STATISTIQUE MATHÉMATIQUE - EXAMEN.

ENSAI - Mastères spécialisés Data Science - 2023/2024 -

ENSAI École nationale de la statistique et de l'analyse de l'information

Le sujet comporte 2 pages. Durée de l'épreuve : 2h. Tous documents autorisés, calculatrices autorisées. Le barème est indicatif et sans engagement. La qualité de la rédaction entre pour une part importante dans la notation.

I. 12 points

On considère un n-échantillon $X = (X_1, ..., X_n)$ d'une v.a. de loi de Bernoulli de paramètre inconnu $\theta \in]0,1[$.

On note $S_n = \sum_{i=1}^n X_i$.

1°. Expliquer succintement pourquoi le modèle est régulier.

2°. Calculer la log-vraisemblance $l(x,\theta)$ de l'échantillon et en déduire l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$ de θ.

3°. Montrer que cet estimateur est fortement consistant et converge en moyenne quadratique. Démontrer que $\widehat{\theta}_n$ est asymptotiquement normal et que

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \rightsquigarrow \mathcal{N}(0, \theta(1-\theta)) \tag{1}$$

4°. On s'intéresse maintenant au paramètre $\lambda = \theta^2$. Démontrer que l'estimateur du maximum de vraisemblance de λ est $\widehat{\lambda}_n = (S_n/n)^2$.

5°. Démontrer que $\widehat{\lambda}_n$ est biaisé et calculer son biais.

6°. Démontrer que $\widehat{\lambda}_n$ est fortement consistant, asymptotiquement normal et préciser la loi limite de $\sqrt{n}(\widehat{\lambda}_n - \lambda)$.

7°. Calculer l'information de Fisher associée à λ, puis démontrer que $\widehat{\lambda}_n$ est asymptotiquement efficace.

8°. On cherche à construire un estimateur sans biais de λ. Pour tout $i = 1, ..., n$, on note $X_{\bullet i} = (X_1, ..., X_{i-1}, X_{i+1}, ..., X_n)$ le vecteur formé des $(n-1)$ observations différentes de la ième. On note $\widehat{\lambda}_n(i)$ l'estimateur du maximum de vraisemblance du modèle associé à $X_{\bullet i}$. Montrer que

$$\widehat{\lambda}_n(i) = \left(\frac{S_n - X_i}{n-1}\right)^2 \tag{2}$$

9°. Calculer $\mathbb{V}(S_n - X_i)$, en déduire $\mathbb{E}[(S_n - X_i)^2]$, puis démontrer que l'estimateur

$$T_n = n\widehat{\lambda}_n - \frac{n-1}{n}\sum_{i=1}^n \widehat{\lambda}_n(i) \tag{3}$$

est sans biais pour λ.

II. 10 points

On considère un n-échantillon $X = (X_1, ..., X_n)$ d'une v.a. dont la densité est

$$f(x) = \frac{3\theta^3}{x^4}\mathbb{1}_{[\theta,+\infty[}(x) \tag{4}$$

avec $\theta > 0$.

1°. Le modèle est-il régulier ? Démontrer que la fonction de répartition $F(x)$ de X_1 s'écrit

$$F(x) = 1 - \left(\frac{\theta}{x}\right)^3 \tag{5}$$

pour $x \geq \theta$, puis que

$$\mathbb{E}[X_1] = \frac{3\theta}{2} \text{ et } \mathbb{V}(X_1) = \frac{3\theta^2}{4}. \tag{6}$$

2°. Soit $M_n = \min(X_1, \ldots, X_n)$. Déterminer la fonction de répartition $F_n(x)$ de M_n. En déduire sa densité, sa moyenne et sa variance.

3°. Déterminer une constante c_n de telle sorte que $\widehat{M}_n = c_n M_n$ soit un estimateur sans biais de θ.

4°. Démontrer que cet estimateur converge en moyenne quadratique et en probabilité.

5°. Peut-on appliquer le théorème de la limite centrale à M_n ? Pourquoi ?

6°. On note $\bar{X}_n$ la moyenne empirique de l'échantillon et on pose $T_n = 2\bar{X}_n/3$. Démontrer que T_n est un estimateur sans biais de θ.

7°. Démontrer qu'il est fortement consistant et asymptotiquement normal. Déterminer la variance de la loi limite.