D. M. Mason, Границы риска для ядерных оценок плотности,
Зап. научн. сем. ПОМИ, 2009, том 363, 66–104

Записки научных
семинаров ПОМИ
Том 363, 2009 г.

D. M. Mason

RISK BOUNDS FOR KERNEL DENSITY ESTIMATORS

Abstract. We use results from probability on Banach spaces and Poissonization techniques to develop sharp finite sample and asymptotic moment bounds for the Lp risk for kernel density estimators. Our results are shown to augment previous work in this area.

1. Introduction

In order to motivate our investigations consider the following minimax result of Wertz (1974). For a given $M > 1$ and $p \geq 1$, let

$$\mathcal{L}_p(M) = \left\{ f : f \text{ is a density on } \mathbb{R} \text{ and } \|f\|_p \leq M \right\}, \tag{1}$$

with $\|f\|_p$ denoting the $L_p(\mathbb{R})$ norm. Wertz (1974) proved for $p > 1$ that for each $n \geq 1$ there exists a density estimator $\widetilde{f}_{0,n}$ based on $X_1, \ldots, X_n$, i.i.d. with density f such that

$$\sup_{f \in \mathcal{L}_p(M)} \left(E \left\| \widetilde{f}_{0,n} - f \right\|_p^p \right)^{1/p} = \inf_{\widetilde{f}_n} \sup_{f \in \mathcal{L}_p(M)} \left(E \left\| \widetilde{f}_n - f \right\|_p^p \right)^{1/p},$$

where $\widetilde{f}_n$ is an arbitrary density estimator of f based on $X_1, \ldots, X_n$. A natural question is the following: assume that f lies in a smooth class of densities $\mathcal{F}$, then do there exist a sequence of constants $a_n \to \infty$ and some $\alpha > 0$ such that

$$a_n \inf_{\widetilde{f}_n} \sup_{f \in \mathcal{L}_p(M) \cap \mathcal{F}} \left(E \left\| \widetilde{f}_n - f \right\|_p^p \right)^{1/p} \to \alpha? \tag{2}$$

This problem remains largely unsolved except for the case $p = 2$ and for certain smooth classes of densities $\mathcal{F}$. Refer to Schipper (1996) and the

---

Research partially supported by NSF Grant DMS–0503908 and SFB 701 University of Bielefeld .

references therein. (Actually, one can infer from Theorem 1 of Devroye (1983) that the limit in (2) is infinite for any sequence of constants $a_n \to \infty$ if the supremum is taken over $f \in \mathcal{L}_p(M)$ instead of over a smooth enough subclass of $\mathcal{L}_p(M)$.)

Much more is known about the less precise problem of showing the existence of constants $0 < \alpha < \beta < \infty$ such that

$$\alpha \leq \liminf_{n\to\infty} a_n \inf_{\widetilde{f}_n} \sup_{f\in\mathcal{L}_p(M)\cap\mathcal{F}} \left(E\left\|\widetilde{f}_n - f\right\|_p^p\right)^{1/p}$$
$$\leq \limsup_{n\to\infty} a_n \inf_{\widetilde{f}_n} \sup_{f\in\mathcal{L}_p(M)\cap\mathcal{F}} \left(E\left\|\widetilde{f}_n - f\right\|_p^p\right)^{1/p} \leq \beta. \tag{3}$$

See Bretagnolle and Huber (1979), Ibragimov and Hasminskii (1980,1982), Efroimovich and Pinsker (1982) and Hasminskii and Ibragimov (1990), as well as the monograph by Devroye and Györfi (1985).

It turns out that approximate solutions in terms of having the proper rate to the precise problem (2), as well as to the coarser problem (3), can often be achieved by sequences of kernel density estimators. In fact, the right side of (3) is bounded by

$$\limsup_{n\to\infty} a_n \sup_{f\in\mathcal{L}_p(M)\cap\mathcal{F}} \left(E\left\|f_n - f\right\|_p^p\right)^{1/p}, \tag{4}$$

where $\{f_n\}$ is any sequence of kernel density estimators. Much of this paper will be concerned with developing bounds on the lim sup in (4).

To fix some notation and assumptions, let $X, X_1, X_2, \ldots$ be i.i.d. with density f. A *kernel density estimator* of f based on $X_1, \ldots, X_n$, $n \geq 1$, is defined to be

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),\ x \in \mathbb{R}, \tag{5}$$

where h_n are positive constants such that

$$h_n \to 0 \text{ and } nh_n \to \infty \text{ as } n \to \infty \tag{h}$$

and K is a kernel satisfying the following condition:

$(K.i)$ K is in $L_2(\mathbb{R})$ and bounded by some constant $0 < \kappa < \infty$.

At times we shall also assume that
$(K.ii)$ $\int_0^\infty \Psi_{K^2}(x)dx < \infty$, where for any measurable function H

$$\Psi_H(x) = \sup_{|y| \geq x} |H(y)|, \quad x \geq 0.$$

Condition $(K.ii)$ is introduced so that when needed we can apply part (c) of Theorem 2 on page 62 of Stein (1970), which says that if

$$\int_0^\infty \Psi_H(x)dx < \infty \tag{6}$$

then

$$H_h * f(z) \to J(H)f(z), \text{ as } h \searrow 0, \text{ for almost all } z \in \mathbb{R}, \tag{7}$$

where for any function H and $h > 0$,

$$J(H) = \int_{\mathbb{R}} H(u)\, du, \tag{8}$$

$$H_h = h^{-1} H\left(\cdot h^{-1}\right) \tag{9}$$

and

$$H_h * f(z) := h^{-1} \int_{\mathbb{R}} H\left(\frac{z-x}{h}\right) f(x)\, dx. \tag{10}$$

Define for any $p \geq 1$

$$\|f_n - Ef_n\|_p = \left(\int_{\mathbb{R}} |f_n(x) - Ef_n(x)|^p dx\right)^{1/p}.$$

Let f_n and f_n' be independent and $f_n' =_d f_n$. Notice that

$$2\left(E\,\|f_n - f\|_p^p\right)^{1/p} = \left(E\,\|f_n - f\|_p^p\right)^{1/p} + \left(E\,\|f_n' - f\|_p^p\right)^{1/p},$$

which by Minkowski's inequality and Jensen's inequality is

$$\geq \left(E\,\|f_n - f_n'\|_p^p\right)^{1/p} \geq \left(E\,\|f_n - Ef_n\|_p^p\right)^{1/p}.$$

Therefore for any $p \geq 1$,

$$\left(E\left\|f_n - f\right\|_p^p\right)^{1/p} \geq 2^{-1}\left(E\left\|f_n - Ef_n\right\|_p^p\right)^{1/p}.$$

Furthermore, by Minkowski's inequality,

$$\left(E\left\|f_n - Ef_n\right\|_p^p\right)^{1/p} + \left\|f - Ef_n\right\|_p \geq \left(E\left\|f_n - f\right\|_p^p\right)^{1/p}.$$

This says that in terms of rates, $2^{-1}\left(E\left\|f_n - Ef_n\right\|_p^p\right)^{1/p}$ provides a lower bound on the L_p risk $\left(E\left\|f_n - f\right\|_p^p\right)^{1/p}$ of the density estimator f_n, and $\left(E\left\|f_n - Ef_n\right\|_p^p\right)^{1/p}$ plus the bias $\left\|f - Ef_n\right\|_p$ term gives an upper bound.

The first goal of this paper is to provide good bounds for the moments $E\left\|f_n - Ef_n\right\|_p^r$ for any $p \geq 1$ and $r \geq 1$. Our main tool will be a moment bound for sums of independent random variables taking values in a Banach space due to Talagrand (1989).

Our second goal will be to study the exact asymptotic behavior of $E\left\|f_n - Ef_n\right\|_p^r$ as $n \to \infty$. Here our tools will be the Poissonization methods developed in Giné, Mason and Zaitsev (2003).

One application of our results will lead to the result that for each $1 \leq p < \infty$, under suitable regularity conditions on the density f, the kernel K and the loss function w, and with Z denoting a standard normal random variable,

$$\lim_{n\to\infty} Ew\left(\frac{\sqrt{nh_n}\left\|f_n - Ef_n\right\|_p}{\|K\|_2\left(E|Z|^p\int_{\mathbb{R}} f^{p/2}(y)\,dy\right)^{1/p}}\right) = w(1). \tag{11}$$

Another application will show that for a suitably defined class of densities $\mathcal{F}_{p/2}$,

$$\limsup_{n\to\infty}\sup_{f\in\mathcal{F}_{p/2}} Ew\left(\sqrt{nh_n}\left\|f_n - Ef_n\right\|_p\right) < \infty,$$

where K, w and p are as in (11). We also discuss conditions on f under which Ef_n can be replaced by f in these last two statements.

We shall not treat the more intricate problem of the derivation of lower bounds in (3). To obtain good bounds requires a considerable amount of ingenuity and special techniques. For instance, Bretagnolle and Huber (1979) base their bounds on a Kullbach information-type inequality, the arguments of Ibragimov and Hasminskii (1980) and Hasminskii and Ibragimov (1990) rely on Fano's inequality, Ibragimov and Hasminskii (1982) use a classic inequality of Hájek and Schipper (1996) utilizes the van Trees inequality.

Our main results are stated and proved in section 2. In section 3 we discuss the relationship of our results to known risk bounds, especially those of Ibragimov and Hasminskii (1980), Hasminskii and Ibragimov (1990) and Bretagnolle and Huber (1979). Moreover, as a by-product of our results we will provide a partial solution to a conjecture of Guerre and Tsybakov (1998).

2. Main results and proofs

We are first interested in finding a good asymptotic bound for

$$E\int_A |f_n(x) - Ef_n(x)|^p \, dx, \tag{12}$$

with A being a measurable set and $p \geq 1$. In the following calculations whenever $\|K\|_p$ appears, we assume that it is finite. Under assumption $(K.i)$ this always holds for $p \geq 2$.

Case 1a. $p > 2$. To handle this case we shall need a fact.

Fact 1. Rosenthal's inequality. *If ξ_i are independent centered random variables, then, for every $p \geq 2$ and $n \in \mathbf{N}$,*

$$E\left|\sum_{i=1}^n \xi_i\right|^p \leq \left(\frac{15p}{\log p}\right)^p \max\left[\left(\sum_{i=1}^n E\xi_i^2\right)^{p/2}, \sum_{i=1}^n E|\xi_i|^p\right]. \tag{13}$$

(This version of Rosenthal's inequality is obtained by symmetrization of the inequality in Theorem 4.1 from Johnson, Schechtman and Zinn (1985).)

By Fact 1 for each $x \in \mathbb{R}$, with $C_p = \left(\frac{15p}{\log p}\right)^p$,

$$E\left(|f_n(x) - Ef_n(x)|^p\right) = E\left|\sum_{i=1}^n \frac{K\left(\frac{x-X_i}{h_n}\right) - EK\left(\frac{x-X_i}{h_n}\right)}{nh_n}\right|^p$$

$$\leq \frac{C_p}{n^{p/2}h_n^p}\left(E\left(K\left(\frac{x-X}{h_n}\right)-EK\left(\frac{x-X}{h_n}\right)\right)^2\right)^{p/2}$$
$$+\frac{C_p}{n^{p-1}h_n^p}E\left|K\left(\frac{x-X}{h_n}\right)-EK\left(\frac{x-X}{h_n}\right)\right|^p,$$

which after some elementary bounds is

$$\leq \frac{2^pC_p}{n^{p/2}h_n^p}\left(\int_{\mathbb{R}} K^2\left(\frac{x-y}{h_n}\right)f(y)\,dy\right)^{p/2}+\frac{2^pC_p}{n^{p-1}h_n^p}\int_{\mathbb{R}}|K|^p\left(\frac{x-y}{h_n}\right)f(y)\,dy. \tag{14}$$

Thus

$$\int_A E\left|\sum_{i=1}^n \frac{K\left(\frac{x-X_i}{h_n}\right)-EK\left(\frac{x-X_i}{h_n}\right)}{nh_n}\right|^p dx$$
$$\leq \frac{2^pC_p}{n^{p/2}h_n^p}\int_A\left(\int_{\mathbb{R}} K^2\left(\frac{x-y}{h_n}\right)f(y)\,dy\right)^{p/2}dx$$
$$+\frac{2^pC_p}{n^{p-1}h_n^p}\int_A\int_{\mathbb{R}}|K|^p\left(\frac{x-y}{h_n}\right)f(y)\,dydx.$$

Now since

$$\int_{\mathbb{R}}\int_{\mathbb{R}}|K|^p\left(\frac{x-y}{h_n}\right)f(y)\,dydx = h_n\|K\|_p^p,$$

we get the bound

$$E\int_A|f_n(x)-Ef_n(x)|^p\,dx$$
$$\leq \frac{2^pC_p}{(nh_n)^{p/2}}\int_A\left(\int_{\mathbb{R}}\frac{1}{h_n}K^2\left(\frac{x-y}{h_n}\right)f(y)\,dy\right)^{p/2}dx+\frac{2^pC_p\|K\|_p^p}{(nh_n)^{p-1}}. \tag{15}$$

Assume now that

$$\int_{\mathbb{R}}(f(y))^{p/2}\,dy<\infty. \tag{16}$$

Condition $(K.ii)$ allows us to apply part (c) of Theorem 2 on page 63 of Stein (1970) to give

$$\int_A \left(\frac{1}{h_n} \int_{\mathbb{R}} K^2 \left(\frac{x-y}{h_n} \right) f(y)\, dy \right)^{p/2} dx \to \|K\|_2^p \int_A (f(y))^{p/2}\, dy,$$

which, since $(nh_n)^{p/2} / (nh_n)^{p-1} \to 0$, implies via (15) that

$$\limsup_{n\to\infty} \left((nh_n)^{p/2} E \int_A |f_n(x) - Ef_n(x)|^p\, dx \right) \le 2^p C_p \|K\|_2^p \int_A (f(y))^{p/2}\, dy.$$

Case 1b. $p = 2$. Obviously,

$$E\left(|f_n(x) - Ef_n(x)|^2 \right) \le \frac{1}{nh_n^2} EK^2 \left(\frac{x-X}{h_n} \right)$$
$$= \frac{1}{nh_n^2} \int_{\mathbb{R}} K^2 \left(\frac{x-y}{h_n} \right) f(y)\, dy.$$

We get as before,

$$\limsup_{n\to\infty} \left(nh_n E \int_A |f_n(x) - Ef_n(x)|^2\, dx \right) \le \|K\|_2^2 \int_A f(y)\, dy.$$

Case 2. $1 \le p < 2$. In this case

$$E \int_A |f_n(x) - Ef_n(x)|^p\, dx \le \int_A \left(E|f_n(x) - Ef_n(x)|^2 \right)^{p/2} dx$$
$$\le \frac{1}{(nh_n)^{p/2}} \int_A \left(\frac{1}{h_n} \int_{\mathbb{R}} K^2 \left(\frac{x-y}{h_n} \right) f(y)\, dy \right)^{p/2} dx$$
$$= \frac{1}{(nh_n)^{p/2}} \int_A \left(\left(K^2\right)_{h_n} * f(y) \right)^{p/2} dy, \tag{17}$$

where for $h > 0$, $\left(K^2\right)_h(\cdot) = h^{-1} K^2(\cdot h^{-1})$.

The following lemma shows that moment conditions on X and K yield a useful bound on $\int_{\mathbb{R}} \left(\left(K^2\right)_{h_h} * f(y) \right)^{p/2} dy$ by choosing $H = K^2 / \|K\|_2^2$. Its statement and proof are based on ideas and results in Chapter 7 in Devroye (1987) and Section 3 of Devroye (1992). See especially, Remark 3 in Devroye (1992).

Lemma 1. *Let H and f be densities on $\mathbb{R}$ and let X have density f and Y have density H. Choose $\frac{1}{s}+\frac{1}{t}=1$, with $t>1$, and $\lambda>t/s$. Further assume that $E|X|^{\lambda}<\infty$ and $E|Y|^{\lambda}<\infty$. Then for any measurable subset A of $\mathbb{R}$ for a finite positive constant $C(\lambda,s)$*

$$\int_A (f(x))^{1/t}\,dx \le C(\lambda,s)\left(E\left[\left(1+|X|^{\lambda}\right)1\{X\in A\}\right]\right)^{1/t} \tag{18}$$

and, with $H_h(\cdot)=h^{-1}H(\cdot h^{-1})$,

$$\int_A (H_h * f(x))^{1/t}\,dx \le C(\lambda,s)\left(E\left[\left(1+|X_h|^{\lambda}\right)1\{X_h\in A\}\right]\right)^{1/t}, \tag{19}$$

where X_h has density $H_h * f$. Moreover, if H satisfies (6) then

$$\limsup_{n\to\infty}\int_A (H_h * f(x))^{1/t}\,dx \le C(\lambda,s)\left(P(A)+E\left[|X|^{\lambda}1\{X\in A\}\right]\right)^{1/t}. \tag{20}$$

Proof. To prove (18) observe that by Hölder's inequality,

$$\int_A (f(x))^{1/t}\,dx \le \left(\int_{\mathbb{R}}\left\{1+|x|^{\lambda}\right\}^{-s/t}dx\right)^{1/s}\left(\int_A\left\{1+|x|^{\lambda}\right\}f(x)dx\right)^{1/t}$$

$$=: C(\lambda,s)\left(E\left[\left(1+|X|^{\lambda}\right)1\{X\in A\}\right]\right)^{1/t}.$$

Next, (19) is a special case of (18). Finally we turn to the proof of (20). Now $X_h =_d X+Y_h$, where Y_h has density H_h and Y_h and X are independent. From this we get that when $\lambda\ge 1$

$$\left(E|X+Y_h|^{\lambda}\right)^{1/\lambda}\le\left(E|X|^{\lambda}\right)^{1/\lambda}+\left(E|Y_h|^{\lambda}\right)^{1/\lambda} \tag{21}$$

and when $0<\lambda<1$

$$E|X+Y_h|^{\lambda}\le E|X|^{\lambda}+E|Y_h|^{\lambda}. \tag{22}$$

By writing

$$E|Y_h|^{\lambda}=\int_{-\infty}^{\infty}|y|^{\lambda}H_h(y)dy=h^{\lambda}E|Y|^{\lambda}, \tag{23}$$

we readily infer from (21), (22) and (23) that for any $0 < \lambda < \infty$,

$$\limsup_{h \searrow 0} E|X + Y_h|^\lambda \le E|X|^\lambda. \tag{24}$$

Furthermore, since we assume condition (6) holds we have by (7) that for almost every y

$$\left(1 + |y|^\lambda\right) H_h(y) \to \left(1 + |y|^\lambda\right) f(y) \text{ as } h \searrow 0.$$

Piecing everything together we conclude from Scheffé's theorem (see exercise 7 on page 862 of Shorack and Wellner (1986)) that

$$E\left[\left(1 + |X_h|^\lambda\right) 1\{X_h \in A\}\right] \to E\left[\left(1 + |X|^\lambda\right) 1\{X \in A\}\right], \text{ as } h \searrow 0. \tag{25}$$

Statement (20) obviously follows from (19) and (25). □

Remark 1. For use later on, observe that from inequality (19) and (23) we get from $X_h =_d X + Y_h$ and the c_r-inequality that

$$E|X_h|^\lambda = E|X + Y_h|^\lambda \le c_\lambda \left(E|X|^\lambda + h^\lambda E|Y|^\lambda\right),$$

where $c_\lambda = 2^{(\lambda-1)\vee 1}$. Thus we get

$$\int_{\mathbb{R}} \left(H_h * f(x)\right)^{1/t} dx \le C(\lambda, s)\left(1 + c_\lambda\left(E|X|^\lambda + h^\lambda E|Y|^\lambda\right)\right)^{1/t}. \tag{26}$$

We summarize these observations in the following proposition.

Proposition 1. *Let K be a kernel satisfying (K.i) and (K.ii). For any $p \ge 2$ such that* (16) *holds, we have for some constant D_p depending only on p*

$$\limsup_{n\to\infty}\left((nh_n)^{p/2} E\int_A |f_n(x) - Ef_n(x)|^p\, dx\right) \le D_p \|K\|_2^p \int_A (f(y))^{p/2}\, dy. \tag{27}$$

Moreover, whenever for a given $1 \le p < 2$, X *and* $H = K^2/\|K\|_2^2$ *the conditions of Lemma* 1 *are satisfied, we have for any measurable subset* $A \subset \mathbb{R}$

$$\limsup_{n\to\infty} \left((nh_n)^{p/2} E \int_A |f_n(x) - Ef_n(x)|^p \, dx \right)$$

$$\le C(\lambda, s) \|K\|_2^p \left(E\left[\left(1 + |X|^\lambda\right) 1\{X \in A\} \right] \right)^{p/2}, \tag{28}$$

where $1/s + p/2 = 1$ *and* λ, Y *and* $C(\lambda, s)$ *are defined as in Lemma* 1.

Note that statement (28) follows from (17) and (20).

We next turn to the task of deriving a useful finite sample bound for

$$E\|f_n - Ef_n\|_p^r = E\left(\int_{\mathbb{R}} |f_n(x) - Ef_n(x)|^p \, dx \right)^{r/p},$$

with $p \ge 1$ and $r \ge 1$.

To do this we shall need an additional fact.

Fact 2. (Theorem 1 of Talagrand (1989)). *If* $\mathbf{B}$ *is a separable Banach space with norm* $\|\cdot\|$, Z_i, $i \in \mathbf{N}$, *are independent mean zero random vectors taking values in* $\mathbf{B}$, *then for a universal constant* $D > 0$ *for all* $r \ge 1$ *and* $n \ge 1$,

$$\left(E\left(\|S_n\|^r\right)\right)^{1/r} \le \frac{Dr}{1 + \log r} \left(E\|S_n\| + \left(E \max_{1\le i\le n} \|Z_i\|^r \right)^{1/r} \right), \tag{29}$$

where $S_n = Z_1 + \cdots + Z_n$.

We get from (29) and the c_r inequality the bound,

$$E\left(\|S_n\|^r\right) \le \frac{D^r 2^{r-1} r^r}{(1 + \log r)^r} \left((E\|S_n\|)^r + E \max_{1\le i\le n} \|Z_i\|^r \right). \tag{30}$$

We shall apply the bound (30) to the random functions

$$Z_i(\cdot) = \frac{K\left(\frac{\cdot - X_i}{h_n}\right) - EK\left(\frac{\cdot - X}{h_n}\right)}{nh_n}, \; i = 1, \ldots, n.$$

As before in the following calculations whenever $\|K\|_p$ appears, we assume that it is finite and when needed that (16) holds. We find by Jensen's inequality that,

$$E\|S_n\|_p = E\left\|f_n - Ef_n\right\|_p$$

$$= E\left(\int_{\mathbb{R}} |f_n(x) - Ef_n(x)|^p\, dx\right)^{1/p} \le \left(E\left\|f_n - Ef_n\right\|_p^p\right)^{1/p}. \tag{31}$$

Case 1a. $p > 2$. By inequality (15) (it also holds when $p = 2$) we see that

$$E\int_{\mathbb{R}} |f_n(x) - Ef_n(x)|^p\, dx$$

$$\le \frac{2^p C_p}{n^{p/2} h_n^p} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} K^2\left(\frac{x-y}{h_n}\right) f(y)\, dy\right)^{p/2} dx + \frac{2^p C_p \left\|K\right\|_p^p}{(nh_n)^{p-1}}. \tag{32}$$

Notice that

$$\left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} K^2\left(\frac{x-y}{h_n}\right) f(y)dy\right)^{p/2} dx\right)^{2/p}$$

$$= \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} K^2\left(\frac{t}{h_n}\right) f(t-x)\, dt\right)^{p/2} dx\right)^{2/p},$$

which by Young's inequality (see page 232 of Folland (1984)) is

$$\le \int_{\mathbb{R}} |K|^2\left(\frac{t}{h_n}\right)\left(\int_{\mathbb{R}} f^{p/2}(x-t)\, dx\right)^{2/p} dt = h_n \left\|K\right\|_2^2 \left(\int_{\mathbb{R}} f^{p/2}(y)\, dy\right)^{2/p}.$$

This gives the bound

$$E\int_{\mathbb{R}} |f_n(x) - Ef_n(x)|^p\, dx \le \frac{2^p C_p \left\|K\right\|_2^p}{(nh_n)^{p/2}} \int_{\mathbb{R}} f^{p/2}(y)\, dy + \frac{2^p C_p \left\|K\right\|_p^p}{(nh_n)^{p-1}}. \tag{33}$$

Therefore for any $r \geq 1$,

$$\left(E\left\|f_n - Ef_n\right\|_p\right)^r \leq \left(\frac{2^p C_p \left\|K\right\|_2^p}{(nh_n)^{p/2}} \int\limits_{\mathbb{R}} f^{p/2}(y)\,dy + \frac{2^p C_p \left\|K\right\|_p^p}{(nh_n)^{p-1}}\right)^{r/p}, \tag{34}$$

which by (32), (34) and the c_r inequality for any $r \geq 1$ is

$$\leq \frac{2^{r/p} 2^r C_p^{r/p}}{(nh_n)^{r/2}} \left(\left\|K\right\|_2^r \left(\int\limits_{\mathbb{R}} f^{p/2}(y)\,dy\right)^{r/p} + \frac{\left\|K\right\|_p^r}{(nh_n)^{r/2-r/p}}\right). \tag{35}$$

Here to derive (35) we used for notational simplicity the rough version of the c_r inequality that says that for all $\gamma > 0$

$$|x+y|^\gamma \leq 2^\gamma \left(|x|^\gamma + |y|^\gamma\right) \tag{36}$$

with $\gamma = r/p$. Next

$$E \max_{1\leq i\leq n} \left(\int\limits_{\mathbb{R}} \left|\frac{1}{nh_n}\left(K\left(\frac{x-X_i}{h_n}\right) - EK\left(\frac{x-X}{h_n}\right)\right)\right|^p dx\right)^{r/p}$$

$$\leq 2^r E \max_{1\leq i\leq n} \left(\int\limits_{\mathbb{R}} \left|\frac{1}{nh_n} K\left(\frac{x-X_i}{h_n}\right)\right|^p dx\right)^{r/p} = \frac{2^r h_n^{r/p} \left\|K\right\|_p^r}{(nh_n)^r}. \tag{37}$$

Inserting the bounds (35) and (37) into (30) we get

$$E\left\|f_n - Ef_n\right\|_p^r = E\|S_n\|_p^r$$

$$\leq \frac{D^r 2^{r-1} r^r}{(1+\log r)^r} \frac{2^{r/p} 2^r C_p^{r/p}}{(nh_n)^{r/2}} \left(\left\|K\right\|_2^r \left(\int\limits_{\mathbb{R}} f^{p/2}(y)\,dy\right)^{r/p} + \frac{\left\|K\right\|_p^r}{(nh_n)^{r/2-r/p}}\right)$$

$$+ \frac{D^r 2^{r-1} r^r}{(1+\log r)^r} \frac{2^r h_n^{r/p} \left\|K\right\|_p^r}{(nh_n)^r},$$

which for an appropriate constant $L_p > 0$ is for all $r \geq 1$ and $n \geq 1$

$$\leq \left(\frac{rL_p}{1+\log r}\right)^r \left[\frac{1}{(nh_n)^{r/2}}\left(\|K\|_2^r \left(\int_{\mathbb{R}} f^{p/2}(y)\,dy\right)^{r/p} + \frac{\|K\|_p^r}{(nh_n)^{r/2-r/p}}\right) + \frac{h_n^{r/p}\|K\|_p^r}{(nh_n)^r}\right]. \tag{38}$$

Case 1b. $p = 2$. In this case

$$E\int_{\mathbb{R}} |f_n(x) - Ef_n(x)|^2\,dx \leq \frac{1}{nh_n^2}\int_{\mathbb{R}} EK^2\left(\frac{x-X}{h_n}\right)dx = \frac{\|K\|_2^2}{nh_n}. \tag{39}$$

Hence,

$$(E\|f_n - Ef_n\|_2)^r \leq \left(E\|f_n - Ef_n\|_2^2\right)^{r/2} \leq \frac{\|K\|_2^r}{(nh_n)^{r/2}}.$$

Now by arguing as in Case 1a, we get that for an appropriate L_2

$$E\|f_n - Ef_n\|_2^r \leq \left(\frac{rL_2}{1+\log r}\right)^r \left[\frac{\|K\|_2^r}{(nh_n)^{r/2}} + \frac{h_n^{r/p}\|K\|_2^r}{(nh_n)^r}\right]. \tag{40}$$

Case 2. $1 \leq p < 2$. In this case we get from inequality (17),

$$E\int_{\mathbb{R}} |f_n(x) - Ef_n(x)|^p\,dx \leq \frac{1}{(nh_n)^{p/2}}\int_A \left(\left(K^2\right)_{h_n} * f(y)\right)^{p/2} dy \tag{41}$$

and from (37), which holds for $p \geq 1$, and (30) that for each $1 \leq p < 2$ there is a constant $L_p > 0$ such that for all $r \geq 1$ and $n \geq 1$

$$E\|f_n - Ef_n\|_p^r$$

$$\leq \left(\frac{rL_p}{1+\log r}\right)^r \left[\frac{\left(\int_{\mathbb{R}} \left(\left(K^2\right)_{h_n} * f(y)\right)^{p/2} dy\right)^{r/p}}{(nh_n)^{r/2}} + \frac{h_n^{r/p}\|K\|_p^r}{(nh_n)^r}\right]. \tag{42}$$

We shall summarize these observations in the following proposition.

Proposition 2. *Let K be a kernel satisfying (K.i). For any $p>2$ such that (16) holds we have for some constant $L_p>0$, and all $r\geq 1$ and $n\geq 1$,*

$$E\left\|f_n-Ef_n\right\|_p^r$$
$$\leq\left(\frac{rL_p}{1+\log r}\right)^r\left[\frac{1}{(nh_n)^{r/2}}\left(\left\|K\right\|_2^r\left(\int_{\mathbb{R}}f^{p/2}(y)\,dy\right)^{r/p}\right.\right.$$
$$\left.\left.+\frac{\left\|K\right\|_p^r}{(nh_n)^{r/2-r/p}}\right)+\frac{h_n^{r/p}\left\|K\right\|_p^r}{(nh_n)^r}\right]. \tag{43}$$

For $p=2$ we have for some constant $L_2>0$, and all $r\geq 1$ and $n\geq 1$,

$$E\left\|f_n-Ef_n\right\|_2^r\leq\left(\frac{rL_2}{1+\log r}\right)^r\left[\frac{\left\|K\right\|_2^r}{(nh_n)^{r/2}}+\frac{h_n^{r/p}\left\|K\right\|_2^r}{(nh_n)^r}\right]. \tag{44}$$

Moreover, whenever $1\leq p<2$, we have for some constant $L_p>0$ and all $r\geq 1$ and $n\geq 1$,

$$E\left\|f_n-Ef_n\right\|_p^r\leq\left(\frac{rL_p}{1+\log r}\right)^r\times$$
$$\left[\frac{\left(\int_{\mathbb{R}}\left(\left(K^2\right)_{h_n}*f(y)\right)^{p/2}dy\right)^{r/p}}{(nh_n)^{r/2}}+\frac{h_n^{r/p}\left\|K\right\|_p^r}{(nh_n)^r}\right], \tag{45}$$

where it is assumed that $\left\|K\right\|_p<\infty$.

Remark 2. Whenever for a given $1\leq p<2$, X and $H=K^2/\left\|K\right\|_2^2$ the conditions of Lemma 1 are satisfied, we have using (45), (42) and (26) that for some constant $L_p>0$ and all $r\geq 1$ and $n\geq 1$,

$$E\left\|f_n-Ef_n\right\|_p^r$$

$$\leq\left(\frac{rL_p}{1+\log r}\right)^r\frac{\left\|K\right\|_2^r\left(C^{1/p}(\lambda,s)\left(1+c_\lambda\left(E|X|^\lambda+h^\lambda E|Y|^\lambda\right)\right)^{1/2}\right)^r}{(nh_n)^{r/2}}$$
$$+\left(\frac{rL_p}{1+\log r}\right)^r\frac{h_n^{r/p}\left\|K\right\|_p^r}{(nh_n)^r},$$

where $1/s + p/2 = 1$ and λ, Y and $C(\lambda, s)$ are defined as in Lemma 1.

Remark 3. Modulo constants depending on p, inequality (33) implies inequality (4.13) of Bretagnolle and Huber (1979) and inequalities (39) and (41) agree with the corresponding parts of their inequalities (4.14) and (4.15).

As an immediate corollary we get.

Corollary 1. *Let K be a kernel satisfying $(K.i)$. For any $p \geq 2$ and $M > 0$ let $\mathcal{L}_{p/2}(M)$ be a class of densities defined as in* (1) *and for $1 \leq p < 2$ let*

$$\mathcal{K}_{p/2}(K, M)$$
$$= \left\{ f : f \text{ is a density on } \mathbb{R} \text{ and } \left\| \left(K^2\right)_h * f \right\|_{p/2} \leq M \text{ for all } 0 < h \leq 1 \right\}. \tag{47}$$

We have for every $t > 0$,

$$\limsup_{n \to \infty} \sup_{f \in \mathcal{F}_{p/2}} E \exp\left(t\sqrt{nh_n} \left\| f_n - Ef_n \right\|_p \right) < \infty, \tag{48}$$

where it is assumed that $\|K\|_p < \infty$ for $1 \leq p < \infty$, $\mathcal{F}_{p/2} = \mathcal{L}_{p/2}(M)$ for $p \geq 2$ and $\mathcal{F}_{p/2} = \mathcal{K}_{p/2}(K, M)$ for $1 \leq p < 2$.

Proof. Clearly by Proposition 2 for $p \geq 1$, we have for some constant $A > 0$ for all $r \geq 1$ and $n \geq 1$

$$(nh_n)^{r/2} \sup_{f \in \mathcal{F}_{p/2}} E \left\| f_n - Ef_n \right\|_p^r \leq \left(\frac{rA}{1 + \log r} \right)^r \left[1 + \frac{1}{(nh_n)^{r/2}} \right]. \tag{49}$$

From this bound, (48) readily follows using the rough version of Stirling's approximation, which says for integers $r \geq 1$ that $r! > (r/e)^r$, and the assumption that $nh_n \to \infty$. □

In the next subsection we shall prove the following result.

Proposition 3. *Let K be a kernel satisfying (K.i) and (K.ii). Under the conditions of Proposition 2 for $p \geq 2$, as $n \to \infty$,*

$$(nh_n)^{p/2} E \left\| f_n - Ef_n \right\|_p^p \to m(p, f, K), \tag{50}$$

and

$$(nh_n)^{p/2} \left\| f_n - Ef_n \right\|_p^p \to_p m(p, f, K), \tag{51}$$

where

$$m(p,f,K) = \|K\|_2^p E|Z|^p \int_{\mathbb{R}} f^{p/2}(y)\,dy, \tag{52}$$

with Z denoting a standard normal random variable. Moreover, the limits (50) *and* (51) *also hold for $1 \le p < 2$ whenever the conditions stated in Remark* 2 *for* (46) *are fulfilled.*

An immediate consequence of Corollary 1 and (51) is the following corollary.

Corollary 2. *Under the conditions of Proposition* 3 *for $p \ge 1$ and any loss function w on $[0,\infty)$ that is continuous at* 1 *and such that for some $\lambda > 0$ and $C > 0$*

$$0 \le w(x) \le C \exp(\lambda x),\ x \in [0, \infty\ ), \tag{53}$$

we have

$$\lim_{n\to\infty} Ew\left(\frac{\sqrt{nh_n}\,\|f_n - Ef_n\|_p}{\|K\|_2\left(E|Z|^p \int_{\mathbb{R}} f^{p/2}(y)\,dy\right)^{1/p}}\right) = w(1). \tag{54}$$

Remark 4. To replace $\|f_n - Ef_n\|_p$ by $\|f_n - f\|_p$ in (50), (51), (52) and (54) requires additional smoothness conditions to control the $L_p(\mathbb{R})$ norm of the bias $f - Ef_n$. Here is a convenient set of conditions, which are detailed in Bretagnolle and Huber (1979) which lead to a good bound for $\|f - Ef_n\|_p$. In addition to $(K.i)$ and $(K.ii)$, assume that

$$\int_{\mathbb{R}} K(x)\,dx = 1. \tag{55}$$

Further assume that for some integer $s \ge 2$,
$(K.iii)$ K is continuous,

$$\int_{\mathbb{R}} u^j K(u)\,du = 0, \quad 1 \le j < s \quad \text{and} \quad \int_{\mathbb{R}} |u|^s |K(u)|\,du < \infty.$$

For kernels K satisfying $(K.iii)$ we define the $s-$Kernel ${}_sK$ for $u \geq 0$,

$${}_sK\left(u\right) = \left(-1\right)^s \int\limits_u^\infty \frac{\left(y-u\right)^{s-1}}{\left(s-1\right)!} K\left(y\right) dy \text{ and } {}_sK\left(-u\right) = -\left(-1\right)^s \ {}_sK\left(u\right).$$

Bretagnolle and Huber (1979) point out that ${}_sK \in L_1\left(\mathbb{R}\right) \cap \mathcal{C}^{(s)}$ and ${}_s\left(K_h\right) = h^s \left({}_sK\right)_h$. (Recall the notation defined in (9).) They show that whenever the density f is s times continuously differentiable with $s-$derivative $f^{(s)} \in L_p\left(\mathbb{R}\right)$, $p \geq 1$, then for all $h > 0$,

$$\left\|f - Ef_n\right\|_p \leq h^s \left\|f^{(s)}\right\|_p \left\|{}_sK\right\|_1. \tag{56}$$

Clearly then by using the inequality

$$\left|\left\|f_n - f\right\|_p - \left\|f_n - Ef_n\right\|_p\right| \leq \left\|f - Ef_n\right\|_p,$$

we see that whenever (56) holds and $\sqrt{nh_n}h_n^s \to 0$, then we can replace $\left\|f_n - Ef_n\right\|_p$ by $\left\|f_n - f\right\|_p$ in (50), (51), (52) and (54).

2.1. Proof of Proposition 3

Before proving Proposition 3 we must gather together some facts.

Fact 3. *Suppose that $\mathcal{H}$ is a finite class of bounded real valued measurable functions H in $L_1\left(\mathbb{R}\right)$ that satisfy (6). Then for any $H \in \mathcal{H}$, (7) holds. Moreover, for all $0 < \varepsilon < 1$, there exist $M, \nu > 0$ and a Borel set C of finite Lebesgue measure $m(C)$ such that*

$$C \subset [-M + \nu, M - \nu], \tag{57}$$

$$\alpha := \int\limits_{|x| > M} f(x)\, dx > 0, \tag{58}$$

$$P\left(C\right) := \int\limits_C f(x)\, dx > 1 - \varepsilon, \tag{59}$$

f is bounded, continuous and bounded away from zero on C, (60)

and uniformly in $H \in \mathcal{H}$,

$$\sup_{z \in C} \left|f * H_h(z) - J(H)\ f(z)\right| \to 0, \text{ as } h \searrow 0. \tag{61}$$

The first statement is just the Stein result cited above. The other statements are proved exactly as in the proof of Lemma 6.1 of Gine, Mason and Zaitsev (2003).

In order to state our next fact let $\xi, \xi_1, \xi_2, \cdots$ be independent, identically distributed random variables satisfying $E\xi = 0$ and $E\xi^2 = 1$, and let Z denote a standard normal random variable. We shall be applying a special case of Theorem 1 of Sweeting (1977). In order to state the particular result that we need, we must first gather together some notation from Sweeting (1977). Assume $E\left|\xi\right|^3 < \infty$ and set

$$\beta_3 = E\left|\xi\right|^3. \tag{62}$$

(The symbol β_3 is defined on page 30, lines -9 to -6, of Sweeting (1977).)

Let Γ denote the class of functions g on $[0, \infty)$ satisfying

(i) $g(0) = 0$ and $g(1) = 1$;

(ii) g is nonnegative and nondecreasing;

(iii) $t/g(t)$ is defined for all $t \in [0, \infty)$ and is nondecreasing.

(The class Γ is defined on page 35, lines 1–5 of Sweeting (1977).)

We shall use the particular function $g \in \Gamma$,

$$g(t) = \min(t, 1), \quad t \in [0, \infty). \tag{63}$$

Let $g \in \Gamma$ and $r \geq 2$ be an integer and define

$$\eta_n = n^{-(r-2)/2} E\left[|\xi|^r g\left(n^{-1/2}|\xi|\right)\right].$$

(The symbol η_n is defined on page 35, line 9, of Sweeting (1977).)

We will always choose g as in (63) and $r \geq 3$, giving

$$\eta_n = n^{-(r-2)/2} E\left[|\xi|^r g\left(n^{-1/2}|\xi|\right)\right] \leq \frac{E|\xi|^r}{n^{(r-2)/2}}. \tag{64}$$

Next let

$$\varepsilon_n = \frac{\beta_3}{\sqrt{n}}. \tag{65}$$

(The symbol ε_n is defined on page 35, line -6, of Sweeting (1977).)

Let ϕ be a fixed Borel measurable function on $\mathbb{R}$. For any $\varepsilon > 0$ and $x \in \mathbb{R}$, define

$$\omega_\phi^\varepsilon(x) = \sup\left\{|\phi(x) - \phi(y)| : |x - y| < \varepsilon\right\}.$$

(The symbol $\omega_\phi^\varepsilon(x)$ is defined at the bottom of page 36 of Sweeting (1977).)

Let $r \geq 3$ be an integer and $g \in \Gamma$. Set

$$h(t) = 1 + t^r g(t),$$

and put

$$\phi^*(x) = h\left(|x|\right)^{-1}\left[\phi(x) - \phi(0)\right]. \tag{66}$$

(The symbol $\phi^*(x)$ is defined on page 37 of Sweeting (1977).)

We will always use the choice with $r \geq 3$,

$$h(t) = 1 + t^r \min(t, 1), \tag{67}$$

i.e., g is as in (63).

Further for any measurable function v on $\mathbb{R}$ denote

$$||v|| = \sup\left\{|v(x)| : x \in \mathbb{R}\right\}.$$

Here is a special case of Theorem 1 of Sweeting (1977) that we will be using.

Fact 4. *Suppose for an integer* $r \geq 3$

$$E|\xi|^r < \infty. \tag{68}$$

Then there exist universal positive constants C_1 *and* C_2 *such that for all measurable functions* ϕ *on* $\mathbb{R}$ *with* ϕ^* *bounded and defined as in* 66 *with* h *as in* (67) *such that*

$$\left| E\phi\left(\frac{\sum_{i=1}^n \xi_i}{\sqrt{n}}\right) - E\phi(Z)\right| \leq C_1\left[||\phi^*||\left(\varepsilon_n + \eta_n\right) + E\omega_\phi^{C_2\varepsilon_n}(Z)\right]. \tag{69}$$

We will be interested in the special case

$$\phi(x) = |x|^p,\ x \in \mathbb{R}, \text{ with } p \geq 1 \text{ an integer.}$$

In this case we choose an integer $r \geq \max(p, 3)$. Thus

$$||\phi^*|| = \sup\left\{\frac{|x|^p}{1 + |x|^r \min(x, 1)} : x \in \mathbb{R}\right\} \leq 1.$$

Whenever $|y - x| \leq C_2\varepsilon_n$, we get using the mean value theorem when $p > 1$ and the triangle inequality when $p = 1$ that for some constant A_p depending on p

$$|\phi(y) - \phi(x)| \leq A_p C_2 \varepsilon_n \left(|x|^{p-1} + |C_2\varepsilon_n|^{p-1}\right).$$

Inserting these bounds into (69) we get

$$\left| E\phi\left(\frac{\sum_{i=1}^{n} \xi_i}{\sqrt{n}}\right) - E\phi(Z)\right| \leq C_1 \left[\varepsilon_n + \eta_n + A_p C_2 \varepsilon_n \left(E\,|Z|^{p-1} + |C_2\varepsilon_n|^{p-1}\right)\right]. \tag{70}$$

We shall need the following special case of Lemma 2.1 of Giné, Mason and Zaitsev (2003). We say that a set D is a (commutative) semigroup if it has a commutative and associative operation, with a zero element. If D is equipped with a σ-algebra $\mathcal{D}$ for which the sum, $+ : (D \times D, \mathcal{D} \otimes \mathcal{D}) \mapsto (D, \mathcal{D})$, is measurable, then we say the $(D, \mathcal{D})$ is a measurable semigroup.

Fact 5. *Let $(D, \mathcal{D})$ be a measurable semigroup; let $Y_0 = 0 \in D$ and let Y, Y_i, $i \in \mathbf{N}$, be independent identically distributed D-valued random variables; for any given $n \in \mathbf{N}$ let η be a Poisson random variable with mean n independent of the sequence $\{Y_i\}$; and let $B \in \mathcal{D}$ be such that $0 <$* $\Pr\{Y \in B\} \leq 1/2$. *Then if $H : D \mapsto \mathbb{R}$ is non-negative and $\mathcal{D}$-measurable,*

$$EH\left(\sum_{i=0}^{n} I(Y_i \in B)Y_i\right) \leq 2EH\left(\sum_{i=0}^{\eta} I(Y_i \in B)Y_i\right). \tag{71}$$

Important special case. We will apply the preceding fact to the semigroup D with the operation sum generated by the point masses δ_{x_i},

$$D = \left\{0, \sum_{i=1}^{n} \delta_{x_i} : \; n \in \mathbf{N}, x_i \in S\right\},$$

where $S = \mathbb{R}$. For A a Borel subset of $\mathbb{R}$ set

$$B = \{\delta_x : x \in A\}.$$

Notice that $I(x \in A) = I(\delta_x \in B)$. Let $(\mathbb{R}, \mathcal{B})$ be the usual Borel measurable space. In this case the σ-algebra $\mathcal{D}$ is generated by the functions $f_{n,A} : \mathbb{R}^n \to D$, by

$$f_{n,A}(x_1, \ldots, x_n) = \sum_{i=1}^{n} I(x_i \in A)\delta_{x_i},$$

$n \in \mathbf{N}$. It is easy to see that for any measurable function $h : \mathbb{R} \mapsto \mathbb{R}$, the map $\mu \mapsto \int h d\mu$ is $\mathcal{D}$-measurable (just note that

$$f_{n,A}^{-1}\left\{\mu \in D, \int h d\mu \le t\right\} = \left\{(x_1, \ldots, x_n) : \sum_{i=1}^{n} I(x_i \in A)h(x_i) \le t\right\}$$

is a measurable subset of $\mathbb{R}^n$). Our functions H will have the general form

$$H\left(\sum_{i=1}^{n} I(\delta_{x_i} \in B)\delta_{x_i}\right) = H\left(\sum_{i=1}^{n} I(x_i \in A)\delta_{x_i}\right)$$

$$= \left(\int_C \left\{\left|\sum_{i=1}^{n} \mathbb{L}\left(\frac{x - x_i}{h}\right) - b(x)\right|^p - c(x)\right\} dx\right)^2, \tag{72}$$

where $p \ge 1$, $\mathbb{L}$ is a measurable bounded function equal to zero off of a compact interval $[-L, L]$, C is a measurable set and A is the Lh-neighborhood of C. In this setup X, $X_i, i \in \mathbf{N}$ will be a sequence of i.i.d. real valued random variables and $Y_i = \delta_{X_i}$, $i \in \mathbf{N}$. Also

$$b(x) = E\mathbb{L}\left(\frac{x - X}{h}\right) \tag{73}$$

and with η a Poisson random variable with mean n independent of X, $X_1, \ldots,$

$$c(x) = E\left(\left|\sum_{i=1}^{\eta} \mathbb{L}\left(\frac{x - X_i}{h}\right) - b(x)\right|^p\right). \tag{74}$$

Now H when considered as a function on $\mathbb{R}^n$ can be shown to be Borel measurable, which in this setup is equivalent to being $\mathcal{D}$-measurable.

We shall also need the following fact, which is Lemma 2.3 of Giné, Mason and Zaitsev (2003).

Fact 6. *If, for each $n \in \mathbf{N}$, $\zeta, \zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$, are independent identically distributed random variables, $\zeta_0 = 0$, and η is a Poisson random variable with mean $\gamma > 0$ and independent of the variables $\{\zeta_i\}_{i=1}^{\infty}$ then, for every $p \geq 2$,*

$$E\left|\sum_{i=0}^{\eta} \zeta_i - \gamma E\zeta\right|^p \leq \left(\frac{15p}{\log p}\right)^p \max\left[\left(\gamma E\zeta^2\right)^{p/2}, \gamma E|\zeta|^p\right]. \tag{75}$$

Moreover, specializing to $\zeta \equiv 1$, we have for every $p \geq 2$,

$$E\left|\eta - \gamma\right|^p \leq \left(\frac{15p}{\log p}\right)^p \max\left[\gamma^{p/2}, \gamma\right]. \tag{76}$$

For any $L > 0$ let

$$\mathbb{L}(u) = K(u) 1\{u \in [-L, L]\}$$

and

$$\overline{\mathbb{L}}(u) = K(u) 1\{u \notin [-L, L]\}.$$

We can write

$$\begin{aligned}
f_n(x) - Ef_n(x) &= \frac{1}{nh_n}\sum_{i=1}^{n}\left(K\left(\frac{x-X_i}{h_n}\right) - EK\left(\frac{x-X_i}{h_n}\right)\right) \\
&= \frac{1}{nh_n}\sum_{i=1}^{n}\left(\mathbb{L}\left(\frac{x-X_i}{h_n}\right) - E\mathbb{L}\left(\frac{x-X_i}{h_n}\right)\right) \\
&+ \frac{1}{nh_n}\sum_{i=1}^{n}\left(\overline{\mathbb{L}}\left(\frac{x-X_i}{h_n}\right) - E\overline{\mathbb{L}}\left(\frac{x-X_i}{h_n}\right)\right) \\
&=: \left(\mathbb{L}_n(x) - E\mathbb{L}_n(x)\right) + \left(\overline{\mathbb{L}}_n(x) - E\overline{\mathbb{L}}_n(x)\right).
\end{aligned}$$

We now have the tools to prove statement (50) of Proposition 3.

Step 1. By Proposition 1 for any $p \geq 1$ there is a constant a_p such that

$$\limsup_{n\to\infty}\left((nh_n)^{p/2} E\left\|\overline{\mathbb{L}}_n - E\overline{\mathbb{L}}_n\right\|_p^p\right) \leq a_p \left\|\overline{\mathbb{L}}\right\|_2^p. \tag{77}$$

Observe that the right side of (77) can be made as small as desired by choosing $L > 0$ large enough.

Step 2. Next by Proposition 1 there is a non-negative measurable function φ_p satisfying $E\varphi_p(X) < \infty$ such that for any measurable subset A of $\mathbb{R}$,

$$\limsup_{n\to\infty}\left((nh_n)^{p/2} E\int_A |\mathbb{L}_n(x) - E\mathbb{L}_n(x)|^p\, dx\right) \le \|\mathbb{L}\|_2^p \int_A \varphi_p(x) f(x)\, dx. \tag{78}$$

Step 3. Let η be a Poission random variable with mean n independent of $X_1, X_2, \ldots$ and set

$$\mathbb{L}_{n,\eta}(x) = \frac{1}{nh_n}\sum_{i=1}^{\eta}\mathbb{L}\left(\frac{x - X_i}{h_n}\right). \tag{79}$$

We see that

$$E\mathbb{L}_{n,\eta}(x) = E\mathbb{L}_n(x) = h_n^{-1}E\mathbb{L}\left(\frac{x-X}{h_n}\right), \tag{80}$$

$$n\ \mathrm{Var}\left(\mathbb{L}_{n,\eta}(x)\right) = h_n^{-2}E\mathbb{L}^2\left(\frac{x-X}{h_n}\right) \tag{81}$$

and

$$n\ \mathrm{Var}\left(\mathbb{L}_n(x)\right) = h_n^{-2}E\mathbb{L}^2\left(\frac{x-X}{h_n}\right) - \left\{h_n^{-1}E\mathbb{L}\left(\frac{x-X}{h_n}\right)\right\}^2. \tag{82}$$

Choose any bounded Borel set C satisfying (60) and (61) with $\mathcal{H} = \{\mathbb{L}, \mathbb{L}^2\}$. Clearly for any such set C,

$$\sup_{x\in C}\left|\sqrt{nh_n\ \mathrm{Var}\left(\mathbb{L}_{n,\eta}(x)\right)} - \sqrt{nh_n\ \mathrm{Var}\left(\mathbb{L}_n(x)\right)}\right|$$

$$\le \sup_{x\in C}\frac{h_n\left(\mathbb{L}_{h_n} * f(x)\right)^2}{\sqrt{(\mathbb{L}^2)_{h_n} * f(x)}} = O(h_n) \tag{83}$$

(see (60), (61), (81) and (82)).

Lemma 2. *Whenever* $h_n \to 0$, $nh_n \to \infty$ *and* C *satisfies* (60) *and* (61) *with* $\mathcal{H} = \{\mathbb{L}, \mathbb{L}^2\}$, *we have for* $p \geq 1$,

$$\lim_{n\to\infty} \int_C \left\{ \left(\sqrt{nh_n}\, E|\mathbb{L}_{n,\eta}(x) - E\mathbb{L}_n(x)| \right)^p - \|\mathbb{L}\|_2^p\, E|Z|^p f^{p/2}(x) \right\} dx = 0 \tag{84}$$

and

$$\lim_{n\to\infty} \int_C \left\{ \left(\sqrt{nh_n}\, E|\mathbb{L}_n(x) - E\mathbb{L}_n(x)| \right)^p - \|\mathbb{L}\|_2^p\, E|Z|^p f^{p/2}(x) \right\} dx = 0. \tag{85}$$

Proof. We will first show that

$$\lim_{n\to\infty} \int_C \Big\{ E\left(\sqrt{nh_n}\, |\mathbb{L}_{n,\eta}(x) - E\mathbb{L}_n(x)| \right)^p - E|Z|^p \left(nh_n \operatorname{Var}(\mathbb{L}_{n,\eta}(x)) \right)^{p/2} \Big\} dx = 0 \tag{86}$$

and

$$\lim_{n\to\infty} \int_C \Big\{ E\left(\sqrt{nh_n}\, |\mathbb{L}_n(x) - E\mathbb{L}_n(x)| \right)^p - E|Z|^p \left(nh_n \operatorname{Var} \mathbb{L}_{n,\eta}(x) \right)^{p/2} \Big\} dx = 0. \tag{87}$$

Let η_1 denote a Poisson random variable with mean 1, independent of $X_1, X_2, \ldots$, and set

$$Y_n(x) = \left[\sum_{j \leq \eta_1} \mathbb{L}\left(\frac{x - X_j}{h_n} \right) - E\mathbb{L}\left(\frac{x - X}{h_n} \right) \right] \Big/ \sqrt{E\mathbb{L}^2\left(\frac{x - X}{h_n} \right)}. \tag{88}$$

Now $\operatorname{Var} Y_n(x) = 1$ and it is readily checked using Fact 6 that for some constant $A > 0$ independent of Y_n and x,

$$E|Y_n(x)|^3 \leq A \frac{h_n^{-3/2}\, E\left| \mathbb{L}\left(\frac{x-X}{h_n} \right) \right|^3}{\left(h_n^{-1}\, E\mathbb{L}^2\left(\frac{x-X}{h_n} \right) \right)^{3/2}} \tag{89}$$

and for any integer $r \geq \max(3, p)$,

$$E|Y_n(x)|^r \leq A \frac{h_n^{-r/2}\, E\left|\mathbb{L}\left(\frac{x-X}{h_n}\right)\right|^r}{\left(h_n^{-1}\, E\mathbb{L}^2\left(\frac{x-X}{h_n}\right)\right)^{r/2}}. \tag{90}$$

Using (61) and (60), which says that for some $\delta > 0$, $f(x) \geq \delta > 0$ for all $x \in C$, we get from (89) and (90) that for all large enough n uniformly in $x \in C$ for some constant $B_0 > 0$,

$$n^{-1/2} \sup_{x \in C} E|Y_n(x)|^3 \leq (nh_n)^{-1/2}\, B_0. \tag{91}$$

and

$$n^{-(r-2)/2} \sup_{x \in C} E|Y_n(x)|^r \leq (nh_n)^{-r/2+1}\, B_0. \tag{92}$$

Let $Y_n^{(1)}(x), \ldots, Y_n^{(n)}(x)$ be i.i.d. $Y_n(x)$. Clearly

$$\frac{\sqrt{n}\,\{\mathbb{L}_{n,\eta}(x) - E\mathbb{L}_n(x)\}}{\sqrt{h_n^{-2}\, E\mathbb{L}^2\left(\frac{x-X}{h_n}\right)}} =_d \frac{\sum_{i=1}^{n} Y_n^{(i)}(x)}{\sqrt{n}}. \tag{93}$$

Therefore by (70), we readily conclude that for some constant D for all large enough n,

$$\sup_{x \in C} \left| \frac{E\left|\sqrt{nh_n}\,\{\mathbb{L}_{n,\eta}(x) - E\mathbb{L}_n(x)\}\right|^p}{\left(h_n^{-1}\, E\mathbb{L}^2\left(\frac{x-X}{h_n}\right)\right)^{p/2}} - E|Z|^p \right|$$

$$\leq D\left(n^{-1/2} \sup_{x \in C} E|Y_n(x)|^3 + n^{-(r-2)/2} \sup_{x \in C} E|Y_n(x)|^r\right). \tag{94}$$

Now by (91) and (92) using $r \geq 3$ in combination with (94) and

$$\sup_{x \in C} \left(h_n^{-1}\, E\mathbb{L}^2\left(\frac{x-X}{h_n}\right)\right)^{p/2} = \sup_{x \in C} \left(nh_n\ \mathrm{Var}(\mathbb{L}_{\eta,n}(x))\right)^{p/2} = O(1), \tag{95}$$

we get then that

$$\left|\int_C \left\{E\left(\sqrt{nh_n}\,|\mathbb{L}_{\eta,n}(x)-Ef_{n,\mathbb{L}}(x)|\right)^p - E|Z|^p\left(nh_n \operatorname{Var}(\mathbb{L}_{\eta,n}(x))\right)^{p/2}\right\}dx\right|$$

$$= O\left(\frac{1}{\sqrt{nh_n}}\right).$$

Similarly one obtains using Fact 4 that

$$\left|\int_C \left\{E\left(\sqrt{nh_n}\,|\mathbb{L}_n(x)-Ef_{n,\mathbb{L}}(x)|\right)^p - E|Z|^p\left(nh_n \operatorname{Var}(\mathbb{L}_n(x))\right)^{p/2}\right\}dx\right|$$

$$= O\left(\frac{1}{\sqrt{nh_n}}\right),$$

which by (83) implies

$$\left|\int_C \left\{E\left(\sqrt{nh_n}|\mathbb{L}_n(x)-Ef_{n,\mathbb{L}}(x)|\right)^p - E|Z|^p\left(nh_n \operatorname{Var}(\mathbb{L}_{\eta,n}(x))\right)^{p/2}\right\}dx\right|$$

$$= O\left(\frac{1}{\sqrt{nh_n}} + h_n\right).$$

Recalling (81), we have

$$nh_n \operatorname{Var}\left(\mathbb{L}_{n,\eta}(x)\right) = h_n^{-1} E\mathbb{L}^2\left(\frac{x-X}{h_n}\right) = E\left(\mathbb{L}^2\right)_{h_n}(x). \qquad (96)$$

Clearly (96) in combination with (61), which implies

$$\sup_{z\in C}\left|\left(\mathbb{L}^2\right)_{h_n} * (z) - \|\mathbb{L}\|_2^2\ f(z)\right| \to 0, \text{ as } h \searrow 0,$$

gives

$$\sup_{z\in C}\left|nh_n \operatorname{Var}(\mathbb{L}_{\eta,n}(x)) - \|\mathbb{L}\|_2^2\ f(z)\right| \to 0, \quad \text{as} \quad h \searrow 0.$$

Lemma 2 now follows by the bounded convergence theorem keeping in mind the properties of C. □

We are now ready to prove statement (50) of Proposition 3. For any $0 < \varepsilon < 1$ choose C as in Fact 3. Now (78) gives

$$\limsup_{n\to\infty}\left((nh_n)^{p/2} E\int_{C^c} |\mathbb{L}_n(x) - E\mathbb{L}_n(x)|^p\,dx\right) \le \|\mathbb{L}\|_2^p \int_{C^c} \varphi_p(x)dx. \tag{97}$$

The right side of (97) can be made a small as desired since $P(C^c)$ can be made arbitrarily small. The same is true for the term

$$\|\mathbb{L}\|_2^p\, E|Z|^p \int_{C^c} f^{p/2}(x)\; dx.$$

(Notice that for $1 \le p < 2$ statement (16) holds by inequality (18).) Since ε can be made arbitrarily small, an elementary argument based on (85) now shows that (50) holds.

We next turn to the proof of (51). Choose any bounded Borel set C satisfying (57) through (61) with $\mathcal{H} = \{\mathbb{L}, \mathbb{L}^2, \mathbb{L}^{2p}\}$. Since $C \subset [-M+\nu, M-\nu]$, we get that $C^{Lh} \subset [-M, M]$ for all $h > 0$ small enough. Moreover we can find a measurable partition $C_1, \ldots, C_k$ of C so that $P\left(C_i^{Lh}\right) \le 1/2$ for $i = 1, \ldots, k$ and all $h > 0$ small enough. We now get from Fact 5 with $C = C_i$ and $A = C_i^{Lh}$ using the fact that $\mathbb{L}$ has support contained in $[-L, L]$ that for $i = 1, \ldots, k$,

$$\begin{aligned}
&E\Bigg[\int_{C_i}\left(\sqrt{nh_n}\,|\mathbb{L}_n(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx \\
&\quad - E\int_{C_i}\left(\sqrt{nh_n}\,|\mathbb{L}_{n,\eta}(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx\Bigg]^2 \\
&\le 2\Bigg[E\int_{C_i}\left(\sqrt{nh_n}\,|\mathbb{L}_{n,\eta}(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx \\
&\quad - E\int_{C_i}\left(\sqrt{nh_n}\,|\mathbb{L}_{n,\eta}(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx\Bigg]^2
\end{aligned}$$

$$= 2Var\left(\int_{C_i}\left(\sqrt{nh_n}\,|\mathbb{L}_{n,\eta}(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx\right). \qquad (98)$$

Set

$$S_n(x) = \sum_{i=1}^{\eta}\mathbb{L}\left(\frac{x - X_i}{h_n}\right) - nE\mathbb{L}\left(\frac{x - X_i}{h_n}\right).$$

Observe that

$$\sqrt{nh_n}\,|\mathbb{L}_{n,\eta}(x) - Ef_{n,\mathbb{L}}(x)| = \frac{1}{\sqrt{nh_n}}|S_n(x)|,$$

and, moreover, $S_n(x)$ and $S_n(y)$ are independent if $|x - y| > 2Lh_n$. Thus for each $i = 1, \ldots, k$,

$$\text{Var}\left(\int_{C_i}\left(\sqrt{nh_n}\,|\mathbb{L}_{n,\eta}(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx\right)$$

$$= \frac{1}{(nh_n)^p}\int_{C_i}\int_{C_i} 1\left(|x - y| \le 2Lh_n\right) cov\left(|S_n(x)|^p, |S_n(y)|^p\right) dxdy.$$

Notice

$$|cov\left(|S_n(x)|^p, |S_n(y)|^p\right)| \le \sqrt{E\,|S_n(x)|^{2p}}\sqrt{E\,|S_n(y)|^{2p}}.$$

Furthermore, by Fact 6

$$E\,|S_n(x)|^{2p} \le \left(\frac{30p}{\log(2p)}\right)^p \max\left[\left(nE\zeta_n^2(x)\right)^p, nE|\zeta_n(x)|^{2p}\right],$$

where

$$\zeta_n(x) = \mathbb{L}\left(\frac{x - X}{h_n}\right).$$

Now

$$\frac{1}{h_n}E\zeta_n^2(x) = \frac{1}{h_n}E\mathbb{L}^2\left(\frac{x - X}{h_n}\right)$$

converges uniformly to $f(x)\int_{\mathbb{R}}\mathbb{L}^2(u)\,du$ on C and

$$\frac{1}{h_n}E\zeta_n^{2p}(x) = \frac{1}{h_n}\mathbb{L}^{2p}\left(\frac{x - X}{h_n}\right)$$

converges uniformly on C to $f(x)\int_{\mathbb{R}} \mathbb{L}^{2p}(u)\,du$. Hence uniformly in $x \in C$ for some constant $B_p(C)$,

$$E|S_n(x)|^{2p} \le B_p(C)\left((nh_n)^p + nh_n\right).$$

Thus uniformly in $i = 1, \dots, k$, for all large enough n using $nh_n \to \infty$,

$$\frac{1}{(nh_n)^p}\int_{C_i}\int_{C_i} 1\left(|x-y| \le 2Lh_n\right) cov\left(|S_n(x)|^p, |S_n(y)|^p\right) dxdy$$

$$\le 2B_p(C)\int_C\int_C 1\left(|x-y| \le 2Lh_n\right) dxdy,$$

which since C is a bounded Lebesgue measurable set is for some constant $D_p(C, L)$

$$\le h_n D_p(C, L).$$

Hence by the above string of inequalities, we see that for each $i = 1, \dots, k$, as $n \to \infty$,

$$Var\left(\int_{C_i}\left(\sqrt{nh_n}\,|\mathbb{L}_{n,\eta}(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx\right) \to 0,$$

from which we infer from (98) that as $n \to \infty$,

$$\int_{C_i}\left(\sqrt{nh_n}|\mathbb{L}_n(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx - E\int_{C_i}\left(\sqrt{nh_n}|\mathbb{L}_{n,\eta}(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx$$

$$\to_P 0.$$

This, in turn, when coupled with (84), (85) and $C_1, \dots, C_k$ being a partition of C implies

$$\int_C\left(\sqrt{nh_n}\,|\mathbb{L}_n(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx \to_P \|\mathbb{L}\|_2^p\, E|Z|^p \int_C f^{p/2}(x)\,dx. \tag{99}$$

Now $P(C^c)$ can be made arbitrarily small and thus, keeping (97) in mind, we see that

$$E\int_{C^c}\left(\sqrt{nh_n}\,|\mathbb{L}_n(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx + \|\mathbb{L}\|_2^p\, E|Z|^p \int_{C^c} f^{p/2}(x)\,dx$$

can be made as small as desired for all large enough n. This observation when combined with (85) allows us to conclude that

$$\int_{\mathbb{R}} \left(\sqrt{nh_n}\,|\mathbb{L}_n(x) - Ef_{n,\mathbb{L}}(x)|\right)^p dx \to_P \|\mathbb{L}\|_2^p\, E|Z|^p \int_{\mathbb{R}} f^{p/2}(x)\, dx.$$

(The necessary argument is given in the proof of Theorem 4.2 of Billingsley (1968).) Recall from Step 1 that

$$\limsup_{n\to\infty} \left((nh_n)^{p/2}\, E\left\|\overline{\mathbb{L}}_n - E\overline{\mathbb{L}}_n\right\|_p^p\right) \le a_p \left\|\overline{\mathbb{L}}\right\|_2^p,$$

which can be made as small as desired by choosing $L > 0$ large enough. In the same way, we can make

$$\left\|\overline{\mathbb{L}}\right\|_2^p E|Z|^p \int_{\mathbb{R}} f^{p/2}(x)\, dx,$$

arbitrarily small. Therefore by a standard argument we conclude that (51) holds. (As above, the argument is in the proof of Theorem 4.2 on Billingsley (1968).) This completes the proof of Proposition 3. □

3. Connections to known risk bounds

In this section we discuss the connections of our results to some known risk bounds of Ibragimov and Hasminskii (1980), Hasminskii and Ibragimov (1990) and Bretagnolle and Huber (1979), and then we use our results to provide a partial solution to a conjecture of Guerre and Tsybakov (1998).

3.1. Connection to results of Ibragimov and Hasminskii

Introduce the following analog of the de la Vallée-Poussin kernel,

$$K_{vp}(x) = \frac{\cos x - \cos 2x}{\pi x^2}, \quad x \in \mathbb{R}. \tag{100}$$

Let f_n be the density estimator based on this kernel. Ibragimov and Hasminskii (1980) and Hasminskii and Ibragimov (1990) have obtained bounds for any $r \ge 1$ and $p \ge 1$ for $E\|f_n - Ef_n\|_p^r$ and $E\|f_n - f\|_p^r$. Their bounds for $E\|f_n - Ef_n\|_p^r$ are of the same order as those given in our Proposition 2 when $p \ge 2$. Refer to Lemma 4 of Ibragimov and Hasminskii (1980).

Specialize now to the following smooth class of densities: for a given $\eta > 0$, $\beta = r + \alpha$, $r = 0, 1, \ldots, 0 < \alpha < 1$, $L > 0$ and $p \geq 1$, let denote $H_p^\beta L$ denote the class of functions $g \in L_p(\mathbb{R})$ with derivatives up to order r and satisfy

$$\left(\int_{\mathbb{R}} \left| g^{(r)}(x+h) - g^{(r)}(x) \right|^p dx \right)^{1/p} \leq L |h|^\alpha, |h| \leq \eta,$$

In the case $\alpha = 1$, the slightly stronger condition is needed, namely,

$$\left(\int_{\mathbb{R}} \left| g^{(r)}(x+h) + g^{(r)}(x-h) - 2g^{(r)}(x) \right|^p dx \right) \leq L |h|, \ |h| \leq \eta.$$

Lemmas 3 and 5 in Ibragimov and Hasminskii (1980) imply that for any density $f \in H_p^\beta L$ and the kernel density estimator f_n based on the kernel (100),

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_{vp}\left(\frac{x - X_i}{h_n} \right),$$

we have

$$\|Ef_n - f\|_p \leq A |h|^\beta. \tag{101}$$

Choose for some $c > 0$ the sequence of bandwidths

$$h_n = \frac{c}{n^{1/(2\beta+1)}}.$$

It follows from (101), Minkowski's inequality and our Corollary 1 that for any $p \geq 2$, $M > 0$ and loss function w satisfying its conditions that

$$\limsup_{n\to\infty} \sup_{f \in H_p^\beta L \cap \mathcal{L}_{p/2}\left((M)^{(p-2)/(p-1)}\right)} Ew\left(n^{\beta/(2\beta+1)} \|f_n - f\|_p \right) < \infty,$$

which since $\|f\|_{p/2} \leq \|f\|_p^{(p-2)/(p-1)}$ implies

$$\limsup_{n\to\infty} \sup_{f \in H_p^\beta L \cap \mathcal{L}_p(M)} Ew\left(n^{\beta/(2\beta+1)} \|f_n - f\|_p \right) < \infty. \tag{102}$$

Statement (102) was proved in Theorem 5 of Ibragimov and Hasminskii (1980).

Ibragimov and Hasminskii (1980) establish in their Theorem 5 that if w is also non-decreasing then for some $\gamma > 0$

$$\liminf_{n\to\infty} \inf_{\widetilde{f}_n} \sup_{f\in H_p^\beta L\cap\mathcal{L}_p(M)} Ew\left(\gamma n^{\beta/(2\beta+1)} \left\|\widetilde{f}_n - f\right\|_p\right) > 0, \tag{103}$$

where $\widetilde{f}_n$ is an arbitrary density estimator of f based on $X_1, \ldots, X_n$. Hasminskii and Ibragimov (1990) show that the rate is different when $1 \le p < 2$. More precisely they prove that for any $r \ge 1$

$$\liminf_{n\to\infty} \inf_{\widetilde{f}_n} \sup_{f\in H_p^\beta L\cap\mathcal{L}_p(M)} n^{\beta r/(q\beta+1)} E\left\|\widetilde{f}_n - f\right\|_p^r > 0, \tag{104}$$

where $q = p/(p-1)$. (For the case $p = 1$ see also Theorem 11 of Ibragimov and Hasminskii (1980).)

3.2. Connection to results of Bretagnolle and Huber

There are similar connections to the work of Bretagnolle and Huber (1979). We shall restrict our discussion to the case $1 \le p < 2$. For $r > 0$, $\beta > 0$ and a positive integer m let

$$F_{r,m,\beta,a} = \{f : \rho_{m,p}(f) \le r\} \subset \mathcal{C}_{[-a,a]}^{(m)}(\beta),$$

where

$$\rho_{m,p}(f) = \left\|f^{(m)}\right\|_p^{p/(2m+1)} \|f\|_{p/2}^{2p/(2m+1)}$$

and $\mathcal{C}_{[-a,a]}^{(m)}(\beta)$ denotes the class of m times continuously differentiable densities f with support contained in $[-a, a]$, $0 < a < \infty$, and satisfying $\left\|f^{(m)}\right\|_p \le \beta$. Applying the construction in their Proposition 3.1 we can find a class $\Theta \subset F_{r,m,\beta,a}$ so that for some $C > 0$,

$$\liminf_{n\to\infty} n^{2p/(2m+1)} \inf_{\widetilde{f}_n} \sup_{f\in\Theta} E\left\|\widetilde{f}_n - f\right\|_p^p \ge Cr, \tag{105}$$

which implies that

$$\liminf_{n\to\infty} n^{2p/(2m+1)} \inf_{\widetilde{f}_n} \sup_{f\in F_{r,m,\beta,a}} E\left\|\widetilde{f}_n - f\right\|_p^p \ge Cr. \tag{106}$$

(We assume that $\beta > 0$ is chosen so that their construction works.) Furthermore, as on page 131 of Bretagnolle and Huber (1979) we can infer that for a suitable positive constant D and for a well chosen sequence of estimators $\widetilde{f}_n$,

$$\limsup_{n\to\infty} n^{2p/(2m+1)} \sup_{f\in F_{r,m,\beta,a}} E\left\|\widetilde{f}_n - f\right\|_p^p \le Dr. \tag{107}$$

Notice that statement (106) implies that for any convex continuous nondecreasing loss function w on $[0,\infty)$

$$\liminf_{n\to\infty} \inf_{\widetilde{f}_n} \sup_{f\in F_{r,m,\beta,a}} Ew\left(n^{2p/(2m+1)}\left\|\widetilde{f}_n - f\right\|_p^p\right) \ge w\left(Cr\right). \tag{108}$$

Now choose any kernel K satisfying $(K.i)$ and (55) and having support contained in $[-b,b]$, $0 < b < \infty$. Applying Lemma 1 with $H = K^2/\left\|K\right\|_2^2$ via inequality (26) we get

$$\int_{\mathbb{R}}\left(\left(K^2\right)_{h_n} * f(x)\right)^{p/2} dx \le \|K\|_2^p\, C(\lambda,s)\left(1+c_\lambda\left(E|X|^\lambda + h^\lambda E|Y|^\lambda\right)\right)^{p/2},$$

where

$$E|X|^\lambda = \int_{\mathbb{R}} |x|^\lambda f(x)\,dx \text{ and } E|Y|^\lambda = \frac{1}{\|K\|_2^2}\int_{\mathbb{R}} |y|^\lambda K^2(y)\,dy.$$

Obviously $E|X|^\lambda \le a^\lambda$ and $E|Y|^\lambda \le b^\lambda$. Thus uniformly in $0 < h \le 1$ and $f \in F_{r,m,\beta,a}$,

$$\left(\int_{\mathbb{R}}\left(\left(K^2\right)_{h_n} * f(x)\right)^{p/2} dx\right)^{2/p} \le \|K\|_2^2\, C(\lambda,s)^{2/p}\left(1+c_\lambda\left(a^\lambda + b^\lambda\right)\right) =: M. \tag{109}$$

Let f_n be a kernel density estimator with kernel K and a sequence of bandwidths satisfying (h). On account of (109) we can apply Corollary 1 to infer that for any such kernel and all $t > 0$,

$$\limsup_{n\to\infty} \sup_{f\in F_{r,m,\beta,a}} E\exp\left(t\sqrt{nh_n}\,\|f_n - Ef_n\|_p\right) < \infty. \tag{110}$$

Assume in addition that K satisfies $(K.iii)$ then for any $f \in F_{r,m,\beta,a}$,

$$\|f - Ef_n\|_p \le h_n^m \beta \, \|_m K\|_1 \,. \tag{111}$$

Now choose for some $c > 0$ the sequence of bandwidths

$$h_n = \frac{c}{n^{1/(2m+1)}}.$$

We see then by (111), Minkowski's inequality and (110) that for all $t > 0$,

$$\limsup_{n\to\infty} \sup_{f\in F_{r,m,\beta,a}} E \exp\left(tn^{m/(2m+1)} \|f_n - f\|_p\right) < \infty. \tag{112}$$

Therefore for any loss function w on $[0,\infty)$ such that for some $\lambda > 0$ and $C > 0$,

$$0 \le w(x^p) \le C \exp(\lambda x), \; x \in [0,\infty), \tag{113}$$

we have

$$\limsup_{n\to\infty} \sup_{f\in F_{r,m,\beta,a}} Ew\left(n^{pm/(2m+1)} \|f_n - f\|_p^p\right) < \infty. \tag{114}$$

3.3. Partial solution to a conjecture of Guerre and Tsybakov

We shall conclude our paper by showing that our results lead to a partial solution to a conjecture of Guerre and Tsybakov (1998). Let $\mathcal{F}(\gamma, L)$ be a class of density functions such that each $f \in \mathcal{F}(\gamma, L)$ admits an analytical continuation to the strip $D_\gamma = \{x + iy : |y| \le \gamma\}$ with $\gamma > 0$ such that $f(x+iy)$ is analytic on the interior of D_γ, bounded on D_γ and for some $L > 0$

$$\int_{\mathbb{R}} |f(x+iy)|^2 \, dx \le L.$$

Clearly, since f is bounded on D_γ, for all $r \ge 1$,

$$\int_{\mathbb{R}} f^r(x) \, dx < \infty. \tag{115}$$

Let

$$\widehat{f}(t) = \int_{\mathbb{R}} \exp(itx) f(x) \, dx,$$

denote the Fourier transform of f. We have for any $f \in \mathcal{F}(\gamma, L)$

$$\frac{1}{2\pi} \int_{\mathbb{R}} \cosh^2(\gamma t) \left| \widehat{f}(t) \right|^2 dt \leq L.$$

(For this and other facts about the class $\mathcal{F}(\gamma, L)$ consult Guerre and Tsybakov (1998) and the references therein.) Introduce the sinc kernel

$$S(t) = \frac{\sin t}{\pi t}, \quad t \in \mathbb{R},$$

were $\frac{\sin 0}{0} := 1$. The kernel S is not in $L_1(\mathbb{R})$ but it is in $L_2(\mathbb{R})$. It satisfies

$$\lim_{A \to \infty} \int_{-A}^{A} S(t)\, dt = 1 \text{ and } \int_{\mathbb{R}} S^2(t)\, dt = \frac{1}{\pi},$$

and its Fourier transform is

$$\widehat{S}(t) = 1\{|t| \leq 1\}.$$

Choose the sequence of bandwidths $h_n = 2\gamma / \log n$ and consider the kernel density estimator

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} S\left(\frac{x - X_i}{h_n}\right).$$

Part of the conjecture in Remark 5 of Guerre and Tsybakov (1998) surmises that for a general class of loss functions w for each $p \geq 2$ and $f \in \mathcal{F}(\gamma, L)$ one has

$$\lim_{n \to \infty} Ew \left(\frac{\sqrt{\frac{2\pi\gamma n}{\log n}} \|f_n - f\|_p}{\left(E|Z|^p \int_{\mathbb{R}} f^{p/2}(y)\, dy\right)^{1/p}} \right) = w(1). \tag{116}$$

Note that since (115) holds for any $f \in \mathcal{F}(\gamma, L)$ and S fulfills $(K.i)$ and $(K.ii)$, our Corollary 2 implies that for any loss function w continuous at 1 and satisfying (53), $p \geq 2$ and $f \in \mathcal{F}(\gamma, L)$,

$$\lim_{n \to \infty} Ew \left(\frac{\sqrt{\frac{2\pi\gamma n}{\log n}} \|f_n - Ef_n\|_p}{\left(E|Z|^p \int_{\mathbb{R}} f^{p/2}(y)\, dy\right)^{1/p}} \right) = w(1). \tag{117}$$

To infer from (117) that (116) also holds it suffices to verify that

$$\sqrt{\frac{n}{\log n}}\,\|f - Ef_n\|_p \to 0. \tag{118}$$

For any $h > 0$ write $S_h(\cdot) = h^{-1} S\left(\cdot h^{-1}\right)$. The bias at each fixed $x \in \mathbb{R}$ based on the kernel S and bandwidth $h > 0$ is

$$b_h(x) = f(x) - S_h * f(x).$$

The argument in the proof of Lemma 1 of Guerre and Tsybakov (1998) shows that the essential supremum of $|b_h|$ satisfies

$$\text{ess sup } |b_h| \le \sqrt{\frac{L}{2\pi\gamma}} \exp(-\gamma/h). \tag{119}$$

Now by Plancherel's theorem

$$\int_{\mathbb{R}} b_h^2(x)\,dx = \int_{\mathbb{R}} (f(x) - S_h * f(x))^2\,dx$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} \left|\widehat{f}(t) - \widehat{S}(ht)\,\widehat{f}(t)\right|^2 dt = \frac{1}{2\pi}\int_{\mathbb{R}} \left|\widehat{f}(t)\,1\{|t| \ge 1/h\}\right|^2 dt$$

$$\le \frac{1}{2\pi}\int_{\mathbb{R}} \cosh^2(\gamma t)\left|\widehat{f}(t)\right|^2 dt \int_{1/h}^{\infty} \frac{1}{\cosh^2(\gamma t)}dt$$

$$\le \frac{L}{\pi}\int_{1/h}^{\infty} \exp(-2\gamma t)\,dt = \frac{L}{2\pi\gamma}\exp(-2\gamma/h). \tag{120}$$

Therefore we see from (119) that (120) that for any $h > 0$ and $p \ge 2$

$$\left(\int_{\mathbb{R}} b_h^p(x)\,dx\right)^{1/p} \le \left((\text{ess sup } |b_h|)^{p-2}\int_{\mathbb{R}} b_h^2(x)\,dx\right)^{1/p}$$

$$\le \sqrt{\frac{L}{2\pi\gamma}}\exp(-\gamma/h). \tag{121}$$

Clearly from inequality (121) we get

$$\sqrt{\frac{n}{\log n}}\,\|f-Ef_n\|_p \le \left(\frac{L}{2\pi\gamma}\right)^{1/2}\frac{1}{\sqrt{\log n}}, \tag{122}$$

which implies (118) and thus (116) for loss functions w as in Corollary 2.

We do not know whether the limit in (116) is uniform for $f \in \mathcal{F}(\gamma, L)$, or even for f in the subclass $L_{p/2}(M) \cap \mathcal{F}(\gamma, L)$, for any $M > 0$. On the other hand, since (122) holds uniformly for $f \in \mathcal{F}(\gamma, L)$, it is easy to combine this with Corollary 1 to show that for any $M > 0$, $2 \le p < \infty$ and loss function w satisfying (53)

$$\limsup_{n\to\infty}\ \sup_{f\in L_{p/2}(M)\cap\mathcal{F}(\gamma,L)} Ew\left(\sqrt{nh_n}\,\|f_n-f\|_p\right) < \infty.$$

Guerre and Tsybakov (1998) also conjectured in their Remark 5 that a minimax result of the form

$$\inf_{\widetilde{f}_n}\sup_{f\in\mathcal{F}} Ew\left(\frac{\sqrt{\frac{2\pi\gamma n}{\log n}}\left\|\widetilde{f}_n-f\right\|_p}{\left(E\,|Z|^p\int_{\mathbb{R}} f^{p/2}(y)\,dy\right)^{1/p}}\right)\to w(1)$$

for a general class of loss functions w, where the infimum is taken over all estimators $\widetilde{f}_n$ of f based on $X_1, \dots, X_n$, i.i.d. with density f and the supremum is over a subclass $\mathcal{F}$ of $\mathcal{F}(\gamma, L)$ with $p \ge 2$. The only result in this direction known to the author is Theorem 2 of Schipper (1996), which says in the case $p = 2$ that

$$\inf_{\widetilde{f}_n}\ \sup_{f\in\mathcal{F}(\gamma,L)} \frac{2\pi\gamma n}{\log n} E\left\|\widetilde{f}_n-f\right\|_2^2 \to 1.$$

Notice that a special case of (116) says that

$$\lim_{n\to\infty}\frac{2\pi\gamma n}{\log n}E\,\|f_n-f\|_2^2 \to 1.$$

For closely related results refer to Theorems 7 and 10 of Ibragimov and Hashminskii (1980).

Dedication and acknowledgement
The author dedicates this paper to Ildar Ibragimov on the occasion of his seventy fifth birthday and to his receipt of the Diploma of *Doctor Honoris Causa* of the Université Victor Segalen Bordeaux 2.

Also the author thanks Boris Levit for introducing him to this area and the remarkable accomplishments of Ibragimov and Hasminskii, as well. He posed the problems that resulted in this paper. He also pointed out the relationship between the author's results and those of Ibragimov and Hasminskii (1980). In addition the author is grateful to Luc Devroye for a number helpful remarks and references to the literature.

References

1. P. Billingsley, *Convergence of Probability Measures.* Wiley, New York (1968).
2. J. Bretagnolle, C. Huber, *Estimation des densités: risque minimax.* — Z. Wahrsch. Verw. Gebiete, **47** (1979), 119–137.
3. L. Devroye, *On arbitrarily slow rates of global convergence in density estimation.* — Z. Wahrsch. Verw. Gebiete, **62** (1983), 475–483.
4. L. Devroye, *A course in density estimation.* Progress in Probability and Statistics, 14. Birkhäuser Boston, Inc., Boston (1987).
5. L. Devroye, *A note on the usefulness of superkernels in density estimation.*— Ann. Statist., **20** (1992), 2037–2056.
6. L. Devroye, L. Györfi, *Nonparametric Density Estimation: The L_1 View.* Wiley, New York (1985).
7. S. Yu. Efroimovich, M. S. Pinsker, *Estimation of square-integrable probability density of a random variable.* — Problems Inform. Transmission, **18** (1982), 175–189.
8. G. B. Folland, *Real Analysis.* Wiley, New York (1984).
9. E. Giné, A. Zaitsev, D. M. Mason, *The L_1- norm density estimator process.* — Ann. Probab., **31** (2003), 719–768.
10. E. Guerre, A. Tsybakov, *Exact asymptotic minimax constants for the estimation of analytical functions in L_p.* — Probab. Theory Related Fields, **112** (1998), 33–51.
11. R. Hasminskii, I. Ibragimov, *On density estimation in the view of Kolmogorov's ideas in approximation theory.* — Ann. Statist., **18** (1990), 999–1010.
12. I. Ibragimov, R. Hasminskii, *On estimation of the density function.* — Zap. Nauch. Sem. LOMI AN SSSR, **98** (1980), 61–85.
13. I. Ibragimov, R. Hasminskii, *An estimate of the density of a distribution belonging to a class of entire functions.* — Teor. Veroyatnost. i Primenen., **27** (1982), 514–524.
14. W. B. Johnson, G. Schechtman, J. Zinn, *Best constants in moment inequalities for linear combinations of independent and exchangeable random variables.* — Ann. Probab., **13** (1985), 234–253.
15. M. Schipper, *Optimal rates and constants in L_2-minimax estimation of probability density functions.* — Math. Methods Statist., **5** (1996), 253–274.
16. G. R. Shorack, J. A. Wellner, *Empirical Processes with Applications to Statistics.* John Wiley & Sons, Inc., New York (1986).

17. E. M. Stein, *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, Princeton, New Jersey (1970).
18. T. J. Sweeting, *Speeds of convergence in the multidimensional central limit theorem.* — Ann. Probab., **5** (1977), 28–41.
19. M. Talagrand, *Isoperimetry and integrability of the sum of independent Banach-space valued random variables.* — Ann. Probab., **17** (1989) 1546–1570.

Statistics Program,
University of Delaware,
Newark, Delaware 19717.

E-mail: davidm@udel.edu

Поступило 1 июня 2008 г.