Sorbonne Université — Année 2020/2021

L3 — Deuxième semestre

Examen de statistique

Exercice 1 : (2 points) Questions de cours :

1. Donner la définition du biais et de l'erreur quadratique moyenne d'un estimateur.
2. Soit $\theta \in \mathbb{R}$ et $\hat{\theta}_n$ un estimateur admettant un moment d'ordre 2. Montrer que

$$\text{EQM}\left(\hat{\theta}_n, \theta\right) = B\left(\hat{\theta}_n, \theta\right)^2 + \mathbb{V}\left[\hat{\theta}_n\right].$$

Exercice 2 : (4 points) Dans chacun des cas suivant, donner la normalité asymptotique de $\hat{\theta}_n$:

1. $\theta > 0$, $\hat{\theta}_n = \frac{1}{\overline{X}_n}$ et

$$\sqrt{n}\left(\overline{X}_n - \frac{1}{\theta}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\theta^2}\right).$$

Fait et refait, on a

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

2. $\theta > 0$, $\hat{\theta}_n = \frac{1}{\overline{X}_n^3}$ et

$$\sqrt{n}\left(\overline{X}_n - \theta^{-1/3}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

On a $g(x) = x^{-3}$ et donc $g'(x) = -3x^{-4}$ et en particulier $g'(\theta^{-1/3}) = -3\theta^{4/3}$ et donc

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \left(0, 9\theta^{8/3}\right).$$

Exercice 3 : Soit $\theta > 0$. On considère une variable aléatoire X ayant pour densité

$$f_\theta(x) = 2e^{2\theta}e^{-2x}\mathbf{1}_{x\geq\theta}$$

Dans tout ce qui suit, on considère $X_1, \ldots, X_n$ des variablles aléatoires i.i.d de même loi que X.

1. Montrer que $\mathbb{E}[X] = \theta + \frac{1}{2}$ et $\mathbb{V}[X] = \frac{1}{4}$.

Version 1 : Pour tout $t < 2$, la fonction génératrice des moments est définie par

$$G_X(t) = e^{2\theta}\int_\theta^{+\infty} 2e^{x(t-2)}dx = e^{2\theta}\frac{2}{t-2}\left[e^{x(t-2)}\right]_\theta^{+\infty} = \frac{-2}{t-2}e^{\theta t}$$

On a $G'_X(t) = \frac{-2\theta}{t-2}e^{\theta t} + \frac{2}{(t-2)^2}e^{\theta t}$, et donc $\mathbb{E}[X] = G'_X(0) = \theta + \frac{1}{2}$. De plus, on a

$$G''_X(t) = \frac{-2\theta^2}{t-2}e^{\theta t} + \frac{2\theta}{(t-2)^2}e^{\theta t} + \frac{2\theta}{(t-2)^2}e^{\theta t} - \frac{4}{(t-2)^3}e^{\theta t}$$

et donc $\mathbb{E}\left[X^2\right] = \theta^2 + \theta + \frac{1}{2}$. Ainsi $\mathbb{V}[X] = \frac{1}{4}$.

Version 2 : A l'aide d'une IPP, on a

$$\begin{aligned}\mathbb{E}[X] = e^{2\theta}\int_\theta^{+\infty} 2xe^{-2x}dx &= e^{\theta}\left(\left[-xe^{-2x}\right]_\theta^{+\infty} + \int_\theta^{+\infty} e^{-2x}dx\right)\\ &= e^{2\theta}\left(\theta e^{-2\theta} + \frac{1}{2}e^{-2\theta}\right)\\ &= \theta + \frac{1}{2}\end{aligned}$$

De la même façon,

$$\begin{aligned}\mathbb{E}\left[X^2\right] = e^{2\theta}\int_\theta^{+\infty} x^2 2e^{-2x}dx = e^{2\theta}\left(\left[-x^2e^{-2x}\right]_\theta^{+\infty} + 2\int_\theta^{+\infty} xe^{-2x}dx\right) &= \theta^2 + \mathbb{E}[X]\\ &= \theta^2 + \theta + \frac{1}{2}.\end{aligned}$$

2. Par la méthode des moments, donner un estimateur $\hat{\theta}_n$ de θ.
   On a $\hat{\theta}_n = \overline{X}_n - 1/2$.
3. L'estimateur $\hat{\theta}_n$ est-il consistant ?
   Par la LGN, on a $\overline{X}_n$ qui converge en probabilité vers $\theta + 1/2$, et donc
   $$\hat{\theta}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \theta.$$
4. L'estimateur $\hat{\theta}_n$ est-il asymptotiquement normal ?
   Par le TLC, on a
   $$\sqrt{n}\left(\overline{X}_n - \theta - 1/2\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{4}\right)$$
   ce qui se réécrit directement,
   $$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{4}\right)$$
5. Montrer que l'estimateur du maximum de vraisemblance est $X_{(1)} = \min_{i=1,\dots,n} X_i$.
   On a
   $$L_{\mathbf{X}}(\theta) = e^{2n\theta}2^n e^{-2n\overline{X}_n}\prod_{i=1}^n \mathbf{1}_{\theta\le X_i}$$

et

$$\prod_{i=1}^{n} \mathbf{1}_{\theta \leq X_i} = 1 \Leftrightarrow \forall i, \theta \leq X_i \Leftrightarrow \theta \leq X_{(1)}$$

et on a donc la vraisemblance

$$L_{\mathbf{X}}(\theta) = e^{2n\theta} 2^n e^{-2n\overline{X}_n} \mathbf{1}_{\theta \leq X_{(1)}}$$

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

6. Montrer que la fonction de répartition F_X de x est définie pour tout x par

$$F_X(x) = \begin{cases} 1 - e^{-2(x-\theta)} & \text{si } x \geq \theta \\ 0 & \text{sinon.} \end{cases}$$

Si $x < \theta$, $\mathbb{P}[X \leq x] = 0$. Si $x \geq \theta$,

$$F_X(x) = e^{2\theta} \int_\theta^x 2e^{-2x} dx = e^{2\theta} \left(-e^{-2x} + e^{-2\theta} \right) = 1 - e^{-2(x-\theta)}$$

7. En déduire la consistance de $X_{(1)}$. On a

$$\mathbb{P}\left[X_{(1)} \leq x\right] = 1 - \mathbb{P}\left[X_{(1)} > x\right] = 1 - \mathbb{P}\left[\forall i, X_i > x\right].$$

Par indépendance, on a

$$\mathbb{P}\left[\forall i, X_i > x\right] = \prod_{i=1}^{n} \mathbb{P}\left[X_i > x\right] = \left(\mathbb{P}\left[X > x\right]\right)^n = e^{-2n(x-\theta)}$$

et donc $\mathbb{P}\left[X_{(1)} \leq x\right] = 1 - e^{-2n(x-\theta)}$. On a donc pour tout $\epsilon > 0$, comme $X_{(1)} \geq \theta$,

$$\mathbb{P}\left[\left|X_{(1)} - \theta\right| \geq \epsilon\right] = \mathbb{P}\left[X_{(1)} - \theta \geq \epsilon\right] = \mathbb{P}\left[X_{(1)} \geq \theta + \epsilon\right] = e^{-2n\epsilon} \xrightarrow[n \to +\infty]{} 0$$

et on a donc convergence en probabilité.

8. Montrer que

$$n\left(X_{(1)} - \theta\right) \sim \mathcal{E}(2)$$

Comme $X_{(1)} \geq \theta$, pour tout $x < 0$, on a $\mathbb{P}\left[n\left(X_{(1)} - \theta\right) \leq x\right] = 0$. Pour tout $x \geq 0$, on a

$$\mathbb{P}\left[n\left(X_{(1)} - \theta\right) \leq x\right] = \mathbb{P}\left[X_{(1)} \leq \theta + \frac{x}{n}\right] = 1 - e^{-2x} = F_{\mathcal{E}(2)}(x).$$