Corrigé Examen Module : Statistique 1

Exercice 1. Soit X une variable aléatoire réelle qui suit une loi exponentielle. Sa fonction de répartition est donnée par :

$$F_X(x) = \begin{cases} 1 - \exp\left(-\frac{1}{\theta}x\right) & \text{si} \quad x \geq 0 \\ 0 & \text{sinon,} \end{cases}$$

où $\theta > 0$ est un paramètre inconnu, et soit un n-échantillon i.i.d. noté $\{X_1, \cdots, X_n\}$ de même loi que X.

1. Montrer que l'espérance de X est telle que $E(X) = \theta$.
2. Soit $T_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, un estimateur de θ. En admettant que $V(X) = \theta^2$, calculer $\mathbb{E}(T_n)$ et $V(T_n)$.
3. Etudier le biais et la variance de cet estimateur. Qu'en déduisez-vous ?

Solution :

1. La densité f de la variable aléatoire X est

$$f(x) = \begin{cases} \frac{1}{\theta}\exp\left(-\frac{1}{\theta}x\right) & \text{si} \quad x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

Donc,

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{+\infty} t f(t) dt = \int_0^{+\infty} t \frac{1}{\theta} \exp\left(-\frac{1}{\theta}t\right) dt \\ &= \frac{1}{\theta}\int_0^{+\infty} t \exp\left(-\frac{1}{\theta}t\right) dt \end{aligned}$$

on fait un changement de variable $y = \frac{1}{\theta}t$ donc $t = y\theta$ et $dt = \theta dy$, on obtient ainsi

$$\begin{aligned} \mathbb{E}(X) &= \theta \int_0^{+\infty} y \exp(y)\, dt \\ &= \theta \frac{\Gamma(2)}{\Gamma(0)} = \theta \end{aligned}$$

2.

$$\begin{aligned} \mathbb{E}(T_n) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= n \\ &= \frac{1}{n} n\left(\mathbb{E}(X)\right), \text{ Indépendance} \\ &= \theta. \end{aligned}$$

Concernant la variance,

$$\begin{aligned} V(T_n) &= V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2} n V(X), \text{ Indépendance} \\ &= \frac{\theta^2}{n} \end{aligned}$$

3. Comme $\mathbb{E}(T_n) = \theta$ donc T_n est un estimateur sans biais, pour la variance en remarque que $V(T_n) \underset{n\longrightarrow\infty}{\longrightarrow} 0$ donc, l'estimateur est convergent.

Exercice 2. Soit une variable aléatoire continue X, de densité :

$$f_\theta(x) = \begin{cases} b - \frac{\theta}{2} & \text{si } -2 < x \leq 0 \\ b + \frac{\theta}{2} & \text{si } 0 < x < 2 \end{cases}$$

avec b une constante à déterminer plus loin. Le paramètre θ est inconnu et $\theta \in]-\frac{1}{2}, \frac{1}{2}[$. On considère un n-échantillon $(X_1, \ldots, X_n)$ pour cette loi (variable).

1. Calculer la valeur de b.
2. Calculez l'espérance de la variable aléatoire X et ensuite déduisez un estimateur pour θ par la méthode des moments. Notons cet estimateur par $\hat{\theta}_n^{(1)}$.
3. Calculez la variance de la variable aléatoire X. Est-elle > 0 pour tout $\theta \in \left]-\frac{1}{2}, \frac{1}{2}\right[$? Déduisez $\text{Var}\left[\hat{\theta}_n^{(1)}\right]$. Etudiez également la convergence et le biais de l'estimateur $\hat{\theta}_n^{(1)}$.
4. Calculez la probabilité $\mathbb{P}[0 < X < 2]$.
5. Soit la variable aléatoire $Y = \mathbb{I}_{0<X<2}$ et le n-échantillon correspondant $Y_i = \mathbb{I}_{0<X_i<2}$ pour $i = 1, \cdots, n$. Quelle est la loi de Y ?
6. Calculez la probabilité $\mathbb{P}[-2 < X < 2]$ et déduisez quelles sont les valeurs possibles pour la variable aléatoire X. Déduisez la relation entre les variables aléatoires $\mathbb{I}_{-2<X\leq 0}$ et $\mathbb{I}_{0<X<2}$.
7. En utilisant le fait que la densité de la variable aléatoire X peut s'écrire :

$$f_\theta(x) = \left(\frac{1}{4} - \frac{\theta}{2}\right)^{\mathbb{I}_{-2<x\leq 0}} \cdot \left(\frac{1}{4} + \frac{\theta}{2}\right)^{\mathbb{I}_{0<x<2}}$$

et la question précédente, écrivez la vraisemblance du n-échantillon $X_1, \cdots, X_n$. Ecrivez cette vraisemblance fonction de $S_n = \sum_{i=1}^{n} \mathbb{I}_{0<X_1<2}$. Trouvez l'estimateur du maximum de vraisemblance de θ.
8. Considérons l'estimateur suivant pour θ : $\hat{\theta}_n^{(2)} = \frac{S_n}{n} - \frac{1}{2}$. Etudiez la convergence, le biais et l'efficacité de $\hat{\theta}_n^{(2)}$.
9. Tenant compte des questions 3) et 8), entre les estimateurs $\hat{\theta}_n^{(1)}$ et $\hat{\theta}_n^{(2)}$ lequel il faut choisir ? Justification.

Solution :

1.
$$\begin{aligned}
\mathbb{E}(X) &= (1-\theta)\int_{-\frac{1}{2}}^{0} x dx + (1+\theta)\int_{0}^{\frac{1}{2}} x dx = \frac{1-\theta}{8} + \frac{1+\theta}{8} = \frac{\theta}{4} \\
\mathbb{E}\left(X^2\right) &= (1-\theta)\int_{-\frac{1}{2}}^{0} x^2 dx + (1+\theta)\int_{0}^{\frac{1}{2}} x^2 dx = \frac{1-\theta}{24} + \frac{1+\theta}{24} = \frac{1}{12} \\
\text{Var}(x) &= \mathbb{E}\left(x^2\right) - \mathbb{E}^2(x) = \frac{1}{12} - \frac{\theta^2}{16} = \frac{-3\theta^2}{48}
\end{aligned}$$

2. $\bar{X}_n = \dfrac{\theta}{4} \Rightarrow \hat{\theta}_n^{(1)} = 4\bar{x}_n$.
3. $\bar{x}_n \xrightarrow{\text{p.s}} \mathbb{E}(X)$, donc $\hat{\theta}_n^{(1)} \longrightarrow 4\mathbb{E}(X) = \theta \Rightarrow \hat{\theta}_n^{(1)}$ fortement convergent.

$$\mathbb{E}\left[\hat{\theta}_u^{(1)}\right] = 4\mathbb{E}\left(\bar{x}_n\right) = 4 \cdot \mathbb{E}(x) = \theta$$

donc $\hat{\theta}_n^{(1)}$ est sans bias. De plus

$$\text{Var}\left(\hat{\theta}_u^{(1)}\right) = 16\frac{\text{Var}(x)}{n} = \frac{4-38^2}{3n}.$$

4. $Y_i \cdot \backsim B(p)$ avec
$$p = \mathbb{P}\left[0 \leq x_i < \frac{1}{2}\right] = \int_0^{1/2} f_\theta(x)dx = \frac{1+\theta}{2}$$
donc
$$S_n = \sum_{i=1}^n y_i \sim B(n,p).$$

5.
$$L_n(\theta) = (1-\theta)^{\sum_{i=1}^n \mathbb{1}_{\{-\frac{1}{2} < x_i < 0\}}} \cdot (1+\theta)^{\sum_{i=1}^n \mathbb{1}_{\{0 \leq x_i < \frac{1}{2}\}}} = (1-\theta)^{n-S_n} \cdot (1+\theta)^{S_n}$$
Mais
$$x = \sum_{i=1}^n \mathbb{1}_{\{-\frac{1}{2}<x_i<0\}} + \sum_{i=1}^n \mathbb{1}_{\{0<x_i<\frac{1}{2}\}} \Rightarrow \sum_{i=1}^n \mathbb{1}_{\{-\frac{1}{2}<x_i<0\}} = n - S_n.$$
Donc
$$\log L_x(\theta) = (n - S_x)\log(1-\theta) + S_n \log(1+\theta).$$
Rq : la fonction for $f_\theta(x)$ est continue, dérivable en θ_1 (pas en x)
$$0 = \frac{\partial}{\partial\theta}\log L_n(\theta) = -\frac{n-S_n}{1-\theta} + \frac{S_n}{1+\theta} \Rightarrow S_n + \theta S_n - n - n\theta + S_n - \theta S_n = 2$$
$$\Rightarrow 2S_n - n - n\theta = 0 \Rightarrow \hat{\theta}_n^{(2)} = 2\frac{S_n}{n} - 1 = 2\vec{Y}_n - 1$$
De plus, pour la dérivée deuxième du logarithme de vraisemblance, en a
$$\frac{\partial^2}{\partial\theta^2}\log L_x(\theta) = -\frac{S_n}{(1+\theta)^2} + \frac{S_n - n}{(1-\theta)^2} \Rightarrow \frac{\partial^2}{\partial\sigma^2}\log L_x(\theta) \leq 0$$
$$S_n \in \{0,1,\ldots,n\} \Rightarrow \hat{\theta}_n^{(2)} \text{ point de max.}$$

6. $\vec{Y}_n \xrightarrow{P.S} \mathbb{E}(y) = p = \frac{1+\theta}{2} => \hat{\theta}_n^{(2)} = 2\bar{y}_n - 1 \underset{n\to\infty}{\xrightarrow{P \cdot s}} \theta \Rightarrow \hat{\theta}_n^{(2)}$ est fortement consistent.
$$\mathbb{E}\left(\bar{Y}_n\right) = \mathbb{E}(y) = p = \frac{1+\theta}{2} \Rightarrow \mathbb{E}\left(\hat{\theta}_n^{(2)}\right) = \theta \Rightarrow \hat{\theta}_n^{(2)} \text{ sans biais.}$$
$$\operatorname{Var}\left(\hat{\theta}_n^{(2)}\right) = 4\operatorname{Var}\left(\bar{y}_n\right) = 4\frac{\operatorname{Var}(y)}{n} = 4\frac{p(1-p)}{n} = \frac{4}{n}\frac{1+\theta}{2} \cdot \frac{1-\theta}{2} = \frac{1-\varepsilon}{n}$$

7.
$$\operatorname{Var}\left(\hat{\theta}_n^{(1)}\right) - \operatorname{Var}\left(\hat{\theta}_n^{(2)}\right) = \frac{4-3\theta^2}{3n} - \frac{1-\theta^2}{n} = \frac{4-3\theta\frac{2}{3}+3\theta}{3n} = \frac{1}{3n} > 0$$
$$\Rightarrow \operatorname{Var}\left(\hat{\theta}_n^{(1)}\right) > \operatorname{Var}\left(\hat{\theta}_n^{(2)}\right) \text{ donc } \hat{\theta}_n^{(1)} \text{ est moins précis que } \hat{\theta}_n^{(2)}$$
$$\Rightarrow \text{on choisit } \hat{\theta}_n^{(2)}.$$