M1 Actuariat, économétrie et statistiques, année 2021–2022.

NOM, Prénom :

Statistique Inférentielle
Contrôle continu

Lundi 17 Octobre

Durée 1h30, documents, téléphone, calculatrice interdits

Le barême (indicatif) prévu est le suivant : 6–7–8 (on tiendra compte (grave) de la présentation et de la clareté des explications):

On rappelle que la densité d'une loi $\Gamma(\alpha, \beta)$ vaut :

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}^+}(x).$$

Exercice 1

Soit $X_1, X_2, ..., X_n$ un échantillon de variables aléatoires de même loi X, où X admet pour densité

$$f_\theta(x) = \frac{2x}{\theta^2} \mathbb{1}_{[0,\theta]}$$

1. Calculer la fonction de répartition de X, son espérance et sa variance.

Soit $x \in \mathbb{R}$, $P(X \leq x) = \int_{-\infty}^{x} f_\theta(t)\,dt = \int_0^x \frac{2t}{\theta^2}\,dt$ où $x \in [0,\theta]$

$$= \left(\frac{x}{\theta}\right)^2$$

$$\Rightarrow F_X(x) = \begin{cases} \left(\frac{x}{\theta}\right)^2 & \text{si } x \in [0,\theta] \\ 0 & \text{Sinon.} \end{cases}$$

$$\mathbb{E}[X] = \int_0^\theta \frac{2x^2}{\theta^2}\,dx = \frac{2}{3}\theta$$

$$V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_0^\theta \frac{2x^3}{\theta^2}\,dx - \frac{4}{9}\theta^2 = \frac{\theta^2}{2} - \frac{4}{9}\theta^2$$

$$= \frac{9\theta^2 - 8\theta^2}{18} = \frac{\theta^2}{18}$$

2. Ah mais vous y avez cru ?? RRroooooo lalala, vous auriez vu vos têtes ! Mais non, bien évidemment, c'est pas un CC, je suis pas encore à ce niveau de fdp. C'est juste un entrainement. On va le faire pendant 1h15/30 et on corrige dans la seconde partie du partie. Du coup, question suivante : Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ ?

(H1) n'est pas vérifiée.

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{2x_i}{\theta^2} \mathbb{1}_{[0,\theta]}(x_i)$$

$$= \frac{2^n}{\theta^{2n}} \prod_{i=1}^{n} x_i \, \mathbb{1}_{(\min x_i \geq 0)} \mathbb{1}_{(\theta \geq \max X_i)}$$

L est maximisé lorsque $\hat{\theta}_n$ est égale au $\max_{i \in [\![1,n]\!]} X_i$

$$\hat{\theta}_n = \max_{i \in [\![1,n]\!]} X_i$$

3. Calculer la fonction de répartition et la densité de $\hat{\theta}_n$.

$$\mathbb{P}(\hat{\theta}_n \leq x) = \mathbb{P}(\max_{i \in [\![1,n]\!]} X_i \leq x) \underset{\text{Vaiid}}{=} \mathbb{P}(X_1 \leq x)^n$$

$$= \begin{cases} \left(\frac{x}{\theta}\right)^{2n} & \text{si } x \in [0,\theta] \\ 0 & \text{Sinon} \end{cases}$$

$$f_{\hat{\theta}_n}(x) = \frac{2n}{\theta^{2n}} x^{2n-1} \mathbb{1}_{[0,\theta]}(x).$$

4. Calculer le biais et la variance $\hat{\theta}_n$. $\hat{\theta}_n$ est-il convergent ?

$$B_{\hat{\theta}_n}(\theta) = \mathbb{E}[\hat{\theta}_n] - \theta$$

$$= \int_0^\theta x \frac{2n}{\theta^{2n}} x^{2n-1} dx - \theta = \int_0^\theta \frac{2n x^{2n}}{\theta^{2n}} dx - \theta$$

$$= \frac{2n}{\theta^{2n}} \times \frac{\theta^{2n+1}}{2n+1} - \theta$$

$$= \frac{2n}{2n+1}\theta - \theta = \frac{-1}{2n+1}\theta \xrightarrow[n\to\infty]{} 0$$

$\Rightarrow$ $\hat{\theta}_n$ asympt sans biais

$$V(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n^2] - \mathbb{E}[\hat{\theta}_n]^2$$

$$= \int_0^\theta \frac{2n x^{2n+1}}{\theta^{2n}} dx - \left(\frac{2n}{2n+1}\right)^2 \theta^2$$

$$= \frac{2n}{\theta^{2n}} \frac{\theta^{2n+2}}{2n+2} - \left(\frac{2n}{2n+1}\right)^2 \theta^2 = \left(\frac{2n}{2n+2} - \frac{4n^2}{(2n+1)^2}\right)\theta^2 = \frac{2n(2n+1)^2 - 4(2n+2)n^2}{(2n+2)(2n+1)^2}\theta^2$$

$$= \frac{2n\theta^2}{(2n+2)(2n+1)^2} \xrightarrow[n\to\infty]{} 0$$

$\Rightarrow$ Convergent

5. Montrer que

$$2n\left(1 - \frac{\hat{\theta}_n}{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{E}(1)$$

$$\mathbb{P}\left(2n\left(1 - \frac{\hat{\theta}_n}{\theta}\right) \le x\right) = \mathbb{P}\left(-\hat{\theta}_n \le \left(\frac{x}{2n} - 1\right)\theta\right)$$

$$= \mathbb{P}\left(\hat{\theta}_n \ge \left(1 - \frac{x}{2n}\right)\theta\right)$$

$$= 1 - \mathbb{P}\left(\hat{\theta}_n \le \left(1 - \frac{x}{2n}\right)\theta\right)$$

$$= 1 - \left(\frac{\left(1 - \frac{x}{2n}\right)\theta}{\theta}\right)^{2n} = 1 - \left(1 - \frac{x}{2n}\right)^{2n}$$

$$= 1 - e^{2n \underbrace{\ln\left(1 - \frac{x}{2n}\right)}_{\sim -\frac{x}{2n}}}$$

$$\Rightarrow \mathbb{P}\left(2n\left(1 - \frac{\hat{\theta}_n}{\theta}\right) \le x\right) \sim \underbrace{1 - e^{-x}}_{\text{f.d.r d'une } \mathcal{E}(1)}$$

D'où $2n\left(1 - \frac{\hat{\theta}_n}{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{E}(1)$.

6. Calculer l'estimateur $\hat{\theta}_M$ par la méthode des moments de θ.

θ est d'une dimens° $\Rightarrow$ 1 moment

$$M_1(\theta) = \mathbb{E}[X] = \int_0^\theta \frac{2x^2}{\theta^2}\,dx = \frac{2}{3}\theta$$

$$m_1 = \overline{X}_n$$

$$\underset{MM}{\Longrightarrow} \quad \frac{2}{3}\hat{\theta}_M = \overline{X}_n \quad \Rightarrow \hat{\theta}_M = \frac{3}{2}\overline{X}_n$$

7. Calculer le biais et la variance de $\hat{\theta}_M$. Est-il convergent ?

$$B_{\hat{\theta}_M}(\theta) = \mathbb{E}[\hat{\theta}_M] - \theta = \mathbb{E}\left[\frac{3}{2}\overline{X}_n\right] - \theta$$

$$= \frac{3}{2n}\sum_{i=1}^n \mathbb{E}[X_i] - \theta = \frac{3}{2} \times \frac{2}{3}\theta - \theta = 0$$

$\Rightarrow$ Sans biais ①

$$V(\hat{\theta}_M) = V\left(\frac{3}{2}\overline{X}_n\right) = \left(\frac{3}{2}\right)^2 \frac{1}{n^2}\sum_{i=1}^n V(X_i) = \left(\frac{3}{2}\right)^2 \frac{1}{n}\frac{\theta^2}{18} = \frac{\theta^2}{8n} \xrightarrow[n\to\infty]{} 0 \text{ ②}$$

① + ② $\Rightarrow$ Convergent.

8. Calculer la normalité asymptotique de $\hat{\theta}_M$.

Ici $\hat{\theta}_M = \frac{3}{2}\bar{X}_n$ $\quad g : x \mapsto \frac{3}{2}x \implies g' = \frac{3}{2}$

On applique le TCL à $X_1, \dots, X_n$

$$\implies \sqrt{n}\,\Big(\bar{X}_n - \underbrace{\tfrac{3}{2}\theta}_{\mathbb{E}[X]}\Big) \longrightarrow \mathcal{N}\Big(0, \underbrace{\tfrac{\theta^2}{18}}_{V(X)}\Big)$$

On applique mtn la méthode Delta au TCL et à g.

$$\implies \sqrt{n}\,\Big(\underbrace{\hat{\theta}_M}_{g(\bar{X}_n)} - \underbrace{\theta}_{g(\frac{3}{2}\theta)}\Big) \longrightarrow \mathcal{N}\Big(0, \underbrace{(g'(\theta))^2\sigma^2(\theta)}_{\left(\frac{3}{2}\right)^2 \times \frac{\theta^2}{18} = \frac{\theta^2}{8}}\Big)$$

9. Quel estimateur entre $\hat{\theta}_n$ et $\hat{\theta}_M$ est préférable au sens du risque quadratique.

$$\mathcal{R}(\hat{\theta}_n) = B_{\hat{\theta}_n}(\theta) + V(\hat{\theta}_n) \sim \frac{1}{n^2}$$

$$\mathcal{R}(\hat{\theta}_M) = B_{\hat{\theta}_M}(\theta) + V(\hat{\theta}_M)$$
$$= \frac{\theta^2}{8n}$$

Exercice 2

On dispose d'un échantillon $X_1, ..., X_n$ i.i.d. issu d'une loi de Poisson de paramètre θ.

1. Ecrire le modèle statistique et montrer qu'il est dominé.

Modèle statistique $\{\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathcal{P}(\theta), \theta \in \mathbb{R}_+^* \subseteq \mathbb{R}\}^n$

Ce modèle est dominé puisque la loi de Poisson est dominé par la mesure de comptage.

2. Calculer l'estimateur $\hat{\theta}_n$ du maximum de vraisemblance de θ.

(H1) à (H4) vérifiées $\Rightarrow L(x_1, \dots, x_n, \theta) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$

$$= \prod_{i=1}^{n} \left(\frac{\theta^{x_i}}{x_i!} e^{-\theta}\right)$$

$$= e^{-n\theta} \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!}$$

$\Rightarrow \ln L = -n\theta + \sum_{i=1}^{n} x_i \ln(\theta) - \ln(x_i!)$

$\frac{\partial}{\partial \theta} \ln L = -n + \frac{1}{\theta} \sum_{i=1}^{n} x_i$ $\quad \hookrightarrow \hat{\theta}_n$ est tq $\frac{1}{\hat{\theta}_n} \sum x_i - n = 0$

$\Rightarrow \boxed{\hat{\theta}_n = \bar{X}_n}$

$\frac{\partial}{\partial \theta^2} \ln L = -\frac{1}{\theta^2} \sum x_i < 0$

$\Rightarrow \hat{\theta}_n \in MV$

3. Montrer que l'estimateur est exhaustif.

$$L(x_1,\dots,x_n,\theta) = \prod \left(\frac{\theta^{x_i}}{x_i!} e^{-\theta}\right) = \underbrace{e^{-n\theta} \prod \theta^{x_i}}_{\varphi(\ ,\theta)} \underbrace{(\prod x_i!)^{-1}}_{\psi(x_1,\dots,x_n)}$$

$$= e^{-n\theta}\, \theta^{\sum x_i} (\prod x_i!)^{-1}$$

$$= \underbrace{e^{-n\theta}\, \theta^{n\hat{\theta}_n}}_{\varphi(\hat{\theta}_n,\theta)} \underbrace{(\prod (x_i!))^{-1}}_{\psi(x_1,\dots,x_n)}$$

$\Rightarrow$ Exhaustif

4. Cet estimateur est-il sans biais ? Convergent ? Efficace ?

$$\mathbb{E}[\hat{\theta}_n] = \frac{1}{n}\sum \mathbb{E}[X_i] = \theta \quad \Rightarrow \text{Sans biais}$$

$$V(\hat{\theta}_n) = \frac{1}{n} V(X_1) = \frac{\theta}{n} \xrightarrow[\infty]{} 0 \quad \longrightarrow \text{Convergent}$$

$$\mathbb{E}\left[\frac{1}{n}\sum\right] = \theta$$

Id

$g = Id$

Efficace $V(X_n) = \dfrac{(g'(\theta))^2}{I_n(\theta)}$

$$I_n(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \ln L\right] = -\mathbb{E}\left[-\frac{1}{\theta^2}\sum X_i\right]$$

$$= \frac{1}{\theta^2}\sum \mathbb{E}[X_i] = \frac{n}{\theta}$$

$$V(\hat{\theta}_n) = \frac{\theta}{n} = \frac{(g'(\theta))^2}{I_n(\theta)}$$

5. Donner la normalité asymptotique de $\hat{\theta}_n$.

On applique le TCL à $X_1, \dots, X_n$

$$\sqrt{n}\,(\underbrace{\overline{X}_n}_{\hat{\theta}_n} - \theta) \longrightarrow \mathcal{N}(0, \theta)$$

Car $\hat{\theta}_n$ est l'EMV de θ et (H1)_(H7) OK

On a $\sqrt{n}\,(\hat{\theta}_n - \theta) \longrightarrow \mathcal{N}(0, I^{-1}(\theta))$

6. (Bonus : hors programme) Donner un intervalle de confiance asymptotique de niveau $1 - \alpha$ de θ.

7. On pose

$$\hat{\theta}^* = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Cet estimateur est-il biaisé ? Quel estimateur entre $\hat{\theta}_n$ et $\hat{\theta}^*$ est préférable au sens du risque quadratique ?

Cet estimateur est sans biais Car $X_1, \ldots, X_n$ iid tq $X_i \sim \mathbb{P}$ et $V(X_i) = \sigma^2 < \infty$

$\Rightarrow \frac{1}{n-1}\sum (X_i - \overline{X}_n)^2$ ESB

$$V(\hat{\theta}^*) = \frac{1}{(n-1)^2}\sum_{i=1}^{n} V((X_i - \overline{X}_n)^2)$$

8. On veut maintenant estimer $\lambda = e^{-\theta} = \mathbb{P}(X=0)$. On pose

$$Y_i = \mathbb{1}_{X_i=0}$$

Donner la loi de Y_1, son espérance et sa variance.

$Y_1 \sim \text{Bern}(\lambda) \Rightarrow \mathbb{E}[Y_1] = \lambda$

$V(Y_1) = \lambda(1-\lambda)$

9. On se propose d'estimer λ par

$$\hat{\lambda}^* = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Cet estimateur est-il convergent ?

$\mathbb{E}[\hat{\lambda}^*] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i] = \frac{1}{n} \times n\lambda = \lambda \Rightarrow$ Sans biais

$V(\hat{\lambda}^*) = \frac{1}{n^2}\sum V(Y_i) = \frac{1}{n}\lambda(1-\lambda) \xrightarrow[n\to\infty]{} 0$

$\Rightarrow$ CV

10. Ecrire la vraisemblance de l'échantillon relativement au paramètre λ et calculer l'estimateur du maximum de vraisemblance $\hat{\lambda}_n$ de λ.

$L(x_1, \ldots, x_n, \lambda) = \lambda^{\sum x_i} e^{-n\lambda} (\prod x_i!)^{-1}$

$\ln L = (\sum x_i)\ln(\lambda) - n\lambda - \sum \ln(x_i!)$

$\frac{\partial}{\partial\lambda}\ln L = \frac{1}{\lambda}\sum x_i - n \Rightarrow \hat{\lambda}_n$ est tq $\frac{\partial}{\partial\lambda}\ln L = 0$

$\Rightarrow \hat{\lambda}_n = \overline{X}_n$

$\frac{\partial}{\partial\lambda^2}\ln L = -\frac{1}{\lambda^2}\sum x_i < 0$ ok EMV

$\mathbb{E}[-\frac{1}{\lambda^2}\sum x_i]$

$\frac{1}{\lambda^2}\sum \mathbb{E}[X_i]$

$\frac{n}{\lambda}$

11. Donner la normalité asymptotique de $\hat{\lambda}_n$.

$$\sqrt{n}\,(\hat{\lambda}_n - \lambda) \longrightarrow \mathcal{N}\left(0, \frac{\lambda}{n}\right)$$

Exercice 3

On suppose qu'au cours de l'année écoulée, le chiffre d'affaire des entreprises ayant un chiffre d'affaire inférieur à $c > 0$ (fixé) peut être représenté par une variable aléatoire X absolument continue de densité :

$$f(x) = \frac{x^{\frac{1}{\theta}-1}}{\theta c^{\frac{1}{\theta}}} \mathbb{1}_{\{0 \leq x < c\}}$$

où $\theta > 0$ est un paramètre, c est supposé connu. Soient $X_1, .., X_n$ n variables aléatoires i. i. d. de même loi que X.

1. Déterminer la loi de $\ln c - \ln X$.

2. Déterminer l'estimateur $\hat{\theta}_n$ du maximum de vraisemblance de θ.

3. L'estimateur $\hat{\theta}_n$ est-il sans biais ? Asympt. sans biais ?

4. L'estimateur $\hat{\theta}_n$ est-il convergent ?

5. L'estimateur $\hat{\theta}_n$ est-il exhaustif ?

6. L'estimateur $\hat{\theta}_n$ est-il efficace? Asympt. efficace ?

7. L'estimateur $\hat{\theta}_n$ est-il asymptotiquement normal ? Si oui, préciser sa variance asymptotique ?

8. On suppose que θ est connu et que c est le paramètre inconnu. Donner l'estimateur $\hat{c}_n$ du maximum de vraisemblance de c.

9. L'estimateur $\hat{c}_n$ est-il exhaustif ?

10. L'estimateur c_n est-il sans biais ?

11. Calculer l'estimateur $\hat{c}_M$ par la méthode des moments.

12. L'estimateur $\hat{c}_M$ est-il sans biais ? Convergent ?

13. Montrer la normalité asymptotique de $\hat{c}_M$ et préciser sa variance asymptotique.

14. On suppose maintenant que c et θ sont inconnus. Calculer l'estimateur du maximum de vraisemblance de (c, θ).