Estimation et Identification statistiques
Contrôle Année 2020-2021

Les exercices sont indépendants.

Exercice 1 *Estimation fréquentiste* (10 points)
On considère une suite de variables aléatoires $Z_1, \ldots, Z_n$, indépendantes suivant toutes une loi de Pareto de mêmes paramètres (θ, a). La densité d'une telle loi est définie sur $[1, \infty[$ et donnée par $f(z) = \frac{a\theta^a}{z^{a+1}}$. On suppose ici $a > 2$ et $\theta > 0$ et on admettra que $E(Z) = \frac{a\theta}{a-1}$ et $V(Z) = E(Z^2) - (E(Z))^2) = \frac{a\theta^2}{(a-1)^2(a-2)}$.
La fonction de répartition est $F(x) = P(Z < x)$ est $F(x) = 1 - \left(\frac{\theta}{x}\right)^a$ pour $x \geq \theta$

1. On cherche à estimer θ en supposant a connu, tout d'abord à l'aide de l'estimateur de la moyenne empirique $Z_m = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Montrer qu'on peut déterminer une variable K tel que l'estimateur KZ_m soit un estimateur sans biais de θ. Calculer sa variance. L'estimateur est il convergent ?
2. On considère à présent l'estimateur $\tilde{Z} = min(Z_1, \ldots, Z_n)$. Déterminer la fonction de répartition de $\tilde{Z}$, $G(x) = P(\tilde{Z} < x)$, à partir de la fonction de répartition de Z. En déduire l'espérance et la variance de la variable aléatoire $\tilde{Z}$. Déterminer la valeur du coefficient K' pour que $K'\tilde{Z}$ soit un estimateur sans biais de θ. Est il convergent ?
3. Comparer les variances des deux estimateurs non biaisés. Quel est le plus efficace ?
4. Il s'agit à présent d'estimer au sens du maximum de vraisemblance le paramètre a de la loi de Pareto. On suppose le paramètre θ de la loi connu dans la suite. Pour estimer a, une première étape consiste à calculer la vraisemblance $L(a; z_1, \ldots, z_n)$, où chaque z_i est une réalisation de Z_i. A partir de l'expression de la vraisemblance, calculer la Log-vraisemblance et montrer que $T(Z) = \frac{n}{\sum_{i=1}^{n} \ln \frac{Z_i}{\theta}}$ est l'estimateur de a au sens du maximum de vraisemblance.
5. Déterminer l'espérance $E(T(Z))$ pour $n \geq 2$ et en déduire un estimateur $T'(Z) = KT(Z)$ qui soit un estimateur sans biais de a.
   REMARQUE : On utilisera les deux résultats suivants :
   (a) Lorsque X est une variable aléatoire suivant une loi de Pareto de paramètres θ et a, la variable aléatoire $Y = \ln\left(\frac{X}{\theta}\right)$ suit une loi exponentielle de paramètre a (ce qui signifie que la densité de Y est $f_Y(t) = ae^{-at}$).
   (b) Une somme de n variables aléatoires suivant une loi exponentielle de paramètre λ suit une loi gamma de paramètres n et λ (ce qui signifie que la densité est $f_X(x) = \frac{\lambda^n}{(n-1)!}x^{n-1}\exp(-\lambda x)\mathbb{1}_{\mathbb{R}^{+*}}(x)$, son espérance est $\frac{n}{\lambda}$ et sa variance $\frac{n}{\lambda^2}$).

Exercice 2 *Estimation bayésienne* (6 points)
On collecte une mesure z relié à un paramètre θ par $z = \theta + b$. Le bruit b a une densité de probabilité $f(b) = \frac{b}{2}$ sur $[0,\ 2]$ et est indépendant de θ. La densité de probabilité *a priori* de θ est $f(\theta) = 2\theta$ sur $[0,\ 1]$.

1. Calculer la densité a posteriori $f_{\theta|z}$ du paramètre θ étant donnée l'observation z. *On ne calculera pas l'expression de $f(z)$ de la loi de Bayes.*
2. Après mesure, on obtient $z = 2.5$. Donner les estimés de θ au sens du risque quadratique minimal, du maximum de vraisemblance a posteriori, du risque en valeur absolue a posteriori et de la moyenne a posteriori.
3. Comparer ces résultats. Quel est selon vous le meilleur estimateur ?

Exercice 3 Un système est modélisé sous la forme $z_n = \theta u_n + b_n$, pour $1 \leq n \leq N$. θ doit être estimé à partir de N mesures, réalisations de v.a.rélles indépendantes $z_1, \ldots, z_N$. Les $b_n, 1 \leq n \leq N$ forment une séquence de v.a. i.i.d. suivant une loi gaussienne de moyenne nulle et de variance σ^2. La séquence $u_n, 1 \leq n \leq N$ est supposée connue (de façon exacte).

1. Calculer $f(Z|\theta)$ avec $Z = [z_1, \ldots, z_N]^T$. On notera $U = [u_1, \ldots, u_n]^T$.
2. Avant l'obtention des mesures, le paramètre θ est supposé compris entre 1 et 5. Donner une méthode permettant de prendre en compte, lors de l'estimation, cette information *a priori*.
3. En déduire $f(\theta|Z)$. *On ne calculera pas l'expression de $f(Z)$ de la loi de Bayes.*
4. Tracer les allures respectives des lois $f(\theta|Z)$, $f(Z|\theta)$ et $f(\theta)$ dans le cas scalaire ($N = 1$).
5. Calculer et comparer le maximum de vraisemblance et le maximum *a posteriori*.

Exercice 4 *Sensibilité aux données aberrantes* (6 points)
Soient $z_1, \ldots, z_N$, N réalisations d'une variable aléatoire θ que l'on cherche à estimer.

1. Exprimer l'estimateur du risque quadratique minimal dans ce cas simple.
2. Établir dans le même contexte l'expression de l'estimateur du risque minimal en moyenne absolue (on distinguera deux cas suivant la parité de N).
3. On suppose qu'une des données est aberrante : suite à un problème de mesure, z_n prend une valeur très éloignée de θ pour un certain $n \in \{1, ..., N\}$. Comparer la robustesse des deux estimateurs précédents à cet écart arbitraire (c'est à dire l'écart entre la valeur estimée en l'absence de donnée aberrante et avec la donnée aberrante). Commenter.