STATISTIQUE MATHÉMATIQUE - EXAMEN.

ENSAI - Mastères spécialisés Data Science - 2021/2022 -

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

Le sujet comporte deux 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. 13 points

On considère un n-échantillon $X_1, \dots, X_n$ d'une v.a. X de loi de Poisson $\mathscr{P}(\lambda)$, où $\lambda > 0$ est un paramètre inconnu. On cherche à estimer

$$\beta = \lambda e^{-\lambda} = \mathbb{P}_\lambda[X = 1] \tag{1}$$

On note $\overline{X}$ la moyenne empirique de l'échantillon. et l'on pose, pour tout $i = 1, \dots, n$,

$$Y_i = \mathbb{1}_{[X_i = 1]} \tag{2}$$

1°. Déterminer l'estimateur du maximum de vraisemblance $\widehat{\lambda}$ de λ et en déduire celui de β, que l'on notera $\widehat{\beta}$.

2°. Étudier les propriétés asymptotiques de $\widehat{\beta}$ (consistance, normalité asymptotique).

3°. Calculer l'information de Fisher associée à β. $\widehat{\beta}$ est-il efficace ? Asymptotiquement efficace ?

4°. À partir d'un estimateur plug-in de β mettant en jeu les v.a. Y_i, déterminer un estimateur VUMSB (sans biais de variance minimale) de β.

5°. On suppose que $\theta = \lambda$ suit une loi *a priori* $\gamma(a, b)$. On rappelle que la densité d'une telle loi est

$$f(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \mathbb{1}_{[0,+\infty[}(\theta) \tag{3}$$

où Γ est la fonction Gamma d'Euler donnée par

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx \tag{4}$$

On rappelle que l'espérance d'une v.a. H de loi $\gamma(a, b)$ est donnée par

$$\mathbb{E}[H] = \frac{a}{b} \tag{5}$$

À partir de la vraisemblance *a posteriori* du modèle, déterminer un estimateur bayésien de θ et étudier ses propriétés.

II. 10 points

On considère un n-échantillon $X_1, \dots, X_n$ d'une v.a. X dont la densité est donnée par

$$f(x) = \frac{x}{2\theta^2} \mathbb{1}_{[0,2\theta]}(x) \tag{6}$$

où θ est un paramètre inconnu que l'on souhaite estimer.

1°. Calculer la vraisemblance $L(x_1, \dots, x_n, \theta)$ du n-échantillon.

2°. Déterminer l'estimateur du maximum de vraisemblance $\widehat{\theta}_1$ de θ. Le modèle est-il régulier ?

3°. À partir de la fonction de répartition F de X, déterminer la loi de $\widehat{\theta}_1$, son espérance, sa variance, son biais et son risque quadratique.

4°. Déterminer la fonction de répartition de $n(\theta - \widehat{\theta}_1)$, puis en déduire la loi limite de cette suite. Montrer que $\widehat{\theta}_1$ est consistant.

5°. Déterminer l'estimateur des moments $\widehat{\theta}_2$ de θ. Étudier la convergence en moyenne quadratique, la consistance forte et la normalité asymptotique.

On suppose que $n = 2m+1$, on note M la médiane de X et $\widehat{M_n}$ la médiane empirique de l'échantillon.

6°. Exprimer $\widehat{M_n}$ en fonction d'une statistique d'ordre de l'échantillon. Déterminer l'expression de M en fonction de θ et en déduire un estimateur plug-in $\widehat{\theta}_3$ de θ.