Faculté Sciences Exactes
Département: Proba-Stat
Année: 2023-2024

Université Djillali Liabès
Licence Mathématiques Appliquées
Module: Statistique Paramétrique

Examen Module : Statistique Paramétrique

Exercice 1 :

Soit un n-échantillon $(X_1, \ldots, X_n)$ où les variables X_i sont i.i.d. de même loi que X avec un espérance mathématique μ et variance σ^2. On dispose de deux estimateurs de μ

$$T_n = \sum_{i=1}^{2} \frac{X_i}{2} \quad \text{et} \quad S_n = \sum_{i=1}^{n} \frac{X_i}{n} - \frac{X_1 - X_2}{2}.$$

— Déterminer lequel des deux estimateurs est le plus efficace.

Exercice 2 :

Soit X une varible aléatoire suivant une loi normal d'espérence mathématique $m = 1$ et d'écart type σ. On veut batir un test pour choisir entre les deux hypothèses : $H_0 : \sigma = 2$ contre $H_1 : \sigma > 2$.

1. Existe-t-il un test uniformément plus puissant (UPP) parmi ceux de seuil $\alpha = 0.05$ pour tester H_0 contre H_1 ?
2. Si oui, Quelle est la région critique de ce test ?
3. Exite-t-il un test UPP au seuil α pour tester les hypothèses $H_0 : \sigma = 2$ contre $H_1 : \sigma \neq 2$?

Exercice 3 :

Considérons une variable aléatoire réelle continue X de densité :

$$f_\theta(x) = \frac{2}{\theta} x e^{-x^2/\theta}; \qquad x > 0,$$

avec θ un paramètre inconnu, $\theta > 0$. Pour cette loi (variable aléatoire), on considère un n-échantillon $(X_1, \ldots, X_n)$.

1. Montrer que la densité f_θ est de type exponentiel.
2. Trouver l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ.
3. On considère la variable aléatoire $Y = X^2$ et son n-échantillon associé $(Y_1, \ldots, Y_n)$.
   — Déterminer la densité de la variable aléatoire Y, et déduire l'espérance et la variance de la variable aléatoire Y.
4. En utilisant les questions précédentes, étudiez le biais, la convergence et l'efficacité de l'estimateur $\hat{\theta}_n$.
5. A l'aide de l'échantillon $(X_1, \ldots, X_n)$, construire un intervalle de confiance pour θ au niveau 0.95.

Solution Examen Statistique Paramétrique S1 2023/2024 :

Exercice n°: 1 : (04 pts)

L'estimateur le plus efficace est celui qui a l'écart quadratique (erreur quadratique) moyen la plus petite.

0.5 pt $E(T_n) = E\left(\frac{1}{2}\sum_{i=1}^{2} X_i\right) = \frac{1}{2}E(X_1+X_2) = \frac{1}{2}\left(E(X_1)+E(X_2)\right) = \frac{1}{2}(\mu+\mu) = \frac{2\mu}{2} = \mu$

0.5 pt $E(S_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{X_1 - X_2}{2}\right) = \frac{1}{n}\sum E(X_i) - \frac{E(X_1)-E(X_2)}{2} = \frac{1}{n}n\mu - \frac{\cancel{\mu-\mu}}{2} = \mu$

alors, les deux estimateurs sont sans biais de μ.

$V(T_n) = V\left(\frac{1}{2}\sum_{i=1}^{2} X_i\right) = \frac{1}{4}\left(V(X_1)+V(X_2)\right) = \frac{1}{4}(\sigma^2+\sigma^2) = \frac{\sigma^2}{2}$ 1 pt

$V(S_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{X_1 - X_2}{2}\right) = \frac{1}{n^2}\sum V(X_i) + \frac{V(X_1)+V(X_2)}{4} = \frac{\sigma^2}{n} + \frac{\sigma^2}{2}$ 1 pt

On déduit : $V(S_n) - V(T_n) = \frac{\sigma^2}{n} + \frac{\sigma^2}{2} - \frac{\sigma^2}{2} = \frac{\sigma^2}{n} > 0$ 1 pt

La variance de S_n est donc supérieure à celle de T_n, ce qui implique que T_n est l'estimateur le plus efficace.

Exercice n°: 2 : (06 pts)

$X \sim N(1, \sigma^2)$

$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-1)^2}, \quad x \in \mathbb{R}, \ \sigma > 0$

1. Existe-t-il un test UPP au seuil $\alpha = 0.05$ pour tester :

$$\begin{cases} H_0 : \sigma = 2 & (\sigma_0) \\ H_1 : \sigma > 2 & (\sigma_1) \end{cases}$$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-1)^2} = e^{-\frac{1}{2\sigma^2}(x-1)^2 - \ln\sigma - \ln(\sqrt{2\pi})}$$

Comme le support de densité est indépendant de σ et f s'écrit sous forme de type exponentiel ($f_X(x) = e^{\alpha(\sigma)a(x) + \beta(\sigma) + B(x)}$) avec 1pt

$\alpha(\sigma) = -\frac{1}{2\sigma^2}$, $a(x) = (x-1)^2$, $\beta(\sigma) = -\ln\sigma$ et $B(x) = -\ln(\sqrt{2\pi})$.

donc, la région critique est donnée par:

$W = \{ \underline{x} \in \mathbb{R}^n \setminus \sum a(x_i) > k_\alpha \}$ si $\alpha(\sigma_2) > \alpha(\sigma_1)$.

$W = \{ \underline{x} \in \mathbb{R}^n \setminus \sum a(x_i) < k_\alpha \}$ si $\alpha(\sigma_2) < \alpha(\sigma_1)$. 1pt.

On a $\alpha(\sigma) = -\frac{1}{2\sigma^2}$ et $\alpha(\sigma_2) = -\frac{1}{2\sigma_2^2} > -\frac{1}{2\sigma_1^2}$ $(\sigma_2 > \sigma_1)$

d'où $W = \{ \underline{x} \in \mathbb{R}^n \setminus \sum_{i=1}^{n} (x_i - 1)^2 > k_\alpha \}$ 1pt

La région critique est indépendant de paramètre σ, alors le test UPP existe.

2) $\alpha = P_{H_0}(W) = P(\sum (x_i - 1)^2 > k_\alpha)$.

Sous H_0, on a $\frac{\sum (x_i - 1)^2}{\sigma_1^2} \sim \chi^2_n$

2pts

$$0.05 = 1 - P_{H_0}\left(\frac{\sum (x_i - 1)^2}{4} < \frac{k_\alpha}{4} \right)$$

$$0.95 = P_{H_0}\left(\chi^2_n < \frac{k_\alpha}{4} \right) \Rightarrow \Phi_{\chi^2_n}\left(\frac{k_\alpha}{4}\right) = 0.95$$

$$k_\alpha = 4\, \Phi^{-1}_{\chi^2_n}(0.95).$$

Donc

$$W = \{ \underline{x} \in \mathbb{R}^n \setminus \sum_{i=1}^{n} (x_i - 1)^2 > 4\, \Phi^{-1}_{\chi^2_n}(0.95) \}.$$

Scanné avec CamScanner

Existe-il un test UPP au seuil α pour tester

$$\begin{cases} H_0 : \sigma = 2 \quad (\sigma_1) \\ H_1 : \sigma \neq 2 \quad (\sigma_2) \end{cases}$$

On a : $W = \left\{ \underline{x} \in \mathbb{R}^n \setminus \sum_{i=1}^{n} (x_i - 1)^2 > k_\alpha \right\}$ si : $\alpha(\sigma_2) > \alpha(\sigma_1)$

$W = \left\{ \underline{x} \in \mathbb{R}^n \setminus \sum (x_i - 1)^2 \cancel{<} k_\alpha \right\}$ si : $\alpha(\sigma_2) < \alpha(\sigma_1)$ 1pt

et comme $\sigma_2 \neq 2$ ($\sigma_2 > 2$ et $\sigma_2 < 2$), donc la région critique dépend le paramètre σ et alors le test UPP n'existe pas.

Exercice n° 3 (10pts)

$$f_\theta(x) = \frac{2}{\theta} x e^{-x^2/\theta} \quad ; x > 0 \ , \ \theta > 0 .$$

1\ Montrer que f_θ est de type exponentiel :

D'abord, le support de f_θ est independant de θ.

$$f_\theta(x) = \frac{2}{\theta} x e^{-x^2/\theta} = e^{-\frac{1}{\theta} x^2 + \ln\left(\frac{2}{\theta}\right) + \ln(x)} = e^{\alpha(\theta) a(x) + \beta(\theta) + B(x)}.$$

avec $\alpha(\theta) = -\frac{1}{\theta}$, $a(x) = x^2$, $\beta(\theta) = \ln\left(\frac{2}{\theta}\right)$ et $B(x) = \ln(x)$.

Donc f_θ est de type exponentiel. 1pt

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

$$L(x_1, \dots x_n ; \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{2}{\theta} x_i e^{-\frac{1}{\theta} x_i^2}$$ 2pts

$$\ln L(x_1, \dots x_n ; \theta) = n \ln(2) - n \ln(\theta) + \sum_{i=1}^{n} \ln x_i - \frac{1}{\theta} \sum x_i^2$$

$$\frac{\partial \ln L(x_1, \dots x_n ; \theta)}{\partial \theta} = 0 \iff -\frac{n}{\theta} + \frac{1}{\theta^2} \sum x_i^2 = 0 \Rightarrow \frac{\sum x_i^2}{\theta} = n \Rightarrow \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

$$\frac{\partial^2 \ln L(x_1 \dots x_n;\theta)}{\partial\theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum x_i^2 = \frac{n}{\theta^2} - \frac{2n}{\theta^2}\frac{\sum x_i^2}{\frac{n}{\theta}} = \frac{n}{\theta^2} - \frac{2n}{\theta^2} = -\frac{n}{\theta^2} <$$

1pt

d'où EMV de θ est $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} x_i^2$

3\ On dispose $Y = X^2$.

Trouver la densité de v.a Y:

$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = F_X(\sqrt{y})$.

$f_Y(y) = [F_X(\sqrt{y})]' = \frac{1}{2\sqrt{y}} f_X(\sqrt{y})$. 1pt

d'où $f_Y(y) = \frac{2}{\theta 2\sqrt{y}}\sqrt{y}\, e^{-\frac{1}{\theta}y}$

$= \frac{1}{\theta} e^{-\frac{1}{\theta}y}$; $y > 0$, $\theta > 0$.

alors $Y \sim \mathcal{E}\left(\frac{1}{\theta}\right)$ et $E(Y) = \theta$, $V(Y) = \theta^2$. 0.5pt

4\ Les propriétés de θ_n:

$E(\hat{\theta}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) = \frac{1}{n}\sum E(X_i^2) = E(Y) = \theta$. 0.5pt

D'où $\hat{\theta}_n$ est sans biais de θ.

$V(\hat{\theta}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) = \frac{1}{n^2}\sum V(X_i^2) = \frac{1}{n}V(Y) = \frac{\theta^2}{n}$. 0.5pt.

Comme $\left.\begin{array}{l} E(\hat{\theta}_n) = \theta \\ V(\hat{\theta}_n) \longrightarrow 0 \end{array}\right\} \Rightarrow \hat{\theta}_n \xrightarrow{P} \theta$

d'où $\hat{\theta}_n$ est un estimateur convergent et converge vers θ.

- $BCR = \frac{1}{I_n(\theta)}$

$I_n(\theta) = -E\left(\frac{\partial^2 \ln L(x_1 \dots x_n;\theta)}{\partial\theta^2}\right)$.

Scanné avec CamScanner

$I_n(\theta) = -E\left(\frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n}(x_i^2)\right) = -\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n}E(x_i^2)$

$= -\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum E(y_i) = -\frac{n}{\theta^2} + \frac{2n\theta}{\theta^3} = -\frac{n}{\theta^2} + \frac{2n}{\theta^2} = \frac{n}{\theta^2}$.

d'où $BCR = \frac{1}{I_n(\theta)} = \frac{\theta^2}{n} = V(\hat{\theta}_n)$.   0.5pt

alors $\hat{\theta}_n$ est un estimateur efficace.

5| Construire un IC pour θ au niveau 0.95.

$1-\alpha$ On a : $Y \sim \mathcal{E}\left(\frac{1}{\theta}\right)$, $\frac{2\sum_{i=1}^{n} Y_i}{\theta} \sim \chi^2_{2n} \Leftrightarrow \frac{2\sum_{i=1}^{n} X_i^2}{\theta} \sim \chi^2_{2n}$.

alors, $\frac{2n\hat{\theta}_n}{\theta} \sim \chi^2_{2n}$.   3pts

$1-\alpha = P\left(k_1 \leq \frac{2n\hat{\theta}_n}{\theta} \leq k_2\right)$; avec.

$k_1 = \Phi^{-1}_{\chi^2_{2n}}(\alpha/2)$ et $k_2 = \Phi^{-1}_{\chi^2_{2n}}(1-\alpha/2)$.

$I.C_{0.95} = \left[\frac{2n\hat{\theta}_n}{k_2}, \frac{2n\hat{\theta}_n}{k_1}\right]$.