ANNALES SCIENTIFIQUES DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Mathématiques

C. HUBER

Estimation fonctionnelle. Risque minimax

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 69, série *Mathématiques*, nº 19 (1981), p. 71-75

<http://www.numdam.org/item?id=ASCFM_1981__69_19_71_0>

ESTIMATION FONCTIONNELLE

RISQUE MINIMAX

C. HUBER

Université de Paris-Nord

1 - Introduction

On donne le risque minimax dans L^2 pour l'estimation d'une fonction réelle de variable réelle dans les cas où cette fonction est :

a) une densité de probabilité

b) la densité spectrale d'un processus Gaussien stationnaire

c) l'intensité d'un processus de Poisson

Dans les trois cas, f est supposée appartenir à un sous ensemble F de $\mathcal{L}^2(\mathbb{R},dx)$ et f_n, $n \in \mathbb{N}$, est une suite d'estimateurs de f, fondée sur la suite d'observations X_n, régie par la probabilité P_f^n, dès que $f_n(X_n,.)$ est mesurelle à valeurs dans un ensemble G contenant F et inclus dans $\mathcal{L}^2(\mathbb{R},dx)$, muni de la tribu de ses boréliens. Pour chaque n, l'ensemble de tous ces estimateurs est noté F_n et le risque de f_n en point f, est, par définition :

$$R(f_n,f) = E_{P_f^n}(||f_n(X_n,.) - f||_2^2) \quad .$$

Les résultats que l'on démontre sont du type suivant :

Il existe a, A et B, constantes ne dépendant que de F, et non de n, telles que :

$$(*) \quad \begin{aligned} &\lim_n \inf_{f_n \in F_n} \sup_{f \in F} R(f_n,f) \geqslant A\, n^{-a} \\ &\exists\, \hat{f}_n \in F_n,\ n \in \mathbb{N} : \lim_n \sup_{f \in F} R(\hat{f}_n,f) \leqslant B\, n^{-a} \end{aligned}$$

Dans chacun des trois cas, les constantes a, A et B sont explicitées en fonction de F.

2 - Résultats

Théorème 1 (1)

Si F est l'ensemble des densités de probabilité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ qui admettent des dérivées jusqu'à l'ordre s $(s \geq 1)$ et dont la dérivée d'ordre s est bornée par b dans $\mathcal{L}^2(\mathbb{R}, dx)$, et si X_n est un n-échantillon d'une variable aléatoire régie par la loi $f\,dx$, le résultat (*) a lieu avec :

$$a = \frac{2s}{2s+1}, \quad A = b^{\frac{1}{2s+1}}\, g_1(s), \quad B = b^{\frac{1}{2s+1}}\, g_2(s)$$

Le rapport des deux constantes g_1 et g_2 est majoré par Cs, où C est une constante universelle :

$$\frac{g_2(s)}{g_1(s)} \leq C\, s$$

(Voir (1) pour une expression explicite de g_1 et de g_2).

Théorème 2 (2)

Si f est la densité spectrale d'un processus gaussien stationnaire, indexé par $\mathbb{N}$, si $F = \{f : \|f^{(s)}\|_2 \leq b\}$, et si X_n consiste en n observations successives $X_1, X_2, \ldots, X_n$ de ce processus, le résultat (*) a lieu avec :

$$a = \frac{2s}{2s+1}, \quad A = \frac{1}{4}\, b^2 \quad \text{et} \quad B = 6\pi(M^2 + k^2)$$

où M et k sont les deux caractéristiques suivantes d'une fonction auxiliaire K : K est un noyau positif défini continu sur $[-1+1]$, symétrique par rapport à 0 et tel que $K(0) = 1$, $\underset{x\varepsilon[-1+1]}{\mathrm{Max}} K(x) = M$ et $\frac{1-K(x)}{|x|^s} \leqslant k$, $x \varepsilon [-1+1]$.

Théorème 3 (3)

Si f est la densité d'un processus de Poisson indexé par $[01]$, si $F=\{f : ||f^{(s)}||_2 < b , 0 \leqslant m < |f| \leqslant M\}$, et X_n une observation sur $[01]$ d'un processus de Poisson de densité nf, le résultat (*) a lieu avec :

$$a = \frac{2s}{2s+1}, \quad A = ||f||_1^{\frac{2s+2}{2s+1}} \; b^{\frac{1}{2s+1}} \; g_1(s)$$

$$B = ||f||_1^{\frac{1}{2s+1}} \; b^{\frac{1}{2s+1}} \; g_2(s)$$

3 - Principe de démonstration

Le résultat (*) se démontre en deux étapes : une minoration du risque maximum, une majoration du risque pour un estimateur particulier. En ce qui concerne la majoration, on choisit, dans les trois cas, un estimateur à noyau et le risque est majoré en considérant séparement le biais et l'aléa de l'estimateur. Une estimation adaptative (1), permet non seulement d'atteindre la bonne vitesse (l'ordre a en n) mais aussi d'optimiser la constante multiplicative.

Principe de minoration

Les trois problèmes ci-dessus admettent la formulation abstraite commune suivante :

La famille H des lois qui régissent l'observation X, à valeurs dans $(E,\mathcal{B})$, est indexée par un espace fonctionnel F. Sur $G \times G$, G contenant F, est définie une pseudo distance D qui sert à définir le risque $R(\hat{f},f)$ d'un estimateur :

$$R(\hat{f},f) = E_f\left[D(\hat{f},f)\right]$$

$\hat{f}$ variant dans $\hat{F}$ ensemble des applications de $(E, \mathcal{B})$ dans G . Les propriétés que l'on exige de D sont :

(i) l'inégalité triangulaire :

$$\exists\, B > 0 \ : \ D(f,g) \leqslant B[D(f,h) + D(h,g)] \qquad f,g,h \in G$$

(ii) la suradditivité :

$$D(f,g) \geqslant \Sigma\, D(f\, 1_{I_j}, g\, 1_{I_j})\ ;\ (I_j)_{j \in J} \quad \text{partition de } \mathbb{R}$$

Supposons qu'on puisse construire dans F un ensemble F_o à 2^N éléments indexé par le cube $C = \{-1, +1\}^N$ d'élément générique $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)$ tel que la propriété suivante de "cube" a lieu :

$$D(f_\varepsilon, f_{\varepsilon'}) \geqslant \Delta \sum_j |\varepsilon_j - \varepsilon'_j|$$

$$h^2(P_{f_\varepsilon}, P_{f_{\varepsilon'}}) \leqslant \delta \sum_j |\varepsilon_j - \varepsilon'_j| \qquad \delta < \frac{1}{4}$$

où Δ et δ sont deux réels positifs et h^2 est le carré de la distance de Hellinger $(h^2(P\, Q) = \int (\sqrt{dP} - \sqrt{dQ})^2)$.

Pour le problème intermédiaire qui consisterait à estimer f dans F_o grâce à un estimateur à valeurs dans F_o , on voit que :

$$(**) \qquad \inf_{\hat{f} \in F_o} \sup_{f \in F_o} R(\hat{f},f) \geqslant \frac{N\Delta}{2} \exp(-8\delta)$$

en commençant par considérer le cas le plus simple où $N = 2$, et en minorant ensuite, pour N quelconque, le risque minimax par le risque Bayesien pour la loi a priori uniforme sur F_o. L'inégalité (**) résulte alors de ce que pour tout couple (P,Q) de probabilités

$$\int dP \wedge dQ \geqslant \frac{1}{2} \left(\int \sqrt{dPdQ}\right)^2 = \frac{1}{2}(1-h^2(P,Q))^2 \geqslant \frac{1}{2} \exp(-4\, h^2(P,Q)) .$$

pourvu que $h^2(P,Q)$ soit inférieur à $1/2$.

Une inégalité analogue en résulte pour le problème d'estimation initial :

$$(***) \qquad \inf_{\hat{f}\in\hat{F}} \ \sup_{f\in F} \ R(\hat{f},f) \geqslant \frac{1}{4B} N \ \Delta \ \exp(-8\delta)$$

Il s'agit donc, dans les trois cas, de construire deux "cubes" en correspondance, dans F_n et dans F de façon à maximiser le minorant de (***) pour chaque n. On voit, d'après le principe de la démonstration, que d'une part d'autres risques que le risque L^2 peuvent etre considérés (risques L^p, pour $p \geqslant 1$,risque de Hellinger $D(f,\hat{f}) = h^2(f,\hat{f})$ etc), et que

d'autre part il permet d'obtenir des résultats non nécessairement asymptotiques, mais valables pour n fini fixé.

(1) J. BRETAGNOLLE et C. HUBER : "Estimation des densités : risque minimax"
Z. für Wahr. 47 , 119-137 , (1979) .

(2) C. HUBER : "Estimation d'une densité spectrale"
Thèse, Orsay (1978) .

(3) C. HUBER : "Estimation de la densité d'un processus de Poisson"
Manuscrit (1979) .