Sorbonne Université
Master Mathématiques et Applications
MU4MA015 : Statistique

14 octobre 2024
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.

Exercice 1 (7 points)
Soit $\theta > 0$ un paramètre inconnu et $X_1, \dots, X_n$ des variables aléatoires i.i.d. suivant une loi exponentielle de paramètre $\frac{1}{\theta}$, c'est-à-dire de densité

$$f_\theta(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \mathbf{1}_{x>0}.$$

1. Rappeler ce que valent $\mathbb{E}[X_1]$ et $\mathbb{V}(X_1)$. En déduire un estimateur $\widehat{\theta}_n$ de θ dont vous justifierez la consistance et la normalité asymptotique.
   Correction. On sait que $\mathbb{E}[X_1] = \theta$ et $\mathbb{V}(X_1) = \theta^2$. Par la méthode des moments, on peut considérer $\widehat{\theta}_n = \overline{X}_n$. Il est fortement consistant par la loi forte des grands nombres. De plus, le TCL donne

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \theta^2\right),$$

   ce qui prouve qu'il est asymptotiquement normal.
2. Soit $\alpha \in]0,1[$. Déterminer un intervalle de confiance de niveau asymptotique $(1-\alpha)$ pour θ.
   Correction. On déduit de la question précédente que

$$\sqrt{n}\left(\frac{\widehat{\theta}_n}{\theta} - 1\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

   Ainsi, en notant $q_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$ le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite, un intervalle de confiance de niveau asymptotique $(1-\alpha)$ est

$$I = \left[\frac{\widehat{\theta}_n}{1+\frac{q_{1-\alpha/2}}{\sqrt{n}}} \;;\; \frac{\widehat{\theta}_n}{1-\frac{q_{1-\alpha/2}}{\sqrt{n}}}\right].$$

   Remarque : On peut aussi appliquer le Lemme de Slutsky pour déduire de la question précédente que

$$\sqrt{n}\frac{\widehat{\theta}_n - \theta}{\widehat{\theta}_n} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1),$$

   d'où l'on déduit l'intervalle

$$J = \left[\widehat{\theta}_n - \frac{q_{1-\alpha/2}\widehat{\theta}_n}{\sqrt{n}} \;;\; \widehat{\theta}_n + \frac{q_{1-\alpha/2}\widehat{\theta}_n}{\sqrt{n}}\right].$$

Ces intervalles sont asymptotiquement équivalents puisque

$$\frac{1}{1 \pm \frac{q_{1-\alpha/2}}{\sqrt{n}}} \underset{n\to\infty}{\sim} 1 \mp \frac{q_{1-\alpha/2}}{\sqrt{n}}.$$

3. Calculer la médiane de la loi de X_1. En déduire un nouvel estimateur $\widetilde{\theta}_n$ de θ et préciser sa normalité asymptotique.
Correction. On a $F_\theta(x) = (1 - e^{-\frac{x}{\theta}})\mathbf{1}_{x\geq 0}$ donc $x_{1/2} = \theta \log 2$, et on choisit donc l'estimateur $\widetilde{\theta}_n = x_{1/2}(n)/\log 2$, où $x_{1/2}(n) = X_{(\lceil \frac{n}{2} \rceil)}$ est la médiane empirique. On a $f_\theta(x_{1/2}) = \frac{1}{2\theta} > 0$, d'où par le résultat de normalité asymptotique de la médiane empirique

$$\sqrt{n}\left(x_{1/2}(n) - \theta \log 2\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \theta^2\right)$$

donc

$$\sqrt{n}\left(\widetilde{\theta}_n - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \left(\frac{\theta}{\log 2}\right)^2\right).$$

4. Lorsque n est grand, entre $\widehat{\theta}_n$ et $\widetilde{\theta}_n$, quel estimateur choisiriez-vous ?
Correction. Puisque $0 < \log 2 < 1$, $\widehat{\theta}_n$ a une plus petite variance asymptotique et est donc préférable.

5. Déterminer l'estimateur du maximum de vraisemblance.
Correction. On a la vraisemblance

$$L_n(\theta) = \frac{1}{\theta^n} \mathrm{e}^{-\frac{1}{\theta} n \overline{X}_n} \prod_{i=1}^n \mathbf{1}_{X_i > 0} = \frac{1}{\theta^n} \mathrm{e}^{-\frac{1}{\theta} n \overline{X}_n},$$

où l'on a pu supprimer le produit d'indicatrices car les X_i sont p.s. strictement positives. On passe à la log-vraisemblance puis à sa dérivée pour en déduire que

$$\ell_n'(\theta) = -\frac{n}{\theta} + \frac{n\overline{X}_n}{\theta^2} \geq 0 \iff \theta \leq \overline{X}_n,$$

ce qui prouve que que $\overline{X}_n$ est l'unique maximiseur, donc l'EMV est $\widehat{\theta}_n = \overline{X}_n$.

6. En remarquant que X_1 suit une loi $\gamma(1, \frac{1}{\theta})$, en déduire la loi de $(n\overline{X}_n)/\theta$.
Correction. On a $\frac{n\overline{X}_n}{\theta} \sim \gamma(n, 1)$.

7. Pour tous paramètres $p > 0$ et $\lambda > 0$, on note $F_{p,\lambda}$ la fonction de répartition d'une loi $\gamma(p, \lambda)$. Celle-ci étant continue et strictement croissante de $[0, +\infty[$ vers $[0, 1[$, elle admet une fonction réciproque $F_{p,\lambda}^{-1}$ de $[0, 1[$ vers $[0, +\infty[$. Proposer un test de taille (non asymptotique) α pour tester

$$H_0 : \theta \leq 3 \qquad \text{contre} \qquad H_1 : \theta > 3.$$

Correction. Il est naturel de chercher un test qui rejette H_0 lorsque $\overline{X}_n > c_\alpha$, où $c_\alpha > 0$ doit être calibré de telle sorte que la taille soit α. On cherche donc c_α tel que

$$\alpha = \sup_{\theta\in[0,3]} \mathbb{P}_\theta\left(\overline{X}_n > c_\alpha\right) = \sup_{\theta\in[0,3]} \mathbb{P}_\theta\left(\frac{n\overline{X}_n}{\theta} > \frac{nc_\alpha}{\theta}\right) = \sup_{\theta\in[0,3]} 1 - F_{n,1}\left(\frac{nc_\alpha}{\theta}\right) = 1 - F_{n,1}\left(\frac{nc_\alpha}{3}\right),$$

la dernière égalité venant de la croissance de $x \mapsto F_{n,1}(x)$. On obtient donc

$$c_\alpha = \frac{3F_{n,1}^{-1}(1-\alpha)}{n}$$

et le test consistant à rejeter H_0 si et seulement si

$$\overline{X}_n > \frac{3F_{n,1}^{-1}(1-\alpha)}{n}$$

est donc de taille α.

Remarque : Pour arriver au même résultat, on peut aussi partir d'un intervalle de confiance (non asymptotique) de niveau $(1-\alpha)$. Puisque le test est unilatère, on cherche un intervalle unilatère de sens opposé à $\Theta_0 =]0, 3]$, c'est-à-dire de la forme $[d_\alpha, +\infty[$. Puisque $\frac{n\overline{X}_n}{\theta} \sim \gamma(n, 1)$, on a pour tout θ

$$1-\alpha = \mathbb{P}_\theta\left(\frac{n\overline{X}_n}{\theta} \leq F_{n,1}^{-1}(1-\alpha)\right) = \mathbb{P}_\theta\left(\theta \geq \frac{n\overline{X}_n}{F_{n,1}^{-1}(1-\alpha)}\right)$$

d'où l'intervalle de confiance

$$I = \left[\frac{n\overline{X}_n}{F_{n,1}^{-1}(1-\alpha)}\ ;\ +\infty\right[$$

valable en particulier pour tout $\theta \leq 3$. Par conséquent un test de niveau α consiste à rejeter H_0 si et seulement si $I \cap \Theta_0 = \emptyset$, c'est-à-dire si et seulement si

$$\overline{X}_n > \frac{3F_{n,1}^{-1}(1-\alpha)}{n}.$$

Le calcul ci-dessus assure qu'il est en fait de taille α.

8. Soit $\overline{x}_n$ une réalisation de $\overline{X}_n$, déterminer la p-valeur $\alpha_0(\overline{x}_n)$ associée.
Correction. Pour la p-valeur, on utilise à nouveau la monotonie de $F_{n,1}$:

$$\begin{aligned}
\alpha_0(\overline{x}_n) &= \inf\{\alpha \in [0,1],\ \text{rejet de } H_0 \text{ au niveau } \alpha\} \\
&= \inf\left\{\alpha \in [0,1],\ \overline{x}_n > \frac{3F_{n,1}^{-1}(1-\alpha)}{n}\right\} \\
&= \inf\left\{\alpha \in [0,1],\ \alpha > 1 - F_{n,1}\left(\frac{n\overline{x}_n}{3}\right)\right\} \\
&= 1 - F_{n,1}\left(\frac{n\overline{x}_n}{3}\right).
\end{aligned}$$

Exercice 2 (3 points)

Soit $\theta \in]0,1[$ et X une variable aléatoire suivant une loi de Rademacher $\mathcal{R}(\theta)$ de paramètre θ, c'est-à-dire prenant les valeurs -1 et $+1$ avec, pour tout $x \in \{-1, 1\}$,

$$\mathbb{P}(X = x) = \theta^{\frac{1+x}{2}}(1-\theta)^{\frac{1-x}{2}}.$$

1. Vérifier que si la variable B suit une loi de Bernoulli de paramètre θ, alors $X = 2B - 1$ suit une loi de Rademacher de paramètre θ. En déduire $\mathbb{E}[X]$ et $\mathbb{V}(X)$.
Correction. D'après la formule donnée, une variable qui suit une loi de Rademacher de paramètre θ prend les valeurs 1 et -1 avec les probabilités respectives θ et $1-\theta$. Or, si B suit une loi de Bernoulli de paramètre θ, elle est à valeurs dans $\{0, 1\}$, donc $X = 2B - 1$ est à valeurs dans $\{-1, 1\}$ avec

$$\mathbb{P}(X = 1) = \mathbb{P}(B = 1) = \theta = 1 - \mathbb{P}(X = -1).$$

On en déduit immédiatement que $\mathbb{E}[X] = 2\theta - 1$ et $\mathbb{V}(X) = 4\theta(1-\theta)$.

2. Le modèle $(\mathcal{R}_\theta)_{0<\theta<1}$ est-il régulier ? Si oui, préciser l'information de Fisher $I(\theta) = I_1(\theta)$ pour une seule donnée.

   Correction. On a $\mathcal{R}_\theta = f_\theta \cdot \mu$ où μ est la mesure de comptage sur $\{-1, 1\}$ et, pour tout $x \in \{-1, 1\}$,
$$f_\theta(x) = \theta^{\frac{1+x}{2}}(1-\theta)^{\frac{1-x}{2}} = e^{\frac{1+x}{2}\log\theta + \frac{1-x}{2}\log(1-\theta)}.$$
   Pour tout $x \in \{-1, 1\}$, la fonction $\theta \mapsto f_\theta(x)$ est C^1 sur $]0, 1[$ donc les deux premiers points de la définition d'un modèle régulier sont vérifiés. De plus, la log-vraisemblance s'écrit
$$\ell_\theta(X) = \frac{1+X}{2}\log\theta + \frac{1-X}{2}\log(1-\theta)$$
   donc le score vaut
$$\ell'_\theta(X) = \frac{1+X}{2\theta} + \frac{X-1}{2(1-\theta)} = \frac{X-(2\theta-1)}{2\theta(1-\theta)}$$
   d'où, puisque $\mathbb{E}[X] = 2\theta - 1$ et $\mathbb{V}(X) = 4\theta(1-\theta)$,
$$\mathbb{E}_\theta\left[\ell'_\theta(X)^2\right] = \frac{1}{\theta(1-\theta)}.$$
   La fonction $\theta \mapsto 1/(\theta(1-\theta))$ étant continue sur $]0, 1[$, le modèle est régulier, d'information de Fisher $I(\theta) = 1/(\theta(1-\theta))$.

3. Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. suivant une loi de Rademacher de paramètre $\theta \in]0, 1[$ inconnu. Proposer un estimateur $\widehat{\theta}_n$ de θ par la méthode des moments. Est-il asymptotiquement efficace ?

   Correction. Puisque $\mathbb{E}_\theta[X] = 2\theta - 1$, on a $\theta = (\mathbb{E}[X] + 1)/2$ donc on peut considérer l'estimateur $\widehat{\theta}_n = (\overline{X}_n + 1)/2$. Par le TCL, on a
$$\sqrt{n}\left(\overline{X}_n - (2\theta - 1)\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, 4\theta(1-\theta))$$
   d'où, en divisant par 2,
$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \theta(1-\theta)) = \mathcal{N}(0, 1/I(\theta)),$$
   ce qui prouve que $\widehat{\theta}_n$ est asymptotiquement efficace.