Feuille 3 : Estimateurs efficaces et intervalles de confiance

Allan Merino

Master 1 Mathématiques, UE Statistiques, Février 2016

Exercice 1

Calculer la borne de Cramer-Rao pour la famille $\mathcal{N}(\mu, 1)$ où μ est inconnu. Même question pour la famille $\mathcal{N}(0, \sigma^2)$ où σ^2 est inconnu. Dans chaque cas, montrer que l'estimateur naturel est efficace.

Exercice 2

Soit la famille des lois de Bernouilli $\mathcal{B}(p)$. Donner la borne de Cramer-Rao pour un estimateur sans biais du rapport $\frac{p}{1-p}$.

Exercice 3

Trouver l'estimateur du maximum de vraisemblance pour le paramètre θ de la loi de Pareto avec a inconnu. Quel est son biais ? Calculer la borne de Cramer-Rao et constater que l'EMV est asymptotiquement efficace.

Exercice 4

On considère une variable aléatoire X positive ou nulle et distribuée sur $\mathbb{R}^+$ suivant la loi Gamma (ou encore loi d'Erlang) de paramètres 3 et θ ($\theta > 0$), c'est-à-dire la loi de densité de probabilité :

$$f(x, \theta) = \frac{x^2 e^{-\frac{x}{\theta}}}{2\theta^3}.$$

1. Calculer $E(X)$, $V(X)$ et montrer que l'estimateur du maximum de vraisemblance du paramètre θ est défini par $\hat{\theta} = \frac{\bar{X}_n}{3}$ où $\bar{X}_n$ désigne la moyenne empirique des valeurs d'un n-échantillon indépendant $(X_1, \ldots, X_n)$ de même loi que X.
2. Montrer que $\hat{\theta}$ est sans biais, convergent et efficace.

Exercice 5

Soit $(X_1, \ldots, X_n)$ un échantillon i.i.d. de loi de Poisson de paramètre $\theta > 0$.

1. Dites en quoi la moyenne et la variance empirique, $\bar{X}_n$ et S_n^2 respectivement, sont des estimateurs sans biais de θ.

2. Montrer que $\bar{X}_n$ est un estimateur effiace. Comparez les variances de $\bar{X}_n$ et S_n^2.

Exercice 6

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur l'intervalle $[0, \theta]$.

1. Calculer, si elles existent, l'espérance et la variance de cette loi.
2. (a) Considérons à présent
$$T_n = \frac{2}{n}\sum_{i=1}^{n} X_i.$$
Calculer le biais et l'erreur quadratique moyenne de T_n (on rappelle que l'erreur quadratique moyenne de T, estimateur de θ, est définie par $E((T-\theta)^2)$).
   (b) Etudier les propriétés asymptotiques de T_n.
3. (a) Montrer que $M_n = max(X_1, \ldots, X_n)$ est l'estimateur du maximum de vraisemblance de θ.
   (b) Calculer F_{M_n}, la fonction de répartition de M_n.
   (c) En déduire le biais et l'erreur quadratique moyenne de M_n.
   (d) Montrer que pour tout $\epsilon > 0$, $\mathbb{P}(|M_n - \theta| \geqslant \epsilon) \to 0$ lorsque $n \to \infty$. Conclure.
   (e) Montrer que pour $x \geqslant 0$, $\mathbb{P}(n(\theta - M_n) \leqslant x) \to 1 - exp(-\frac{x}{\theta})$ lorsque $n \to \infty$. Que signifie ce résultat ?
4. Quel est le meilleur estimateur de θ : T_n ou M_n ?

Exercice 7

Soit $(\Omega, \mathscr{A}, m)$ un espace mesuré. Soit $T : (\Omega, \mathscr{A}) \to \mathbb{R}^d$ une application mesurable et $f : \Theta \to \mathbb{R}^d$ telles que
$$(\forall \theta \in \Theta), \int_\Omega exp(< f(\theta), T(\omega) >) dm(\omega) \in \mathbb{R}_+^*.$$
On définit ϕ par $\phi(\theta) = ln\left(\int_\Omega exp(< f(\theta), T(\omega) >) dm(\omega)\right)$. La famille $(\mathbb{P}_\theta)_{\theta\in\Theta}$ de probabilités définies par $\mathbb{P}_\theta = p_\theta . m$, avec $p_\theta(\omega) = exp(< f(\theta), T(\omega) > -\phi(\theta))$ constitue sur $(\Omega, \mathscr{A})$ un modèle statistique appelé modèle exponentiel.

1. (Résultat préliminaire) Soit $(\alpha_1, \alpha_2) \in \mathbb{R}^2$, $\alpha_1 < \alpha_2$. Montrer que
$$(\forall \alpha \in]\alpha_1, \alpha_2[), |u|e^{\alpha u} \leqslant A(e^{\alpha_1 u} + e^{\alpha_2 u}) \quad (\forall u \in \mathbb{R}),$$
où A est une constante positive.
2. Soit un modèle exponentiel de la forme $\mathbb{P}_\theta(d\omega) = p_\theta(\omega) dm(\omega)$, avec Θ un ouvert de $\mathbb{R}^p$. Posons $f = Id$ (donc $d = p$). Montrer que T est un estimateur sans biais de $g(\theta) = grad\phi(\theta) = \left(\frac{\partial\phi(\theta)}{\partial\theta_1}, \ldots, \frac{\partial\phi(\theta)}{\partial\theta_p}\right)$.
3. Montrer que la matrice de covariance $K_T(\theta)$ a pour terme général $\frac{\partial^2\phi(\theta)}{\partial\theta_i\partial\theta_j}$.
4. Montrer que T est de dispersion minimale.

$\boxed{\textbf{Exercice 8}}$

Soit $p \in \Theta = [0,1]$ et $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathscr{B}(1,p)$. Posons $F_n = \bar{X}_n$

1. Montrer que $0 \leqslant \sqrt{p(1-p)} \leqslant \frac{1}{2}$.
2. En déduire que
$$\left]F_n - \frac{z_{1-\frac{\alpha}{2}}}{2\sqrt{n}}; F_n + \frac{z_{1-\frac{\alpha}{2}}}{2\sqrt{n}}\right[$$
est un intervalle de confiance approximatif de niveau au moins $1-\alpha$ de p (où $\alpha \in]0,1[$).

$\boxed{\textbf{Exercice 9}}$

Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi normale $\mathscr{N}(\mu, \sigma^2)$ avec $\mu \in \mathbb{R}$ et $\sigma > 0$. Soit $\alpha \in]0,1[$. Montrer que si μ est inconnu, alors l'intervalle de confiance symétrique de niveau $1-\alpha$ pour σ^2 est

$$\left]\frac{(n-1)\tilde{\mathrm{S}}_{\mathrm{n}}^2}{k_{n-1,1-\frac{\alpha}{2}}}; \frac{(n-1)\tilde{\mathrm{S}}_{\mathrm{n}}^2}{k_{n-1,\frac{\alpha}{2}}}\right[,$$

où $k_{m,p}$ est le p-quantile d'une loi $\chi^2(m)$.

$\boxed{\textbf{Exercice 10}}$

Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathbb{P}_\theta, \theta \in \Theta$. Trouver des intervalles de confiance asymptotiques de niveau $1-\alpha$ pour le paramètre θ dans les cas suivants :

1. $\Theta = [0,1]$ et $\mathbb{P}_\theta = Ber(\theta)$ (loi de Bernouilli),
2. $\Theta = \mathbb{R}^+$ et $\mathbb{P}_\theta = \mathscr{P}(\theta)$ (loi de Poisson),
3. $\Theta = \mathbb{R}_+^*$ et $\mathbb{P}_\theta = \mathscr{U}([0,\theta])$ (loi uniforme).