Sorbonne Université
Master Mathématiques et Applications
MU4MA015 : Statistique

7 octobre 2022
Aucun document autorisé
Durée : 1 heure

Contrôle Continu

Rappels

- Il sera tenu compte du soin apporté à la présentation et à la rédaction. Si vous n'arrivez pas à démontrer un résultat, vous pouvez l'admettre pour la suite de l'exercice.
- On notera comme d'habitude $X_{(1)} = \min_{1\leq i\leq n} X_i$.

Soit $\theta > 0$ un paramètre inconnu et X de densité

$$f_\theta(x) = \frac{3\theta^3}{x^4}\mathbf{1}_{[\theta,+\infty[}(x).$$

Dans la suite, on suppose disposer d'un échantillon $(X_1, \ldots, X_n)$ i.i.d. de même loi que X.

1. Calculer $\mathbb{E}_\theta[X]$ et $\mathbb{E}_\theta[X^2]$. En déduire un estimateur $\widehat{\theta}_n$ de θ par la méthode des moments. Démontrer qu'il est fortement consistant et asymptotiquement normal.
Correction. Pour $k < 3$, $\mathbb{E}[X^k] = \frac{3}{3-k}\theta^k$. D'où $\mathbb{E}[X] = \frac{3}{2}\theta$, $\mathbb{E}[X^2] = 3\theta^2$ et $\mathrm{Var}(X) = \frac{3}{4}\theta^2$. La loi forte des grands nombres implique que $\widehat{\theta}_n := \frac{2}{3}\overline{X}_n$ est fortement consistant. Le Théorème Central Limite montre que $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\theta^2}{3}\right)$.

2. Soit $0 < \alpha < 1$. Déduire de la question précédente un intervalle de confiance de niveau asymptotique $(1-\alpha)$ pour θ.
Correction. Première solution : $\sqrt{3n}\left(\frac{\widehat{\theta}_n}{\theta} - 1\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ puis, en notant $q := \Phi^{-1}(1-\alpha/2)$,
$$IC^1_{1-\alpha} = \left[\frac{\widehat{\theta}_n}{1+q/\sqrt{3n}}\ ;\ \frac{\widehat{\theta}_n}{1-q/\sqrt{3n}}\right].$$
Seconde solution :
$$\sqrt{3n}\,\frac{\widehat{\theta}_n - \theta}{\widehat{\theta}_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$
d'où
$$IC^2_{1-\alpha} = \left[\widehat{\theta}_n - \frac{q\widehat{\theta}_n}{\sqrt{3n}}\ ;\ \widehat{\theta}_n + \frac{q\widehat{\theta}_n}{\sqrt{3n}}\right].$$

3. Proposer un test de niveau asymptotique α pour décider entre $H_0 : \theta = 1$ et $H_1 : \theta \neq 1$.
Correction. Première solution :
$$T_1(\mathbf{X}) = \mathbf{1}_{1\notin IC^1_{1-\alpha}} = \mathbf{1}_{|\widehat{\theta}_n - 1| > q/\sqrt{3n}}.$$

Seconde solution :
$$T_2(\mathbf{X}) = \mathbf{1}_{1 \notin IC^2_{1-\alpha}} = \mathbf{1}_{|\widehat{\theta}_n - 1| > q\widehat{\theta}_n/\sqrt{3n}}.$$

4. Calculer la fonction de répartition F_θ de X. En déduire la médiane de la loi de X.
Correction. $\forall x \in \mathbb{R}, F_\theta(x) = \left(1 - \frac{\theta^3}{x^3}\right) \mathbf{1}_{[\theta,\infty]}(x)$. Puisque F_θ réalise une bijection de $]\theta, \infty[$ dans $]0, 1[$, on trouve $x_{1/2} = 2^{1/3}\theta$.
5. En déduire un estimateur $\widetilde{\theta}_n$ de θ. Établir sa normalité asymptotique.
Correction. On pose $\widetilde{\theta}_n = 2^{-1/3}x_{1/2}(n)$. On a $f_\theta(x_{1/2}) = \frac{3}{2^{4/3}\theta} > 0$ et $\sqrt{n}(x_{1/2}(n) - x_{1/2} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{2^{2/3}\theta^2}{9}\right)$. Par suite, $\sqrt{n}(\widetilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\theta^2}{9}\right)$.
6. Entre $\widehat{\theta}_n$ et $\widetilde{\theta}_n$, quel estimateur choisissez-vous ?
Correction. $\frac{\theta^2}{3} > \frac{\theta^2}{9}$ quel que soit $\theta > 0$ donc $\widetilde{\theta}_n$ est préférable.
7. Déterminer l'estimateur du maximum de vraisemblance $\check{\theta}_n$ de θ.
Correction. $\forall \theta' > 0, L_n(\theta') = \frac{3^n \theta'^{3n}}{\prod_{i=1}^n X_i^4} \mathbf{1}_{]0,X_{(1)}]}(\theta')$, fonction croissante sur $]0, X_{(1)}]$ puis nulle après $X_{(1)}$. Donc l'EMV est $\check{\theta}_n = X_{(1)}$.
8. Pour tout $t > 0$, calculer $\mathbb{P}(n(X_{(1)} - \theta) > t)$, puis $\lim_{n\to\infty} \mathbb{P}(n(X_{(1)} - \theta) > t)$. En déduire que $n(\check{\theta}_n - \theta)$ converge en loi vers une loi exponentielle dont on précisera le paramètre.
Correction. D'après le calcul de la fonction de répartition F_θ ci-dessus, $\forall t > 0$,
$$\mathbb{P}(n(X_{(1)} - \theta) > t) = \left(1 + \frac{t}{\theta n}\right)^{-3n} \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-3t/\theta},$$
donc $\forall t > 0$,
$$\mathbb{P}(n(X_{(1)} - \theta) \leq t) \underset{n\to\infty}{\longrightarrow} 1 - \mathrm{e}^{-3t/\theta}.$$
Puisque $n(X_{(1)} - \theta)$ est une variable aléatoire positive, il s'ensuit que $\forall t \in \mathbb{R}$
$$\mathbb{P}(n(X_{(1)} - \theta) \leq t) \underset{n\to\infty}{\longrightarrow} \left(1 - \mathrm{e}^{-3t/\theta}\right) \mathbf{1}_{t\geq 0},$$
donc $n(\check{\theta}_n - \theta)$ converge en loi vers une loi exponentielle de paramètre $3/\theta$.
9. Entre $\widehat{\theta}_n$, $\widetilde{\theta}_n$ et $\check{\theta}_n$, quel estimateur choisissez-vous ?
Correction. Ainsi $\check{\theta}_n$ converge à vitesse $1/n$ vers θ, tandis que les deux précédents estimateurs convergent à vitesse $1/\sqrt{n}$. L'EMV $\check{\theta}_n$ est donc préférable.