

Examen de statistique

Exercice 1 : On considère des variables aléatoires i.i.d $X_1, \dots, X_n$ ayant pour densité f_θ définie pour tout $x \in \mathbb{R}$ par

$$f_\theta(x) = \frac{3\theta^3}{x^4} \mathbf{1}_{x \geq \theta}.$$

avec $\theta > 0$.

1. Montrer que

$$\mathbb{E}[X] = \frac{3}{2}\theta \quad \text{et} \quad \mathbb{V}[X] = \frac{3}{4}\theta^2.$$

2. En déduire un estimateur de θ , que l'on notera θ_n . Est-il consistant ?
3. Donner sa normalité asymptotique.
4. Soit $\alpha \in (0, 1)$, construire un intervalle de confiance asymptotique de θ de niveau $1 - \alpha$.
5. Soit $\theta_0 > 0$, en déduire un test asymptotique de niveau α pour tester

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta \neq \theta_0.$$

6. Montrer que l'estimateur du maximum de vraisemblance est $\hat{\theta}_n = X_{(1)}$ où $X_{(1)} = \min X_i$.
7. Donner la convergence en loi de $n(\hat{\theta}_n - \theta)$.
8. Construire un intervalle de confiance de niveau $1 - \alpha$ de la forme

$$IC_{1-\alpha}(\theta) = [\hat{\theta}_n c_\alpha; \hat{\theta}_n]$$

9. Soit $\theta_0 > 0$, construire un test de taille α pour tester

$$H_0 : \theta \leq \theta_0 \quad \text{contre} \quad \theta > \theta_0.$$

10. Soit $x_{(1)}$ une réalisation de $X_{(1)}$. Calculer la p -valeur associée. On prendra la convention $\inf(\emptyset) = 0$.

Correction :

1. On a

$$\mathbb{E}[X] = \int_{\theta}^{+\infty} \frac{3\theta^3}{x^3} dx = \left[-\frac{3\theta^3}{2x^2} \right]_{\theta}^{+\infty} = \frac{3}{2}\theta.$$

De plus, on a

$$\mathbb{E}[X^2] = \int_{\theta}^{+\infty} \frac{3\theta^3}{x^2} dx = \left[-\frac{3\theta^3}{x} \right]_{\theta}^{+\infty} = 3\theta^2.$$

On obtient donc

$$\mathbb{V}[X] = 3\theta^2 - \frac{9}{4}\theta^2 = \frac{3}{4}\theta^2.$$

2. On considère donc l'estimateur $\theta_n = \frac{2}{3}\bar{X}_n$. Par la loi des grands nombres, on a

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{3}{2}\theta$$

et on obtient donc la convergence de θ_n (en multipliant par 2/3).

3. Le TCL nous donne

$$\sqrt{n} \left(\bar{X}_n - \frac{3}{2}\theta \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \frac{3}{4}\theta^2 \right),$$

et en multipliant par 2/3, on obtient

$$\sqrt{n} (\theta_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \frac{1}{3}\theta^2 \right).$$

4. On peut réécrire la normalité asymptotique comme

$$\sqrt{3n} \frac{\theta_n - \theta}{\theta} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Option 1 : On a

$$\sqrt{3n} \frac{\theta_n - \theta}{\theta_n} = \frac{\theta}{\theta_n} \sqrt{3n} \frac{\theta_n - \theta}{\theta}.$$

Comme la fonction $g : x \mapsto x^{-1}$ est continue en θ , par le théorème de continuité, on a

$$g(\theta_n) = \frac{1}{\theta_n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} g(\theta) = \frac{1}{\theta},$$

et en particulier

$$\frac{\theta}{\theta_n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 1.$$

Par le théorème de Slutsky, on obtient

$$\sqrt{3n} \frac{\theta_n - \theta}{\theta_n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

et donc l'intervalle de confiance asymptotique

$$IC_{1-\alpha}(\theta) = \left[\theta_n \pm q_{1-\alpha/2} \frac{\theta_n}{\sqrt{3n}} \right]$$

Option 2 : On peut remarquer

$$\sqrt{3n} \frac{\theta_n - \theta}{\theta} = \sqrt{3n} \left(\frac{\theta_n}{\theta} - 1 \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

On note $q_{1-\alpha/2}$ le quantile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite, et on a

$$\begin{aligned} 1 - \alpha &= \lim_{n \rightarrow +\infty} \mathbb{P} \left[-q_{1-\alpha/2} \leq \sqrt{3n} \left(\frac{\theta_n}{\theta} - 1 \right) \leq q_{1-\alpha/2} \right] = \lim_{n \rightarrow +\infty} \mathbb{P} \left[\frac{1 - \frac{q_{1-\alpha/2}}{\sqrt{3n}}}{\frac{\theta_n}{\theta}} \leq \frac{1}{\theta} \leq \frac{1 + \frac{q_{1-\alpha/2}}{\sqrt{3n}}}{\frac{\theta_n}{\theta}} \right] \\ &= \lim_{n \rightarrow +\infty} \mathbb{P} \left[\frac{\theta_n}{1 - \frac{q_{1-\alpha/2}}{\sqrt{3n}}} \geq \theta \geq \frac{\theta_n}{1 + \frac{q_{1-\alpha/2}}{\sqrt{3n}}} \right] \end{aligned}$$

la dernière égalité n'étant vraie que pour n suffisamment grand, i.e. tel que $1 - \frac{q_{1-\alpha/2}}{\sqrt{3n}} \geq 0$.

5. Le test asymptotique basé sur l'intervalle de confiance précédent revient à

$$\text{"On rejette } H_0 \text{"} \iff \theta_0 \notin IC_{1-\alpha}(\theta) \iff |\theta_n - \theta_0| > \frac{q_{1-\alpha/2}}{\sqrt{3n}} \theta_n.$$

On obtient donc la zone de rejet

$$ZR = \left\{ |\theta_n - \theta_0| > \frac{q_{1-\alpha/2}}{\sqrt{3n}} \theta_n \right\}.$$

6. On a la vraisemblance

$$L_n(\theta) = \prod_{i=1}^n f_\theta(X_i) = \prod_{i=1}^n \frac{3\theta^3}{X_i^4} \mathbf{1}_{X_i \geq \theta} = 3^n \theta^{3n} \mathbf{1}_{\theta \leq X_{(1)}} \prod_{i=1}^n \frac{1}{X_i^4}$$

qui est nulle sur $(X_{(1)}, +\infty)$ et croissant sur $(0, X_{(1)})$ et le maximum est donc atteint en $X_{(1)}$.

7. On a, avec les calculs usuels, pour tout $x \geq 0$,

$$\begin{aligned} \mathbb{P} \left[n \left(X_{(1)} - \theta \right) \leq x \right] &= \mathbb{P} \left[X_{(1)} \leq \theta + \frac{x}{n} \right] = 1 - \mathbb{P} \left[X_{(1)} \geq \theta + \frac{x}{n} \right] \\ &= 1 - \left(\mathbb{P} \left[X_1 \geq \theta + \frac{x}{n} \right] \right)^n \end{aligned}$$

De plus, on a

$$\mathbb{P} \left[X_1 \geq \theta + \frac{x}{n} \right] = \int_{\theta+x/n}^{+\infty} \frac{3\theta^3}{u^4} du = \left[-\frac{\theta^3}{u^3} \right]_{\theta+x/n}^{+\infty} = \left(\frac{\theta}{\theta + x/n} \right)^3 = \left(\frac{1}{1 + x/\theta n} \right)^3 = \left(1 + \frac{x}{\theta n} \right)^{-3}.$$

On obtient donc

$$\mathbb{P} \left[n \left(X_{(1)} - \theta \right) \leq x \right] = 1 - \left(1 + \frac{x}{\theta n} \right)^{-3n} = 1 - \exp \left(-3n \log \left(1 + \frac{x}{\theta n} \right) \right) \xrightarrow[n \rightarrow +\infty]{} 1 - e^{-\frac{3}{\theta} x}$$

et on obtient donc

$$n \left(X_{(1)} - \theta \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{E} \left(\frac{3}{\theta} \right).$$

8. On cherche c_α tel que

$$\begin{aligned} 1 - \alpha = \mathbb{P} \left[X_{(1)} c_\alpha \leq \theta \right] &= 1 - \mathbb{P} \left[X_{(1)} \geq \frac{\theta}{c_\alpha} \right] \iff \mathbb{P} \left[X_{(1)} \geq \frac{\theta}{c_\alpha} \right] = \alpha \\ &\iff \left(\mathbb{P} \left[X_1 \geq \frac{\theta}{c_\alpha} \right] \right)^n = \alpha \\ &\iff c_\alpha^{3n} = \alpha \\ &\iff c_\alpha = \alpha^{1/3n}. \end{aligned}$$

On obtient donc l'intervalle de confiance

$$IC_{1-\alpha}(\theta) = \left[X_{(1)} \alpha^{1/3n}; X_{(1)} \right].$$

9. Comme $X_{(1)}$ peut être aussi petit que l'on veut, on cherche c_α tel que

$$\alpha = \sup_{\theta \leq \theta_0} \mathbb{P}_\theta \left[X_{(1)} \geq c_\alpha \right] = \sup_{\theta \leq \theta_0} \left(\frac{\theta}{c_\alpha} \right)^{3n} = \left(\frac{\theta_0}{c_\alpha} \right)^{3n} \iff c_\alpha = \theta_0 \alpha^{-1/3n}.$$

On a donc la zone de rejet

$$ZR = \left\{ X_{(1)} \geq \theta_0 \alpha^{-1/3n} \right\}.$$

10. On a

$$p\text{-valeur} = \inf \left\{ \alpha \in [0, 1], x_{(1)} \geq \theta_0 \alpha^{-1/3n} \right\} = \inf \left\{ \alpha \in [0, 1], \alpha \geq \left(\frac{\theta_0}{x_{(1)}} \right)^{3n} \right\} = \begin{cases} \left(\frac{\theta_0}{x_{(1)}} \right)^{3n} & \text{si } \theta_0 \leq x_{(1)} \\ 0 & \text{sinon} \end{cases}$$

Exercice 2 : On considère des variables aléatoires $X_1, \dots, X_n$ i.i.d avec $X_1 \sim \mathcal{N}(0, \theta)$ et $\theta > 0$.

1. Montrer que l'estimateur du maximum de vraisemblance est

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

2. Donner la loi de $\frac{n\hat{\theta}_n}{\theta}$.

3. Soit $\alpha \in (0, 1)$. En déduire un intervalle de confiance de niveau $1 - \alpha$ de θ .

4. Soit $\theta_0 > 0$, en déduire un test de niveau α pour tester

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta \neq \theta_0.$$

Correction :

1. On a la vraisemblance

$$L_n(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{X_i^2}{2\theta}} = (2\pi\theta)^{-n/2} e^{-\frac{1}{2\theta} \sum_{i=1}^n X_i^2}$$

et la log-vraisemblance

$$\ell_n(\theta) = -\frac{n}{2} \log(\theta) - \frac{n}{2} \log(2\pi) - \frac{1}{2\theta} \sum_{i=1}^n X_i^2$$

et donc

$$\ell'_n(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^n X_i^2 = \frac{-n\theta + \sum_{i=1}^n X_i^2}{2\theta^2} \geq 0 \iff \theta \leq \frac{1}{n} \sum_{i=1}^n X_i^2$$

et on obtient donc le résultat.

2. Comme les $X_i/\sqrt{\theta}$ sont i.i.d et suivent des loi normales centrées réduites, on a

$$\frac{n\hat{\theta}_n}{\theta} = \sum_{i=1}^n \left(\frac{X_i}{\sqrt{\theta}} \right)^2 \sim \chi_n^2.$$

3. En notant $k_{\alpha/2}$ et $k_{1-\alpha/2}$ les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ d'une chi deux à n degrés de liberté, on a

$$1 - \alpha = \mathbb{P} \left[k_{\alpha/2} \leq \frac{n\hat{\theta}_n}{\theta} \leq k_{1-\alpha/2} \right] = \mathbb{P} \left[\frac{k_{\alpha/2}}{n\hat{\theta}_n} \leq \frac{1}{\theta} \leq \frac{k_{1-\alpha/2}}{n\hat{\theta}_n} \right] = \mathbb{P} \left[\frac{n\hat{\theta}_n}{k_{\alpha/2}} \geq \theta \geq \frac{n\hat{\theta}_n}{k_{1-\alpha/2}} \right]$$

et on obtient l'intervalle de confiance

$$IC_{1-\alpha}(\theta) = \left[\frac{n\hat{\theta}_n}{k_{1-\alpha/2}}; \frac{n\hat{\theta}_n}{k_{\alpha/2}} \right].$$

4. Le test consiste en

$$\text{"On rejette } H_0 \text{"} \iff \theta_0 \notin IC_{1-\alpha}(\theta)$$

et on obtient la zone de rejet

$$ZR = \left\{ \theta_0 < \frac{n\hat{\theta}_n}{k_{1-\alpha/2}} \right\} \cup \left\{ \theta_0 > \frac{n\hat{\theta}_n}{k_{\alpha/2}} \right\}.$$

Rappel sur le test du rapport de vraisemblance : Soit $\mathbf{X} = (X_1, \dots, X_n)$ un échantillon et on suppose que X_1 admet une densité f_θ avec $\theta \in \Theta$. Soit $\theta_0, \theta_1 \in \Theta$, on souhaite tester

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta = \theta_1.$$

On considère le rapport de vraisemblance

$$\Phi(\mathbf{X}) = \frac{L_{\mathbf{X}}(\theta_1)}{L_{\mathbf{X}}(\theta_0)} = \frac{\prod_{i=1}^n f_{\theta_1}(X_i)}{\prod_{i=1}^n f_{\theta_0}(X_i)}.$$

Le test (de niveau α) de rapport de vraisemblance consiste à construire une zone de rejet de la forme

$$ZR = \left\{ \frac{\prod_{i=1}^n f_{\theta_1}(X_i)}{\prod_{i=1}^n f_{\theta_0}(X_i)} > c_\alpha \right\} \quad \text{avec} \quad \mathbb{P}_{\theta_0} \left[\frac{\prod_{i=1}^n f_{\theta_1}(X_i)}{\prod_{i=1}^n f_{\theta_0}(X_i)} > c_\alpha \right] = \alpha.$$

Exercice 3 : On considère des variables aléatoires i.i.d $X_1, \dots, X_n$ admettant pour densité

$$f_\theta(x) = \frac{\theta}{x^{\theta+1}} \mathbf{1}_{x \geq 1}$$

avec $\theta > 0$.

1. Montrer que $\log(X_1) \sim \mathcal{E}(\theta)$.
2. On admettra que si on a $Y_1, \dots, Y_n \sim \mathcal{E}(\theta)$, alors $\theta \sum_{i=1}^n Y_i$ suit une loi gamma de paramètre $n, 1$ dont on connaît les quantiles. On notera γ_α le quantile d'ordre α de cette loi. Soit $\theta_0 > \theta_1 > 0$. Tester au niveau α

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta = \theta_1.$$

Correction :

1. Pour tout $x \geq 0$, on a

$$\mathbb{P}[\log(X_1) \leq x] = \mathbb{P}[X_1 \leq e^x] = \int_1^{+\infty} \frac{\theta}{t^{\theta+1}} dt = \left[\frac{-1}{t^\theta} \right]_1^{e^x} = 1 - e^{-\theta x}.$$

2. On a le rapport de vraisemblance

$$\Phi(\mathbf{X}) = \frac{\prod_{i=1}^n \theta_1 X_i^{-1-\theta_1}}{\prod_{i=1}^n \theta_0 X_i^{-1-\theta_0}} = \frac{\theta_1^n}{\theta_0^n} \prod_{i=1}^n X_i^{\theta_0 - \theta_1}.$$

On cherche donc c_α t.q

$$\begin{aligned}
 \alpha &= \mathbb{P}_{\theta_0} \left[\frac{\theta_1^n}{\theta_0^n} \prod_{i=1}^n X_i^{\theta_0 - \theta_1} \geq c_\alpha \right] = \mathbb{P}_{\theta_0} \left[\prod_{i=1}^n X_i^{\theta_0 - \theta_1} \geq \frac{\theta_0^n}{\theta_1^n} c_\alpha \right] \\
 &= \mathbb{P}_{\theta_0} \left[(\theta_0 - \theta_1) \sum_{i=1}^n \log(X_i) \geq \log \left(\frac{\theta_0^n}{\theta_1^n} c_\alpha \right) \right] \\
 &= \mathbb{P}_{\theta_0} \left[\sum_{i=1}^n \log(X_i) \geq \frac{1}{\theta_0 - \theta_1} \log \left(\frac{\theta_0^n}{\theta_1^n} c_\alpha \right) \right] \\
 &= \mathbb{P}_{\theta_0} \left[\theta_0 \sum_{i=1}^n \log(X_i) \geq \frac{\theta_0}{\theta_0 - \theta_1} \log \left(\frac{\theta_0^n}{\theta_1^n} c_\alpha \right) \right]
 \end{aligned}$$

ce que l'on peut réécrire comme

$$1 - \mathbb{P}_{\theta_0} \left[\theta_0 \sum_{i=1}^n \log(X_i) \leq \frac{\theta_0}{\theta_0 - \theta_1} \log \left(\frac{\theta_0^n}{\theta_1^n} c_\alpha \right) \right] = 1 - \alpha$$

i.e on a

$$\frac{\theta_0}{\theta_0 - \theta_1} \log \left(\frac{\theta_0^n}{\theta_1^n} c_\alpha \right) = \gamma_{1-\alpha}$$

et donc

$$c_\alpha = \frac{\theta_1^n}{\theta_0^n} \exp \left(\frac{\theta_0 - \theta_1}{\theta_0} \gamma_{1-\alpha} \right).$$

On obtient donc la zone de rejet

$$ZR = \left\{ \frac{\theta_1^n}{\theta_0^n} \prod_{i=1}^n X_i^{\theta_0 - \theta_1} \geq \frac{\theta_1^n}{\theta_0^n} \exp \left(\frac{\theta_0 - \theta_1}{\theta_0} \gamma_{1-\alpha} \right) \right\} = \left\{ \sum_{i=1}^n \log(X_i) \geq \frac{\theta_0 - \theta_1}{\theta_0} \gamma_{1-\alpha} \right\}.$$