Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Acad. Sci. Paris, Ser. I 340 (2005) 59–62
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Inégalités d'oracle pour l'estimation d'une densité de probabilité

Philippe Rigollet

Laboratoire de probabilités et modèles aléatoires, UMR CNRS 7599, université Paris 6, 4, place Jussieu, case 188, 75252 Paris cedex 05, France

Reçu le 24 mai 2004 ; accepté après révision le 3 novembre 2004

Disponible sur Internet le 16 décembre 2004

Présenté par Paul Deheuvels

Résumé

Nous étudions le problème de l'estimation d'une densité de probabilité dans $L_2(\mathbb{R})$. A partir d'une formulation du risque quadratique intégré dans le domaine des fréquences de Fourier, nous montrons qu'il est proche du risque ℓ_2 dans le modèle de suite gaussienne. En appliquant alors une version modifiée de la méthode Stein par blocs, nous obtenons une inégalité d'oracle sur les estimateurs linéaires monotones et une inégalité d'oracle sur les estimateurs à noyau. Comme conséquence, l'estimateur proposé est adaptatif au sens minimax exact (i.e. à la constante près) sur la famille de classes de Sobolev. ***Pour citer cet article : Ph. Rigollet, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

Abstract

Oracle inequalities for probability density estimations. We study the problem of the nonparametric estimation of a probability density in $L_2(\mathbb{R})$. Expressing the mean integrated squared error in the Fourier domain, we show that it is close to the mean squared error in the Gaussian sequence model. Then, applying a modified version of Stein's blockwise method, we obtain a linear monotone oracle inequality and a kernel oracle inequality. As a consequence, the proposed estimator is sharp minimax adaptive (i.e. up to a constant) on a scale of Sobolev classes of densities. ***To cite this article: Ph. Rigollet, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

1. Introduction

Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de densité de probabilité inconnue $p \in L_2(\mathbb{R})$. Nous étudions l'estimation de p à partir de l'échantillon $\mathbb{X}^n = (X_1, \ldots, X_n)$. Soit $\hat{p}_n$ un estimateur de p, nous mesurons sa performance par son *risque quadratique intégré* défini par $\mathbb{E}_p\|\hat{p}_n - p\|^2$, où $\mathbb{E}_p$ désigne l'espérance par rapport à la loi de l'échantillon $\mathbb{X}^n$ et $\|\cdot\|$ désigne la norme usuelle sur $L_2(\mathbb{R})$. Pour toute fonction h de $L_2(\mathbb{R})$, notons $\mathcal{F}[h]$ sa *transformée de Fourier*. Soit $\varphi(\omega) \triangleq \mathcal{F}[p](\omega)$ la *fonction caractéristique* associée à la variable X_1 et

Adresse e-mail : rigollet@ccr.jussieu.fr (Ph. Rigollet).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2004.11.009

$\varphi_n(\omega) \triangleq \frac{1}{n}\sum_{k=1}^n \mathrm{e}^{\mathrm{i}\omega X_k}$ la *fonction caractéristique empirique* associée à l'échantillon $\mathbb{X}^n$. Pour une *fonction de poids* $\lambda \in L_2(\mathbb{R}) \cap L_1(\mathbb{R})$, définissons l'*estimateur linéaire* $\hat{p}_\lambda$ de p par la formule d'inversion de Fourier :

$$\hat{p}_\lambda(x) \triangleq \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}\omega x}\hat{\varphi}_\lambda(\omega)\,\mathrm{d}\omega, \quad \text{où } \hat{\varphi}_\lambda(\omega) \triangleq \varphi_n(\omega)\lambda(\omega). \tag{1}$$

On remarque alors que les estimateurs à noyau définis par

$$\hat{p}_{n,h}(x) \triangleq \frac{1}{n}\sum_{i=1}^n K_h(X_i - x), \quad \text{où } K_h(x) \triangleq \frac{1}{h}K\left(\frac{x}{h}\right), \tag{2}$$

sont des estimateurs linéaires dont la fonction de poids est $\mathcal{F}[K_h]$, pourvu que $\mathcal{F}[K] \in L_2(\mathbb{R}) \cap L_1(\mathbb{R})$. Par la formule de Plancherel, on a

$$2\pi\mathbb{E}_p\|\hat{p}_\lambda - p\|^2 = \Delta_n\big(\lambda, |\varphi|^2\big) - \frac{1}{n}\int_{\mathbb{R}} \big|\varphi(\omega)\big|^2\big|\lambda(\omega)\big|^2\,\mathrm{d}\omega, \tag{3}$$

où

$$\Delta_n\big(\lambda, |\varphi|^2\big) \triangleq \int_{\mathbb{R}} \left(\big|1-\lambda(\omega)\big|^2\big|\varphi(\omega)\big|^2 + \frac{1}{n}\big|\lambda(\omega)\big|^2\right)\mathrm{d}\omega. \tag{4}$$

Comme dans [4,1], on remarque alors que $\Delta_n(\lambda, |\varphi|^2)$ que nous appellerons Δ-risque est similaire à l'erreur ℓ_2 d'un estimateur linéaire dans le modèle de suite gaussienne (voir [8], Ch. 3) en transposant la formule dans le cas continu et en remplaçant la variance du bruit par $\frac{1}{n}$.

2. Application de la méthode de Stein par blocs à l'estimation de densité

Pour construire l'estimateur de Stein par blocs, nous suivons la méthode de modèle de suite gaussienne de [8], Ch. 3 et [2]. La fonction de poids est obtenue par la minimisation d'un estimateur asymptotiquement sans biais du Δ-risque sur une certaine classe de fonctions de poids contenue dans la classe $\mathcal{H}_0$ des fonctions de poids de $L_1(\mathbb{R})$ à valeurs dans $[0,1]$ (donc $\mathcal{H}_0 \subset L_2(\mathbb{R})$). En effet, le risque quadratique intégré correspondant à une fonction de poids de $L_2(\mathbb{R})$ qui n'est pas dans $\mathcal{H}_0$ peut être amélioré uniformément sur l'ensemble des densités de $L_2(\mathbb{R})$ par projection sur $[0,1]$ (cf. [3]). En remplaçant $|\varphi(\omega)|^2$ par un estimateur asymptotiquement sans biais, $[|\varphi_n(\omega)|^2 - \frac{1}{n}]$, dans (4), on se ramène à la minimisation en λ de

$$l_n(\lambda) \triangleq \int_{\mathbb{R}} \left(\big[\big|\lambda(\omega)\big|^2 - 2\lambda(\omega)\big]\big|\varphi_n(\omega)\big|^2 + \frac{2}{n}\lambda(\omega)\right)\mathrm{d}\omega.$$

Soit alors la classe $\mathcal{H}^*$ des fonctions de poids constantes par « blocs » :

$$\mathcal{H}^* \triangleq \left\{\lambda \in \mathcal{H}_0 : \lambda(\omega) = \sum_{j=0}^J t_j \mathbb{1}_{\{\omega \in B_j\}},\ 0 \leqslant t_j \leqslant 1,\ j = 0,\dots,J\right\},$$

où $\{B_j\}_{j=0}^J$ est une partition de $[-\Omega_n, \Omega_n]$, $\Omega_n > 0$, telle que $B_0 = (-b_0, b_0)$, $b_0 > 0$ et $\forall 1 \leqslant j \leqslant J$, $B_j = -B'_j \cup B'_j$, $B'_j \triangleq [b_{j-1}, b_j)$, $-B'_j \triangleq (-b_j, -b_{j-1}]$, $0 < b_{j-1} < b_j$. La minimisation de l_n sur la classe $\mathcal{H}^*$ conduit à définir l'*estimateur de Stein par blocs sur le système* $\{B_j\}_{j=0}^J$, $\hat{p}_{\tilde{\lambda}}$, par la fonction de poids dépendant des données :

$$\tilde{\lambda}(\omega) \triangleq \sum_{j=0}^J \left(1 - \frac{|B_j|}{n\int_{B_j}|\varphi_n(\omega)|^2\,\mathrm{d}\omega}\right)_+ \mathbb{1}_{\{\omega \in B_j\}} \quad \text{où } |B_j| = \int_{B_j}\mathrm{d}\omega.$$

Théorème 2.1. *Supposons que $\|p\|^2 \leqslant G < \infty$ et que chacun des B_j vérifie $1 \leqslant |B_j| \leqslant 4n$. Alors, pour toute suite $(\mu_n)_n$ vérifiant $\mu_n \to \infty$, $n \to \infty$, il existe des constantes $C > 0$ et $D = D(G) > 0$ indépendantes de toute autre variable, telles que pour n vérifiant $C < \mu_n$, l'estimateur de Stein par blocs sur le système $\{B_j\}_{j=0}^J$ vérifie l'inégalité d'oracle*

$$\mathbb{E}_p\|\hat{p}_{\tilde{\lambda}} - p\|^2 \leqslant \frac{1}{1 - C\mu_n^{-1}}\Big(\min_{\lambda \in \mathcal{H}^*} \mathbb{E}_p\|\hat{p}_\lambda - p\|^2 + JD\frac{(\log n)^4\mu_n}{n}\Big). \tag{5}$$

La démonstration de ce résultat fait appel à des lemmes de [4,5]. Moyennant une spécification correcte des B_j, il est possible d'obtenir une inégalité d'oracle sur la classe des fonctions de poids monotones :

$$\mathcal{H}_{\text{mon}}^\infty \triangleq \big\{\lambda \in \mathcal{H}_0\colon\ \lambda\big(|\omega_1|\big) \leqslant \lambda\big(|\omega_2|\big),\ |\omega_2| \leqslant |\omega_1|\big\}.$$

Soit $\Omega_n = n^\alpha(\log n)^2$, $\alpha \geqslant 1/2$ et soit $(\nu_n)_n$ une suite déterministe tendant vers l'infini. Le système de *blocs faiblement géométriques* (*BFG*) est alors défini par :

$$|B_0| = \nu_n, \quad |B_j| = (1 + \nu_n^{-1})^j\nu_n, \quad j = 1, 2, \ldots, J-1, \quad |B_J| = \Omega_n - \sum_{j=0}^{J-1}|B_j|,$$

où $J \triangleq \min\{m\colon \sum_{j=0}^m (1 + \nu_n^{-1})^j\nu_n \geqslant \Omega_n\}$. On montre que le système BFG vérifie $J \leqslant C'(\log n)\nu_n$, $\forall n \geqslant n_0$, pour un entier n_0 et une constante $C'(n_0, \alpha) > 0$. L'estimateur de Stein sur le système BFG est appelé *estimateur de* Stein-*BFG*.

Théorème 2.2. *Soit $\Omega_n = n(\log n)^2$. Supposons que $\|p\|^2 \leqslant G < \infty$ et que chacun des B_j vérifie $1 \leqslant |B_j| \leqslant 4n$. Alors, pour toute suite $(\mu_n)_n$ vérifiant $\mu_n \to \infty$, $n \to \infty$, il existe des constantes $C > 0$ et $D = D(G) > 0$ indépendantes de toute autre variable, telles que pour $n \geqslant n_1$ où n_1 ne dépend que de $(\mu_n)_n$, l'estimateur de Stein par blocs sur le système $\{B_j\}_{j=0}^J$ vérifie l'inégalité d'oracle*

$$\mathbb{E}_p\|\hat{p}_{\tilde{\lambda}} - p\|^2 \leqslant \frac{1}{1 - C\mu_n^{-1}}\left((1 + \nu_n^{-1})\big(1 + (\log n)^{-1}\big)\min_{\lambda \in \mathcal{H}_{\text{mon}}^\infty} \mathbb{E}_p\|\hat{p}_\lambda - p\|^2 + \frac{\tau_n(D)}{2\pi}\right), \tag{6}$$

où $\tau_n(D) \triangleq \frac{D}{n}\nu_n\mu_n(\log n)^5$.

Soit la famille de classes de densités de probabilités de Sobolev $\{\Theta(\beta, Q),\ \beta > 1/2,\ Q > 0\}$ définies par

$$\Theta(\beta, Q) \triangleq \left\{p : \mathbb{R} \to \mathbb{R}_+,\ \int_{\mathbb{R}} p(x)\,\mathrm{d}x = 1,\ \int_{\mathbb{R}} \big|\varphi(\omega)\big|^2|\omega|^{2\beta}\,\mathrm{d}\omega \leqslant Q\right\}.$$

En utilisant la minoration du risque de Golubev [4] et en prenant le supremum sur $p \in \Theta(\beta, Q)$ dans les deux membres de (6), avec par exemple $\mu_n = \nu_n = \log(\log n)$ pour la majoration du risque, on obtient le corollaire suivant.

Corollaire 2.3. *L'estimateur de Stein-BFG $\hat{p}_{\tilde{\lambda}}$ est adaptatif au sens minimax exact sur la famille de classes de Sobolev $\{\Theta(\beta, Q),\ \beta > 1/2,\ Q > 0\}$, i.e. pour tous $\beta > 1/2$, $Q > 0$,*

$$\sup_{p \in \Theta(\beta, Q)} \mathbb{E}_p\|\hat{p}_{\tilde{\lambda}} - p\|^2 = \big(1 + \mathrm{o}(1)\big)\inf_{T_n}\sup_{p \in \Theta(\beta, Q)} \mathbb{E}_p\|T_n - p\|^2, \quad n \to \infty,$$

où l'infimum est pris sur tous les estimateurs de p.

3. Une inégalité d'oracle pour les estimateurs à noyau

Définissons $\mathcal{K}_0$ comme la classe des noyaux $K \in L_2(\mathbb{R})$ qui admettent une version de la transformée de Fourier symétrique autour de 0, décroissante sur $\mathbb{R}_+$ et à valeurs dans $[0,1]$ de sorte que $\mathcal{F}[K] \in \mathcal{H}_{\mathrm{mon}}^{\infty}$. D'après le Théorème 2.2 on a directement le théorème suivant :

Théorème 3.1. *Supposons que $\|p\|^2 \leqslant G < \infty$. Pour toute suite $(\mu_n)_n$ vérifiant $\mu_n \to \infty,\ n \to \infty$, il existe des constantes $C > 0$ et $D' = D'(G) > 0$ indépendantes de toute autre variable, telles que pour n vérifiant $C < \mu_n$, l'estimateur de Stein-BFG $\hat{p}_{\tilde{\lambda}}$ vérifie l'inégalité d'oracle*

$$\mathbb{E}_p\|\hat{p}_{\tilde{\lambda}} - p\|^2 \leqslant \frac{1}{1 - C\mu_n^{-1}}\left((1+\nu_n^{-1})\big(1+(\log n)^{-1}\big)\inf_{K\in\mathcal{K}_0}\inf_{h>0}\mathbb{E}_p\|\hat{p}_{n,h} - p\|^2 + \frac{\tau_n(D')}{2\pi}\right). \tag{7}$$

Le lemme suivant donne une minoration pour le terme principal dans le membre de droite en (7).

Lemme 3.2. *Soit $p \in L_2(\mathbb{R})$ une densité de probabilité et $K \in L_1(\mathbb{R})$ un noyau symétrique tel que $\int K(x)\,\mathrm{d}x = 1$ et vérifiant l'une des deux conditions suivantes*

(i) *K est positif,*
(ii) *Il existe $s \geqslant 1$ entier et $\delta > 0$ tels que $\int |t|^{2s+\delta}|K(t)|\,\mathrm{d}t < \infty$, $\alpha_k \triangleq \int t^k K(t)\,\mathrm{d}t = 0,\ k = 1, 2, \ldots, 2s-1$ et $(-1)^{s+1}\alpha_{2s} > 0$.*

Alors, il existe des constantes $c > 0$ (dépendant de p) et $0 < a < 1$, toutes deux dépendant de K, telles que

$$\inf_{h>0}\mathbb{E}_p\|\hat{p}_{n,h} - p\|^2 \geqslant cn^{-a}. \tag{8}$$

Corollaire 3.3. *Soit $K \in \mathcal{K}_0$ un noyau vérifiant les hypothèses du Lemme* 3.2. *Alors, pour toute densité $p \in L_2(\mathbb{R})$* fixée *l'estimateur de Stein-BFG $\hat{p}_{\tilde{\lambda}}$ vérifie l'inégalité d'oracle suivante* :

$$\mathbb{E}_p\|\hat{p}_{\tilde{\lambda}} - p\|^2 \leqslant \big(1 + \mathrm{o}(1)\big)\inf_{h>0}\mathbb{E}_p\|\hat{p}_{n,h} - p\|^2, \quad n \to \infty. \tag{9}$$

Ici $\mathrm{o}(1)$ *dépend de K et de p.*

Le Corollaire 3.3 est comparable au résultats de Stone [7] et Wegkamp [9]. A la différence de ceux-ci, il est valable pour des noyaux dont le support n'est pas compact (e.g. le noyau gaussien et ceux de Fejer et de Silverman) et reste valable pour une densité p non bornée. Les démonstrations des résultats de cette Note figurent dans [6].

Références

[1] L.L. Boiko, G.K. Golubev, How to improve the nonparametric density estimation in S-PLUS, Problems of Information Transmission 36 (2000) 354–361.
[2] L. Cavalier, A.B. Tsybakov, Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation, Math. Methods Statist. 10 (2001) 247–282.
[3] D.B.H. Cline, Admissible kernel estimators of a multivariate density, Ann. Statist. 16 (1988) 1421–1427.
[4] G.K. Golubev, Nonparametric estimation of smooth probability densties in L_2, Problems of Information Transmission 28 (1992) 44–54.
[5] G.K. Golubev, B.Y. Levit, Distribution function estimation: adaptive smoothing, Math. Methods Statist. 5 (1996) 383–403.
[6] P. Rigollet, Adaptive density estimation using Stein's blockwise method, Prépublication LPMA 913, 2004, http://www.proba.jussieu.fr/mathdoc/textes/PMA-913.pdf
[7] C.J. Stone, An asymptotically optimal window selection rule for kernel density estimates, Ann. Statist. 12 (1984) 1285–1297.
[8] A.B. Tsybakov, Introduction à l'estimation non paramétrique, Springer, 2004.
[9] M.H. Wegkamp, Quasi-universal bandwidth selection for kernel density estimators, Canad. J. Statist. 27 (1999) 409–420.