ANNALES DE L'I. H. P., SECTION B

JEAN BRETAGNOLLE
JAN MIELNICZUK

On asymptotic minimaxity of the adaptative kernel estimate of a density function

Annales de l'I. H. P., section B, tome 25, nº 2 (1989), p. 143-152
<http://www.numdam.org/item?id=AIHPB_1989__25_2_143_0>

Ann. Inst. Henri Poincaré,

Vol. 25, n° 2, 1989, p. 143-152. *Probabilités et Statistiques*

On asymptotic minimaxity of the adaptative kernel estimate of a density function

by

Jean BRETAGNOLLE and Jan MIELNICZUK (*)

Université de Paris-Sud U.A. 743, C.N.R.S.,
Statistique Appliquée, Mathématiques, Bât. n° 425,
91405 Orsay Cedex, France

ABSTRACT. — The adaptative kernel estimate of a density function with a data dependent bandwidth arising naturally as an empirical counterpart of MSE-optimal parameter is dealt with. It is shown to be asymptotically minimax in a family of densities having bounded second derivative in a neighbourhood of 0. As a by-product a simple proof of Farell's [4] result on minimax bounds for density estimates is obtained.

Key words : Data dependent bandwidth, kernel estimator, mean square error, minimax estimate.

RÉSUMÉ. — On considère l'estimateur à noyau (adaptatif) d'une densité construit pour minimiser asymptotiquement le risque quadratique en un point. On prouve qu'il est minimax dans la famille des densités ayant une dérivée seconde bornée au voisinage du point. Du même coup on trouve une preuve simple des résultats de Farrell [4] basée sur l'inégalité de Cramer-Rao.

Mots clés : Estimateur à noyau, fenêtre adaptative, risque quadratique, estimateur minimax.

Classification A.M.S. : 62 G 05, 62 G 20.

(*) On leave from Institute of Computer Science, Polish Academy of Sciences Warsaw, Poland. Supported by the grant of the French government, No. 93572.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be independent real random variables with a common distribution function F and a density f with respect to Lebesgue measure on $\mathbb{R}$. Consider the problem of nonparametric estimation of $f(0)$ by means of kernel estimate $\hat{f}_h(0)$ (Parzen [7]):

$$\hat{f}_h(x) = (nh)^{-1} \sum_{i=1}^{n} K((x - X_i)/h) \tag{1.1}$$

where kernel K integrates to 1 and $h = h(n)$ is a smoothing parameter (bandwidth). The natural bandwidth's choice is provided by computation of the mean square error $R(f, \hat{f}_h) = E_f(f(0) - \hat{f}_h(0))^2$; under standard assumptions on K, f, f'':

$$R(f, \hat{f}_h) = \frac{f(0)}{nh} \int K^2(y)\, dy + [f''(0)]^2 h^4 \left[\int K(y)\, y^2\, dy \right]^2 / 4 + o(1/nh + h^4) \tag{1.2}$$

If $f(0) f''(0) \neq 0$ minimizing main terms of (1.2) yields h_0 satisfying

$$nh_0^5 = \frac{f(0)}{[f''(0)]^2} \int K^2(y)\, dy \Big/ \left[\int K(y)\, y^2\, dy \right]^2$$

$$n^{4/5} R(f, \hat{f}_{h_0}) = \frac{5}{4} \left\{ [f(0)]^4 [f''(0)]^2 \times \left[\int K^2(y)\, dy \right]^4 \left[\int K(y)\, y^2\, dy \right]^2 \right\}^{1/5} + o(1) \tag{1.3}$$

Replacing $f(0)$ and $f''(0)$ by some preliminary estimates we get a natural empirical counterpart $\hat{h}_{0,n}$ of h_0 with the following property of the resulting adaptive estimate (Woodroofe [11])

$$R(f, \hat{f}_{\hat{h}_{0,n}}) / R(f, \hat{f}_{h_0}) \to 1 \quad \text{as } n \to \infty \tag{1.4}$$

(Recall that h_0 and $\hat{h}_{h_0}$ depend on n.)

The problem of finding suitable modification of (1.1) which for fixed density f will yield smaller mean square error attracted attention of many authors. The proposed methods consist in bias reduction by considering of so-called Parzen kernels (Parzen [7]), geometric extrapolation (Tarell and Scott [10]) or choosing appropriately constructed bandwidths depending on data and point at which density is estimated (Abramson [1]). Such procedures usually yield an estimate which is not a bona fide density function although there exist methods to tackle with this problem (*cf.* Gajek [5]). All these results, however, leave unanswered the question about possible improvement or kernel estimate uniformly over reasonable family of densities. For example, Farrell's [4] result shows that it is not possible to get a better rate than $n^{-4/5}$ when considering the family $\mathscr{F}$ of densities having bounded second derivative in a neighbourhood of 0 but improvement of a constant is *a priori* possible. For the related problems in the area of global estimation we refer to Bretagnolle and Huber [3]. We show however that the answer to this question is negative, namely that the adaptive kernel estimator is asymptotically minimax in $\mathscr{F}$. The method is to construct a parametric family $\{f_\theta\}$ of densities intrinsically joined with the kernel estimate $\hat{f}_h$, for which $\{f_\theta\} \subset \mathscr{F}$ and

$$\sup_{f \in \{f_\theta\}} \mathrm{R}(f, f_n)/\mathrm{R}(f, \hat{f}_h)$$

can be minorized for an arbitrary estimate f_n of $f(0)$. As a by-product, Farrell's result on a minimax bound for density estimates is obtained.

Note. — Our referee pointed out to our attention two recent preprints from Brown and Farrell [12] and from Donoho and Liu [13]: Estimating $f(0)$ in various families (slightly different of ours) they compare minimax lower bounds (on all estimators) to minimax lower bounds on kernel estimators or linear procedures. Brown and Farrell give tables for *finite samples.*

2. THE CONSTRUCTION

From now on we will impose the following conditions on K and $h(n)$:

(a) kernel K is a function integrating to 1;

(b) $\{x : \mathrm{K}(x) \neq 0\} \subset [-\mathrm{A}, \mathrm{A}]$ for some $0 < \mathrm{A} < \infty$;

(c) $\alpha = \sup_x |\mathrm{K}(x)| = \mathrm{K}(0)$;

(d) $\beta = \int \mathrm{K}^2(y)\, dy \leqq \alpha$;

(e) $h(n) \to 0$ and $nh(n) \to \infty$.

Let us observe that if K is *nonnegative* function such that (a) and (c) hold, then (d) is automatically satisfied.

Let us denote f_0 an arbitrary smooth density equal to 1 on $[-1/4, 1/4]$ and put $\mathrm{K}_h(.) = h^{-1}\mathrm{K}(./h)$. From now on we will assume that $4\mathrm{A}h \leqq 1$ so that the support of K_h is contained in $[-1/4, 1/4]$. We define e_h by

$$1 + \lambda e_h(x) = \mathrm{K}_h(-x) \qquad \text{where} \quad \lambda^2 h = \beta - h \quad (x \in \mathbb{R}) \tag{2.1}$$

(we center and reduce K_h wrt f_0), thus

$$\mathrm{E}_f(e_h) = 0, \qquad \mathrm{E}_f(e_h^2) = 1 \tag{2.2}$$

Moreover, using (*c*) we get

$$\| e_h \|_\infty \leqq (\alpha + h)/\lambda h. \tag{2.3}$$

Let θ be some real number. If we assume $\rho_h(\theta) = |\theta| (\alpha + h)/\lambda h \leqq 1/2$, $f_{\theta, h} = f_0 (1 + \theta e_h)$ is a density. We shall use parametric families $\mathscr{F}_{\Theta, h} = \{ f_{\theta, h} \mid \theta \in \Theta \}$ where $\sup_\Theta \rho_h(\theta) \leqq 1/2$.

The density $f_{\theta, h}$ can be represented in the equivalent way

$$f_{\theta, h}(x) = f_0(x)(1 + \theta (h/(\beta - h))^{1/2} (K_h(-x) - 1)).$$

Let $f_n = f_n(X_1, \ldots, X_n)$ denote arbitrary estimate of $f(0)$. The results are based on the following

LEMMA. — *Let r_0 be $(\alpha - \beta)/\beta$. Taking at stage n $\Theta_n = [-\Theta_{\max}, \Theta_{\max}]$ where $n \Theta_{\max}^2 = T > 1$, we have*

$$\varliminf_n \inf_{f_n} \sup_{\theta \in \Theta_n} \frac{R(f_\theta, f_n)}{R(f_\theta, \hat{f}_h)} \geqq \begin{cases} 1 - (r_0^2 T)^{-1} & \text{for } r_0 > 0 \\ (1 - T^{-1/2})^2 & \text{for } r_0 = 0 \end{cases} \tag{2.4}$$

Proof. — First we observe that, since $nh(n) \to \infty$ by condition (*e*), the condition on $\rho_h(\theta)$ is asymptotically fulfilled as $\lim_n \sup_{\theta \in \Theta_n} \rho_h(\theta) = 0$.

Further

$$E_\theta \hat{f}_h(0) = \int (1 + \lambda e_h(x)) f_0(x)(1 + \theta e_h(x))\, dx = 1 + \lambda\theta,$$

thus the bias $E_\theta \hat{f}_h(0) - f_\theta(0) = \theta(\beta - \alpha)/\lambda h$.

Moreover

$$n \operatorname{Var}_\theta \hat{f}_h(0) = E_\theta (1 + \lambda e_h - (1 + \lambda\theta))^2$$
$$= \lambda^2 \int f_0(x)(1 + \theta e_h(x))(e_h^2(x) + \theta^2 - 2\theta e_h(x))\, dx$$

and straightforward computations using (2.1), (2.2) and (2.3) give:

$$n \operatorname{Var}_\theta \hat{f}_h(0) = \lambda^2 (1 + O(\rho + \theta^2)). \tag{2.5}$$

Thus using variance-squared bias decomposition of $R(f_\theta, \hat{f}_h)$ we have

$$R(f_\theta, \hat{f}_h) \leqq (\theta(\beta - \alpha)/\lambda h)^2 + \lambda^2 n^{-1}(1 + O(\rho)).$$

On the other hand, putting $\varphi(\theta) = E_\theta(f_n(0))$ and using Cramer-Rao's bound,

$$R(f_\theta, f_n) \geqq (\varphi(\theta) - (1 + \theta(\alpha - h)/\lambda h))^2 + (\varphi'(\theta))^2 / n I(\theta),$$

where

$$I(\theta) = \int f_\theta(x) \left[\frac{\partial}{\partial\theta} (\log f_\theta(x)) \right]^2 dx$$

$$= \int f_0(x)\, e_h^2(x)\, (1+\theta\, e_h(x))^{-1}\, dx = 1 + O(\rho)$$

(Cramer-Rao's conditions are fullfilled since all expressions are polynomial *wrt* θ). Denoting by $R(\hat{f}_n/f_h)$ the ratio of the respective mean square errors we have

$$(1+O(\rho+\theta^2))\, R(f_n/\hat{f}_h) \geqq \frac{n(\varphi(\theta)-(1+\theta(\alpha-h)/\lambda h))^2 + (\varphi'(\theta))^2}{n\theta^2(\lambda-((\alpha-h)/\lambda h))^2+\lambda^2}.$$

Setting $\varphi(\theta) = 1+\lambda\theta+\lambda Y(\theta)$, $b=\beta-h$, $a=\alpha-h$, we get, using (2.1):

$$(1+O(\rho+\theta^2))\, R(f_n/\hat{f}_h)$$
$$\geqq \{n(\theta(b-a)+bY)^2 + b^2(1+Y')^2\}/\{n\theta^2(b-a)^2+b^2\} \quad (2.6)$$

Substituting $x=\theta n^{1/2}$ and

$$y(x) = -n^{1/2}\, Y(\theta) \qquad \text{and} \qquad r=(a-b)/b$$

we obtain

$$(1+O(\rho+\theta^2))\, R(f_n/\hat{f}_h) \geqq \frac{\{(y+rx)^2+(1-y')^2\}}{\{r^2x^2+1\}} \quad (2.7)$$

Since $\limsup_{n} \sup_{\theta} (\rho(\theta)+\theta^2) = 0$ then denoting by τ_y the supremum of the above expression on $[-T^{1/2}, T^{1/2}]$ it is enough to show that τ_y is bounded from below by the expressions given in (2.4). Let us proceed by contradiction. Assume that $\tau_y \leqq 1$ and consider R defined by:

$$R(t) = \{(ty+rx)^2+(1-ty')^2\}/\{r^2x^2+1\}.$$

Since $R(t)$ is convex, with $R(1) \leqq \tau_y \leqq 1 \leqq R(0)$, $R'(0) \leqq 0$, thus $rxy-y' \leqq 0$. Whence putting $y(x) = z(x)\exp(rx^2/2)$ we obtain that z is nondecreasing. Observe futher that $\bar{y}(x) = -y(-x)$ provides the same bound τ. Thus, by convexity, τ is greater that the bound obtained for $(y+\bar{y})/2$. In other words, we can assume y an odd function, whet, together with $z(0)=0$ implies $y(x) \geqq 0$ for $0 \leqq x \leqq T^{1/2}$.

Consider first the case $r_0 > 0$. Note that $r > r_0$. Since y is positive on $[0, T^{1/2}]$ it is sufficient to verify that inequality holds at $x=T^{1/2}$ where (using $y \geqq 0$) it is implied by $r_0^2 T \geqq (1+r_0^2 T)(1-(r_0^2 T)^{-1})$.

For $r=r_0=0$ we argue as follows: if $y(T^{1/2}) \leqq 1-T^{-1/2}$ then in view of $y(0)=0$ there exists $s \in [0, T^{1/2}]$ such that $y'(s) \leqq T^{-1/2}-1/T$ and $1-y'(s) \geqq 1-T^{-1/2}$, □

3. RESULTS

Let us define by $\Phi(\gamma, L)\,(\gamma = k+\alpha,\ 0<\alpha\leqq 1)$ the class of functions possessing k-th order derivative ($k\geqq 1$) such that for $-1/4\leqq x,\ y\leqq 1/4$

$$|f^{(k)}(x)-f^{(k)}(y)|\leqq L\,|x-y|^{\alpha}. \qquad (3.1)$$

We will show that the lemma yields in straightforward manner the minimax bounds on density estimates (Farrell [4]), see also theorem 5.1 in Ibragimov, Hasminskii [6]).

COROLLARY 1.

$$\varliminf_{n} \operatorname{Inf}_{f_n} \operatorname{Sup}_{f\in\Phi(\gamma, L)} n^{2\gamma/(2\gamma+1)} R(f_n, f) > 0.$$

Proof. — In order to prove the corollary 1 let us consider an arbitrary, nonuniform kernel satisfying conditions $(a)-(d)$ and such that Hölder condition (3.1) is satisfied by $K^{(k)}$ with constant $C(K)$ and the power α. It is easy to see that in this case f_θ defined by (2.3) satisfy

$$|f_\theta^{(k)}(x)-f_\theta^{(k)}(y)|\leqq \theta\, C(K)\,|x-y|^{\alpha}\, h^{-k-1/2-\alpha}(\beta-h)^{-1/2}.$$

Thus if $\varepsilon>h$ and

$$L^2=\theta^2 C^2(K)/h^{2\gamma+1}(\beta-h) \qquad (3.2)$$

then $f_\theta\in\Phi(\beta, L)$. Under assumption $nh^{2\gamma+1}=C$ the condition (3.2) is equivalent to the following condition on the range of θ: $\theta\in[-\Theta_{\max}, \Theta_{\max}]$ where $n\,\Theta_{\max}=L^2 C(\beta-\varepsilon)/C^2(K)$. Observe also that the proof of the lemma yields for sufficiently large n

$$\inf_\theta R(f_\theta, \hat{f}_h)\geqq(\beta-\varepsilon)\, n^{-2\gamma/(2\gamma+1)}\, C^{-1/(2\gamma+1)}$$

Thus using (2.4) we obtain ($r_0>0$)

$$(1+o(1))\, R(f_\theta, f_n)\, n^{2\gamma/(2\gamma+1)}\, C^{1/(2\gamma+1)}(\beta-\varepsilon)^{-1}\geqq 1-C^2(K)/(r^2 L^2 C(\beta-\varepsilon)).$$

Whence we obtain that $\varliminf_{n} \sup_{\theta} R(f_\theta, f_n)\, n^{2\gamma/(2\gamma+1)}$ is bounded from below by

$$(1-a/C)\, C^{-1/(2\gamma+1)}(\beta-\varepsilon) \qquad \text{where} \quad a=C^2(K)/r\,L^2(\beta-\varepsilon).$$

It is easy to check that the maximum of the above expression is attained at $C=a(2\gamma+1)$ and is equal to

$$L(\varepsilon)=\left(\frac{2\gamma+1}{2\gamma+2}\right)\left(\frac{r\,L^2(\beta-\varepsilon)}{(2\gamma+2)\,C^2(K)}\right)^{1/(2\gamma+1)}(\beta-\varepsilon).$$

Since $\varepsilon>0$ can be chosen arbitrarily this gives the bound equal to $L(0)$, for $\gamma=2$ $L(0)=(5/6)\,(r\,L^2\,\beta^6/6\,C^2(K))^{1/5}$. □

Let $\|.\|$ denote supremum norm on $[-1/4, 1/4]$. We will prove.

COROLLARY 2. — *Let $\mathcal{F}$ be a family of twice differentiable densities such that $\|f''\| < \infty$. Let* K *be nonuniform kernel satisfying conditions (a)—(d) such that* K″ *exists and is bounded. Let $\hat{f}(n, \mathrm{K}, \mathrm{C})$ be the kernel estimate of $f(0)$ with $nh^5 = \mathrm{C}$.*

There exists a constant D(K) *depending only on* K *such that*

$$\varliminf_{n} \inf_{f_n} \sup_{\mathcal{F}} \mathrm{R}(f, f_n)/\mathrm{R}(f, \hat{f}(\mathrm{K}, n, \mathrm{C})) \geqq 1 - \mathrm{D}(\mathrm{K})/\mathrm{C}.$$

Proof. — Put $\delta = \sup |\mathrm{K}''|$ and let range of θ be defined by $n\Theta_{\mathrm{Max}}^2 = \mathrm{C}\beta\delta^{-2}$. Since $\theta^2/h = o(1)$ for such choice of θ and h, we have $\|f_\theta\| \leqq 2$ for sufficiently large n. Moreover, it is easy to see that

$$\|f''^2\| = n\theta^2\delta^2/(\beta - h)\, nh^5 = n\theta^2\delta^2/\beta\mathrm{C} + o(1).$$

Thus $\|f''^2\| \leqq 2$ for sufficiently large n. The relation (2.4) yields the conclusion of the corollary with bound greater than $1 - \delta^2/(r^2\beta)\,\mathrm{C}$. □

The Corollary 2 should be compared with the following result. Let $\alpha \in [\alpha_0, \alpha_1]$ where $0 < \alpha_0 < \alpha_1 < \infty$. S_0 will stand for an interval $[-s_0, s_0]$ and $\mathcal{F}_\alpha$ consists of densities of the following form:

(*a*) $f(x) = \alpha + \beta x + r(x)$, $x \in \mathbb{R}$;

(*b*) $|r(x)| \leqq mx^2$, $x \in \mathrm{S}_0$.

It is easy to see that $\mathcal{F}_\alpha$ is nonempty for every choice of $r(x)$ satisfying (*b*) if

$$\int_{\mathrm{S}_0} (\alpha_1 + mx^2)\, dx \leqq 1$$

and

$$|\beta| \leqq \beta_1 \quad \text{and} \quad \alpha_0 - \beta_1 s_0 - ms_0^2 \geqq 0.$$

Put $\mathcal{F} = \bigcup \mathcal{F}_\alpha$, summation being over $\alpha \in [\alpha_0, \alpha_1]$. It is known (*cf.* Sacks and Ylvisacker [9] and Sacks and Strawderman [8]), respectively that the kernel estimate f_n^* based on Epanechnikov kernel and $h(n) = \mathrm{A}(\alpha_1, m)\, n^{-1/5}$) is asymptotically minimax in $\mathcal{F}$ against the class of kernel estimates $\hat{f}_n$ with deterministic bandwidths, namely

$$(\inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathrm{R}(f, \hat{f}_n)) / \sup_{f \in \mathcal{F}} \mathrm{R}(f, f_n^*) \to 1$$

but it is not asymptotically minimax against all possible choices of f_n *i. e.* there exists an estimate $\tilde{f}_n$ such that

$$\inf_{n \geqq n_0} \inf_{f \in \mathcal{F}} \mathrm{R}(f, f_n^*)/\mathrm{R}(f, \tilde{f}_n) \geqq 1 + a$$

for some $a > 0$.

The link of this result with our Corollary 2 is the following. It is clear that under the imposed assumptions $\{f_\theta\} \subset \mathcal{F}$ for appropriate choice of

s_0, α_1 and m. Thus taking in Corollary 2 the Epanechnikov kernel for K and $h(n)=A(\alpha_1, m)\,n^{-1/5}$ we obtain that there exists θ_0 such that

$$R(f_{\theta_0}, \tilde{f}_n)/R(f_{\theta_0}, f_n^*) \geqq 1-c$$

for some small positive number c. At the same time

$$\inf_{\theta} R(f_\theta, f_n^*)/R(f_\theta, \tilde{f}_n) \geqq 1+a.$$

Thus Sacks and Strawderman's result shows that it is not possible to replace the lower bound in the Corollary 2 by 1. However we will show that the adaptive kernel estimate is asymptotically minimax among all estimates f_n.

Put $\gamma=\int K(y)\,y^2\,dy$ and consider the following adaptive kernel estimate (*cf.* Woodroofe [11]). Let H be symmetric two times differentiable kernel with a compact support and let $\tilde{f}_\delta(0)$ be a kernel estimate of $f(0)$ defined in (1.1) pertaining to the kernel H and bandwidth sequence $\delta(n)$. Denote by $\tilde{f}_\delta''(0)$ its second derivative

$$\tilde{f}_\delta''(0)=n^{-1}\delta^{-3}\sum_{i=1}^{n} H''(-X_i/\delta)$$

The adaptive bandwidth is defined by the following equation [*cf.* (1.3)]. If

$$n\tilde{h}^5=(\beta/\gamma^2)(\tilde{f}(0)/\tilde{f}''^2(0)),$$

then (3.3)

$$n\hat{h}^5=(c_n \vee n\tilde{h}^5) \wedge C_n$$

where $c_n \to 0$, $C_n \to \infty$.
We prove that $\hat{f}_h$ cannot be improved uniformly on $\mathscr{F}$.

COROLLARY 3. — *Assume that* K *is symmetric and* $|H''|$ *and* $|K''|$ *are bounded. Let* $\delta(n)=A\,n^{-1/5+\varepsilon}$ *for some positive* A *and* ε *and assume that* $c_n=n^{-a}$, $C_n=n^b$ *where* $a>0$, $0<11\,b/5<6\,\varepsilon$. *Let* $\hat{h}$ *be defined by* (3.3). *Then*

$$\varliminf_{n} \inf_{f_n} \sup_{f \in \mathscr{F}} R(f, f_n)/R(f, \hat{f}_{\hat{h}}) \geqq 1.$$

Proof. — Let $h_M(h_m)$ be defined by the equality $nh_M^5=C_n\,(nh_m^5=c_n)$. Consider the parametric family defined in Corollary 2 with C replaced by C_n. Put $\hat{f}_{h_M}=\hat{f}(n, K, C_n)$. By Corollary 2

$$\varliminf_{n} \inf_{f_n} \sup_{f \in \mathscr{F}} R(f, f_n)/R(f, \hat{f}_{h_M}) \geqq 1 \quad \text{as } C_n \to \infty.$$

To enlighten notations, we set $\varphi_n=\hat{f}_{\hat{h}}$.

Now it sufficies to show that

$$\lim_n \sup_\theta |\mathrm{R}(f_\theta, \varphi_n) - \mathrm{R}(f_\theta, \hat{f}_{h_\mathrm{M}})| / \mathrm{R}(f_\theta, \hat{f}_{h_\mathrm{M}}) = 0$$

which follows from

$$\lim_n \sup_\theta \mathrm{E}(\varphi_n - \hat{f}_{h_\mathrm{M}})^2 / \mathrm{R}(f_\theta, \hat{f}_{h_\mathrm{M}}) = 0. \tag{3.4}$$

Proof of (3.4). — Observe that in view of (2.5)

$$\inf_\theta \mathrm{R}(f_\theta, \hat{f}_{h_\mathrm{M}}) \geqq c\, \mathrm{C}_n^{-1/5} n^{-4/5} \quad \text{for some } c>0;$$

thus it suffices to show that

$$\mathrm{E}_\theta(\varphi_n - \hat{f}_{h_\mathrm{M}})^2 = o(\mathrm{C}_n^{-1/5} n^{-4/5}) \tag{3.5}$$

uniformly in θ. Since $|\mathrm{K}|$ is bounded by α we have

$$\mathrm{E}_\theta(\varphi_n - \hat{f}_{h_\mathrm{M}})^2 \leqq 4\alpha^2 h_m^{-2} \mathrm{P}_\theta(\hat{h} \neq h_\mathrm{M}).$$

Further

$$\mathrm{P}_\theta(\hat{h} \neq h_\mathrm{M}) \leqq \mathrm{P}_\theta(\tilde{f}(0) < 1/2) + \mathrm{P}_\theta(\tilde{f}''(0)^2 > (2\mathrm{C}_n)^{-1})$$

and observe that twofold integration by parts yields

$$\mathrm{E}_\theta \tilde{f}''_\delta(0) = \delta^{-1} \int f''_\theta(y)\, \mathrm{H}(y/\delta)\, dy$$

and substituting $f''_\theta(x) = (\beta - h_\mathrm{M})^{-1/2} \theta h_\mathrm{M}^{-5/2} \mathrm{K}''(x/h_\mathrm{M})$ for small x we obtain that the above integral is of the form

$$(\beta - h)^{-1/2} \theta h_\mathrm{M}^{-3/2} \delta^{-1} \int \mathrm{K}''(y)\, \mathrm{H}(y h_\mathrm{M}/\delta)\, dy.$$

Expanding H around 0 and using the fact that $\int \mathrm{K}''(y)\, dy = 0$ and K″ is symmetric we obtain that the above integral is of order $(\mathrm{C}_n h_\mathrm{M}/n)^{1/2} \delta^{-3}$. Thus it is easy to see that a condition

$$\mathrm{C}_n^{2+1/5}/n^{6\varepsilon} \to 0$$

implies that

$$\sup_\theta (\mathrm{E}_\theta \tilde{f}''(0))^2 = o(\mathrm{C}_n^{-1}). \tag{3.6}$$

Analogously we obtain that

$$\mathrm{E}_\theta \tilde{f}_\delta(0) = 1 + O(\theta h_\mathrm{M}^{5/2}/\delta^3) = 1 + o(1) \tag{3.7}$$

provided that

$$\mathrm{C}_n^2 = o(n^{4/5+6\varepsilon}).$$

Observe further that $\tilde{f}_\delta$ and $\tilde{f}''_\delta$ are sums of i. i. d. *r. v's* and

$$|\tilde{f}_\delta| = O(\delta^{-1}), \qquad \mathrm{Var}_\theta \tilde{f}_\delta = O((n\delta)^{-1})$$
$$|\tilde{f}''_\delta| = O(\delta^{-3}), \qquad \mathrm{Var}_\theta \tilde{f}''_\delta = O((n\delta^5)^{-1}).$$

Thus Benett inequality (Benett [2]) together with (3.6) and (3.7) yields

$$P_\theta(\hat{h} \neq h_M) \leqq C\{\exp(-C_1 n\delta) + \exp(-C_2 n\delta^5/C_n)\}$$

for some $C, C_1, C_2 > 0$. Thus (3.4) is satisfied provided that $C_n^{1/5} n^{6/5} c_n^{-2/5} \exp(-n^{5\varepsilon}/C_n) = o(1)$ which is obvious in view of the imposed assumptions. □

Remark. – Let us devide the sample into three parts of size m, m, $n-m$ respectively, where $m = o(n)$ and let $\tilde{f}_\delta(\tilde{f}''_\delta)$ denote now the pilot estimate of the density (the second derivative) based on the first (the second) subsample. Denote by f_n^* the kernel estimate based on the third subsample with $\hat{h}$ defined as in (3.3). The similar reasoning to that presented above shows that Corollary 3 is also true for f_n^* under suitable growth conditions imposed on c_n and C_n.

REFERENCES

[1] I. S. ABRAMSON, On Bandwidth Variation in Kernel Estimates a Square Root Law, *Ann. Stat.*, Vol. **10**, 1982, pp. 1217-1223.

[2] G. BENETT, Probability Inequalities for the Sum of Independent Random Variables, *J. Amer. Statist.*, Vol. **57**, 1962, pp. 33-45.

[3] J. BRETAGNOLLE and C. HUBER, Estimation des densités : risque minimax, *Z. Wachrscheinlichkeitstheorie und verw. Gebiete*, Vol. **47**, 1979, pp. 113-137.

[4] R. H. FARRELL, On the Best Obtainable Rates of Convergence in Estimation of a Density Function at a Point, *Ann. Math. Statist.*, Vol. **43**, 1972, pp. 170-180.

[5] L. GAJEK, On Improving Density Functions Wchich are not Bona Fide Functions, *Ann. Statist.*, Vol. **14**, 1986, pp. 1612-1618.

[6] I. A. IBRAGIMOV and R. Z. HASMINSKII, *Statistical Estimation, Asymptotic Theory*, Springer, 1981.

[7] E. PARZEN, On Estimation of Probability Density and Mode, *Ann. Math. Statist.*, Vol. **33**, 1962, pp. 1065-1076.

[8] J. SACKS and W. STRAWDERMAN, Improvements on Linear Minimax Estimates. In *Statistical Decision Theory and Related Topics* 3, Vol. **2**, S. S. Gupta and J. O. Berger Eds., New York, Academic Press, 1982, pp. 287-304.

[9] J. SACKS and D. YLVISACKER, Asymptotically Optimal Kernels for Density Estimation at a Point, *Ann. Statist.*, Vol. **9**, 1981, pp. 334-346.

[10] G. R. TARELL and D. W. SCOTT, On Improving Convergence Rates for Nonnegative Kernel Density Estimators, *Ann. Statist.*, Vol. **8**, 1980, pp. 1160-1163.

[11] M. WOODROOFE, On Chosing a Delta Sequence, *Ann. Math. Statist.*, Vol. **41**, 1970, pp. 1665-1671.

[12] L. D. BROWN and R. H. FARRELL, *A Lower Bound for the Risk in Estimating the Value of a Probability Density*, Preprint Math. Dt Cornell University, 1987.

[13] D. L. DONOHO and R. L. LIU, *Geometrizing Rates of Convergence, III*, Preprint Dt of Statistics, University of California, 1987.

(*Manuscrit reçu le* 11 *avril* 1988)
(*corrigé le* 28 *novembre* 1988.)