=1}^n\log(X_i) \Leftrightarrow \hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n\log(X_i)\,.$$

On a alors $\mathbb{E}_\theta\left[\hat{\theta}_n\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta[\log(X_i)] = \mathbb{E}_\theta[\log(X_1)] = \theta$ d'après la question 3. $\hat{\theta}_n$ est sans biais.

5. (1 point) Calculez l'information de Fisher $I_n(\theta)$. En déduire que $\hat{\theta}_n$ est un estimateur efficace de $\theta$.

$$I_n(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log L(\theta;X)\right] = -\mathbb{E}_\theta\left[\frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^n\log(X_i)\right] = -\frac{n}{\theta^2} + \frac{2n}{\theta^3}\mathbb{E}_\theta[\log(X_1)] = \frac{n}{\theta^2}.$$

Par ailleurs

$$\mathrm{Var}_\theta\left[\hat{\theta}_n\right] = \mathrm{Var}_\theta\left(\frac{1}{n}\sum_{i=1}^n\log(X_i)\right) = \frac{\mathrm{Var}_\theta(\log(X_1))}{n} = \frac{\theta^2}{n} = \frac{1}{I_n(\theta)}$$

$\hat{\theta}_n$ est donc un estimateur efficace de $\theta$. En particulier, il est de variance minimale parmi les estimateurs sans biais.

| $\alpha$ | 2,5 % | 5 % | 95 % | 97,5 % |
|---|---|---|---|---|
| Quantile de $\mathcal{N}(0,1)$ | -1.96 | -1.64 | 1.64 | 1.96 |
| Quantile de $\hat{\theta}_n/\theta$ | 0.65 | 0.70 | 1.35 | 1.43 |

Table 1 – Quantiles de la loi $\mathcal{N}(0,1)$ et de la loi de $\hat{\theta}_n/\theta$ pour $n = 25$.

6. (1,5 points) Montrez que $\sqrt{n}\left(\frac{\hat{\theta}_n}{\theta} - 1\right)$ converge en loi vers la loi $\mathcal{N}(0,1)$ quand $n$ tend vers $+\infty$. En déduire un intervalle de confiance pour $\theta$ de niveau asymptotique 95%. On pourra utiliser les quantiles de la loi $\mathcal{N}(0,1)$ donnés dans la Table 1.
Les variables $\log(X_i)$ sont i.i.d de carré intégrable. Par le théorème de la limite centrale, $\sqrt{n}\left(\hat{\theta}_n - \mathbb{E}_\theta(\log(X))\right)$ converge en loi vers la loi $\mathcal{N}(0, \mathrm{Var}_\theta(\log(X)) = \theta^2)$. Autrement dit, $\frac{\sqrt{n}}{\theta}\left(\hat{\theta}_n - \theta\right)$ converge en loi vers la loi $\mathcal{N}(0,1)$. D'après les quantiles de la loi $\mathcal{N}(0,1)$,

$$\begin{aligned}\mathbb{P}_\theta\left[\left|\sqrt{n}\left(\frac{\hat{\theta}_n}{\theta} - 1\right)\right| \leq 1,96\right] \simeq 0.95 \quad &\Leftrightarrow \quad \mathbb{P}_\theta\left[1 - \frac{1,96}{\sqrt{n}} \leq \frac{\hat{\theta}_n}{\theta} \leq 1 + \frac{1,96}{\sqrt{n}}\right] \simeq 0.95 \\ &\Leftrightarrow \quad \mathbb{P}_\theta\left[\frac{\hat{\theta}_n}{1 + \frac{1,96}{\sqrt{n}}} \leq \theta \leq \frac{\hat{\theta}_n}{1 - \frac{1,96}{\sqrt{n}}}\right] \simeq 0.95,\end{aligned}$$

dès que $n \geq (1,96)^2$.

7. (1,5 points) Montrez que si $X$ est de densité $f_\theta$, $\log(X)/\theta$ est de loi exponentielle de paramètre 1. En déduire que $\hat{\theta}_n/\theta$ est un pivot pour l'estimation de $\theta$.
Soit $h$ une fonction mesurable bornée.

$$\mathbb{E}_\theta\left[h\left(\frac{\log(X)}{\theta}\right)\right] == \int_1^{+\infty} h\left(\frac{\log(x)}{\theta}\right) x^{-(1+\frac{1}{\theta})}\frac{dx}{\theta}.$$

On pose $u = \frac{\log(x)}{\theta} \Leftrightarrow x = e^{\theta u}$. On obtient

$$\mathbb{E}_\theta\left[h\left(\frac{\log(X)}{\theta}\right)\right] = \int_0^{+\infty} h(u)e^{-(\theta+1)u}e^{\theta u}\,du = \int_0^{+\infty} h(u)e^{-u}\,du.$$

Donc $\log(X)/\theta \sim \mathcal{E}(1)$. $\frac{\hat{\theta}_n}{\theta} = \frac{1}{n}\sum_{i=1}^n \frac{\log(X_i)}{\theta}$. $n\frac{\hat{\theta}_n}{\theta}$ est donc la somme de $n$ variables aléatoires indépendantes de loi $\mathcal{E}(1)$ ' (i.e ;. une loi $\Gamma(n,1)$. La loi de $\hat{\theta}_n/\theta$ ne dépend pas donc pas de $\theta$.

8. (0,5 points) Pour $n = 25$, la Table 1 donne quelques quantiles de la loi de $\hat{\theta}_n/\theta$. En déduire un intervalle de confiance pour $\theta$ de niveau 95%.
D'après la table 1,

$$\mathbb{P}_\theta\left[0.65 \leq \frac{\hat{\theta}_n}{\theta} \leq 1.43\right] = 0.95 \Leftrightarrow \mathbb{P}_\theta\left[\frac{\hat{\theta}_n}{1.43} \leq \theta \leq \frac{\hat{\theta}_n}{0.65}\right] = 0.95$$