DOI: 10.1214/EJP.v11-354

Electronic Journal of Probability

Vol. 11 (2006), Paper no. 33, pages 844–859.

Journal URL
http://www.math.washington.edu/~ejpecp/

Weighted uniform consistency of kernel density estimators with general bandwidth sequences

Dony, Julia* and Einmahl, Uwe†
Department of Mathematics
Free University of Brussels (VUB)
Pleinlaan 2
B-1050 Brussels, Belgium
e-mail: jdony@vub.ac.be, ueinmahl@vub.ac.be

Abstract

Let $f_{n,h}$ be a kernel density estimator of a continuous and bounded d-dimensional density f. Let $\psi(t)$ be a positive continuous function such that $\|\psi f^{\beta}\|_{\infty} < \infty$ for some $0 < \beta < 1/2$. We are interested in the rate of consistency of such estimators with respect to the weighted sup-norm determined by ψ. This problem has been considered by Giné, Koltchinskii and Zinn (2004) for a deterministic bandwidth h_n. We provide "uniform in h" versions of some of their results, allowing us to determine the corresponding rates of consistency for kernel density estimators where the bandwidth sequences may depend on the data and/or the location.

Key words: kernel density estimator, weighted uniform consistency, convergence rates, uniform in bandwidth, empirical process

AMS 2000 Subject Classification: Primary 60B12, 60F15, 62G07.

Submitted to EJP on April 24 2006, final version accepted September 8 2006.

*Research supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen)

†Research partially supported by an FWO-Vlaanderen Grant

1 Introduction

Let $X, X_1, X_2, \dots$ be i.i.d. $\mathbb{R}^d$-valued random vectors and assume that the common distribution of these random vectors has a bounded Lebesgue density function, which we shall denote by f. A kernel K will be any measurable positive function which satisfies the following conditions:

$$(K.i) \qquad \int_{\mathbb{R}^d} K(s)ds = 1,$$

$$(K.ii) \qquad \|K\|_\infty := \sup_{x \in \mathbb{R}^d} |K(x)| = \kappa < \infty.$$

The kernel density estimator of f based upon the sample $X_1, \dots, X_n$ and bandwidth $0 < h < 1$ is defined as follows,

$$f_{n,h}(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - t}{h^{1/d}}\right), \quad t \in \mathbb{R}^d.$$

Choosing a suitable bandwidth sequence $h_n \to 0$ and assuming that the density f is continuous, one obtains a strongly consistent estimator $\hat{f}_n := f_{n,h_n}$ of f, i.e. one has with probability 1, $\hat{f}_n(t) \to f(t), t \in \mathbb{R}^d$. There are also results concerning uniform convergence and convergence rates. For proving such results one usually writes the difference $\hat{f}_n(t) - f(t)$ as the sum of a probabilistic term $\hat{f}_n(x) - \mathbb{E}\hat{f}_n(t)$ and a deterministic term $\mathbb{E}\hat{f}_n(t) - f(t)$, the so-called bias. The order of the bias depends on smoothness properties of f only, whereas the first (random) term can be studied via empirical process techniques as has been pointed out by Stute and Pollard (see [11, 12, 13, 10]), among other authors.

After the work of Talagrand [14], who established optimal exponential inequalities for empirical processes, there has been some renewed interest in these problems. Einmahl and Mason [3] looked at a large class of kernel type estimators including density and regression function estimators and determined the precise order of uniform convergence of the probabilistic term over compact subsets. Giné and Guillou [5] (see also Deheuvels [1]) showed that if K is a "regular" kernel, the density function f is bounded and h_n satisfies among others the regularity conditions

$$\frac{\log(1/h_n)}{\log\log n} \longrightarrow \infty \quad \text{and} \quad \frac{nh_n}{\log n} \longrightarrow \infty,$$

one has with probability 1,

$$\|\hat{f}_n - \mathbb{E}\hat{f}_n\|_\infty = O\left(\sqrt{\frac{|\log h_n|}{nh_n}}\right). \tag{1}$$

Moreover, this rate cannot be improved.

Recently, Giné, Koltchinskii and Zinn (see [8]) obtained refinements of these results by establishing the same convergence rate for density estimators with respect to weighted sup-norms.

Under additional assumptions on the bandwidth sequence and the density function, they provided necessary and sufficient conditions for stochastic and almost sure boundedness for the quantity

$$\sqrt{\frac{nh_n}{|\log h_n|}} \sup_{t \in \mathbb{R}^d} |\psi(t)\{\hat{f}_n(t) - \mathbb{E}\hat{f}_n(t)\}|.$$

Results of this type can be very useful when estimating integral functionals of the density f (see for example Mason [9]). Suppose for instance that we want to estimate $\int_{\mathbb{R}^d} \phi(f(t))dt < \infty$ where $\phi : \mathbb{R} \to \mathbb{R}$ is a measurable function. Then a possible estimator would be given by $\int_{\mathbb{R}^d} \phi(f_{n,h}(t))dt$. Assuming that ϕ is Lipschitz and that $\int_{\mathbb{R}^d} f^{\beta}(t)dt =: c_\beta < \infty$ for some $0 < \beta < 1/2$, one can conclude that for some constant $D > 0$,

$$\left| \int_{\mathbb{R}^d} \phi(f_{n,h}(t))dt - \int_{\mathbb{R}^d} \phi(\mathbb{E}f_{n,h}(t))dt \right| \le Dc_\beta \sup_{t \in \mathbb{R}^d} |f^{-\beta}(t)\{f_{n,h}(t) - \mathbb{E}f_{n,h}(t)\}|,$$

and we see that this term is of order $\sqrt{|\log h|/nh}$. For some further related results, see also Giné, Koltchinskii and Sakhanenko [6, 7].

In practical applications the statistician has to look at the bias as well. It is well known that if one chooses small bandwidth sequences, the bias will be small whereas the probabilistic term which is of order $O(\sqrt{|\log h_n|/nh_n})$, might be too large. On the other hand, choosing a large bandwidth sequence will increase the bias. So the statistician has to balance both terms and typically, one obtains bandwidth sequences which depend on some quantity involving the unknown distribution. Replacing this quantity by a suitable estimator, one ends up with a bandwidth sequence depending on the data $X_1, \ldots, X_n$ and, in some cases, also on the location x. There are many elaborate schemes available in the statistical literature for finding such bandwidth sequences. We refer the interested reader to the article by Deheuvels and Mason [2] (especially Sections 2.3 and 2.4) and the references therein. Unfortunately, one can no longer investigate the behavior of such estimators via the aforementioned results, since they are dealing with density estimators based on deterministic bandwidth sequences.

To overcome this difficulty, Einmahl and Mason [4] introduced a method allowing them to obtain "uniform in h" versions of some of their earlier results as well as of (1). These results are immediately applicable for proving uniform consistency of kernel–type estimators when the bandwidth h is a function of the location x or the data $X_1, \ldots, X_n$.

It is natural then to ask whether one can also obtain such "uniform in h" versions of some of the results by Giné, Koltchinskii and Zinn [8]. We will answer this in the affirmative by using a method which is based on a combination of some of their ideas with those of Einmahl and Mason [4].

In order to formulate our results, let us first specify what we mean by a "regular" kernel K. First of all, we will assume throughout that K is compactly supported. Rescaling K if necessary, we can assume that its support is contained in $[-1/2, 1/2]^d$. Next consider the class of functions

$$\mathcal{K} = \left\{ K((\cdot - t)/h^{1/d}) : h > 0, t \in \mathbb{R}^d \right\}.$$

For $\epsilon > 0$, let $\mathcal{N}(\epsilon, \mathcal{K}) = \sup_Q \mathcal{N}(\kappa\epsilon, \mathcal{K}, d_Q)$, where the supremum is taken over all probability measures Q on $(\mathbb{R}^d, \mathcal{B})$, d_Q is the $L_2(Q)$-metric and, as usual, $\mathcal{N}(\epsilon, \mathcal{K}, d_Q)$ is the minimal number of balls $\{g : d_Q(g, g') < \epsilon\}$ of d_Q -radius ϵ needed to cover $\mathcal{K}$. We assume that $\mathcal{K}$ satisfies the following uniform entropy condition:

$$(K.iii) \qquad \text{for some } C > 0 \text{ and } \nu > 0: \quad \mathcal{N}(\epsilon, \mathcal{K}) \leq C\epsilon^{-\nu}, \quad 0 < \epsilon < 1.$$

Van der Vaart and Wellner [15] provide a number of sufficient conditions for $(K.iii)$ to hold. For instance, it is satisfied for general $d \geq 1$, whenever $K(x) = \phi(p(x))$, with $p(x)$ being a polynomial in d variables and ϕ a real valued function of bounded variation. Refer also to condition (K) in [8].

Finally, to avoid using outer probability measures in all of our statements, we impose the following measurability assumption:

$$(K.iv) \qquad \mathcal{K} \text{ is a pointwise measurable class.}$$

With "pointwise measurable", we mean that there exists a countable subclass $\mathcal{K}_0 \subset \mathcal{K}$ such that we can find for any function $g \in \mathcal{K}$ a sequence of functions $g_m \in \mathcal{K}_0$ for which $g_m(z) \to g(z), z \in \mathbb{R}^d$. This condition is discussed in van der Vaart and Wellner [15] and in particular it is satisfied whenever K is right continuous. The following assumptions were introduced by Giné, Koltchinskii and Zinn [8]. Note that we need slightly less regularity since we will not determine the precise limiting constant or limiting distribution. In the following we will denote the sup-norm on $\mathbb{R}^d$ by $|\cdot|$.

Assumptions on the density. Let $B_f := \{t \in \mathbb{R}^d : f(t) > 0\}$ be the positivity set of f, and assume that B_f is open and that the density f is bounded and continuous on B_f. Further, assume that

$$(D.i) \qquad \forall\, \delta > 0,\ \exists\, h_0 > 0 \text{ and } 0 < c < \infty \text{ such that } \forall\, x, x + y \in B_f,$$
$$c^{-1} f^{1+\delta}(x) \leq f(x + y) \leq c f^{1-\delta}(x), \qquad |y| \leq h_0,$$

$$(D.ii) \qquad \forall\, r > 0, \text{ set } F_r(h) := \{(x, y) : x + y \in B_f, f(x) \geq h^r, |y| \leq h\}, \text{ then}$$
$$\lim_{h\to 0} \sup_{(x,y)\in F_r(h)} \left| \frac{f(x + y)}{f(x)} - 1 \right| = 0.$$

Assumptions on the weight function ψ.

$$(W.i) \qquad \psi : B_f \to \mathbb{R}^+ \text{ is positive and continuous,}$$

$$(W.ii) \qquad \forall\, \delta > 0,\ \exists\, h_0 > 0 \text{ and } 0 < c < \infty \text{ such that } \forall\, x, x + y \in B_f \text{ and}$$
$$c^{-1} \psi^{1-\delta}(x) \leq \psi(x + y) \leq c\psi^{1+\delta}(x), \qquad |y| \leq h_0,$$

$$(W.iii) \qquad \forall\, r > 0, \text{ set } G_r(h) := \{(x, y) : x + y \in B_f, \psi(x) \leq h^{-r}, |y| \leq h\}, \text{ then}$$
$$\lim_{h\to 0} \sup_{(x,y)\in G_r(h)} \left| \frac{\psi(x + y)}{\psi(x)} - 1 \right| = 0.$$

Extra assumptions. For $0 < \beta < 1/2$, assume that

$$(WD.i) \qquad \|f^{\beta}\psi\|_{\infty} = \sup_{t \in B_f} |f^{\beta}(t)\psi(t)| < \infty,$$

$$(WD.ii) \qquad \forall\, r > 0, \quad \lim_{h \to 0} \sup_{(x,y) \in G_r(h)} \left| \frac{f(x+y)}{f(x)} - 1 \right| = 0.$$

A possible choice for the weight function would be $\psi = f^{-\beta}$ in which case the last assumptions follow from the corresponding one involving the density. For some discussion of these conditions and examples, see page 2573 of Giné, Koltchinskii and Zinn [8].

Now, consider two decreasing functions

$$a_t := a(t) = t^{-\alpha} L_1(t) \quad \text{and} \quad b_t := b(t) := t^{-\mu} L_2(t), \quad t > 0,$$

where $0 < \mu < \alpha < 1$ and L_1, L_2 are slowly varying functions. Further define the functions

$$\begin{aligned} \lambda(t) &:= \sqrt{t a_t |\log a_t|}, \quad t > 0, \\ \lambda_n(h) &:= \sqrt{nh|\log h|}, \quad n \geq 1,\ a_n \leq h \leq b_n, \end{aligned}$$

and it is easy to see that the function λ is regularly varying at infinity with positive exponent $0 < \eta := \frac{1-\theta}{2} < 1/2$ for some $0 < \theta < 1$. Finally, we assume that $\lambda(t)$ is strictly increasing ($t > 0$).

Theorem 1.1. *Assume that the above hypotheses are satisfied for some $0 < \beta < 1/2$, and that we additionally have*

$$\limsup_{t \to \infty} t \mathbb{P}\left\{\psi(X) > \lambda(t)\right\} < \infty. \tag{2}$$

Then it follows that

$$\Delta_n := \sup_{a_n \leq h \leq b_n} \sqrt{\frac{nh}{|\log h|}}\, \|\psi(f_{n,h} - \mathbb{E} f_{n,h})\|_{\infty}$$

is stochastically bounded.

Note that if we choose $a_n = b_n = h_n$ we re-obtain the first part of Theorem 2.1 in Giné, Koltchinskii and Zinn [8]. They have shown that assumption (2) is necessary for this part of their result if $B_f = \mathbb{R}^d$ or $K(0) = \kappa$. Therefore this assumption is also necessary for our Theorem 1.1.

Remark. Choosing the estimator f_{n,h_n} where $h_n \equiv H_n(X_1, \ldots, X_n; x) \in [a_n, b_n]$ is a general bandwidth sequence (possibly depending on x and the observations $X_1, \ldots, X_n$) one obtains that

$$\|\psi(f_{n,h_n} - \mathbb{E} f_{n,h_n})\|_{\infty} = O_{\mathbb{P}}(\sqrt{|\log a_n|/n a_n}). \tag{3}$$

Indeed, due to the monotonicity of the function $h \to nh/|\log h|, 0 < h < 1$ we can infer from the stochastic boundedness of Δ_n that for all $\epsilon > 0$ and large enough n, there is a finite constant C_ϵ such that

$$\mathbb{P}\left\{ \sup_{a_n \le h \le b_n} \|\psi(f_{n,h} - \mathbb{E}f_{n,h})\|_\infty > C_\epsilon \sqrt{\frac{|\log a_n|}{n a_n}} \right\} \le \epsilon,$$

which in turn trivially implies (3). Note that this is exactly the same stochastic order as for the estimator f_{n,a_n} where one uses the deterministic bandwidth sequence a_n.

Theorem 1.2. *Assume that the above hypotheses are satisfied for some $0 < \beta < 1/2$, and that we additionally have*

$$\int_1^\infty \mathbb{P}\{\psi(X) > \lambda(t)\}\, dt < \infty. \tag{4}$$

Then we have with probability one,

$$\limsup_{n\to\infty} \sup_{a_n \le h \le b_n} \sqrt{\frac{nh}{|\log h|}} \|\psi(f_{n,h} - \mathbb{E}f_{n,h})\|_\infty \le C, \tag{5}$$

where C is a finite constant.

Remark. If we consider the special case $a_n = b_n$, and if we use the deterministic bandwidth sequence $h_n = a_n$, we obtain from the almost sure finiteness of Δ_n that for the kernel density estimator $\hat{f}_n = f_{n,h_n}$, with probability one,

$$\limsup_{n\to\infty} \frac{\|\psi(\hat{f}_n - \mathbb{E}\hat{f}_n)\|_\infty}{\sqrt{nh_n/|\log h_n|}} \le C < \infty.$$

Moreover we can apply Proposition 2.6 of Giné, Koltchinskii and Zinn [8], and hence the latter implies assumption (4) to be necessary for (5) if $B_f = \mathbb{R}^d$ or $K(0) > 0$.
Furthermore, with the same reasoning as in the previous remark following the stochastic boundedness result, Theorem 1.2 applied to density estimators f_{n,h_n} with general (stochastic) bandwidth sequences $h_n \equiv H_n(X_1, \ldots, X_n; x) \in [a_n, b_n]$ leads to the same almost sure order $O(\sqrt{|\log a_n|/na_n})$ as the one one would obtain by choosing a deterministic bandwidth sequence $h_n = a_n$.

We shall prove Theorem 1.1 in Section 2 and the proof of Theorem 1.2 will be given in Section 3. In both cases we will bound Δ_n by a sum of several terms and we show already in Section 2 that most of these terms are almost surely bounded. To do that, we have to bound certain binomial probabilities, and use an empirical process representation of kernel estimators. So essentially, there will be only one term left for which we still have to prove almost sure boundedness, which will require the stronger assumption (4) in Theorem 1.2.

2 Proof of Theorem 1.1

Throughout this whole section we will assume that the general assumptions specified in Section 1 as well as condition (2) are satisfied. Moreover, we will assume without loss of generality that

$\|f^\beta \psi\|_\infty \le 1$.

Recall that we have for any $t \in B_f$ and $a_n \le h \le b_n$,

$$\begin{aligned}&\sqrt{\frac{nh}{|\log h|}}\psi(t)\{f_{n,h}(t) - \mathbb{E}f_{n,h}(t)\}\\ &\qquad = \frac{\psi(t)}{\lambda_n(h)}\sum_{i=1}^n K\left(\frac{X_i - t}{h^{1/d}}\right) - \frac{n\psi(t)}{\lambda_n(h)}\mathbb{E}K\left(\frac{X-t}{h^{1/d}}\right). \qquad (6)\end{aligned}$$

We first show that the last term with the expectation can be ignored for certain t's. To that end we need the following lemma.

Lemma 2.1. *For $a_n \le h \le b_n$ and for large enough n, we have for all $t \in B_f$,*

$$\frac{n\psi(t)}{\lambda_n(h)}\mathbb{E}K\left(\frac{X-t}{h^{1/d}}\right) \le \gamma_n + 2\kappa\sqrt{\frac{nh}{|\log h|}}f(t)\psi(t),$$

where $\gamma_n \to 0$.

Proof. For any $r > 0$, we can split the centering term as follows in two parts:

$$\begin{aligned}\frac{n\psi(t)}{\lambda_n(h)}\mathbb{E}K\left(\frac{X-t}{h^{1/d}}\right) &= \frac{nh\psi(t)}{\lambda_n(h)}\int_{[-1/2,1/2]^d} K(u)f(t+uh^{1/d})\,du\\ &\le \frac{\kappa nh\psi(t)}{\lambda_n(h)}\sup_{\substack{|u|\le 1/2\\ t+uh^{1/d}\in B_f}} f(t+uh^{1/d})I_{\{f(t)\le h^r\}}\\ &\quad + \frac{\kappa nh\psi(t)}{\lambda_n(h)}\sup_{\substack{|u|\le 1/2\\ t+uh^{1/d}\in B_f}} f(t+uh^{1/d})I_{\{f(t)> h^r\}}\\ &=: \gamma_n(t,h) + \xi_n(t,h).\end{aligned}$$

Now take $0 < \delta < 1-\beta$ and choose $\tau > 0$ such that

$$\sup_{a_n\le h\le b_n}\frac{h^{\tau(1-\beta-\delta)}}{(nh)^{-1}\lambda_n(h)} \longrightarrow 0. \qquad (7)$$

Note that such a $\tau > 0$ exists, since the denominator does not converge faster to zero than a negative power of n, as does $h \in [a_n, b_n]$. We now study both terms $\xi_n(t,h)$ and $\gamma_n(t,h)$ for the choice $r = \tau$. For $\delta > 0$ chosen as above, there are $h_0 > 0, c < \infty$ such that for $x, x+y \in B_f$ with $|y| \le h_0$,

$$c^{-1}f^{1+\delta}(x) \le f(x+y) \le cf^{1-\delta}(x). \qquad (8)$$

Moreover, for the choice of $\tau > 0$ we obtain by condition $(D.ii)$ that for all h small enough and $x \in B_f$ with $f(x) \ge h^\tau$,

$$f(x+y) \le 2f(x), \quad |y| \le h^{1/d}. \qquad (9)$$

Therefore, in view of (9) and recalling the definition of $\lambda_n(h)$, we get for $t \in \mathbb{R}^d$ that

$$\xi_n(t,h) \le 2\kappa\sqrt{\frac{nh}{|\log h|}}f(t)\psi(t). \qquad (10)$$

Finally, using condition $(WD.i)$ in combination with (7) and (8), it's easy to show that

$$\sup_{t\in\mathbb{R}^d}\sup_{a_n\le h\le b_n}\gamma_n(t,h)=:\gamma_n\longrightarrow 0,$$

finishing the proof of the lemma. □

To simplify notation we set

$$\Delta_n:=\sup_{a_n\le h\le b_n}\sqrt{\frac{nh}{|\log h|}}\,\|\psi(f_{n,h}-\mathbb{E}f_{n,h})\|_\infty\,,$$

and set for any function $g:\mathbb{R}^d\to\mathbb{R}$ and $C\subset\mathbb{R}^d$, $\|g\|_C:=\sup_{t\in C}|g(t)|$. We start by showing that choosing a suitable $r>0$ it will be sufficient to consider the above supremum only over the region

$$A_n:=\{t\in B_f:\psi(t)\le b_n^{-r}\}\subset\mathbb{R}^d. \tag{11}$$

Lemma 2.2. *There exists an $r>0$ such that with probability one,*

$$\sup_{a_n\le h\le b_n}\sqrt{\frac{nh}{|\log h|}}\|\psi(f_{n,h}-\mathbb{E}f_{n,h})\|_{\mathbb{R}^d\setminus A_n}\longrightarrow 0.$$

Proof. Choose $r>0$ sufficiently large so that, eventually, $b_n^r\le n^{-2}$. Note that $\psi(t)>b_n^{-r}$ implies that $f(t)\le b_n^{r/\beta}$, and consequently we get that $f(t)\psi(t)\le f(t)^{1-\beta}\le b_n^{r(1/\beta-1)}$, such that for $\beta<1/2$ this last term is bounded above by n^{-2} for large n. Recalling Lemma 2.1 we can conclude that

$$\sup_{a_n\le h\le b_n}\sqrt{\frac{nh}{|\log h|}}\|\psi\mathbb{E}f_{n,h}\|_{\mathbb{R}^d\setminus A_n}\longrightarrow 0,$$

and it remains to be shown that with probability one,

$$Y_n:=\sup_{a_n\le h\le b_n}\sqrt{\frac{nh}{|\log h|}}\|\psi f_{n,h}\|_{\mathbb{R}^d\setminus A_n}\longrightarrow 0.$$

It is obvious that

$$\mathbb{P}\{Y_n\ne 0\}\le\sum_{i=1}^n\mathbb{P}\{d(X_i,A_n^c)\le b_n\},$$

where as usual $d(x,A)=\inf_{y\in A}|x-y|, x\in\mathbb{R}^d$. Then, since $\psi(s)>b_n^{-r}$ implies by $(W.ii)$ that $\psi(t)\ge c^{-1}b_n^{-r(1-\delta)}$ for n large enough, $|s-t|\le b_n$ and $\delta>0$, due to our choice of r, it is possible to find a small $\delta>0$ such that, eventually, $\psi(t)\ge\lambda(n^3)$. Hence, it follows using (2) that

$$\mathbb{P}\{Y_n\ne 0\}\le n\mathbb{P}\{\psi(X)\ge\lambda(n^3)\}=O(n^{-2}),$$

which via Borel-Cantelli implies that with probability one, $Y_n=0$ eventually. □

We now study the remaining part of the process Δ_n, that is

$$\Delta_n' := \sup_{a_n \le h \le b_n} \sqrt{\frac{nh}{|\log h|}} \left\| \psi(f_{n,h} - \mathbb{E} f_{n,h}) \right\|_{A_n}.$$

We will handle the uniformity in bandwidth over the region A_n by considering smaller intervals $[h_{n,j}, h_{n,j+1}]$, where we set

$$h_{n,j} := 2^j a_n, \quad n \ge 1, \ j \ge 0.$$

The following lemma shows that a finite number of such intervals is enough to cover $[a_n, b_n]$.

Lemma 2.3. *If $l_n := \max\{j : h_{n,j} \le 2b_n\}$, then for n large enough, $l_n \le 2\log n$ and $[a_n, b_n] \subset [h_{n,0}, h_{n,l_n}]$.*

Proof. Suppose $l_n > 2\log n$, then there is a $j_0 > 2\log n$ such that $h_{n,j_0} \le 2b_n$, and hence this j_0 satisfies $4^{\log n} n^{-\alpha} L_1(n) < h_{n,j_0} \le 2n^{-\mu} L_2(n)$. Consequently, we must have $n \le 2n^{\alpha-\mu} L_2(n)/L_1(n)$, which for large n is impossible given that L_2/L_1 is slowly varying at infinity. The second part of the lemma follows immediately after noticing that $h_{n,0} = a_n$ and $b_n \le h_{n,l_n}$. □

For each $j \ge 0$, split A_n into the regions

$$A^1_{n,j} := \left\{ t \in A_n \ : \ f(t)\psi(t) \le \epsilon_n^{1-\beta} \sqrt{\frac{|\log h_{n,j+1}|}{nh_{n,j+1}}} \right\},$$

$$A^2_{n,j} := \left\{ t \in A_n : 0 < \psi(t) \le \epsilon_n^{-\beta} \left(\frac{nh_{n,j+1}}{|\log h_{n,j+1}|} \right)^{\beta/2(1-\beta)} \right\},$$

where we take $\epsilon_n = (\log n)^{-1}, n \ge 2$. Note that if $f\psi > L$, by condition $(WD.i)$, $\psi \le L^{-\beta/(1-\beta)}$, implying that for all $j \ge 0$, the union of $A^1_{n,j}$ and $A^2_{n,j}$ equals A_n. With (6) in mind, set for $0 \le j \le l_n - 1$ and $i = 1, 2$

$$\Delta^{(i)}_{n,j} := \sup_{h_{n,j} \le h \le h_{n,j+1}} \sqrt{\frac{nh}{|\log h|}} \left\| \psi(f_{n,h} - \mathbb{E} f_{n,h}) \right\|_{A^i_{n,j}},$$

$$\Phi^{(i)}_{n,j} := \sup_{t \in A^i_{n,j}} \sup_{h_{n,j} \le h \le h_{n,j+1}} \frac{\psi(t)}{\lambda_n(h)} \sum_{i=1}^n K\left(\frac{X_i - t}{h^{1/d}} \right),$$

$$\Psi^{(i)}_{n,j} := \sup_{t \in A^i_{n,j}} \sup_{h_{n,j} \le h \le h_{n,j+1}} \frac{n\psi(t)}{\lambda_n(h)} \mathbb{E} K\left(\frac{X - t}{h^{1/d}} \right).$$

In particular, we have

$$\Delta^{(i)}_{n,j} \le \Phi^{(i)}_{n,j} + \Psi^{(i)}_{n,j}, \quad i = 1, 2,$$

and from Lemma 2.1 and the definition of $A^1_{n,j}$, it follows that we can ignore the centering term $\Psi^{(1)}_{n,j}$. Hence, we get that

$$\Delta_n' \le \left(\delta_n + \max_{0 \le j \le l_n - 1} \Phi^{(1)}_{n,j} \right) \vee \max_{0 \le j \le l_n - 1} \Delta^{(2)}_{n,j}, \tag{12}$$

with $\delta_n \to 0$, and we will prove stochastic boundedness of Δ'_n by showing it for both $\max_{0\le j\le l_n-1}\Phi^{(1)}_{n,j}$ and $\max_{0\le j\le l_n-1}\Delta^{(2)}_{n,j}$. Therefore, set

$$\lambda_{n,j} := \lambda_n(h_{n,j}) = \sqrt{2^j}\sqrt{na_n|\log 2^j a_n|}, \quad j\ge 0,$$

and note that $\lambda_{n,j} \ge \lambda(n2^j)$. Let's start with the first term, $\Phi^{(1)}_{n,j}$. We clearly have for $0\le j\le l_n-1$ that

$$\Phi^{(1)}_{n,j} \le \kappa \sup_{t\in A^1_{n,j}} \frac{\psi(t)}{\lambda_{n,j}}\sum_{i=1}^n I\{|X_i-t|\le h_{n,j}^{1/d}\} =: \kappa\Lambda_{n,j}.$$

For $k=1,\dots,n$, set $B_{n,j,k} := A^1_{n,j}\cap\{t : |X_k-t|\le h_{n,j}^{1/d}\}$, then it easily follows that

$$\Lambda_{n,j} = \max_{1\le k\le n}\sup_{t\in B_{n,j,k}} \frac{\psi(t)}{\lambda_{n,j}}\sum_{i=1}^n I\{|X_i-t|\le h_{n,j}^{1/d}\}.$$

Recall from (11) that $\psi(t)\le b_n^{-r}\le h_{n,j}^{-r}$ on A_n for $0\le j\le l_n-1$. Then it follows from conditions $(W.iii)$ and $(WD.ii)$ that there is a ρ small such that $(1-\rho)\psi(t)\le\psi(s)\le(1+\rho)\psi(t)$ and $f(s)\le(1+\rho)f(t)$ if $|s-t|\le h_{n,j}^{1/d}$. In this way we obtain for $t\in A^1_{n,j}, |s-t|\le h_{n,j}^{1/d}$ and large enough n that for a positive constant $C_1>1$,

$$\psi(t)\le C_1\psi(s) \quad\text{and}\quad f(s)\psi(s)\le C_1\epsilon_n^{1-\beta}\sqrt{\frac{|\log h_{n,j+1}|}{nh_{n,j+1}}}.$$

Hence, we can conclude that

$$\Lambda_{n,j}\le C_1\max_{1\le k\le n}\frac{\psi(X_k)}{\lambda_{n,j}}\sum_{i=1}^n I\{|X_i-X_k|\le 2h_{n,j}^{1/d}\}I\{X_k\in\tilde A^1_{n,j}\}, \tag{13}$$

where $\tilde A^1_{n,j} := \{t : f(t)\psi(t)\le C_1\epsilon_n^{1-\beta}\sqrt{|\log h_{n,j+1}|/nh_{n,j+1}}\}$, and it follows that

$$\max_{0\le j\le l_n-1}\Lambda_{n,j}\le C_1\max_{1\le k\le n}\frac{\psi(X_k)}{\lambda(n)} + C_1\max_{0\le j\le l_n-1}\max_{1\le k\le n}\frac{\psi(X_k)}{\lambda_{n,j}}M_{n,j,k}I\{X_k\in\tilde A^1_{n,j}\}, \tag{14}$$

where $M_{n,j,k} := \sum_{i=1}^n I\{|X_i-X_k|\le 2h_{n,j}^{1/d}\}-1$. Note that the first term is stochastically bounded by assumption (2). Thus in order to show that $\max_{0\le j\le l_n-1}\Phi^{(1)}_{n,j}$ is stochastically bounded, it is enough to show that this is also the case for the second term in (14). As a matter of fact, it follows from the following lemma that this term converges to zero in probability.

Lemma 2.4. *We have for $1\le k\le n$ and $\epsilon>0$,*

$$\max_{0\le j\le l_n-1}\mathbb{P}\{\psi(X_k)M_{n,j,k}I\{X_k\in\tilde A^1_{n,j}\}\ge\epsilon\lambda_{n,j}\} = O(n^{-1-\eta}),$$

where $\eta>0$ is a constant depending on α and β only.

Proof. Given $X_k = t$, $M_{n,j,k}$ has a Binomial$(n-1, \pi_{n,j}(t))$ distribution, where $\pi_{n,j}(t) := \mathbb{P}\{|X-t| \le 2h_{n,j}^{1/d}\}$. Furthermore, since for large enough n, $\psi(t) \le C_1 b_n^{-r} \le b_n^{-r-1}$ on A_n, it follows for $c > 1$ and large n that $f(s)/f(t) \le c, |s-t| \le b_n^{1/d}$, so that

$$\pi_{n,j}(t) \le 4^d c h_{n,j} f(t).$$

Using the fact that the moment-generating function $\mathbb{E}\exp(sZ)$ of a Binomial(n,p)-variable Z is bounded above by $\exp(npe^s)$, we can conclude that for $t \in \tilde{A}^1_{n,j}$ and any $s > 0$,

$$\begin{aligned}
p_{n,j}(t) &:= \mathbb{P}\left\{\psi(X_k)M_{n,j,k} \ge \epsilon\lambda_{n,j} \| X_k = t\right\} \\
&\le \exp\left(c4^d n h_{n,j} f(t) e^s - \frac{\epsilon s \lambda_{n,j}}{\psi(t)}\right) \\
&\le \exp\left(\frac{\lambda_{n,j}}{\psi(t)}(C_2\epsilon_n^{1-\beta}e^s - \epsilon s)\right), \quad s > 0,\ t \in \tilde{A}^1_{n,j}.
\end{aligned}$$

Choosing $s = \log(1/\epsilon_n)/2 = \log\log n/2$, we obtain for some n_0 (which is independent of j) that

$$p_{n,j}(t) \le \exp\left(-\frac{\epsilon\lambda_{n,j}\log\log n}{3\psi(t)}\right), \quad n \ge n_0,\ t \in \tilde{A}^1_{n,j}.$$

Setting $\tilde{B}_{n,j} := \{t \in \tilde{A}^1_{n,j} : \psi(t) \le \lambda_{n,j}/\log n\}$, it's obvious that for any $\tilde{\eta} > 0$,

$$\max_{0 \le j \le l_n - 1} \sup_{t \in \tilde{B}_{n,j}} p_{n,j}(t) = O(n^{-\tilde{\eta}}). \tag{15}$$

Next, set $\tilde{C}_{n,j} := \tilde{A}^1_{n,j} \backslash \tilde{B}_{n,j} = \{t \in \tilde{A}^1_{n,j} : \lambda_{n,j}/\log n < \psi(t)\}$, then using once more the fact that $\psi \le f^{-\beta}$, we have that $\psi f \le (\log n/\lambda_{n,j})^{1+\theta}$ on this set, where $\theta = \beta^{-1} - 2 > 0$. By Markov's inequality, we then have for $t \in \tilde{C}_{n,j}$,

$$\begin{aligned}
p_{n,j}(t) &\le 4^d c\epsilon^{-1} n h_{n,j} f(t)\psi(t)/\lambda_{n,j} \\
&\le 4^d c\epsilon^{-1}(\log n)^{1+\theta}\lambda_{n,j}^{-\theta}/|\log h_{n,j}| \\
&\le 4^d c'\epsilon^{-1}\left(\frac{\log n}{na_n}\right)^{\theta/2}, \quad t \in \tilde{C}_{n,j}.
\end{aligned} \tag{16}$$

Further, note that by regular variation, $\lambda_{n,j}/\log n \ge \lambda_{[n(\log n)^{-\gamma}],j}$ for some $\gamma > 0$. Therefore, we have from (2) that

$$\mathbb{P}\{\psi(X_k) \ge \lambda_{n,j}/\log n\} = O\left((\log n)^\gamma/n\right), \quad k = 1, \ldots, n.$$

Combining this with (15) and (16), we find that

$$\begin{aligned}
&\max_{0 \le j \le l_n - 1} \mathbb{P}\{\psi(X_k)M_{n,j,k} I\{X_k \in \tilde{A}^1_{n,j}\} \ge \epsilon\lambda_{n,j}\} \\
=\ & \max_{0 \le j \le l_n - 1} \left\{\int_{\tilde{B}_{n,j}} p_{n,j}(t)f(t)dt + \int_{\tilde{C}_{n,j}} p_{n,j}(t)f(t)dt\right\} \\
\le\ & O(n^{-\tilde{\eta}}) + O\left((\log n/na_n)^{\theta/2}\right)\mathbb{P}\{\psi(X) \ge \lambda_{n,j}/\log n\} \\
=\ & O\left(n^{-1-\frac{\theta}{2}(1-\alpha)}(\log n)^{\gamma+\frac{\theta}{2}}L_1(n)^{-\frac{\theta}{2}}\right) \\
\le\ & O(n^{-1-\frac{\theta}{3}(1-\alpha)}),
\end{aligned}$$

proving the lemma. $\square$

It is now clear that $\max_{0\le j\le l_n-1}\Phi^{(1)}_{n,j}$ is stochastically bounded under condition (2), and it remains to be shown that this is also the case for $\max_{0\le j\le l_n-1}\Delta^{(2)}_{n,j}$.

Let α_n be the empirical process based on the i.i.d sample $X_1,\ldots,X_n$. Then we have for any measurable bounded function $g:\mathbb{R}^d\to\mathbb{R}$,

$$\alpha_n(g):=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(g(X_i)-\mathbb{E}g(X_1)\right).$$

For $0\le j\le l_n-1$, consider the following class of functions defined by

$$\mathcal{G}_{n,j}:=\left\{\psi(t)K\left(\frac{\cdot-t}{h^{1/d}}\right)\ :\ t\in A^2_{n,j},h_{n,j}\le h\le h_{n,j+1}\right\},$$

then obviously,

$$\left\|\sqrt{n}\alpha_n\right\|_{\mathcal{G}_{n,j}}\ge\lambda_{n,j}\Delta^{(2)}_{n,j},$$

where as usual $\|\sqrt{n}\alpha_n\|_{\mathcal{G}_{n,j}}=\sup_{g\in\mathcal{G}_{n,j}}|\sqrt{n}\alpha_n(g)|$. To show stochastic boundedness of $\Delta^{(2)}_{n,j}$, we will use a standard technique for empirical processes, based on a useful exponential inequality of Talagrand [14], in combination with an appropriate upper bound of the moment quantity $\mathbb{E}\left\|\sum_{i=1}^n\varepsilon_i g(X_i)\right\|_{\mathcal{G}_{n,j}}$, where $\varepsilon_1,\ldots,\varepsilon_n$ are independent Rademacher random variables, independent of $X_1,\ldots,X_n$.

Lemma 2.5. *For each $j=0,\ldots,l_n-1$, the class $\mathcal{G}_{n,j}$ is a VC-class of functions with envelope function*

$$G_{n,j}:=\kappa\epsilon_n^{-\beta}\left(\frac{nh_{n,j+1}}{|\log h_{n,j+1}|}\right)^{\beta/2(1-\beta)}$$

that satisfies the uniform entropy condition

$$\mathcal{N}\left(\epsilon,\mathcal{G}_{n,j}\right)\le C\epsilon^{-\nu-1},\quad 0<\epsilon<1,$$

where C and ν are positive constants (independent of n and j).

Proof. Consider the classes

$$\begin{aligned}\mathcal{F}_{n,j}&=\left\{\psi(t)\ :\ t\in A^2_{n,j}\right\},\\ \mathcal{K}_{n,j}&=\left\{K\left(\frac{\cdot-t}{h^{1/d}}\right)\ :\ t\in A^2_{n,j},\ h_{n,j}\le h\le h_{n,j+1}\right\},\end{aligned}$$

with envelope functions $F_{n,j}:=\epsilon_n^{-\beta}\left(\frac{nh_{n,j+1}}{|\log h_{n,j+1}|}\right)^{\beta/(2(1-\beta)}$ and κ respectively. Then $\mathcal{G}_{n,j}\subset\mathcal{F}_{n,j}\mathcal{K}_{n,j}$ and it follows from our assumptions on K that $\mathcal{K}_{n,j}$ is a VC-class of functions. Furthermore, it is easy to see that the covering number of $\mathcal{F}_{n,j}$, which we consider as a class of constant functions, can be bounded above as follows :

$$\mathcal{N}\left(\epsilon\sqrt{Q(F^2_{n,j})},\mathcal{F}_{n,j},d_Q\right)\le C_1\epsilon^{-1},\quad 0<\epsilon<1.$$

Since $\mathcal{K}_{n,j}$ is a VC-class, we have for some positive constants ν and $C_2 < \infty$ that

$$\mathcal{N}\left(\epsilon\kappa, \mathcal{K}_{n,j}, d_Q\right) \le C_2 \epsilon^{-\nu}.$$

Thus, the conditions of lemma A1 in Einmahl and Mason [3] are satisfied, and we obtain the following uniform entropy bound for $\mathcal{G}_{n,j}$:

$$\mathcal{N}\left(\epsilon, \mathcal{G}_{n,j}\right) \le C\epsilon^{-\nu-1}, \quad 0 < \epsilon < 1,$$

proving the lemma. $\square$

Now, observe that for all $t \in A^2_{n,j} \subset A_n$ and $h_{n,j} \le h \le h_{n,j+1}$, we have by condition $(W.iii)$ for large n,

$$\begin{aligned} \mathbb{E}\left[\psi^2(t) K^2\left(\frac{X-t}{h^{1/d}}\right)\right] &\le 2\mathbb{E}\left[\psi^2(X) K^2\left(\frac{X-t}{h^{1/d}}\right)\right] \\ &= 2\int_{\mathbb{R}^d} \psi^2(x) f(x) K^2((x-t)/h^{1/d}) dx. \end{aligned}$$

Recalling that $\left\|\psi f^{\beta}\right\|_\infty \le 1$, we see that this integral is bounded above by

$$2h_{n,j+1}\ \|f\|_\infty^{1-2\beta}\, \|K\|_2^2 =: C_\beta h_{n,j+1}.$$

As the exponent $\beta/2(1-\beta)$ in the definition of $G_{n,j}$ is strictly smaller than $1/2$, it is easily checked that by choosing the β in Proposition A.1 of Einmahl and Mason [3] to be equal to $G_{n,j}$, and $\sigma^2_{n,j} = C_\beta h_{n,j+1}$, there exists an $n_0 \ge 1$ so that the assumptions of Proposition A.1 in Einmahl and Mason [3] are satisfied for all $0 \le j \le l_n - 1$ and $n \ge n_0$. Therefore, we can conclude that

$$\mathbb{E}\|\sum_{i=1}^n \varepsilon_i g(X_i)\|_{\mathcal{G}_{n,j}} \le C'\sqrt{n h_{n,j} \log n}, \quad n \ge n_0,\ 0 \le j \le l_n - 1,$$

where C' is a positive constant depending on α, β, ν and C only (where the β is again the one from condition $(WD.i)$). Moreover, as for $0 \le j \le l_n - 1$ we have $|\log h_{n,j}| \ge |\log b_n| \sim \mu \log n$, we see that for some $n_1 \ge n_0$,

$$\mathbb{E}\|\sum_{i=1}^n \varepsilon_i g(X_i)\|_{\mathcal{G}_{n,j}} \le C''\lambda_{n,j}, \quad 0 \le j \le l_n - 1. \tag{17}$$

Recalling that $\Delta^{(2)}_{n,j} \le \|\sum_{i=1}^n \varepsilon_i g(X_i)\|_{\mathcal{G}_{n,j}}/\lambda_{n,j}$ it follows from Markov's inequality that the variables $\Delta^{(2)}_{n,j}$ are stochastically bounded for all $0 \le j \le l_n - 1$. However, to prove that the maximum of these variables is stochastically bounded too, we need to use more sophisticated tools. One of them is the inequality of Talagrand [14] mentioned above. (For a suitable version, refer to Inequality A.1 in [3].) Employing this inequality, we get that

$$\mathbb{P}\left\{\max_{1\le m\le n}\left\|\sqrt{m}\alpha_m\right\|_{\mathcal{G}_{n,j}} \ge A_1\left(\mathbb{E}\|\sum_{i=1}^n \varepsilon_i g(X_i)\|_{\mathcal{G}_{n,j}} + x\right)\right\}$$

$$\leq\ 2\left[\exp\left(-\frac{A_2x^2}{n\sigma^2_{n,j}}\right)+\exp\left(-\frac{A_2x}{G_{n,j}}\right)\right],$$

where A_1, A_2 are universal constants. Next, recall that $\sigma^2_{n,j} = 2C_\beta h_{n,j}$ and that $G_{n,j} \leq c\epsilon_n^{-\beta}\sqrt{nh_{n,j}/|\log h_{n,j}|}$, then choosing $x = \rho\lambda_{n,j}$ ($\rho > 1$), we can conclude from the foregoing inequality and (17) that for large n,

$$\begin{aligned}
\mathbb{P}&\left\{\|\sqrt{n}\alpha_n\|_{\mathcal{G}_{n,j}} \geq A_1(C''+\rho)\lambda_{n,j}\right\}\\
&\leq\ 2\left[\exp\left(-\frac{A_2\rho^2}{2C_\beta}\frac{\lambda^2_{n,j}}{nh_{n,j}}\right)+\exp\left(-A_2\rho\frac{\lambda_{n,j}}{G_{n,j}}\right)\right]\\
&\leq\ 4\exp\left(-\frac{A_2\rho^2}{2C_\beta}|\log h_{n,j}|\right),
\end{aligned}\tag{18}$$

where we used the fact that $\inf_{0\leq j\leq l_n-1}\lambda_{n,j}/(G_{n,j}|\log h_{n,j}|) \to \infty$ as $n \nearrow \infty$. Finally, since $\|\sqrt{n}\alpha_n\|_{\mathcal{G}_{n,j}} \geq \lambda_{n,j}\Delta^{(2)}_{n,j}$, we just showed that

$$\mathbb{P}\left\{\max_{0\leq j<l_n}\Delta^{(2)}_{n,j} \geq M\right\} \leq \sum_{j=0}^{l_n-1}\mathbb{P}\left\{\|\sqrt{n}\alpha_n\|_{\mathcal{G}_{n,j}} \geq \lambda_{n,j}M\right\} \leq 4n^{-2},\tag{19}$$

provided we choose $M \geq A_1(C''+\sqrt{5\mu C_\beta/A_2})$ and n is large enough. It's now obvious that $\max_{0\leq j\leq l_n-1}\Delta^{(2)}_{n,j}$ is stochastically bounded, which, in combination with (14) and the result in lemma 2.4 proves Theorem 1.1. □

3 Proof of Theorem 1.2

In view of Lemma 2.2 it is sufficient to prove that under assumption (4), we have with probability one that

$$\limsup_{n\to\infty}\Delta'_n \leq M',$$

for a suitable positive constant $M' > 0$. Recalling relation (12), we only need to show that for suitable positive constants M'_1, M'_2,

$$\limsup_{n\to\infty}\max_{0\leq j\leq l_n-1}\Phi^{(1)}_{n,j} \leq M'_1, \quad a.s,\tag{20}$$

and

$$\limsup_{n\to\infty}\max_{0\leq j\leq l_n-1}\Delta^{(2)}_{n,j} \leq M'_2, \quad a.s.\tag{21}$$

The result in (21) follows easily from (19) and the Borel-Cantelli lemma, and as is shown below, it turns out that (20) holds with $M'_1 = 0$, i.e this term goes to zero. Recall now from (14) that

$$\max_{0\leq j\leq l_n-1}\Phi^{(1)}_{n,j} \leq C_1\kappa\max_{1\leq k\leq n}\frac{\psi(X_k)}{\lambda(n)}$$

$$+ \; C_1\kappa \max_{0\le j\le l_n-1} \max_{1\le k\le n} \frac{\psi(X_k)}{\lambda_{n,j}} M_{n,j,k} I\{X_k \in \widetilde{A}^1_{n,j}\},$$

where $M_{n,j,k} = \sum_{i=1}^n I\{|X_i - X_k| \le 2h_{n,j}^{1/d}\} - 1$. From condition (4) and the assumption on a_n we easily get that with probability one, $\psi(X_k)/\lambda(n) \to 0$, and consequently we also have that $\max_{1\le k\le n} \psi(X_k)/\lambda(n) \to 0$, finishing the study of the first term. To simplify notation, set

$$Z_n := \max_{0\le j\le l_n-1} \max_{1\le k\le n} \frac{\psi(X_k)}{\lambda_{n,j}} M_{n,j,k} I\{X_k \in \widetilde{A}^1_{n,j}\},$$

take $n_k = 2^k, k \ge 1$, and set $h'_{k,j} := h_{n_k,j}$ and $l'_k := l_{n_{k+1}}$. Then note that

$$\max_{n_k\le n\le n_{k+1}} Z_n \le \max_{0\le j< l'_k} \max_{1\le i\le n_{k+1}} \frac{\psi(X_i)}{\lambda_{n_k,j}} M'_{k,j,i} I\{X_i \in A'_{k,j}\},$$

where $M'_{k,j,i} = \sum_{m=1}^{n_{k+1}} I\{|X_m - X_i| \le 2h'^{1/d}_{k,j}\} - 1$ and $A'_{k,j} = \{t : f(t)\psi(t) \le C_1\epsilon_{n_k}^{1-\beta} \sqrt{|\log h'_{k,j}|/n_k h'_{k,j}}\}$, and after some minor modifications, we obtain similarly to Lemma 2.4 that for $\epsilon > 0$,

$$\mathbb{P}\left\{\max_{n_k\le n\le n_{k+1}} Z_n \ge \epsilon\right\} = O\left(l'_k n_k^{-\eta'}\right), \quad \eta' > 0,$$

which implies again via Borel-Cantelli that $Z_n \to 0$ almost surely, proving (20) with $M'_1 = 0$. □

Acknowledgements. The authors thank the referee for a careful reading of the manuscript. Thanks are also due to David Mason for some useful suggestions.

References

[1] Deheuvels, P. (2000). Uniform limit laws for kernel density estimators on possibly unbounded intervals. *Recent advances in reliability theory (Bordeaux, 2000),* 477–492, Stat. Ind. Technol., Birkhäuser Boston, MA. MR1783500

[2] Deheuvels, P. and Mason, D.M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.* **7**(3), 225–277. MR2111291

[3] Einmahl, U. and Mason, D.M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13**(1), 1–37. MR1744994

[4] Einmahl, U. and Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33**(3), 1380–1403. MR2195639

[5] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.* **38**(6), 907–921.

[6] Giné, E., Koltchinskii, V. and Sakhanenko, L. (2003). Convergence in distribution of self-normalized sup-norms of kernel density estimators. *High dimensional probability, III (Sandjberg, 2002), Progr. Probab.,* **55**, 241–253. Birkhauser, Basel.

[7] Giné, E,; Koltchinskii, V. and Sakhanenko, L. (2004). Kernel density estimators: convergence in distribution for weighted sup-norms. *Probab. Theory Related Fields,* **130**(2), 167–198

[8] Giné, E. Koltchinskii, V. and Zinn, J. (2004). Weighted uniform consistency of kernel density estimators. *Ann. Probab.* **32**(3), 2570–2605.

[9] Mason, D.M. (2003). Representations for integral functionals of kernel density estimators. *Austr. J. Stat.,* **32**(1,2), 131–142.

[10] Pollard, D. (1984). *Convergence of stochastic processes.* Springer Series in Statistics. Springer-Verlag, New York. MR762984

[11] Stute, W. (1982). A law of the logarithm for kernel density estimators. *Ann. Probab.,* **10**(2), 414–422. MR647513

[12] Stute, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.,* **10**(1), 86–107. MR637378

[13] Stute, W. (1984). The oscillation behavior of empirical processes: the multivariate case. *Ann. Probab.,* **12**(2), 361–379. MR735843

[14] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22**(1), 28–76. MR1258865

[15] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. Springer-Verlag, New York. MR1385671