

Feuille de TD 3 : IC et information de Fisher

Exercice 1 (Information de Fisher et inégalité de Cramér-Rao)

Soit X une variable aléatoire de densité $f(x, \theta)$ paramétrée par $\theta \in \Theta$. On considère une statistique $T(X)$ telle que $E[T(X)] = \psi(\theta)$ avec ψ dérivable. On suppose que f est de classe C^3 et qu'on peut inverser les symboles d'intégration et de dérivation dans les expressions

$$\int \frac{\partial f(x, \theta)}{\partial \theta} dx, \quad \int T(x) \frac{\partial f(x, \theta)}{\partial \theta} dx \quad \text{et} \quad \int \frac{\partial^2 f(x, \theta)}{\partial \theta^2} dx.$$

1. On note V le score défini par

$$V = \frac{\partial}{\partial \theta} \log f(X, \theta).$$

Calculer $E[V]$ et montrer que $\text{Cov}(V, T) = \psi'(\theta)$.

2. Montrer que

$$\text{Var}(V) = -E \left[\frac{\partial^2}{\partial \theta^2} \log f(X, \theta) \middle| \theta \right].$$

On note cette quantité $\mathcal{I}(\theta)$ et on l'appelle *information de Fisher*.

3. Montrer la borne de Cramér-Rao (ou inégalité de l'information) :

$$\text{Var}(T) \geq \frac{\psi'(\theta)^2}{\mathcal{I}(\theta)}.$$

4. On considère l'estimateur du maximum de vraisemblance $\hat{\theta}$. On suppose qu'on est dans un cadre qui assure que $\hat{\theta}$ est consistant, et que $\theta \in \mathring{\Theta}$. On note ℓ la fonction de log-vraisemblance. Montrer que pour un certain $\tilde{\theta} \in]\theta, \hat{\theta}[$, on a

$$\frac{\partial \ell}{\partial \theta}(\theta) + (\hat{\theta} - \theta) \frac{\partial^2 \ell}{\partial \theta^2}(\tilde{\theta}) = 0$$

puis, en notant $\mathcal{I}_1$ l'information de Fisher associée à une observation, donner une hypothèse sur la densité f qui permette de montrer que

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \mathcal{I}_1(\theta)^{-1})$$

Exercice 2 (EMV et modèle gaussien)

Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ où $\mu \in \mathbb{R}$ et $\sigma^2 > 0$ sont des paramètres inconnus.

1. Calculer l'information de Fisher associée à (μ, σ^2) .
2. Calculer l'estimateur du maximum de vraisemblance $(\hat{\mu}, \hat{\sigma}^2)$ de (μ, σ^2) . Montrer que $\hat{\sigma}^2$ est biaisé. Quelle est sa loi ? Vérifie-t-il la borne de Cramér-Rao ?
3. Proposer un estimateur sans biais de σ^2 . Montrer qu'il a un plus grand risque quadratique que l'estimateur du maximum de vraisemblance. Comparer sa variance à la borne de Cramér-Rao.

Exercice 3 (Loi de Poisson)

Soit $X_1, \dots, X_n$ des variables aléatoires i.i.d suivant une loi de Poisson de paramètre λ et

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

1. Montrer que $\bar{X}_n$ est un estimateur sans biais et asymptotiquement normal de λ .
2. Montrer que S_n^2 est un estimateur sans biais de λ et montrer sa normalité asymptotique.
On rappelle que $\mathbb{E}[(X_1 - \lambda)^4] = \lambda + 3\lambda^2$.
3. Quel estimateur privilégier ?
4. Montrer que
 - (a) $\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\bar{X}_n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$
 - (b) $\sqrt{n} \frac{\bar{X}_n - \lambda}{S_n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$
 - (c) $\sqrt{n}(g(\bar{X}_n) - g(\lambda)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$ pour un bon choix de g .
5. Déterminer les intervalles de confiance correspondant, lequel est le meilleur ?

Exercice 4 (Loi uniforme)

Soit $X_1, \dots, X_n$ un échantillon de loi uniforme sur $[0, \theta]$.

1. Le modèle est-il régulier ?
2. Par la méthode des moments, proposer un estimateur de θ . Montrer sa consistance et sa normalité asymptotique. En déduire un intervalle de confiance asymptotique de niveau $1 - \alpha$.
3. Soit $\theta_0 > 0$, proposer un test de niveau asymptotique α pour tester

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta \neq \theta_0.$$

4. Calculer l'estimateur du maximum de vraisemblance et calculer son risque quadratique.
5. Calculer la loi limite de $n(\theta - \hat{\theta}_n^{\text{MV}})$.
6. Déterminer $c_{\alpha,n}$ tel que $[X_{(n)}, c_{\alpha,n}X_{(n)}]$ soit un intervalle de confiance de niveau $1 - \alpha$.
7. Calculer la médiane de X_1 et en déduire un nouvel estimateur de θ . Donner sa normalité asymptotique ainsi qu'un nouvel intervalle de confiance asymptotique.
8. Quel intervalle choisiriez-vous ?
9. Soit $\lambda_0 > 0$. En déduire un test de niveau α pour tester $H_0 : \lambda = \lambda_0$ contre $H_1 : \lambda \neq \lambda_0$.