

Maîtrises Math-Mass 2001-2002

Partiel - 13 novembre 2001

9h - 11h30

Problème Propriétés de la moyenne empirique d'un échantillon gaussien.

Soit $n \geq 1$ quelconque. On considère un n -échantillon (X_1, \dots, X_n) de variables gaussiennes de loi $\mathcal{N}(\theta, 1)$, avec $\theta \in \mathbb{R}$. On note $\bar{X}(n) = \frac{1}{n} \sum_{i=1}^n X_i$ la moyenne empirique. Le but de ce problème est d'établir quelques-unes de ses propriétés en tant qu'estimateur de θ . La fonction de coût considérée est comme toujours la fonction de coût quadratique.

1) Premières propriétés.

- Préciser le modèle statistique $(\Omega^{(n)}, \mathcal{A}^{(n)}, \{\mathbb{P}_\theta^{(n)}, \theta \in \Theta\})$ associé à l'échantillon (X_1, \dots, X_n) .
- Montrer que $\bar{X}(n)$ est une statistique exhaustive, minimale et complète.
- Montrer que $\bar{X}(n)$ est un estimateur sans biais, de risque constant.
- Montrer que le modèle statistique est un modèle régulier, et que $\bar{X}(n)$ est un estimateur régulier, efficace, et uniformément de variance minimale parmi les estimateurs sans biais. Calculer l'information de Fisher $I_n(\theta)$ associée.

2) On veut ici montrer que $\bar{X}(n)$ est un estimateur admissible.

- Rappeler ce que signifie un estimateur meilleur qu'un autre, et ce qu'est un estimateur admissible. En déduire qu'un estimateur meilleur que $\bar{X}(n)$ est nécessairement biaisé.
- Soit T un estimateur quelconque de θ , de carré intégrable. On note $b(\theta) = \mathbb{E}_{\mathbb{P}_\theta^{(n)}}[T] - \theta$ son biais, et $R(\theta, T)$ son risque quadratique. Montrer que cet estimateur est régulier, et que b est dérivable (*il est conseillé de l'admettre dans un premier temps*). A l'aide de l'inégalité de Cramer-Rao, établir l'inégalité suivante :

$$R(\theta, T) \geq b^2(\theta) + \frac{(1 + b'(\theta))^2}{n}$$

c) On suppose T meilleur que $\bar{X}(n)$. Montrer qu'alors on a l'inégalité

$$b^2(\theta) + \frac{(1 + b'(\theta))^2}{n} \leq \frac{1}{n}.$$

En déduire que b est une fonction bornée et décroissante sur \mathbb{R} , et qu'elle admet donc des limites finies en $-\infty$ et en $+\infty$. Montrer que si ces limites ne sont pas nulles, alors b n'est pas bornée.

d) Conclusion.

3) Montrer que $\bar{X}(n)$ est un estimateur minimax.

4) On considère désormais le modèle statistique d'échantillonnage $(\Omega, \mathcal{A}, \{\mathbb{P}_\theta, \theta \in \Theta\})$ avec $\Theta = \mathbb{R}$, associé à la famille de variables gaussiennes indépendantes $(X_n)_{n \geq 1}$ de loi $\mathcal{N}(\theta, 1)$, et on veut déterminer la robustesse de la famille d'estimateurs $(\bar{X}(n), n \geq 1)$. Cette notion a été introduite pour caractériser des estimateurs peu sensibles à des modifications légères du modèle statistique de base.

Soient $(T_n)_{n \geq 1}$ une suite d'estimateurs qui converge \mathbb{P}_θ -p.s. vers θ . Supposons qu'en réalité les variables $(X_n)_{n \geq 1}$ ne soient pas exactement gaussiennes, mais suivent plutôt une loi du type

$$\mu_{\theta,t,\nu} = (1-t)\mathcal{N}(\theta, 1) + t\nu$$

avec ν mesure de probabilité inconnue et $t \in [0, 1]$ "proche" de 0. Autrement dit, pour toute fonction borélienne bornée, quel que soit i ,

$$\mathbb{E}[f(X_i)] = (1-t) \int_{-\infty}^{+\infty} f(u) \frac{e^{-\frac{(u-\theta)^2}{2}}}{\sqrt{2\pi}} du + t\nu(f)$$

On note $\mathbb{P}_\theta^{(\nu,t)}$ la loi correspondante sur Ω . Si elle existe, on note $T_\infty(\theta, \nu, t)$ la limite $\mathbb{P}_\theta^{(\nu,t)}$ -p.s. de $(T_n)_{n \geq 1}$. On dit que la suite d'estimateurs $(T_n)_{n \geq 1}$ est robuste si et seulement si le taux d'accroissement $\frac{T_\infty(\theta, \nu, t) - \theta}{t}$ est bornée uniformément sur l'ensemble des lois ν à support compact quand t tend vers 0, ou, de façon équivalente et sous réserve de dérivabilité, si et seulement si

$$\sup_{x \in \mathbb{R}} \left| \frac{\partial T_\infty}{\partial t}(\theta, \delta_x, 0) \right| < +\infty$$

avec δ_x la masse de Dirac en x .

- a) On suppose $T_n = \bar{X}(n)$. Montrer que $\bar{X}_\infty(\theta, \delta_x, t)$ existe quel que soit $x \in \mathbb{R}$ et $t \in [0, 1]$, puis que $(\bar{X}(n))_{n \geq 1}$ n'est pas robuste.
- b) Soit $\phi(x)$ la fonction de répartition de la loi gaussienne centrée réduite, et Y une gaussienne centrée réduite indépendante de X_1 sous \mathbb{P}_θ . Montrer que $\phi(X_1) = \mathbb{E}_{\mathbb{P}_\theta}[1_{Y \leq X_1} | X_1]$ \mathbb{P}_θ -p.s. En déduire

$$\theta = \sqrt{2}\phi^{-1}(\mathbb{E}_{\mathbb{P}_\theta}[\phi(X_1)])$$

- c) A partir de l'égalité précédente, et à l'aide de la méthode de substitution, construire une nouvelle famille d'estimateurs de θ , notée $(U_n)_{n \geq 1}$, dont on montrera qu'elle est fortement consistante sous \mathbb{P}_θ et qu'elle converge $\mathbb{P}_\theta^{(\nu,t)}$ -p.s. quels que soient ν et t .
- d) Montrer que $(U_n)_{n \geq 1}$ est robuste.