

Formulaire

Loi	Notation	Densité
Exponentielle	$\mathcal{E}(\lambda)$	$f(x \lambda) = \lambda e^{-\lambda x} \mathbb{1}_{x>0}$
Gamma	$\mathcal{G}a(a, b)$	$f(x a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbb{1}_{x>0}$
Laplace	$\mathcal{L}(\mu, b)$	$f(x \mu, b) = \frac{1}{2b} \exp \left[-\frac{ x - \mu }{b} \right]$
Poisson	$\mathcal{P}oi(\lambda)$	$f(x \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \mathbb{1}_{x \in \mathbb{N}}$

French – English Lexicon

- échantillon : *sample*
- famille exponentielle : *exponential family*
- espace naturel des paramètres : *natural parameter space*
- *i.i.d.* : *independent and identically distributed*
- statistique libre : *ancillary statistic*
- statistique exhaustive : *sufficient statistic*
- statistique complète : *complete statistic*
- vraisemblance : *likelihood*

Statistiques d'ordre Étant données $X_1, \dots, X_n$ *i.i.d.* de densité f et de fonction de répartition F , les densités de $\min(X_1, \dots, X_n)$ et $\max(X_1, \dots, X_n)$ sont respectivement

$$f_1(x) = n[1 - F(x)]^{n-1} f(x) \quad \text{et} \quad f_n(x) = nF(x)^{n-1} f(x).$$

Exercice 1

..... / 8.5

Dans cet exercice il vous est demandé de donner la ou les bonnes réponses. Seules les réponses justifiées seront validées, au prorata du nombre de bonnes réponses. Il n'y a pas de points négatifs, mais toute réponse fausse conduit à une note nulle.

1. Soit la densité définie sur $\mathbb{R}$ par $f(x | \alpha, \theta) = \alpha \theta^{-\alpha} x^{\alpha-1}$ pour $0 \leq x \leq \theta$. Alors $f(\cdot | \alpha, \theta)$:

- (a) forme une famille exponentielle de dimension 2, (d) forme une famille exponentielle uniquement
(b) forme une famille exponentielle de dimension 1, lorsque α est fixé.
(c) ne forme pas une famille exponentielle,

(c) Le support de la densité dépend de θ inconnu. Elle ne peut donc pas former une famille exponentielle.

2. Soit un jeu de données $\mathbf{x}$ correspondant à des observation d'un modèle statistique de fonction de répartition F . On dispose d'un échantillon bootstrap `mstar` de la médiane de $(X_1^*, \dots, X_n^*)$ pour $X_1^*, \dots, X_n^*$ *i.i.d.* suivant la fonction de répartition empirique $\hat{F}_n$. Quel code permet d'obtenir un intervalle de confiance bootstrap empirique (*empirical bootstrap confidence interval*) de la médiane de la loi F au niveau 95% ?

- (a) `quantile(mstar, c(.05, .95))`
- (b) `median(x) - quantile(mstar - median(x), c(.95, .05))`
- (c) `quantile(mstar, c(.025, .975))`
- (d) `median(x) - quantile(mstar - median(x), c(.975, .025))`

(d) Pour $\alpha \in [0, 1]$, un intervalle de confiance bootstrap empirique au niveau $1 - \alpha$ de la médiane de F est donné par $\left[\text{median}(\mathbf{x}) - q_{1-\alpha/2}, \text{median}(\mathbf{x}) - q_{\alpha/2} \right]$, où $q_{\alpha/2}$ et $q_{1-\alpha/2}$ sont les quantiles de $\text{median}(X_1^*, \dots, X_n^*) - \text{median}(\mathbf{x})$ d'ordre $\alpha/2$ et $1 - \alpha/2$. En prenant $\alpha = 5\%$, on obtient le résultat.

3. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables *i.i.d.* de loi uniforme sur $[0, \theta]$, avec $\theta > 0$. Parmi les variables suivantes, donner celle(s) qui sui(ven)t asymptotiquement la loi $\mathcal{N}(0, 1)$.

- (a) $\sqrt{3n} \left(1 - \theta / (2\bar{X}_n) \right)$,
- (b) $2\sqrt{3n} \left(1 - \theta / \bar{X}_n \right)$,
- (c) $\sqrt{n} \left(1 - (2\bar{X}_n) / \theta \right)$,
- (d) $\frac{\sqrt{n}}{3} \left(1 - (2\bar{X}_n) / \theta \right)$.
- (e) Aucune

(a) $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires *i.i.d.* de variance finie : $\text{Var}[X_1] = \theta^2/12$. Le théorème Central Limite permet donc d'écrire

$$\sqrt{n} \left(\bar{X}_n - \frac{\theta}{2} \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N} \left(0, \frac{\theta^2}{12} \right).$$

La méthode delta avec la fonction $g : x \mapsto 1/x$, dérivable sur $\mathbb{R}$ avec $g'(\theta/2) = -4/\theta^2 \neq 0$, donne

$$\sqrt{n} \left(\frac{1}{\bar{X}_n} - \frac{2}{\theta} \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N} \left(0, \frac{16}{\theta^4} \frac{\theta^2}{12} \right) \equiv \mathcal{N} \left(0, \frac{4}{3\theta^2} \right) \Leftrightarrow \sqrt{3n} \left(\frac{\theta}{2\bar{X}_n} - 1 \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

(c, d) suivent asymptotiquement une loi normale mais dont les variances ne sont pas égales à 1.

4. Soit X qui suit la loi de Laplace $\mathcal{L}(\mu, b)$ où le paramètre de position $\mu \in \mathbb{R}$ est connu. L'information de Fisher apportée par $T(X) = |X - \mu|$ sur b est :

- (a) μb^2 ,
- (b) $(2 - b^2)/b^4$,
- (c) $1/b^2$,
- (d) $(b - 2\mu)/b^3$.

(c) $f(x | \theta)$ forme une famille exponentielle de statistique naturelle $T(X)$. $T(X)$ est donc exhaustive, et l'information de Fisher apportée par $T(X)$ sur b et la même que celle apportée par X sur b . La densité est deux fois différentiable et on a

$$\frac{\partial^2}{\partial b^2} \log f(x | b) = \frac{1}{b^2} - \frac{2|x - \mu|}{b^3} \quad \text{et} \quad I_{T(X)}(b) = I_X(b) = -\mathbb{E}_b \left[\frac{\partial^2}{\partial b^2} \log f(X | b) \right] = -\frac{1}{b^2} + \frac{2}{b^3} \mathbb{E}_b [|X - \mu|]$$

On obtient une forme canonique en posant $\theta = -1/b$. La fonction de partition associée est alors $a(\theta) = -\theta/2$, avec $\theta \in \mathbb{R}_*$. On en déduit que

$$\mathbb{E}_b [|X - \mu|] = -\frac{\partial}{\partial \theta} \log a(\theta) = -\frac{1}{\theta} = b.$$

5. Soient X_1, X_2, X_3, X_4 i.i.d. suivant la loi de Poisson $\mathcal{P}(\lambda)$, avec $\lambda \in \mathbb{R}_+^*$. On pose

$$T(X_1, \dots, X_4) = \log \left(1 + \sum_{i=1}^4 X_i \right) \quad \text{et} \quad R(X_1, \dots, X_4) = \log \left[\exp \left(-\frac{X_1 + X_2}{2} \right) + \frac{1}{2} \exp(X_4 - X_3) \right].$$

- (a) $T(X_1, \dots, X_4)$ est exhaustive, (c) $R(X_1, \dots, X_4)$ est exhaustive,
(b) $T(X_1, \dots, X_4)$ n'est pas exhaustive, (d) $R(X_1, \dots, X_4)$ n'est pas exhaustive.

(a, d) La loi de Poisson forme une famille exponentielle de statistique naturelle $S(X) = X$. On en déduit que $S(X_1, \dots, X_4) = \sum_{i=1}^4 X_i$ est une statistique exhaustive minimale. $T(X)$ étant une transformation bijective de $S(X_1, \dots, X_4)$ elle est également exhaustive.

Raisonnons par l'absurde. Supposons que $R(X_1, \dots, X_4)$ est exhaustive. Comme $S(X_1, \dots, X_4)$ est exhaustive minimale, il existe une fonction g telle que $S(X_1, \dots, X_4) = g[R(X_1, \dots, X_4)]$. Or on a $R(0, 0, 1, 1) = R(0, 0, 4, 4)$, donc $g[R(0, 0, 1, 1)] = g[R(0, 0, 4, 4)]$. Mais, $g[R(0, 0, 1, 1)] = S(0, 0, 1, 1) = 2$ et $g[R(0, 0, 4, 4)] = S(0, 0, 4, 4) = 8$. Absurde !

Remarque On aurait également pu montrer que le rapport de vraisemblance en ces points dépend de λ :

$$\frac{L(\lambda; 0, 0, 1, 1)}{L(\lambda; 0, 0, 4, 4)} = \frac{(4!)^2}{\lambda^2}.$$

6. Soient X_1, X_2, X_3, X_4 i.i.d. suivant la loi $\mathcal{N}(\mu, \sigma^2)$, avec $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}_+^*$ inconnus. On note $\text{med}(X_1, \dots, X_4)$ la médiane de $X_1, \dots, X_4$. Parmi les statistiques suivantes, lesquelles sont libres ?

- (a) X_1/X_2 , (d) $\sum_{i=1}^4 (X_i - X_1)^2$,
(b) $X_1 - X_2$, (e) $\frac{4 \times \text{med}(X_1, \dots, X_4) - \sum_{i=1}^4 X_i}{\max(X_1, \dots, X_4) - \min(X_1, \dots, X_4)}$.
(c) $(X_1 - X_2)/(X_3 - X_4)$,

Soit Y_1, Y_2, Y_3, Y_4 i.i.d. de loi $\mathcal{N}(0, 1)$. Pour $k \in \llbracket 1, 4 \rrbracket$, on peut écrire $X_k = \mu + \sigma Y_k$.

(c, e) On a

$$\frac{X_1 - X_2}{X_3 - X_4} = \frac{Y_1 - Y_2}{Y_3 - Y_4} \quad \text{et} \quad \frac{4 \times \text{med}(X_1, \dots, X_4) - \sum_{i=1}^4 X_i}{\max(X_1, \dots, X_4) - \min(X_1, \dots, X_4)} = \frac{4 \times \text{med}(Y_1, \dots, Y_4) - \sum_{i=1}^4 Y_i}{\max(Y_1, \dots, Y_4) - \min(Y_1, \dots, Y_4)}.$$

Comme la loi de $Y_1, \dots, Y_4$ ne dépend pas de μ et σ , il en est de même pour toute loi jointe de $(Y_1, \dots, Y_4)$. Donc les statistiques précédentes sont libres.

(a, b, d) On a

$$\frac{X_1}{X_2} = \frac{Y_1 + \mu/\sigma}{Y_2 + \mu/\sigma}, \quad X_1 - X_2 = \sigma(Y_1 - Y_2), \quad \text{et} \quad \sum_{i=1}^4 (X_i - X_1)^2 = \sigma^2 \sum_{i=1}^4 (Y_i - Y_1)^2.$$

Ces statistiques sont des transformations de $Y_1, \dots, Y_4$ dépendant de μ et/ou σ . Comme la loi de $Y_1, \dots, Y_4$ ne dépend pas de μ et σ , les statistiques ne peuvent pas être libres.

7. Soit X de densité $f(\cdot | \theta)$, $\theta \in \mathbb{R}$. Soient $T(X)$ une statistique exhaustive pour θ et $S(X)$ une statistique libre non constante telle que $S(X) = g(T(X))$, avec g une fonction mesurable. Alors nécessairement,

- | | |
|--|---|
| (a) $T(X)$ est complète, | (d) $S(X)$ est minimale exhaustive, |
| (b) $T(X)$ est minimale exhaustive, | (e) $T(X)$ est libre, |
| (c) $T(X)$ et $S(X)$ sont indépendantes, | (f) Aucune de ces propositions n'est vraie. |

(f) est la bonne réponse !

(a) La statistique ne peut pas être complète. En effet, $S(X)$ étant libre, on a

$$\mathbb{E}_\theta[g(T(X))] = \mathbb{E}_\theta[S(X)] = c, \quad \text{avec } c \text{ une constante indépendante de } \theta.$$

On en déduit que $\mathbb{P}_\theta[g(T(X)) - c = 0] = \mathbb{P}_\theta[S(X) = c] \neq 1$ car $S(X)$ n'est pas constante. Il existe donc une fonction $\phi : x \mapsto g(x) - c$ non nulle $\mathbb{P}_\theta$ presque sûrement pour tout $\theta \in \mathbb{R}$ telle que $\mathbb{E}_\theta[\phi(T(X))] = 0$. Donc $T(X)$ n'est pas complète.

(b, d) ne peuvent être vérifiées que si on exprime $T(X)$ et $S(X)$ comme fonction de toute autre statistique exhaustive. Ce n'est pas le cas ici.

(c) n'est pas vraie car $S(X)$ est une transformation mesurable de $T(X)$.

(e) n'est pas vraie car $T(X)$ est exhaustive.

8. Soit $X_1, \dots, X_n$ variables aléatoires *i.i.d.* suivant la loi $\mathcal{E}(e^\theta)$, $\theta \in \mathbb{R}$. On pose pour $k \in \llbracket 1, n \rrbracket$ et $\delta \in \mathbb{R}_+^*$ connu, $Y_k = \mathbb{1}_{\{X_k \leq \delta\}}$. On observe $(y_1, \dots, y_{n-1}) = (1, \dots, 1)$ et $y_n = 0$. L'estimateur du maximum de vraisemblance de θ est

- | | |
|-------------------------------|---------------------------------------|
| (a) $\log(\delta)/(n-1)$, | (c) $\log[\log(n)/\delta]$, |
| (b) $-\log[(n-1)/n]/\delta$, | (d) $\log\{-\log[(n-1)/n]/\delta\}$. |

(c) $Y_1, \dots, Y_n$ sont *i.i.d.* (transformation mesurable de variables *i.i.d.*) suivant la loi de Bernoulli de paramètre $p = \mathbb{P}[X_1 \leq \delta] = 1 - \exp(-\delta e^\theta)$. La vraisemblance s'écrit donc

$$L(\theta; y_1, \dots, y_n) = \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i} = \left[1 - \exp(-\delta e^\theta)\right]^{n-1} \exp(-\delta e^\theta).$$

La log-vraisemblance est dérivable sur $\mathbb{R}$ et on a

$$\frac{\partial}{\partial \theta} \log L(\theta; y_1, \dots, y_n) = \frac{-\delta e^\theta}{1 - \exp(-\delta e^\theta)} \left[1 - n \exp(-\delta e^\theta)\right].$$

La log-vraisemblance admet donc un unique point critique

$$\hat{\theta}_n = \log \left[\frac{\log(n)}{n} \right].$$

La densité de Y_1 forme une famille exponentielle régulière, ce point critique est donc l'unique estimateur du maximum de vraisemblance de θ .

9. Soient $X_1, \dots, X_n$ *i.i.d.* suivant la loi de densité $f(x | \theta) = \exp[-(x - \theta)] \mathbb{1}\{x \geq \theta\}$. Alors :

- (a) $\min(X_1, \dots, X_n)$ est l'estimateur du maximum de vraisemblance de θ et il est sans biais, (c) $\max(X_1, \dots, X_n)$ est l'estimateur du maximum de vraisemblance de θ et il est sans biais,
 (b) $\min(X_1, \dots, X_n) - \frac{1}{n}$ est un estimateur sans biais de θ , (d) $\frac{n}{n-1} \min(X_1, \dots, X_n)$ est un estimateur sans biais de θ .

(b) La vraisemblance s'écrit

$$L(\theta; x_1, \dots, x_n) = \exp \left[- \sum_{i=1}^n (x_i - \theta) \right] \mathbb{1}_{\{\min(x_1, \dots, x_n) \geq \theta\}}.$$

On en déduit alors que l'estimateur du maximum de vraisemblance est $\hat{\theta}_n = \min(X_1, \dots, X_n)$. La réponse (c) ne peut donc pas être correcte. En utilisant la loi du min on obtient

$$\hat{\theta}_n = \int_{\theta}^{+\infty} x n e^{-n(x-\theta)} dx = \theta + \frac{1}{n}.$$

On en déduit que $\hat{\theta}_n - 1/n$ est un estimateur sans biais de θ . Les réponses (a) et (c) ne sont donc pas vraies car ces estimateurs sont biaisés.

Exercice 2

...../11.5

La loi de Maxwell de paramètre $b \in \mathbb{R}_+^*$ admet pour densité sur $\mathbb{R}_+^*$

$$f(x | b) = \sqrt{\frac{2}{\pi}} \frac{x^2}{b^3} \exp \left(-\frac{x^2}{2b^2} \right) \mathbb{1}_{\{x > 0\}}.$$

Dans cet exercice, on suppose que **le paramètre b est inconnu**. On note pour $n \in \mathbb{N}$, $X_1, \dots, X_n$ un ensemble de variables aléatoires *i.i.d.* suivant $f(\cdot | b)$

1. Montrer que $\{f(\cdot | b); b \in \mathbb{R}_+^*\}$ forme une famille exponentielle. La famille ainsi définie est-elle sous forme canonique? Si non, donner la forme canonique. La famille est-elle minimale? Régulière?

On a bien une famille exponentielle car la densité s'écrit

$$f(x | b) = c(b)h(x) \exp [\eta(b)T(x)]$$

avec

$$c(b) = \frac{1}{b^3} \sqrt{\frac{2}{\pi}}, \quad h(x) = x^2 \mathbb{1}_{\{x > 0\}}, \quad \eta(b) = -\frac{1}{2b^2}, \quad \text{et} \quad T(x) = x^2.$$

La famille n'est pas sous forme canonique. Pour obtenir la forme canonique, on pose $\theta = \eta(b)$ et on obtient alors

$$f(x | \theta) = a(\theta)h(x) \exp(\theta T(x)), \quad \text{avec} \quad a(\theta) = \frac{4\sqrt{-\theta^3}}{\sqrt{\pi}}.$$

La statistique étant de dimension 1, la famille est minimale. Par ailleurs, l'espace naturel des paramètres est donné par

$$\Theta = \left\{ \theta \in \mathbb{R} \mid \int_{\mathbb{R}} h(x) \exp(\theta T(x)) dx = \int_0^{+\infty} x^2 \exp(\theta x^2) dx < \infty \right\}.$$

La fonction $x \mapsto x^2 \exp(\theta x^2)$ est intégrable en $+\infty$ lorsque $\theta < 0$. D'où $\Theta =]-\infty, 0[$. On aurait également pu obtenir ce résultat en utilisant la caractérisation de la forme canonique à savoir

$$\Theta = \left\{ \theta \in \mathbb{R} \left| a(\theta) = \frac{4\sqrt{-\theta^3}}{\sqrt{\pi}} > 0 \right. \right\} =]-\infty, 0[.$$

Dans tous les cas, on obtient que la famille est régulière.

2. Montrer qu'il existe un unique estimateur du maximum de vraisemblance de b donné par

$$\hat{b}_n = \sqrt{\frac{1}{3n} \sum_{i=1}^n X_i^2}.$$

La log-vraisemblance s'écrit

$$\ell(b; x_1, \dots, x_n) = \sum_{i=1}^n \log f(x_i | b) = -3n \log b - \frac{1}{2b^2} \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \log \left(\sqrt{\frac{2}{\pi}} x_i \right).$$

Elle est dérivable sur $\mathbb{R}_+^*$ et

$$\frac{\partial}{\partial b} \ell(b; x_1, \dots, x_n) = -\frac{3n}{b} + \frac{1}{b^3} \sum_{i=1}^n x_i^2 = -\frac{3n}{b} \left(b^2 - \frac{1}{3n} \sum_{i=1}^n x_i^2 \right).$$

La log-vraisemblance admet donc un unique point critique dans $\mathbb{R}_+^*$ donnée par

$$\hat{b}_n = \sqrt{\frac{1}{3n} \sum_{i=1}^n X_i^2}.$$

Comme $f(\cdot | b)$ forme une famille exponentielle régulière, ce point critique est l'unique estimateur du maximum de vraisemblance de b . Alternativement, on aurait pu conclure en faisant un tableau de signe.

3. Montrer que $\hat{b}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} b$.

La statistique naturelle d'une famille exponentielle admet un moment d'ordre 1 et on a

$$\mathbb{E} [X_1^2] = -\frac{\partial}{\partial \theta} \log[a(\theta)] = -\frac{\partial}{\partial \theta} \log \left[\frac{4\sqrt{-\theta^3}}{\sqrt{\pi}} \right] = -\frac{3}{2\theta} = 3b^2.$$

$(X_n^2)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires *i.i.d.* comme transformation mesurable des $(X_n)_{n \in \mathbb{N}^*}$ qui sont *i.i.d.*. La loi forte des grands nombres donne alors

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 3b^2.$$

la fonction $x \mapsto \sqrt{x/3}$ étant continue sur $\mathbb{R}_+^*$, on en déduit que $\hat{b}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} b$.

4. Calculer l'information de Fisher apportée par X_1 sur b et celle apportée par X_1 sur le paramètre de la forme canonique.

La log-vraisemblance est deux fois différentiable et on a

$$\frac{\partial^2}{\partial b^2} \ell(b; x_1) = \frac{\partial}{\partial b} \left[-\frac{3}{b} + \frac{1}{b^3} x_1^2 \right] = \frac{3}{b^2} - \frac{3}{b^4} x_1^2.$$

On obtient alors l'information de Fisher apportée par X_1 sur b

$$I_{X_1}(b) = -\mathbb{E} \left[\frac{3}{b^2} - \frac{3}{b^4} X_1^2 \right] = -\frac{3}{b^2} + \frac{3}{b^4} \times 3b^3 = \frac{6}{b^2}.$$

On en déduit que l'information de Fisher apportée par X_1 sur $\theta = \eta b$ est

$$I_{X_1}(\theta) = \frac{\partial}{\partial \theta} \eta^{-1}(\theta) I_{X_1}(\eta^{-1}(\theta)) \frac{\partial}{\partial \theta} \eta^{-1}(\theta), \quad \text{avec} \quad b = \eta^{-1}(\theta) = \sqrt{-\frac{1}{2\theta}}.$$

On a donc

$$I_{X_1}(\theta) = \frac{3}{2\theta^2}.$$

5. En déduire qu'il existe une suite $(c_n)_{n \in \mathbb{N}^*}$, dont l'expression est une fonction de $X_1, \dots, X_n$ uniquement, telle que

$$c_n (\hat{b}_n - b) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

La famille exponentielle étant régulière, le théorème fondamental de la statistique donne

$$\sqrt{n} (\hat{b}_n - b) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N} \left(0, I_{X_1}^{-1}(b) \right) \equiv \mathcal{N} \left(0, \frac{b^2}{6} \right). \quad (1)$$

On a vu à la question 3. que $\hat{b}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} b$. La fonction $x \mapsto \sqrt{6}/x$ étant continue sur $\mathbb{R}_+^*$, on a

$$\frac{\sqrt{6}}{\hat{b}_n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \frac{\sqrt{6}}{b},$$

et le théorème de Slutsky permet alors de conclure

$$\frac{\sqrt{6n}}{\hat{b}_n} (\hat{b}_n - b) \xrightarrow[n \rightarrow \infty]{d} \frac{\sqrt{6}}{b} \mathcal{N} \left(0, \frac{b^2}{6} \right) \equiv \mathcal{N}(0, 1).$$

6. Montrer qu'il existe une fonction h bijective sur $\mathbb{R}_+^*$ telle que pour $\alpha \in]0, 1[$,

$$\mathbb{P} \left[\sqrt{n} \left| h(\hat{b}_n) - h(b) \right| \leq q_{1-\alpha/2} \right] \xrightarrow[n \rightarrow \infty]{} 1 - \alpha,$$

où $q_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi normale $\mathcal{N}(0, 1)$.

Pour obtenir le résultat souhaité, il suffit de montrer qu'il existe une fonction h inversible telle que

$$\sqrt{n} \left[h(\hat{b}_n) - h(b) \right] \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

Partant de (1), pour tout fonction h inversible, la méthode delta donne

$$\sqrt{n} \left[h(\hat{b}_n) - h(b) \right] \xrightarrow[n \rightarrow \infty]{d} \mathcal{N} \left(0, [h'(b)]^2 \frac{b^2}{6} \right).$$

Il suffit donc de montrer qu'il existe une fonction h inversible telle que

$$[h'(b)]^2 \frac{b^2}{6} = 1 \iff h'(b) = \frac{\sqrt{6}}{b}.$$

Il suffit donc de prendre $h : x \mapsto \sqrt{6} \log x$.

7. Montrer que $\hat{b}_n$ est une statistique complète et exhaustive pour b .

$T(X) = X^2$ est la statistique naturelle associée à une famille exponentielle régulière. On en déduit que $S(X_1, \dots, X_n) = \sum_{i=1}^n X_i^2$ est une statistique complète et exhaustive pour b . Comme $\hat{b}_n$ est une fonction bijective de $S(X_1, \dots, X_n)$, c'est également une statistique complète et exhaustive pour b .

8. Montrer que pour tout $c \in \mathbb{R}_+^*$, cX_1 admet pour densité $f(\cdot \mid bc)$. En déduire que $(X_1/b)^2$ suit la loi gamma de paramètre $(3/2, 1/2)$ (c.f. Formulaire).

Indication. $2\Gamma(3/2) = \sqrt{\pi}$.

Pour tout $c \in \mathbb{R}_+^*$, $x \mapsto cx$ réalise un $\mathcal{C}^1$ -difféomorphisme de $\mathbb{R}_+^*$ dans $\mathbb{R}_+^*$ et la densité de cX_1 est

$$f\left(\frac{x}{c} \mid b\right) \frac{1}{c} = \sqrt{\frac{2}{\pi}} \frac{x^2}{(bc)^3} \exp\left(-\frac{x^2}{2(bc)^2}\right) \mathbf{1}_{\{x>0\}} = f(x \mid bc).$$

$\phi : x \mapsto x^2$ réalise un $\mathcal{C}^1$ -difféomorphisme de $\mathbb{R}_+^*$ dans $\mathbb{R}_+^*$ et la densité de $(X_1/b)^2$ est

$$\begin{aligned} g(x) &= f\left(\phi^{-1}(x) \mid 1\right) \left| \frac{d}{dx} \phi^{-1}(x) \right| = \sqrt{\frac{2}{\pi}} x \exp\left(-\frac{x}{2}\right) \mathbf{1}_{\{\sqrt{x}>0\}} \times \frac{1}{2\sqrt{x}} \\ &= \left(\frac{1}{2}\right)^{\frac{3}{2}} \frac{1}{\Gamma(3/2)} x^{1/2} \exp\left(-\frac{x}{2}\right) \mathbf{1}_{x>0}. \end{aligned}$$

Il s'agit de la densité de la loi Gamma de paramètre $(3/2, 1/2)$.

9. On admet que pour tout $c \in \mathbb{R}_+^*$, $c \sum_{i=1}^n X_i^2$ suit la loi Gamma de paramètre $(3n/2, 1/(2cb^2))$. Calculer alors le biais de l'estimateur $\hat{b}_n$ par rapport à b . En déduire un estimateur sans biais de b , noté $\hat{\beta}_n$, s'exprimant comme une fonction mesurable de $\hat{b}_n$.

En utilisant le résultat de l'énoncé, on obtient

$$\hat{b}_n^2 = \frac{1}{3n} \sum_{i=1}^n X_i^2 \sim \text{Gamma}\left(\frac{3n}{2}, \frac{3n}{2b^2}\right).$$

Le théorème de transfert permet alors d'écrire que

$$\begin{aligned} \mathbb{E}[\hat{b}_n] &= \int_0^{+\infty} \sqrt{x} \left(\frac{3n}{2b^2}\right)^{\frac{3n}{2}} \frac{1}{\Gamma(3n/2)} x^{\frac{3n}{2}-1} \exp\left(-\frac{3n}{2b^2}x\right) dx \\ &= \underbrace{\sqrt{\frac{2b^2}{3n}} \frac{\Gamma(3n/2 + 1/2)}{\Gamma(3n/2)}}_{=1} \int_0^{+\infty} \sqrt{x} \left(\frac{3n}{2b^2}\right)^{\frac{3n+1}{2}} \frac{1}{\Gamma(3n/2 + 1/2)} x^{\frac{3n+1}{2}-1} \exp\left(-\frac{3n}{2b^2}x\right) dx. \end{aligned}$$

On en déduit que le biais de l'estimateur $\hat{b}_n$ par rapport à b est

$$\mathbb{E}[\hat{b}_n] - b = b \left[\sqrt{\frac{2}{3n}} \frac{\Gamma(3n/2 + 1/2)}{\Gamma(3n/2)} - 1 \right] \quad \text{et} \quad \hat{\beta}_n = \sqrt{\frac{3n}{2}} \frac{\Gamma(3n/2)}{\Gamma(3n/2 + 1/2)} \hat{b}_n.$$

11. Montrer que $\hat{\beta}_n$ est l'unique estimateur sans biais de b de variance minimale (UMVUE).

$\hat{\beta}_n$ est un estimateur sans biais de b . De plus la statistique $\hat{b}_n$ est une statistique complète et exhaustive pour b . Le théorème de Lehmann-Scheffé permet de conclure que $\mathbb{E}[\hat{\beta}_n | \hat{b}_n]$ est l'unique estimateur sans biais de b de variance minimale. Or $\hat{\beta}_n$ est une fonction mesurable de $\hat{b}_n$. On en déduit donc que $\mathbb{E}[\hat{\beta}_n | \hat{b}_n] = \hat{\beta}_n$.

12. Montrer que pour toute fonction g linéaire par rapport à chacune de ses coordonnées, $g(X_1, \dots, X_n)/\hat{b}_n$ est une statistique indépendante de $\hat{b}_n$.

D'après la question 8., si $X_1, \dots, X_n$ suivent la loi $f(\cdot | b)$ alors $Y_1 = X_1/b, \dots, Y_n = X_n/b$ suivent la loi $f(\cdot | 1)$. Toute fonction mesurable (indépendante de b) de $Y_1, \dots, Y_n$ admettra donc une distribution indépendante de b . On en déduit en particulier que

$$\frac{g(X_1, \dots, X_n)}{\hat{b}_n} = \frac{bg(Y_1, \dots, Y_n)}{b\sqrt{\frac{1}{3n} \sum_{i=1}^n Y_i^2}} = \frac{g(Y_1, \dots, Y_n)}{\sqrt{\frac{1}{3n} \sum_{i=1}^n Y_i^2}}$$

est une statistique libre. Par ailleurs, $\hat{b}_n$ étant exhaustive et complète, le théorème de Basu assure que $g(X_1, \dots, X_n)/\hat{b}_n$ et $\hat{b}_n$ sont indépendantes.