Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Acad. Sci. Paris, Ser. I 340 (2005) 623–626
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Estimation de la densité de probabilité améliorée par pre-testing

Armel Fabrice Yode

L.A.T.P., université de Provence, 39, rue F. Joliot Curie, 13453 Marseille cedex 13, France

Reçu le 28 février 2004 ; accepté après révision le 8 mars 2005

Disponible sur Internet le 18 avril 2005

Présenté par Paul Deheuvels

Résumé

Nous considérons le problème d'estimation non-paramétrique de la densité de probabilité multidimensionnelle. Grâce au concept de risque minimax avec normalisation aléatoire introduit par Lepski [Math. Methods Statist. 8 (1999) 441–486], en considérant une hypothèse « plausible » que la densité se decompose en produit de densités marginales, nous construisons un estimateur qui peut être adaptatif et dont la qualité dépendant de l'observation est meilleure que celle de l'estimation minimax $n^{-\frac{\beta}{2\beta+d}}$ avec une probabilité controlée. ***Pour citer cet article : A.F. Yode, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

Abstract

Improvement in the accuracy of multidimensional probability density estimate by pre-testing. We consider the non-parametric problem of multidimensional probability density estimate. Using concept of minimax risk with random normalizing factor introduced by Lepski [Math. Methods Statist. 8 (1999) 441–486], by considering an independence hypothesis, we build an estimator which can be adaptive and whose accuracy, depending on the observation, is better than the minimax estimate, $n^{-\frac{\beta}{2\beta+d}}$, with prescribed confidence level. ***To cite this article: A.F. Yode, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

1. Introduction

Nous considérons l'expérience statistique engendrée par l'observation $X^n = (X_1, \ldots, X_n)$ où $X_i = (X_i^{(1)}, \ldots, X_i^{(d)})$, $i = 1, \ldots, n$, $d > 2$, sont des vecteurs aléatoires indépendants identiquement distribués ($i.i.d$) de loi commune admettant la densité f inconnue. Nous supposons que f est une fonction à support compact appartenant à un certain espace fonctionnel Σ.

Pour tout estimateur $\tilde{f}_n = \tilde{f}_n(X^n)$ de f sur Σ, le risque maximal normalisé par la suite positive $\varphi_n(\Sigma) \to 0$, $n \to +\infty$, est défini par

Adresse e-mail : yode@cmi.univ-mrs.fr (A.F. Yode).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2005.03.012

$$R_n\big(\tilde{f}_n, \Sigma, \varphi_n(\Sigma)\big) = \sup_{f\in\Sigma} E_f^n\big(\varphi_n^{-1}(\Sigma)\|\tilde{f}_n - f\|_2\big)^q, \quad q>0,\ \|f\|_2 = \Bigg(\int\limits_{[0,1]^d} f^2(x)\,\mathrm{d}x\Bigg)^{1/2}. \tag{1}$$

Soit $\mathcal{M}_n$, l'ensemble de tous les estimateurs possibles de f sur Σ. La suite $\varphi_n(\Sigma)$ est appelée vitesse optimale de convergence sur Σ en considérant le risque (1) si

(i) $\liminf_{n\to+\infty} \inf_{\tilde{f}_n\in\mathcal{M}_n} R_n(\tilde{f}_n, \Sigma, \varphi_n(\Sigma)) > 0$,
(ii) il existe $\bar{f}_n \in \mathcal{M}_n$ dit asymptotiquement optimal tel que $\limsup_{n\to+\infty} R_n(\bar{f}_n, \Sigma, \varphi_n(\Sigma)) < +\infty$.

Étant donné $\bar{f}_n$ et $\varphi_n(\Sigma)$, en utilisant (ii), pour tout $0<\gamma<1$, il existe une constante $M(\gamma)>0$ telle que

$$\inf_{f\in\Sigma} P_f^n\big\{\|\bar{f}_n - f\|_2 \leqslant M(\gamma)\varphi_n(\Sigma)\big\} \geqslant 1-\gamma, \quad \text{pour } n \text{ assez grand.} \tag{2}$$

Ainsi, avec un niveau de confiance d'au moins $1-\gamma$, la région de confiance pour f est la boule de centre $\bar{f}_n$ et de rayon $M(\gamma)\varphi_n(\Sigma)$. Le rayon de cette région est décrit par $\varphi_n(\Sigma)$. Ainsi, $\varphi_n(\Sigma)$ est considérée comme la qualité d'estimation. Peut-on trouver une normalisation qui tend plus vite vers 0 que $\varphi_n(\Sigma)$? C'est à cette question que nous tenterons de répondre dans cette Note. Il est clair que la réponse est négative dans le cadre de la théorie minimax d'après (i). Cependant, la considération d'autres types de risques peut conduire à une amélioration éventuelle de la qualité d'estimation. Nous utiliserons le concept de risque minimax avec normalisation aléatoire initié par Lepski [4] pour répondre à cette question.

2. Le problème

2.1. Motivations

Soit $\Sigma_d(\beta, L)$, $\beta = m+\tau$, $m\in\mathbb{N}$, $\tau\in(0,1]$, $L>0$, l'espace de Hölder isotrope sur $[0,1]^d$ i.e. l'ensemble des fonctions f à support compact $[0,1]^d$, m fois continûment différentiables dans chaque direction telles que pour tous $Z^{(l)} = (x_1,\dots,x_i^{(l)},\dots,x_d)\in[0,1]^d$, $l=1,2$,

$$\sup_{(x_1,\dots,x_{i-1},x_{i+1},\dots,x_d)} \left|\frac{\partial^m f}{\partial x_i^m}\big(Z^{(1)}\big) - \frac{\partial^m f}{\partial x_i^m}\big(Z^{(2)}\big)\right| \leqslant L\big|x_i^{(1)} - x_i^{(2)}\big|^\tau, \quad i=1,\dots,d.$$

Soient l'ensemble

$$\Sigma = \Bigg\{f:\mathbb{R}^d\to\mathbb{R}\colon\ f>0,\ \int\limits_{\mathbb{R}^d} f(x)\,\mathrm{d}x = 1,\ f\in\Sigma_d(\beta,L),\ \|f\|_\infty\leqslant S\Bigg\}, \quad S>0, \tag{3}$$

et le sous-ensemble de Σ

$$\Sigma_0 = \Bigg\{f\in\Sigma\colon\ \exists f_k:\mathbb{R}\to\mathbb{R},\ f_k>0,\ \int\limits_{\mathbb{R}} f_k(x_k)\,\mathrm{d}x_k = 1,\ k=1,\dots,d,\ f(x) = \prod_{k=1}^d f_k(x_k)\Bigg\},$$

où $\|f\|_\infty = \sup_{x\in[0,1]^d}|f(x)|$. Soient les estimateurs

$$\bar{f}_n(x) = \frac{1}{n\vartheta_n^d}\sum_{i=1}^n K\bigg(\frac{x-X_i}{\vartheta_n}\bigg), \quad \bar{f}_n^{(0)}(x) = \prod_{k=1}^d \bar{f}_{kn}(x_k), \quad x\in[0,1]^d, \tag{4}$$

où $\bar{f}_{kn}(x_k) = \frac{1}{nb_n}\sum_{i=1}^n K_*(\frac{x_k - X_i^{(k)}}{b_n})$, $X_i^{(k)}$ est la k-ième composante de X_i ; les fonctions K, K_* sont Lipschitziennes, à supports compacts sur $\mathbb{R}^d$ et $\mathbb{R}$ respectivement ; $\int_{\mathbb{R}^d} u_1^{a_1}\cdots u_d^{a_d}K(u)\,\mathrm{d}u = 0$ pour tout $(a_1,\dots,a_d)\in$

$\mathbb{N}^d$ tel que $\sum_{i=1}^d a_i \leqslant m$, $\int_{\mathbb{R}} u^l K_*(u)\,\mathrm{d}u = 0$ pour tout $l \in \mathbb{N}$ tel que $l \leqslant m$; les suites $\vartheta_n = C_1 n^{-\frac{1}{2\beta+d}}$ et $b_n = C_2 n^{-\frac{1}{2\beta+1}}$ où $C_1 > 0$, $C_2 > 0$. L'estimateur $\bar{f}_n$ est asmptotiquement optimal sur Σ et atteint la vitesse de convergence $\varphi_n(\Sigma) = n^{-\frac{\beta}{2\beta+d}}$. L'estimateur $\bar{f}_n^{(0)}$ atteint la vitesse $\varphi_n(\Sigma_0) = n^{-\frac{\beta}{2\beta+1}}$ sur Σ_0 i.e. $\limsup_{n\to+\infty} R_n(\bar{f}_n^{(0)}, \Sigma_0, \varphi_n(\Sigma_0)) < +\infty$.

Nous remarquons que la qualité d'estimation $\varphi_n(\Sigma)$ depend du facteur d'inflation d lié à la dimension. Pour de grandes valeurs de d, $\varphi_n(\Sigma)$ tend très lentement vers 0. Cependant, sous l'hypothèse d'indépendance, la qualité d'estimation devient $\varphi_n(\Sigma_0)$ et on ne paie pas d'effet dimensionnel au moins asymptotiquement. De ce fait, nous considérons l'hypothèse $H_0 : f \in \Sigma_0$. Une première idée pour résoudre ce problème est l'estimation adaptative (Lepski [4], Hoffmann [1], Hoffmann et Lepski [3]). La qualité d'estimation dépend dans ce cas de la densité inconnue i.e. de l'information si $f \in \Sigma_0$ ou non. Cette information ne peut être obtenue à l'issue de l'observation. D'où la deuxième idée qui est le centre d'intérêt de ce travail et qui est basée sur les travaux de Lepski [4]. Elle consiste à tester $H_0 : f \in \Sigma_0$ et à utiliser de façon optimale la réponse du test pour améliorer la qualité d'estimation $\varphi_n(\Sigma)$ si l'hypothèse est vraie sans que celle-ci soit dégradée quand l'hypothèse est fausse.

2.2. Risque minimax avec normalisation aléatoire

Soit la famille de normalisations aléatoires

$$\Omega_n = \bigl\{\rho_n \in \bigl(0, \varphi_n(\Sigma)\bigr]\colon\ \rho_n \text{ est mesurable par rapport à } X^n\bigr\}. \tag{5}$$

Pour tout $\rho_n \in \Omega_n$, pour tout $\tilde{f}_n \in \mathcal{M}_n$, introduisons le risque $R_n^{(r)}(\tilde{f}_n, \Sigma, \rho_n) = \sup_{f\in\Sigma} E_f^n(\rho_n^{-1}\|\tilde{f}_n - f\|_2)^q$. Du point de vue mathématique, l'objectif est de trouver $\rho_n^* \in \Omega_n$, $f_n^* \in \mathcal{M}_n$ tels que : $\limsup_{n\to+\infty} R_n^{(r)}(f_n^*, \Sigma, \rho_n^*) < +\infty$, la probabilité de l'événement $\{\rho_n^* \ll \varphi_n(\Sigma)\}$ soit controlée sur Σ_0 i.e. $\liminf_{n\to+\infty} \inf_{f\in\Sigma_0} P_f^n\{\rho_n^* \ll \varphi_n(\Sigma)\} > 0$ et f_n^* soit adaptatif. Soient $0 < \delta < 1$ et $0 < \alpha_n < 1 - \delta$ pour tout n, une suite fixée.

Définition 2.1 (*Hoffmann et Lepski* [3]). La caractéristique de $\rho_n \in \Omega_n$ est la suite

$$x_n(\rho_n) = \inf\Bigl\{x \in \bigl(0, \varphi_n(\Sigma)\bigr]\colon\ \inf_{f\in\Sigma_0} P_f^n\{\rho_n \leqslant x\} \geqslant 1 - \alpha_n\Bigr\}.$$

La notion de caractéristique permet de comparer les normalisations aléatoires.

Définition 2.2 (*Hoffmann et Lepski* [3]). La normalisation $\rho_n^* \in \Omega_n$ est dite α_n-optimale par rapport à Σ_0 si les conditions suivantes sont vérifiées

(1) pour tout $\rho_n \in \Omega_n$ telle que $\lim_{n\to\infty} x_n(\rho_n)/x_n(\rho_n^*) = 0$ alors $\liminf_{n\to\infty} \inf_{\tilde{f}_n\in\mathcal{M}_n} R_n^{(r)}(\tilde{f}_n, \Sigma, \rho_n) = +\infty$,

(2) il existe un estimateur f_n^* dit α_n-adaptatif tel que $\limsup_{n\to\infty} R_n^{(r)}(f_n^*, \Sigma, \rho_n^*) < +\infty$.

Ce nouveau concept a permis à Lepski [4] d'améliorer la qualité d'estimation d'un signal dans le modèle du bruit blanc gaussien unidimensionnel et à Hoffmann [1] celle du coéfficient de diffusion en dimension 1. Une extension de ce concept est proposée par Hoffmann et Lepski [3] et est utilisée pour la recherche de variables significatives dans le cas de la regression multiple.

2.3. Résultats

Introduisons l'estimateur à noyau défini par $\hat{f}_n(x) = \frac{1}{nh_n^d}\sum_{i=1}^n K(\frac{x-X_i}{h_n})$, où $h_n = C(n^{-1}\sqrt{\log(\frac{2}{\alpha_n})})^{\frac{2}{4\beta+d}}$, $C = C(\beta, d, K)$ est une constante strictement positive.

Soient la statistique

$$T_n = \|\hat{f}_n - \bar{f}_n^{(0)}\|_2^2 - \frac{1}{n^2 h_n^{2d}} \sum_{i=1}^n \int_{[0,1]^d} K^2\biggl(\frac{x - X_i}{h_n}\biggr)\mathrm{d}x, \tag{6}$$

l'événement $\mathcal{A}_n = \{T_n \leqslant (\lambda\varphi_n(\alpha_n))^2\}$ où $\varphi_n(\alpha_n) = (n^{-1}\sqrt{\log\frac{2}{\alpha_n}})^{\frac{2\beta}{4\beta+d}}$, $\lambda = 2^{\frac{12\beta+d}{4\beta+d}} L_0^{\frac{d}{4\beta+d}} \Upsilon^{\frac{\beta}{4\beta+d}}$ et posons

$$\rho_n^* = \begin{cases} \varphi_n(\alpha_n) & \text{si } \mathcal{A}_n \text{ est vrai,} \\ \varphi_n(\Sigma) & \text{si } \mathcal{A}_n^c \text{ est vrai,} \end{cases} \qquad f_n^* = \begin{cases} \bar{f}_n^{(0)} & \text{si } \mathcal{A}_n \text{ est vrai,} \\ \bar{f}_n & \text{si } \mathcal{A}_n^c \text{ est vrai,} \end{cases}$$

où $\mathcal{A}_n^c$ est le complémentaire de $\mathcal{A}_n$, $\bar{f}_n^{(0)}$ et $\bar{f}_n$ sont définis par (4). Fixons les constantes positives A_*, B_* et r^* telles que

$$A_* > \frac{8S\|K_*\|^2(\beta+2)}{2\beta+1}, \qquad B_* > \frac{4S\|K\|_2^2(4\beta+3d+4)}{(4\beta+d)},$$
$$r^* < \min\left\{1, \frac{4\beta+3d+4}{2(4\beta+d)} - \frac{B_*}{8S\|K\|_2^2}, \frac{\beta+2}{2\beta+1} - \frac{A_*}{S\|K_*\|_2^2}\right\}.$$

Théorème 2.3. *Soient $q \geqslant 2$, $d > 2$, $\beta > \frac{d}{4}$ et $\alpha_n = O_n(n^{-r^*})$. Alors ρ_n^* est une normalisation aléatoire α_n-optimale et f_n^* est un estimateur α_n-adaptatif par rapport à Σ_0. En particulier,*

$$\limsup_{n\to\infty} R_n^{(r)}(f_n^*, \Sigma, \rho_n^*) \leqslant M_*, \tag{7}$$

où $M_* = 2^q L_0^{\frac{dq}{2\beta+d}} S^{\frac{\beta q}{2\beta+d}} \|K\|_2^{\frac{2\beta q}{2\beta+d}} + (\lambda(1+\sqrt{\frac{q}{r^*}})^{1/2} + \frac{1}{\sqrt{2}} + 2^{\frac{6\beta+d}{4\beta+d}} L_0^{\frac{d}{8\beta+d}} \Upsilon^{\frac{\beta}{4\beta+d}})^q$.

Remarque 1. Les constantes A_*, B_* et r^* proviennent du problème de test résolu dans Yodé [5].
La condition $\beta > \frac{d}{4}$ implique que $nh_n^d \to +\infty$, $n \to +\infty$ qui est une condition classique pour l'estimateur à noyau $\hat{f}_n$. La contrainte $d > 2$ apparaît dans la démonstration du Théorème 3.1 (Yodé [5]) et est liée au choix de la statistique (6) qui est construite selon le concept de risques minimax avec normalisations aléatoires (Lepski [4]). Nous croyons qu'un autre choix de test par exemple le test du χ^2 (Ingster [2]) ou en utilisant des techniques de calculs plus fines, on pourrait inclure le cas $d = 2$ dans nos résultats de test. Selon notre construction, la caractéristique de ρ_n^*, d'après le Théorème 3.1 de Yodé [5], est $x_n(\rho_n^*) = \varphi_n(\alpha_n) \gg \varphi_n(\Sigma_0)$.

Théorème 2.4. *Sous les conditions du Théorème* 2.3*, si $\alpha_n = O_n(\varphi_n^q(\Sigma_0))$ alors f_n^* est un estimateur adaptatif.*

Remarque 2. D'après (7), nous obtenons une interprétation statistique du type (2). Ainsi, si l'hypothèse H_0 est vraie, nous obtenons une région de confiance plus fine que celle fournie par l'estimation minimax avec la probabilité plus grande que $1 - \alpha_n$. Par définition, $\rho_n^* \leqslant \varphi_n(\Sigma)$. Alors f_n^* est asymptotiquement optimal. En d'autres termes, nous conservons les acquis de l'estimation minimax. De plus, f_n^* peut être adaptatif.

Remerciements

Je remercie le Professeur Oleg V. Lepski pour son aide et ses encouragements.

Références

[1] M. Hoffmann, On estimating the diffusion coefficient : parametric versus nonparametric, Ann. Inst. H. Poincaré 37 (2001) 339–372.
[2] Yu.I. Ingster, Asymptotically minimax testing of the hypothesis of independence, Zap. Nauchn. Seminar. LOMI 153 (1986) 60–72. Translation in J. Soviet. Math. 44 (1989) 466–476.
[3] M. Hoffmann, O.V. Lepski, Random rates in anisotropic regression, Ann. Statist. 30 (2002) 325–396.
[4] O.V. Lepski, How to improve the accuracy of estimation, Math. Methods Statist. 8 (1999) 441–486.
[5] A.F. Yodé, Test d'indépendance non-paramétrique, C. R. Math. Acad. Sci. Paris 336 (11) (2003) 955–958.