

TD n°1 : Statistique

Exercice 1

On considère la variable aléatoire “durée d’attente à un feu rouge”. La durée maximale d’attente à ce feu rouge est notée θ , paramètre inconnu strictement positif. On observe un échantillon t_1, \dots, t_n de taille n , où t_i désigne la durée d’attente observée pour le $i^{\text{ème}}$ individu. On fait l’hypothèse que les variables aléatoires T_i associées aux observations t_i sont indépendantes et de même loi uniforme sur $[0, \theta]$ i.e. $T_i \sim U[0, \theta]$.

1) Représenter le graphe de la densité de la loi $U[0, \theta]$ et préciser ses paramètres moyenne et variance.

2) On désire estimer le paramètre θ . Déterminer le biais et la variance de la statistique $\bar{T} = \frac{1}{n} \sum_{i=1}^n T_i$. Montrer que $\hat{\theta}_1 = 2\bar{T}$ est un estimateur sans biais et convergent en probabilité de θ .

3) Montrer que l’estimateur du maximum de vraisemblance de θ est $Y_n = \sup_i T_i$.

a) En utilisant l’équivalence des événements suivante $Y_n < y \iff T_i < y, \forall i = 1, \dots, n$, calculer la fonction de répartition de Y_n . En déduire sa densité et calculer $E[Y_n]$ et $Var Y_n$

b) Montrer que la statistique $\hat{\theta}_2 = \frac{n+1}{n} Y_n$ est un estimateur sans biais et convergent en probabilité de θ .

4) Lequel des deux estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$ choisiriez-vous pour estimer θ ?

Exercice 2

La durée de fonctionnement d’un matériel électrique est représentée par une variable aléatoire réelle X suivant une loi de Weibull de densité :

$$f(x; \theta, \lambda) = \frac{\lambda}{\theta} x^{\lambda-1} \exp\left\{-\frac{x^\lambda}{\theta}\right\} \quad x > 0$$

avec $\theta > 0$ et $\lambda > 0$. On suppose que λ est connu.

1) Déterminer la loi de $U = X^\lambda$ puis calculer $E(X^\lambda)$ et $Var(X^\lambda)$.

2) On considère un échantillon (X_1, \dots, X_n) de même loi que X . Calculer l’estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ . Cet estimateur est-il sans biais ? convergent ? efficace ? Calculer son erreur quadratique moyenne.

Exercice 3

Soient X_1, \dots, X_n , n variables aléatoires indépendantes de même loi de densité :

$$f(x) = \beta e^{\beta(\alpha-x)} \mathbf{1}_{[\alpha, +\infty[}(x)$$

1) α étant connu, déterminer l’estimateur du maximum de vraisemblance de $\omega = 1/\beta$ noté $\hat{\omega}$. Vérifier qu’il est sans biais et convergent. Montrer enfin que $\hat{\omega}$ est l’estimateur efficace de ω .

2) β étant connu, étudier l'estimateur du maximum de vraisemblance de α noté $\hat{\alpha}$. On admet que la densité de probabilité de $\hat{\alpha}$ est :

$$f(u) = n\beta e^{n\beta(\alpha-u)} \mathbf{1}_{[\alpha, +\infty[}(u)$$

En s'aidant de ce qui a été fait à la première question, déterminer le biais et la variance de $\hat{\alpha}$. En déduire un estimateur sans biais et convergent de α . Déterminer $E \left[\frac{\partial^2 \ln f(X_1, \dots, X_n; \alpha)}{\partial \alpha^2} \right]$. Que dire de l'efficacité de $\hat{\alpha}$?

3) β étant connu, déterminer l'estimateur de α obtenu à l'aide de la méthode des moments noté $\bar{\alpha}$. Comparer les deux estimateurs $\bar{\alpha}$ et $\hat{\alpha}$.

Exercice 4

1) Soient X_1, \dots, X_n , n variables aléatoires indépendantes suivant des lois de Poisson de paramètre λ . Déterminer l'estimateur du maximum de vraisemblance du paramètre λ . Est-il sans biais, convergent, efficace ?

2) Même question lorsque X_1, \dots, X_n sont n variables aléatoires de loi de Poisson de paramètres $\lambda_j = j\lambda$, $j \in \{1, \dots, n\}$

Exercice 1

1) La densité de T_i est

$$f(t_i; \theta) = \begin{cases} \frac{1}{\theta} & \text{si } t_i \in [0, \theta] \\ 0 & \text{sinon} \end{cases}$$

On a de plus (voir tables ou calculs élémentaires)

$$E[T_i] = \frac{\theta}{2} \text{ et } \text{var}[T_i] = \frac{\theta^2}{12}$$

2)

$$\begin{aligned} E[\bar{T}] &= E[T_1] = \frac{\theta}{2} \\ \text{var}[\bar{T}] &= \frac{\text{var}[T_1]}{n} = \frac{\theta^2}{12n} \end{aligned}$$

Donc

$$\begin{aligned} E[\hat{\theta}_1] &= 2E[\bar{T}] = \theta \\ \text{var}[\hat{\theta}_1] &= 4\text{var}[\bar{T}] = \frac{\theta^2}{3n} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

$\hat{\theta}_1$ est donc un estimateur non biaisé et convergent de θ .

3) La vraisemblance de t_1, \dots, t_n est

$$f(t_1, \dots, t_n; \theta) = \begin{cases} \frac{1}{\theta^n} & \text{si } t_i \in [0, \theta], \forall i = 1, \dots, n \\ 0 & \text{sinon} \end{cases}$$

Maximiser la vraisemblance de t_1, \dots, t_n revient à maximiser $\frac{1}{\theta^n}$ sous les contraintes $t_i \in [0, \theta], \forall i = 1, \dots, n$. Puisque $\frac{1}{\theta^n}$ est une fonction décroissante de θ et que les contraintes impose $t_i \leq \theta, \forall i = 1, \dots, n$, on en déduit

$$Y_n = \sup_i T_i$$

a) On a

$$\begin{aligned} P[Y_n < y] &= P[T_1 < y \text{ et } T_2 < y \dots \text{ et } T_n < y] \\ &= \prod_{k=1}^n P[T_k < y] \\ &= P[T_1 < y]^n \\ &= \begin{cases} \left(\frac{y}{\theta}\right)^n & \text{si } y \in [0, \theta] \\ 1 & \text{si } y \geq \theta \\ 0 & \text{si } y \leq 0 \end{cases} \end{aligned}$$

La densité de Y_n est donc

$$f(y) = \begin{cases} n \frac{y^{n-1}}{\theta^n} & \text{si } y \in [0, \theta] \\ 0 & \text{sinon} \end{cases}$$

On en déduit

$$E[Y_n] = \int_0^\theta n \frac{y^n}{\theta^n} dy = \frac{n\theta}{n+1}$$

$$E[Y_n^2] = \int_0^\theta n \frac{y^{n+1}}{\theta^n} dy = \frac{n\theta^2}{n+2} \text{ et donc } \text{var}[Y_n] = \frac{n\theta^2}{(n+2)(n+1)^2}$$

$\hat{\theta}_2 = \frac{n+1}{n} Y_n$ est donc un estimateur non biaisé de θ . Puisque

$$\text{var}[\hat{\theta}_2] = \left(\frac{n+1}{n}\right)^2 \text{var}[Y_n] = \frac{\theta^2}{n(n+2)} \xrightarrow{n \rightarrow \infty} 0$$

$\hat{\theta}_2 = \frac{n+1}{n} Y_n$ est un estimateur convergent de θ .

4) Puisque la variance de $\hat{\theta}_2$ est plus faible que la variance de $\hat{\theta}_1$ (au moins pour n grand car $\text{var}[\hat{\theta}_2] \simeq \frac{\theta^2}{n^2}$ et $\text{var}[\hat{\theta}_1] = \frac{\theta^2}{3n}$), on choisira l'estimateur $\hat{\theta}_2$.

Exercice 2

1) un changement de variables élémentaire conduit à

$$f(u) = \begin{cases} \frac{1}{\theta} \exp\left(-\frac{u}{\theta}\right) & \text{pour } u > 0 \\ 0 & \text{si } u \leq 0 \end{cases}$$

Des calculs simples ou les tables de la loi Gamma permettent d'obtenir

$$E[U] = \theta \text{ et } \text{var}[U] = \theta^2$$

2) La vraisemblance de t_1, \dots, t_n est

$$\begin{aligned} f(x_1, \dots, x_n; \theta) &= \prod_{k=1}^n \left[\frac{\lambda}{\theta} x_k^{\lambda-1} \exp\left\{-\frac{x_k^\lambda}{\theta}\right\} \right] \quad x_i > 0 \\ &= \left(\frac{\lambda}{\theta}\right)^n \left[\prod_{k=1}^n x_k \right]^{\lambda-1} \exp\left\{-\frac{1}{\theta} \sum_{k=1}^n x_k^\lambda\right\} \end{aligned}$$

et son logarithme

$$\ln f(x_1, \dots, x_n; \theta) = n \ln \lambda - n \ln \theta + (\lambda - 1) \ln \left[\prod_{k=1}^n x_k \right] - \frac{1}{\theta} \sum_{k=1}^n x_k^\lambda$$

Puisque le domaine de définition de la vraisemblance est indépendant de θ , la recherche du maximum de vraisemblance de θ se fait comme suit

$$\begin{aligned} \frac{\partial \ln f(x_1, \dots, x_n; \theta)}{\partial \theta} &\geq 0 \Leftrightarrow -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{k=1}^n x_k^\lambda \geq 0 \\ &\Leftrightarrow \theta \leq \frac{1}{n} \sum_{k=1}^n x_k^\lambda \end{aligned}$$

d'où

$$\hat{\theta}_n = \frac{1}{n} \sum_{k=1}^n X_k^\lambda$$

En remarquant que $\hat{\theta}_n = \frac{1}{n} \sum_{k=1}^n U_k$, on a

$$\begin{aligned} E[\hat{\theta}_n] &= E[U_1] = \theta \\ \text{var}[\hat{\theta}_n] &= \frac{\text{var}U_1}{n} = \frac{\theta^2}{n} \end{aligned}$$

Donc $\hat{\theta}_n$ est un estimateur sans biais et convergent de θ . La borne de Cramer-Rao d'un estimateur non biaisé de θ est

$$\begin{aligned} BCR(\theta) &= \frac{-1}{E\left[\frac{\partial^2 \ln f(x_1, \dots, x_n; \theta)}{\partial \theta^2}\right]} \\ &= \frac{-1}{\frac{n}{\theta^2} - \frac{2}{\theta^3} E\left[\sum_{k=1}^n X_k^\lambda\right]} \\ &= \frac{-1}{\frac{n}{\theta^2} - \frac{2}{\theta^3} n\theta} = \frac{\theta^2}{n} \end{aligned}$$

Puisque $\text{var}[\hat{\theta}_n] = BCR(\theta)$, $\hat{\theta}_n$ est l'estimateur efficace de θ . Son erreur quadratique moyenne est

$$E\left[(\hat{\theta}_n - \theta)^2\right] = \text{var}[\hat{\theta}_n] + \text{biais}^2[\hat{\theta}_n] = \frac{\theta^2}{n}$$

Exercice 3

1) La vraisemblance de x_1, \dots, x_n est

$$\begin{aligned} f(x_1, \dots, x_n; \omega) &= \prod_{k=1}^n \left[\frac{1}{\omega} \exp\left\{-\frac{\alpha - x_k}{\omega}\right\} \right] \quad x_k > \alpha \\ &= \left(\frac{1}{\omega}\right)^n \exp\left\{-\frac{1}{\omega} \left(\sum_{k=1}^n x_k - n\alpha\right)\right\} \end{aligned}$$

et son logarithme

$$\ln f(x_1, \dots, x_n; \omega) = -n \ln \omega + -\frac{1}{\omega} \left(\sum_{k=1}^n x_k - n\alpha\right)$$

Puisque le domaine de définition de la vraisemblance est indépendant de ω , la recherche du maximum de vraisemblance de ω se fait comme suit

$$\begin{aligned} \frac{\partial \ln f(x_1, \dots, x_n; \omega)}{\partial \omega} &\geq 0 \Leftrightarrow -\frac{n}{\omega} + \frac{1}{\omega^2} \left(\sum_{k=1}^n x_k - n\alpha\right) \geq 0 \\ \Leftrightarrow \omega &\leq \frac{1}{n} \sum_{k=1}^n x_k - \alpha \end{aligned}$$

d'où

$$\hat{\omega} = \frac{1}{n} \sum_{k=1}^n X_k - \alpha$$

On montre aisément que $U_k = X_k - \alpha$ suit une loi exponentielle de densité

$$f(u) = \begin{cases} \frac{1}{\omega} \exp\left(-\frac{u}{\omega}\right) & \text{pour } u > 0 \\ 0 & \text{si } u \leq 0 \end{cases}$$

Donc comme dans l'exercice précédent

$$\begin{aligned} E[\hat{\omega}] &= E[U_1] = \omega \\ \text{var}[\hat{\omega}] &= \frac{\text{var}U_1}{n} = \frac{\omega^2}{n} \end{aligned}$$

Donc $\hat{\omega}$ est un estimateur sans biais et convergent de ω . La borne de Cramer-Rao d'un estimateur non biaisé de ω est

$$BCR(\omega) = \frac{-1}{E\left[\frac{\partial^2 \ln f(x_1, \dots, x_n; \omega)}{\partial \omega^2}\right]} = \frac{\omega^2}{n}$$

Puisque $\text{var}[\hat{\omega}] = BCR(\omega)$, $\hat{\omega}$ est l'estimateur efficace de ω .

2) voir correction dans le livre p. 224.

Exercice 4

voir solution dans le livre p. 207.