Test (2011) 20: 72–94
DOI 10.1007/s11749-010-0188-0

ORIGINAL PAPER

A general result on the uniform in bandwidth consistency of kernel-type function estimators

David M. Mason · Jan W.H. Swanepoel

Received: 23 October 2009 / Accepted: 8 March 2010 / Published online: 24 March 2010

Abstract We develop a general theorem to prove the uniform in bandwidth consistency of kernel-type function estimators. This method unifies the approaches in some other recent papers. We show how to apply our results to kernel distribution function estimators and the smoothed empirical process. The results are applicable to establish strong uniform consistency of data-driven bandwidth kernel-type function estimators.

Keywords Kernel estimation · Distribution function · Uniform in bandwidth

Mathematics Subject Classification (2000) 60F15 · 62G07 · 62G08

1 A general uniform in bandwidth result

We present a general method based on empirical process techniques to prove uniform in bandwidth consistency of kernel-type function estimators. It is a distillation of some recent results by Einmahl and Mason (2005) and Dony et al. (2006), whose work was motivated by original groundwork by Nolan and Marron (1989). Our main theoretical result, the Theorem below, may be viewed as a notational reformulation and generalization of Theorem 2 of Dony et al. (2006). However, we greatly extend its scope and applicability. Our featured applications will be to kernel distribution function estimators (kdfe's) as well as to the smoothed empirical process. Both of

David M. Mason's research was partially supported by NSF Grant DMS–0503908 and Jan W.H. Swanepoel's was partially supported by NRF of South Africa.

D.M. Mason (✉)
Statistics Program, University of Delaware, Newark, DE 19716, USA
e-mail: davidm@udel.edu

J.W.H. Swanepoel
North-West University, Potchefstroom Campus, Potchefstroom, South Africa

these statistics have attracted a considerable amount of attention in the literature, some of which we cite below.

We shall show how our results are useful to establish strong uniform consistency of data-driven bandwidth kdfe's. For more applications refer to Einmahl and Mason (2005), Dony et al. (2006) and Dony and Mason (2008).

Here is our basic setup. Let $X, X_1, X_2, \dots$ be i.i.d. random variables from a probability space $(\Omega, \mathcal{A}, P)$ to a measure space $(S, \mathcal{S})$. In the following, $\|\cdot\|_\infty$ denotes the supremum norm on $\ell^\infty(S)$, the space of bounded real valued measurable functions on S. Let $\mathcal{G}$ denote a class of measurable real valued functions g of $(u, h) \in S \times (0, 1]$. We shall assume that $\mathcal{G}$ satisfies:

(G.i) $\sup_{0<h\le 1} \sup_{g\in\mathcal{G}} \|g(\cdot, h)\|_\infty =: \kappa < \infty$.

Assume that there exists a constant $C > 0$ such that for all $h \in (0, 1]$

(G.ii) $\sup_{g\in\mathcal{G}} \mathbb{E}[g^2(X, h)] \le Ch$.

We shall also assume that the class of functions $\mathcal{G}$ satisfies the following uniform entropy condition:

(F.i) For some $\gamma \ge 0$, $C_0 > 0$ and $\nu_0 > 0$, $N(\epsilon, \mathcal{G}_\gamma) \le C_0\epsilon^{-\nu_0}$, $0 < \epsilon < 1$, where

$$\mathcal{G}_\gamma = \left\{g_\gamma(\cdot, h) : g_\gamma(\cdot, h) = h^\gamma g(\cdot, h),\ \text{for } g \in \mathcal{G} \text{ and } h \in (0, 1]\right\}.$$

We shall see that the introduction of the parameter $\gamma \ge 0$ will permit us to treat the smoothed empirical process.

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

(F.ii) $\mathcal{G}$ is a pointwise measurable class.

(For the definitions of pointwise measurable and of $N(\epsilon, \mathcal{G}_\gamma)$, see the Appendix, where we use κ as our envelope function.)

For any $n \ge 1$, $g \in \mathcal{G}$ and $0 < h < 1$, define

$$g_{n,h} := n^{-1} \sum_{i=1}^{n} g(X_i, h).$$

Here is our main result. Its proof is deferred to Sect. 4.

Theorem *Assuming* (G.i), (G.ii), (F.i) *and* (F.ii), *we have for* $c > 0, 0 < h_0 < 1$, *w.p.* 1,

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n} \le h \le h_0} \sup_{g\in\mathcal{G}} \frac{\sqrt{n}|g_{n,h} - \mathbb{E}g_{n,h}|}{\sqrt{h(|\log h| \vee \log\log n)}} =: A(c) < \infty. \tag{1.1}$$

Remark 1 It is straightforward to modify the proof of the Theorem to show that it remains true when (F.i) is substituted by the bracketing condition:

(F′.i) For some $C_0 > 0$ and $\nu_0 > 0$, $N_{[\ ]}(\epsilon, \mathcal{F}, L_2(P)) \le C_0\epsilon^{-\nu_0}$, $0 < \epsilon < 1$.

Refer to p. 270 of van der Vaart (1998) for the definition of $N_{[\,]}(\epsilon, \mathcal{F}, L_2(P))$. All that one must do is replace the use of Proposition A.1 in the Appendix by Lemma 19.34 of van der Vaart (1998).

Remark 2 The results in Einmahl and Mason (2005) on the uniform in bandwidth consistency of kernel density and regression function (bounded case) estimators can be readily derived from our Theorem. For example, in the former case one simply takes $\mathcal{G} = \{k((x - \cdot)/h^{1/r}) : \ 0 < h \le 1, \ x \in \mathbb{R}^r\}$, for an integer $r \ge 1$, where $k(\cdot)$ is the kernel function of the kernel density estimator, which is assumed to be bounded, in $L_1(\mathbb{R}^r)$ and satisfy $\int_{\mathbb{R}^r} k(x)\,dx = 1$. In addition, assume that the underlying distribution function $F(\cdot)$ has a bounded density. Under these assumptions, it is readily verified that (G.ii) is satisfied. ((G.i) holds trivially.) We also assume that (F.i), with $\gamma = 0$, and (F.ii) hold. In this case for each $g \in \mathcal{G}$, $g_{n,h} = hf_{n,h^{1/r}}(x)$, where $f_{n,h^{1/r}}(x)$ is the kernel density estimator

$$f_{n,h^{1/r}}(x) = \sum_{i=1}^{n} k\big((x - X_i)\big/h^{1/r}\big)/(nh). \tag{1.2}$$

Our Theorem also implies Theorem 2 of Dony et al. (2006), which concerns a general class of regression function estimators. The application of our general result is not confined to these classical nonparametric function kernel estimators. We shall not rederive results for such estimators here, and instead we shall focus on the uniform in bandwidth consistency of kdfe's and the smoothed empirical process. We shall soon see that these form classes of estimators that are different from the classical kernel density and regression function estimators.

2 Kernel distribution function estimators

In this section we shall apply our Theorem to the standard kdfe as well as to a transformed kdfe. Hereafter we discuss the strong consistency of data-driven bandwidth kdfe's.

2.1 Standard kdfe

As above, let $X, \ X_1, \ X_2, \ldots$ be i.i.d. with distribution function F. Suppose that F satisfies the Lipschitz condition

$$\big|F(x) - F(y)\big| \le C|x - y|, \quad \text{for all } x, y \in \mathbb{R}, \ \text{some } 0 < C < \infty. \tag{2.1}$$

Let K be a distribution function on $\mathbb{R}$ and consider the standard kdfe of F based on $X_1, \ldots, X_n$:

$$\widehat{F}_{n,h}(x) = n^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}.$$

We shall assume that

$$\int_{-\infty}^{\infty} |z| \, dK(z) < \infty. \tag{2.2}$$

As usual, F_n will denote the empirical distribution function

$$F_n(x) = n^{-1} \sum_{i=1}^{n} 1\{X_i \le x\}, \quad x \in \mathbb{R},$$

where $1\{\cdot\}$ is the indicator function. By results in Sect. 5 of Nolan and Pollard (1987), the class of functions

$$\mathcal{K} = \left\{ K\left(\frac{x - \cdot}{h}\right) : 0 < h \le 1, x \in \mathbb{R} \right\}$$

satisfies (F.i), and trivially also (F.ii). Moreover, the class

$$\mathcal{G} = \left\{ K\left(\frac{x - \cdot}{h}\right) - 1\{\cdot \le x\} : 0 < h \le 1, x \in \mathbb{R} \right\}$$

is also readily shown to fulfill (F.i) and (F.ii).

Towards verifying (G.ii), observe that for each $0 < h \le 1$ and $x \in \mathbb{R}$,

$$\mathbb{E}\left(K\left(\frac{x - X}{h}\right) - 1\{X \le x\} \right)^2$$
$$= \mathbb{E}K^2\left(\frac{x - X}{h}\right) - 2\mathbb{E}\left(K\left(\frac{x - X}{h}\right) 1\{X \le x\} \right) + F(x).$$

Now

$$\mathbb{E}K^2\left(\frac{x - X}{h}\right) = \int_{-\infty}^{\infty} K^2\left(\frac{x - y}{h}\right) dF(y),$$

which after integrating by parts and the change of variables $z = (x - y)/h$,

$$= \int_{-\infty}^{\infty} F(x - hz) \, dK^2(z).$$

Also,

$$\mathbb{E}\left(K\left(\frac{x - X}{h}\right) 1\{X \le x\} \right) = \int_{-\infty}^{x} K\left(\frac{x - y}{h}\right) dF(y),$$

which by integration by parts and the change of variables $z = (x - y)/h$,

$$= K(0)F(x) + \int_{0}^{\infty} F(x - hz) \, dK(z).$$

Thus

$$\mathbb{E}\left(K\left(\frac{x - X}{h}\right) - 1\{X \le x\} \right)^2$$

$$= \big(1 - 2K(0)\big)F(x) - 2\int_0^\infty F(x - hz)\,dK(z) + \int_{-\infty}^\infty F(x - hz)\,dK^2(z)$$
$$= -2\int_0^\infty \big(F(x - hz) - F(x)\big)\,dK(z) + \int_{-\infty}^\infty \big(F(x - hz) - F(x)\big)\,dK^2(z),$$

which by the Lipschitz assumption in (2.1), the inequality

$$\int_{-\infty}^\infty |z|\,dK^2(z) \le 2\int_{-\infty}^\infty |z|\,dK(z),$$

and (2.2) is

$$\le hC\left(2\int_0^\infty z\,dK(z) + \int_{-\infty}^\infty |z|\,dK^2(z)\right) \le h4C\int_{-\infty}^\infty |z|\,dK(z) < \infty.$$

Thus (G.ii) holds. Condition (G.i) is satisfied trivially. Therefore we can apply our Theorem to infer the following useful result.

Proposition 1 *Assume that* F *satisfies* (2.1) *and* K *fulfills* (2.2). *Then for* $c > 0$, $0 < h_0 < 1$, *w.p.* 1, *for some constant* $0 < A(c) < \infty$,

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n} \le h \le h_0} \sup_{x\in\mathbb{R}} \frac{\sqrt{n}|\widehat{F}_{n,h}(x) - F_n(x) - (\mathbb{E}\widehat{F}_{n,h}(x) - F(x))|}{\sqrt{h(|\log h| \vee \log\log n)}} = A(c). \quad (2.3)$$

Define the classical empirical process by

$$\alpha_n(x) = \sqrt{n}\big\{F_n(x) - F(x)\big\}, \quad x \in \mathbb{R}. \quad (2.4)$$

Notice that (2.3) readily implies the following result.

Corollary 1 *Under the assumptions of Proposition* 1, *for any sequence of positive constants* $0 < b_n < 1$ *satisfying* $b_n \to 0$ *and* $b_n \ge (\log n)^{-1}$, *w.p.* 1,

$$\sup_{\frac{c\log n}{n} \le h \le b_n} \frac{\sup_{x\in\mathbb{R}} |\sqrt{n}\{\widehat{F}_{n,h}(x) - \mathbb{E}\widehat{F}_{n,h}(x)\} - \alpha_n(x)|}{\sqrt{\log\log n}} = O\big(\sqrt{b_n}\big). \quad (2.5)$$

From (2.5) we obtain a uniform in bandwidth extension of the results derived by Boos (1986) in his Theorems 1 and 2.

Corollary 2 *Under the assumptions of Proposition* 1, *if* $b_n \to 0$, $b_n \ge (\log n)^{-1}$ *and*

$$\sup_{\frac{c\log n}{n} \le h \le b_n} \sup_{x\in\mathbb{R}} \sqrt{n}\big|\mathbb{E}\widehat{F}_{n,h}(x) - F(x)\big| = O\big(\sqrt{b_n \log\log n}\big), \quad (2.6)$$

then

$$\sup_{\frac{c\log n}{n} \le h \le b_n} \sqrt{n}\|\widehat{F}_{n,h} - F_n\|_\infty = O\big(\sqrt{b_n \log\log n}\big) \quad a.s. \quad (2.7)$$

By (2.10) below, (2.6) holds if $\sqrt{n}b_n = O(\sqrt{b_n \log\log n})$. Also, from (2.13) below it follows that (2.6) holds if $\sqrt{n}b_n^2 = O(\sqrt{b_n \log\log n})$ and the more restrictive conditions (2.11) and (2.12) below are imposed on F and K respectively.

We can readily derive from (2.5) a uniform in bandwidth Finkelstein-type functional law of the iterated logarithm [FLIL] for the sequence of classes of random functions

$$\mathcal{F}_n = \left\{ \frac{\eta_{n,h}(\cdot)}{\sqrt{2\log\log n}} : \frac{c\log n}{n} \le h \le b_n \right\},$$

where $\eta_{n,h}(\cdot)$ is the *centered kdfe process*

$$\eta_{n,h}(\cdot) = \sqrt{n}\left\{\widehat{F}_{n,h}(\cdot) - \mathbb{E}\widehat{F}_{n,h}(\cdot)\right\}.$$

Towards formulating our uniform FLIL, let $B[0,1]$ denote the space of bounded functions φ on $[0,1]$ equipped with supremum norm $\|\varphi\|_{[0,1]} = \sup_{t\in[0,1]} |\varphi(t)|$, and $B[\mathbb{R}]$ denote the space of bounded functions on $\mathbb{R}$ equipped with supremum norm $\|\varphi\|_{\mathbb{R}} = \sup_{t\in\mathbb{R}} |\varphi(t)|$. Further, let

$$H = \left\{ \varphi : \varphi(0) = \varphi(1) = 0, \varphi \text{ is absolutely continuous and } \int_0^1 \left(\varphi'(s)\right)^2 ds \le 1 \right\}.$$

For all $\varepsilon > 0$, let

$$H_F^{\varepsilon} = \left\{ \phi : \phi \in B[\mathbb{R}] \text{ and } \inf_{\varphi\in H} \|\varphi \text{ o } F - \phi\|_{\mathbb{R}} \le \varepsilon \right\}$$

and for each $\varphi \in H$, let

$$B_{\varepsilon}(\varphi) = \left\{ \phi : \phi \in B[\mathbb{R}] \text{ and } \|\varphi \text{ o } F - \phi\|_{\mathbb{R}} \le \varepsilon \right\}.$$

Finkelstein (1971) proved the following FLIL (now called the compact LIL) for the empirical process $\alpha_n(\cdot)$: w.p. 1, for all $\varepsilon > 0$ there exists an $N > 0$ such that for all $n \ge N$,

$$\frac{\alpha_n(\cdot)}{\sqrt{2\log\log n}} \in H_F^{\varepsilon}$$

and for every $\varphi \in H$ there exists a subsequence $\{n_k\}$ such that for all $k \ge 1$,

$$\frac{\alpha_{n_k}(\cdot)}{\sqrt{2\log\log n_k}} \in B_{\varepsilon}(\varphi).$$

Clearly it is routine to combine (2.5) with the Finkelstein FLIL to infer the following FLIL.

Corollary 3 *Under the assumptions of Proposition 1, for any sequence of positive constants $0 < b_n < 1$ satisfying $b_n \to 0$ and $b_n \ge (\log n)^{-1}$, w.p. 1, for all $\varepsilon > 0$ there exists an N such that $\mathcal{F}_n \subset H_F^{\varepsilon}$ for all $n \ge N$ and for every $\varphi \in H$ there exists a subsequence $\{n_k\}$ such that for all $k \ge 1$,*

$$\frac{\eta_{n_k,h}(\cdot)}{\sqrt{2\log\log n_k}} \in B_{\varepsilon}(\varphi), \quad \textit{uniformly in } \frac{c\log n_k}{n_k} \le h \le b_{n_k}. \tag{2.8}$$

Notice that, in particular, Corollary 3 implies the following uniform in bandwidth Chung (1949) LIL:

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n}\le h\le b_n} \sup_{x\in\mathbb{R}} \frac{\sqrt{n}|\widehat{F}_{n,h}(x)-\mathbb{E}\widehat{F}_{n,h}(x)|}{\sqrt{2\log\log n}} = \frac{1}{2}, \quad \text{a.s.} \tag{2.9}$$

Remark 3 By imposing further regularity conditions on b_n we can replace the $\mathbb{E}\widehat{F}_{n,h}(x)$ centering term by $F(x)$ in (2.5), (2.8) and (2.9). Observe that

$$\mathbb{E}\widehat{F}_{n,h}(x) - F(x)$$

equals, after integrating by parts and the change of variables $z = (x-y)/h$,

$$\int_{-\infty}^{\infty} \big(F(x-hz) - F(x)\big)\, dK(z).$$

Therefore using (2.1) and (2.2) we see that

$$\sup_{x\in\mathbb{R}} \big|\mathbb{E}\widehat{F}_{n,h}(x) - F(x)\big| \le hC \int_{-\infty}^{\infty} |z|\, dK(z) < \infty,$$

which gives

$$\sup_{\frac{c\log n}{n}\le h\le b_n} \sup_{x\in\mathbb{R}} \big|\mathbb{E}\widehat{F}_{n,h}(x) - F(x)\big| = O(b_n). \tag{2.10}$$

Thus for any sequences $0 < b_n < 1$ such that $\sqrt{n}b_n/\sqrt{\log\log n} = o(1)$, we can replace $\mathbb{E}\widehat{F}_{n,h}(x)$ by $F(x)$ in (2.5), (2.8) and (2.9).

If further we assume that

$$F \text{ has a bounded density with a uniformly bounded derivative} \tag{2.11}$$

and

$$\int_{-\infty}^{\infty} z\, dK(z) = 0, \qquad \int_{-\infty}^{\infty} z^2\, dK(z) < \infty, \tag{2.12}$$

then it is easy to show by Taylor's theorem that

$$\sup_{\frac{c\log n}{n}\le h\le b_n} \sup_{x\in\mathbb{R}} \big|\mathbb{E}\widehat{F}_{n,h}(x) - F(x)\big| = O\big(b_n^2\big). \tag{2.13}$$

In this case, if $\sqrt{n}b_n^2/\sqrt{\log\log n} = o(1)$, we can replace $\mathbb{E}\widehat{F}_{n,h}(x)$ by $F(x)$.

Remark 4 Swanepoel (1987) was the first to establish under various regularity conditions a Finkelstein-type FLIL for the sequence of *kdfe processes*

$$\frac{\sqrt{n}\{\widehat{F}_{n,h}(\cdot) - F(\cdot)\}}{\sqrt{2\log\log n}}.$$

More recently Giné and Nickl (2009) have obtained under smoothness conditions a FLIL for the special case of the *kdfe process* when $\widehat{F}_{n,h}(x)=\int_{-\infty}^{x} f_{n,h}(y)\,dy$, where $f_{n,h}$ is the kernel density estimator defined in Remark 2. When the kernel k used to define $f_{n,h}$ is a bounded density, Corollary 3 can be shown to imply their FLIL by choosing $K(x)=\int_{-\infty}^{x} k(y)\,dy$. They derive their FLIL via a general exponential inequality for $\sqrt{n}\|\widehat{F}_{n,h}-F_n\|_\infty$ and they do not assume that the kernel k is a density.

Remark 5 Since the empirical process $\alpha_n(\cdot)$ converges weakly to the Brownian bridge $B(F(\cdot))$, we see from (2.5) that for any sequence of positive constants $0<b_n<1$ satisfying $b_n\to 0$, $b_n\geq(\log n)^{-1}$ and $\sqrt{b_n\log\log n}=o(1)$ that the *centered kdfe process* $\eta_{n,h}(\cdot)$ converges weakly to $B(F(\cdot))$, uniformly in $\frac{c\log n}{n}\leq h\leq b_n$, in the sense that

$$\sup_{\frac{c\log n}{n}\leq h\leq b_n}\sup_{x\in\mathbb{R}}\left|\eta_{n,h}(x)-\alpha_n(x)\right|=o_p(1). \tag{2.14}$$

2.2 Transformed kdfe

Swanepoel and Van Graan (2005) have recently proposed a new kdfe based on a nonparametric transformation of the data. Their estimator is defined by

$$\widetilde{F}_{n,h}(x)=n^{-1}\sum_{i=1}^{n}K\left(\frac{\widehat{F}_{n,h}(x)-\widehat{F}_{n,h}(X_i)}{h}\right).$$

It was shown by these authors that the asymptotic bias and mean squared error of $\widetilde{F}_{n,h}(x)$ are considerably smaller than those of the standard kernel estimator $\widehat{F}_{n,h}(x)$.

We shall now show in Proposition 2 (the proof is deferred to Sect. 4) that modified versions of the results in (2.3)–(2.14) continue to hold if $\widehat{F}_{n,h}(x)$ is replaced by $\widetilde{F}_{n,h}(x)$ and the supremum is taken for $a_n\leq h\leq b_n$ instead of $\frac{c\log n}{n}\leq h\leq b_n$ for appropriate a_n and b_n.

Proposition 2 *Assume*:

(A) *K has a continuous density k which has support in a compact interval with continuous derivative k' such that*

$$\int_{-\infty}^{\infty} xk(x)\,dx=0.$$

(B) *F has derivatives f and f' which exist everywhere on $(-\infty,\infty)$ with $\|f\|_\infty<\infty$ and $\|f'\|_\infty<\infty$.*

(C) *$\{a_n\}$ and $\{b_n\}$ are sequences of real numbers such that $0<a_n<b_n\leq 1$, satisfying as $n\to\infty$,*

(C.i) $\frac{na_n^2\log\log n}{(\log n)^2}\to\infty$ *and* (C.ii) $\frac{nb_n^4}{\log\log n}\to 0$.

Then (2.3)–(2.14) hold with $\widehat{F}_{n,h}(x)$ replaced by $\widetilde{F}_{n,h}(x)$ and the supremum is taken for $a_n\leq h\leq b_n$ instead of $\frac{c\log n}{n}\leq h\leq b_n$.

2.3 Applications to the strong consistency of data-driven bandwidth kdfe's

The smoothed kdfe has been investigated by many authors, who have found that to use it in practice a data-driven choice of h is required. A number of suggestions for choosing h depending on the data have been made in the literature for the standard kernel estimator $\widehat{F}_{n,h}(x)$. The proposals by Sarda (1993), Altman and Léger (1995), Bowman et al. (1998), and Polansky and Baker (2000) are analogs of the "leave-one-out" and "plug-in" methods, which have been widely used in density estimation. These data-based bandwidth sequences, $\widehat{h}_n$, satisfy w.p. 1 for all large enough $n \geq 1$, the following inequality

$$an^{-1/3} \leq \widehat{h}_n \leq bn^{-1/3}, \tag{2.15}$$

for some constants $0 < a < b < \infty$. Under the assumptions (2.1) and (2.2), it immediately follows from (2.9) and (2.10) that for any $c > 0$ and sequence of constants $0 < b_n < 1$ satisfying $b_n \geq (\log n)^{-1}$ and $\sqrt{n} b_n / \sqrt{\log \log n} = o(1)$,

$$\sup_{\frac{c \log n}{n} \leq h \leq b_n} \|\widehat{F}_{n,h} - F\|_\infty \to 0, \quad \text{a.s. as } n \to \infty,$$

which in turn implies that

$$\lim_{n\to\infty} \|\widehat{F}_{n,\widehat{h}_n} - F\|_\infty = 0, \quad \text{a.s.}$$

Further, if we assume condition (2.11) and the first equality in (2.12), it follows from (2.9) and (2.13) that, for any sequence of constants $0 < b_n < 1$ such that $b_n \geq (\log n)^{-1}$ and $\sqrt{n} b_n^2 / \sqrt{\log \log n} = o(1)$,

$$\sup_{\frac{c \log n}{n} \leq h \leq b_n} \frac{\sqrt{n}\|\widehat{F}_{n,h} - F\|_\infty}{\sqrt{\log \log n}} = O(1), \quad \text{a.s. as } n \to \infty,$$

which implies that

$$\frac{\sqrt{n}\|\widehat{F}_{n,\widehat{h}_n} - F\|_\infty}{\sqrt{\log \log n}} = O(1), \quad \text{a.s. as } n \to \infty, \tag{2.16}$$

since, w.p. 1, (2.15) holds for all large n, for some constants $0 < a < b < \infty$.

As far as the estimator $\widetilde{F}_{n,h}(x)$ is concerned, it is clear from the proof of Proposition 2 that under conditions (A) and (B) we have

$$\lim_{n\to\infty} \sup_{a_n \leq h \leq b_n} \|\widetilde{F}_{n,h} - F\|_\infty = 0, \quad \text{a.s.,}$$

for any sequences $0 < a_n < b_n < 1$ satisfying $b_n \to 0$ and $na_n / \log n \to \infty$. Swanepoel and Van Graan (2005) derived an optimal bandwidth selector $\widetilde{h}_n$ for $\widetilde{F}_{n,h}(x)$ that satisfies w.p. 1, for all large enough $n \geq 1$, the inequality

$$an^{-1/7} \leq \widetilde{h}_n \leq bn^{-1/7}, \tag{2.17}$$

for some constants $0 < a < b < \infty$. Consequently, we deduce that

$$\lim_{n\to\infty} \|\widetilde{F}_{n,\widetilde{h}_n} - F\|_\infty = 0, \quad \text{a.s.}$$

Furthermore, suppose that condition (B) in Proposition 2 is replaced by (B′): F has four bounded derivatives on $(-\infty, \infty)$. Also, assume that (C.ii) is replaced by (C′.ii): $\frac{nb_n^{12}}{\log\log n} \to 0$. Under conditions (A), (B′), (C.i) and (C′.ii), it then follows from the proof of Proposition 2 and (2.17) that (2.16) also holds if $\widehat{F}_{n,\widehat{h}_n}$ is replaced by $\widetilde{F}_{n,\widetilde{h}_n}$.

3 Further applications

We shall begin by showing that the oscillation result of Serfling (1982) and Boos (1986) can be derived from our Theorem. Their papers were in a sense the motivation for our work. Next we consider the smoothed empirical process.

3.1 An oscillation result

Assume that F satisfies the Lipschitz condition in (2.1). Consider the class of functions

$$\mathcal{G} = \left\{1\left\{\cdot \in (x, x+th]\right\} : 0 \le t \le 1, 0 < h \le 1, x \in \mathbb{R}\right\}.$$

Notice that for each integer $n \ge 1$,

$$n^{-1}\sum_{i=1}^{n} 1\left\{X_i \in (x, x+th]\right\} = F_n(x+th) - F_n(x).$$

It is well known that the class $\mathcal{G}$ satisfies (F.i) and (F.ii). Also, by the Lipschitz assumption (2.1),

$$\mathbb{E}1^2\left\{X \in (x, x+th]\right\} = F(x+th) - F(x) \le Cth \le Ch.$$

Thus we clearly see that (G.i) and (G.ii) hold. Therefore by our Theorem for $c > 0$, $0 < h_0 < 1$, w.p. 1, for some constant $0 < A(c) < \infty$,

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n} \le h \le h_0} \sup_{x\in\mathbb{R}} \sup_{t\in[0,1]} \frac{|\alpha_n(x+th) - \alpha_n(x)|}{\sqrt{h(|\log h| \vee \log\log n)}} = A(c), \tag{3.1}$$

where $\alpha_n(\cdot)$ is the classical empirical process defined in (2.4). Now choose any sequence of positive constants b_n such that

$$b_n \to 0 \quad \text{and} \quad nb_n/\log n \to \infty. \tag{3.2}$$

Obviously from (3.1) we get for any $c > 0$,

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \sup_{t\in[0,1]} \frac{|\alpha_n(x+tb_n) - \alpha_n(x)|}{\sqrt{b_n(|\log b_n| \vee \log\log n)}}$$

$$= \limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \sup_{|y|\le b_n} \frac{|\alpha_n(x+y)-\alpha_n(x)|}{\sqrt{b_n(|\log b_n| \vee \log\log n)}} \le A(c), \quad \text{a.s.}$$

This of course implies the oscillation result

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \sup_{|y|\le b_n} \frac{|\alpha_n(x+y)-\alpha_n(x)|}{\sqrt{b_n}\log n} < \infty, \quad \text{a.s.,} \tag{3.3}$$

which was proved by Serfling (1982) in the uniform case and Boos (1986) in the case when F is Lipschitz. Under additional growth conditions on b_n, Stute (1982) established an exact version of this result. It is easy to formulate and derive a general $\mathbb{R}^r$, $r \ge 1$, version of these results where the X takes values in $\mathbb{R}^r$ and the intervals $(x, x+th]$ are replaced by rectangles $(x_1, x_1 + t_1 h^{1/r}] \times \cdots \times (x_r, x_r + t_r h^{1/r}]$. For closely related results along this line refer to Stute (1984).

3.2 Smoothed empirical process

Let H be a nonnegative nonincreasing bounded right-continuous function defined on $[0, \infty)$ such that

$$\int_{\mathbb{R}^r} H\big(|z|\big)\, dz = 1, \tag{3.4}$$

for an integer $r \ge 1$, where $|z|$ stands for the Euclidean norm of $z \in \mathbb{R}^r$. Define the kernel k on $\mathbb{R}^r$ by

$$k(z) = H\big(|z|\big), \quad z \in \mathbb{R}^r. \tag{3.5}$$

Let $\mathcal{G}$ be a class of measurable real valued functions defined on $\mathbb{R}^r$ such that for some $M > 0$,

$$\sup_{g\in\mathcal{G}} \sup_{z\in\mathbb{R}^r} \big|g(z)\big| \le M. \tag{3.6}$$

Assume that $\mathcal{G}$ is pointwise measurable and that with the envelope function constantly equal to M, for some $C_0 > 0$ and $\nu_0 > 0$,

$$N(\epsilon, \mathcal{G}) \le C_0 \epsilon^{-\nu_0}, \quad 0 < \epsilon < 1. \tag{3.7}$$

Define the class of real valued measurable functions $\varGamma$ on $\mathbb{R}^r$ by

$$\varGamma = \left\{\varphi_{g,h} : \varphi_{g,h}(\cdot) = \frac{1}{h^r}\int_{\mathbb{R}^r} g(y) k\left(\frac{y-\cdot}{h}\right) dy, g \in \mathcal{G}, h \in (0,1]\right\}. \tag{3.8}$$

Now let X, X_1, $X_2 \dots$ be i.i.d., taking values in $\mathbb{R}^r$ with density f. Notice that for each integer $n \ge 1$ and $\varphi_{g,h} \in \varGamma$,

$$\frac{1}{n}\sum_{i=1}^{n} \varphi_{g,h}(X_i) = \int_{\mathbb{R}^r} g(y) f_{n,h}(y)\, dy,$$

where $f_{n,h}(y)$ is the kernel density estimator (see (1.2) in Remark 2):

$$f_{n,h}(y) = \frac{1}{nh^r}\sum_{i=1}^{n} k\left(\frac{y - X_i}{h}\right). \tag{3.9}$$

As in Yukich (1992), van der Vaart (1994) and Giné and Nickl (2008), the *centered smoothed empirical process* indexed by $\mathcal{G}$ with kernel k is defined to be

$$\alpha_{n,h}(g) := \sqrt{n}\int_{\mathbb{R}^r} g(y)\left\{f_{n,h}(y) - \mathbb{E}f_{n,h}(y)\right\}dy, \quad g \in \mathcal{G}. \tag{3.10}$$

We shall first study the behavior of the difference between $\alpha_{n,h}(g)$ and the usual empirical process indexed by $\mathcal{G}$:

$$\alpha_n(g) := \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}g(X)\right\}, \quad g \in \mathcal{G}. \tag{3.11}$$

Assume that for some constants $A > 0$ and $\rho > 0$, for all $z \in \mathbb{R}^r$ and $0 < h \leq 1$,

$$\sup_{g\in\mathcal{G}} \mathbb{E}\big(g(X + hz) - g(X)\big)^2 \leq Ah|z|^\rho \tag{3.12}$$

and

$$\int_{\mathbb{R}^r} |y|^\rho k(y)\,dy < \infty. \tag{3.13}$$

Remark 6 Notice that (3.12) is satisfied by any class $\mathcal{G}$ of measurable real valued functions g defined on $\mathbb{R}^r$ such that for some $D > 0$ and $\rho \geq 1$,

$$\sup_{g\in\mathcal{G}}\big(g(x + y) - g(x)\big)^2 \leq D|y|^\rho, \quad x, y \in \mathbb{R}^r.$$

It also holds for the class of indicators of lower quadrants in $\mathbb{R}^r$,

$$\mathcal{G} = \left\{1\left\{\cdot \in R(x)\right\} : x \in \mathbb{R}^r\right\}, \tag{3.14}$$

where

$$R(x) = (-\infty, x_1] \times \cdots \times (-\infty, x_r], \quad x \in \mathbb{R}^r,$$

whenever the distribution function F of X is Lipschitz. To see this notice that

$$\begin{aligned}&\mathbb{E}\big(1\{X \in R(x + y)\} - 1\{X \in R(x)\}\big)^2\\ &\quad\leq F\big(x_1 + |y_1|, \ldots, x_r + |y_r|\big) - F\big(x_1 - |y_1|, \ldots, x_r - |y_r|\big),\end{aligned}$$

which, since F is assumed to be Lipschitz, is for some $D > 0$, $\leq D|y|$. We see that the class in (3.14) gives as a special case the centered smoothed empirical process

$$\alpha_{n,h}(x) = \sqrt{n}\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_r}\left\{f_{n,h}(y) - \mathbb{E}f_{n,h}(y)\right\}dy, \quad x \in \mathbb{R}^r. \tag{3.15}$$

Now, introduce the class of functions

$$\overline{\Gamma} = \{\varphi_{g,h} - g : g \in \mathcal{G}, h \in (0,1]\}.$$

This class obviously fulfills (G.i) with the bound $\kappa = 2M$, where M is as in (3.6). Notice that

$$\begin{aligned}\mathbb{E}\big(\varphi_{g,h}(X) - g(X)\big)^2 &= \mathbb{E}\left(\int_{\mathbb{R}^r} \frac{1}{h^r}\big(g(y) - g(X)\big)k\left(\frac{y-X}{h}\right)dy\right)^2 \\ &= \mathbb{E}\left(\int_{\mathbb{R}^r} \big(g(X+hz) - g(X)\big)k(z)\,dz\right)^2,\end{aligned}$$

which by the Cauchy–Schwarz inequality, (3.4), (3.12), and (3.13) is

$$\le hA \int_{\mathbb{R}^r} |y|^\rho k(y)\,dy < \infty.$$

Thus the class $\overline{\Gamma}$ satisfies (G.ii), too. Since $\mathcal{G}$ satisfies (F.i), it is readily checked that the class of functions $\mathcal{G}_r = \{h^r g : g \in \mathcal{G}$ and $h \in (0,1]\}$ does as well. Proposition A.2 in the Appendix says that the class of functions $\Gamma_r = \{h^r\varphi_{g,h} : g \in \mathcal{G}, h \in (0,1]\}$ fulfills (F.i). From this we readily conclude that

$$\overline{\Gamma}_r = \{h^r\varphi_{g,h} - h^r g : g \in \mathcal{G}, h \in (0,1]\}$$

also does. Further, since $\mathcal{G}$ satisfies (F.ii), the class $\overline{\Gamma_r}$ does as well.

We can now apply our theorem with $\gamma = r$ to infer the following result.

Proposition 3 *Let k be a kernel defined on $\mathbb{R}^r$ that satisfies (3.4) and (3.5). Let $\mathcal{G}$ be a pointwise measurable class of real valued functions defined on $\mathbb{R}^r$ fulfilling (3.6) and (3.7). Assume, in addition, that (3.12) and (3.13) hold. Then for any $c > 0$ and $0 < h_0 < 1$, w.p. 1, for some constant $0 < A(c) < \infty$,*

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n} \le h \le h_0} \sup_{g\in\mathcal{G}} \frac{|\alpha_{n,h}(g) - \alpha_n(g)|}{\sqrt{h(|\log h| \vee \log\log n)}} = A(c). \tag{3.16}$$

Finally, we point out that the empirical process $\alpha_n(g)$ indexed by a class of functions $\mathcal{G}$ fulfilling (3.6) and (3.7) converges weakly to a Brownian bridge indexed by $\mathcal{G}$ (consult pp. 141–142 of van der Vaart and Wellner 1996) and it satisfies the compact LIL as well (see Theorem 9 on p. 609 of Ledoux and Talagrand 1989). Therefore a uniform in $\frac{c\log n}{n} \le h \le b_n$ weak convergence theorem and a compact LIL for $\alpha_{n,h}(\cdot)$ can be inferred from (3.16) just as we did above from (2.3) for the centered kdfe process $\eta_{n,h}(\cdot)$. We should mention here that for positive sequences $h_n \to 0$ and classes $\mathcal{G}$ that satisfy much less restrictive assumptions than (G.i) and (G.ii), the process $\alpha_{n,h_n}(\cdot)$ converges weakly to a Brownian bridge indexed by $\mathcal{G}$. For details refer to Giné and Nickl (2008) and the references therein.

Remark 7 The discussion in Remark 3 above regarding the replacement of the centering in the kdfe process $\eta_{n,h}(\cdot)$ carries over verbatim to the process $\alpha_{n,h}(g)$. Analogous statements hold concerning the replacement of the centering $\int_{\mathbb{R}^r} g(y)\mathbb{E} f_{n,h}(y)dy$ in $\alpha_{n,h}(g)$ by $\mathbb{E}g(X)$. It is all a matter of deriving uniform bounds on the bias

$$\int_{\mathbb{R}^r} g(y)\mathbb{E} f_{n,h}(y)\,dy - \mathbb{E}g(X).$$

For an in-depth study of bounds on this bias term, confer Giné and Nickl (2008).

4 Proofs

4.1 Proof of theorem

Let α_n be the empirical process based on the sample $X_1,\ldots,X_n$; i.e., if $\varphi : S \to \mathbb{R}$, we have

$$\alpha_n(\varphi) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\varphi(X_i) - \mathbb{E}\varphi(X)\big),$$

whenever $\mathbb{E}\varphi(X)$ is finite and meaningful. Notice that in this notation

$$g_{n,h} - \mathbb{E}g_{n,h} = \frac{1}{\sqrt{n}}\alpha_n\big(g(\cdot,h)\big),$$

so we get that for any $n \geq 1$ and $0 < h \leq 1$,

$$\sup_{g\in\mathcal{G}} \frac{\sqrt{n}|g_{n,h} - \mathbb{E}g_{n,h}|}{\sqrt{h(|\log h| \vee \log\log n)}} = \sup_{g\in\mathcal{G}} \frac{|\alpha_n(g(\cdot,h))|}{\sqrt{h(|\log h| \vee \log\log n)}}.$$

We first note that by (G.ii)

$$\mathbb{E}\big[h^{2\gamma} g^2(X,h)\big] = \mathbb{E}\big[g_\gamma^2(X,h)\big] \leq Ch^{1+2\gamma}. \tag{4.1}$$

Set for $j \geq 0$ and $c > 0$,

$$h_{j,n} = \big(2^j c \log n\big)/n$$

and

$$\mathcal{G}_{j,n} = \big\{g_\gamma(\cdot,h) : g \in \mathcal{G}, h_{j,n} \leq h \leq h_{j+1,n}\big\}.$$

Clearly by (4.1) for $h_{j,n} \leq h \leq h_{j+1,n}$,

$$\mathbb{E}\big[g_\gamma^2(X,h)\big] \leq C(2h_{j,n})^{1+2\gamma} =: \sigma_{j,n}^2. \tag{4.2}$$

(From this point on the proof follows closely the lines of that of Theorem 2 in Dony et al. 2006.) We shall use Proposition A.1 in the Appendix to bound

$$\mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i\varphi(X_i)\right\|_{\mathcal{G}_{j,n}}.$$

(For a functional Ψ defined on a class of functions $\mathcal{F}$, $\|\Psi\|_{\mathcal{F}}$ denotes $\sup_{\varphi\in\mathcal{F}}|\Psi(\varphi)|$.) To that end we note that each $\mathcal{G}_{j,n}$ satisfies (A.1) of the proposition with $G=\beta=\kappa(2h_{j,n})^{\gamma}$ and (A.3) with $\sigma^2=\sigma^2_{j,n}$. Further, since $\mathcal{G}_{j,n}\subset\mathcal{G}_{\gamma}$, we see by (F.i) that each $\mathcal{G}_{j,n}$ also fulfills (A.2). Finally to see that (A.4) holds, observe that

$$\sup_{g\in\mathcal{G}_{j,n}}\|g_{\gamma}\|_{\infty}\le(2h_{j,n})^{\gamma}\kappa,$$

which by keeping in mind that $\sigma^2_{j,n}=C(2h_{j,n})^{1+2\gamma}$ is for large enough n and all $j\ge 0$,

$$\le\frac{1}{2\sqrt{\nu+1}}\sqrt{n\sigma^2_{j,n}/\log((\kappa(2h_{j,n})^{\gamma})\vee 1/\sigma_{j,n})}.$$

Now by applying Proposition A.1 we get for some $D_1>0$ and $D_2>0$ for all large enough n and $j\ge 0$,

$$\mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i\varphi(X_i)\right\|_{\mathcal{G}_{j,n}}\le D_1\sqrt{n(h_{j,n})^{1+2\gamma}\left|\log(D_2h_{j,n})\right|}. \tag{4.3}$$

Let for large enough n

$$l_n:=\max\{j:h_{j,n}\le 2h_0\},$$

then a little calculation shows that

$$l_n\sim\frac{\log(\frac{nh_0}{c\log n})}{\log 2}. \tag{4.4}$$

For $k\ge 1$, set $n_k=2^k$, and let

$$c_{j,k}:=\sqrt{n_k(h_{j,n_k})^{1+2\gamma}\left(|\log D_2h_{j,n_k}|\vee\log\log n_k\right)},\qquad j\ge 0.$$

Recalling (4.2) and applying Inequality A.1 in the Appendix with

$$M=(2h_{j,n_k})^{\gamma}\kappa\quad\text{and}\quad\sigma^2_{\mathcal{G}}=\sigma^2_{\mathcal{G}_{j,n_k}}\le C(2h_{j,n_k})^{1+2\gamma}=:D_0(h_{j,n_k})^{1+2\gamma},$$

we get for any $t>0$,

$$\mathbb{P}\left\{\max_{n_{k-1}\le n\le n_k}\left\|\sqrt{n}\alpha_n\right\|_{\mathcal{G}_{j,n_k}}\ge A_1(D_1c_{j,k}+t)\right\}$$
$$\le 2\left[\exp\left(-A_2t^2/\left(D_0n_k(h_{j,n_k})^{1+2\gamma}\right)\right)+\exp\left(-A_2t/\left(\kappa(2h_{j,n_k})^{\gamma}\right)\right)\right].$$

Set for any $\rho>1$, $j\ge 0$ and $k\ge 1$,

$$p_{j,k}(\rho):=\mathbb{P}\left\{\max_{n_{k-1}\le n\le n_k}\left\|\sqrt{n}\alpha_n\right\|_{\mathcal{G}_{j,n_k}}\ge A_1(D_1+\rho)c_{j,k}\right\}.$$

As we have $c_{j,k}/\sqrt{n_k(h_{j,n_k})^{1+2\gamma}} \geq \sqrt{\log\log n_k}$, we readily obtain, for $j \geq 0$,

$$p_{j,k}(\rho) \leq 2\left[\exp\left(-\frac{\rho^2 A_2}{D_0}\log\log n_k\right) + \exp\left(-\frac{\sqrt{c}\rho A_2}{\kappa}\sqrt{\log n_k \log\log n_k}\right)\right],$$

which for $\gamma = \frac{A_2}{D_0} \wedge \frac{\sqrt{c}A_2}{\kappa}$ implies

$$p_{j,k}(\rho) \leq 4\exp(-\rho\gamma\log\log n_k).$$

Thus

$$P_k(\rho) := \sum_{j=0}^{l_{n_k}-1} p_{j,k}(\rho) \leq 4 l_{n_k}(\log n_k)^{-\rho\gamma},$$

which by (4.4) is for all large k and large enough $\rho > 1$:

$$P_k(\rho) \leq 8(\log n_k)^{1-\rho\gamma} = 8\left(\frac{1}{k\log 2}\right)^{\rho\gamma-1} \leq k^{-2}.$$

Notice that by definition of l_n, for large k,

$$2h_{l_{n_k},n_k} = h_{l_{n_k}+1,n_k} \geq 2h_0,$$

which implies that we have for $n_{k-1} \leq n \leq n_k$,

$$\left[\frac{c\log n}{n}, h_0\right] \subset \left[\frac{c\log n_k}{n_k}, h_{l_{n_k},n_k}\right].$$

Thus for all large enough k and $n_{k-1} \leq n \leq n_k$,

$$A_k(\rho) := \left\{\max_{n_{k-1}\leq n\leq n_k} \sup_{g\in\mathcal{G}} \sup_{\frac{c\log n}{n}\leq h\leq h_0} \frac{\sqrt{n}|g_{n,h} - \mathbb{E}g_{n,h}|}{\sqrt{h(|\log h| \vee \log\log n)}} > 2A_1(D_1+\rho)\right\}$$

$$\subset \bigcup_{j=0}^{l_{n_k}-1}\left\{\max_{n_{k-1}\leq n\leq n_k} \left\|\sqrt{n}\alpha_n\right\|_{\mathcal{G}_{j,n_k}} \geq A_1(D_1+\rho)c_{j,k}\right\}.$$

It follows now for large enough ρ that

$$\mathbb{P}(A_k(\rho)) \leq P_k(\rho) \leq k^{-2},$$

which by the Borel–Cantelli lemma implies our Theorem.

4.2 Proof of Proposition 2

Throughout the proof, without loss of generality, we shall assume that k has support in $[-1, 1]$. By applying a Taylor series expansion, we have

$$\widetilde{F}_{n,h}(x) = \int_{-\infty}^{\infty} K\left(\frac{\widehat{F}_{n,h}(x) - \widehat{F}_{n,h}(y)}{h}\right) dF_n(y)$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} K\left(\frac{F(x)-F(y)}{h}\right) dF_n(y) \\
&\quad + \frac{1}{h}\int_{-\infty}^{\infty} \left[\left(\widehat{F}_{n,h}(x)-\widehat{F}_{n,h}(y)\right)-\left(F(x)-F(y)\right)\right] k \\
&\quad \times \left(\frac{F(x)-F(y)}{h}\right) dF_n(y) \\
&\quad + \frac{1}{2h^2}\int_{-\infty}^{\infty} \left[\left(\widehat{F}_{n,h}(x)-\widehat{F}_{n,h}(y)\right)-\left(F(x)-F(y)\right)\right]^2 k'(\Delta)\, dF_n(y) \\
&=: I_{n1}(x)+I_{n2}(x)+I_{n3}(x),
\end{aligned}
$$

where Δ is between $h^{-1}[\widehat{F}_{n,h}(x)-\widehat{F}_{n,h}(y)]$ and $h^{-1}[F(x)-F(y)]$. Note that we suppress the dependence of $I_{n1}(x)$, $I_{n2}(x)$ and $I_{n3}(x)$ on h.

Consider $I_{n2}(x)$. Write

$$
\begin{aligned}
&\left[\left(\widehat{F}_{n,h}(x)-\widehat{F}_{n,h}(y)\right)-\left(F(x)-F(y)\right)\right] \\
&\quad = \left[\left(\widehat{F}_{n,h}(x)-F_n(x)\right)-\left(\widehat{F}_{n,h}(y)-F_n(y)\right)\right] \\
&\qquad + \left\{\left[F_n(x)-F(x)\right]-\left[F_n(y)-F(y)\right]\right\} \\
&\quad =: \left[Q_{n1}(x)-Q_{n1}(y)\right]+\left\{Q_{n2}(x,y)\right\}.
\end{aligned}
\tag{4.5}
$$

Let $U_1, U_2, \ldots$ denote a sequence of i.i.d. uniform $(0,1)$ random variables. For each $n \geq 1$ write $\overline{F}_n$ for the empirical distribution function based on the first n of these random variables. Then, as a sequence of processes, $Q_{n2}(x,y)$ is equal in distribution to

$$
\left[\overline{F}_n\left(F(x)\right)-F(x)\right]-\left[\overline{F}_n(F(y))-F(y)\right]. \tag{4.6}
$$

Therefore in our treatment of the almost sure limiting behavior of $Q_{n2}(x,y)$ as $n \to \infty$ we can assume that it has the form (4.6).

From (3.3) and the conditions (C) we get, after a little change of notation, that

$$
\sup_{a_n \leq h \leq b_n} \sup_{\{(x,y):|F(x)-F(y)|\leq h\}} \left[\frac{n}{h\log n}\right]^{1/2} \left|Q_{n2}(x,y)\right| = O(1), \quad \text{a.s.} \tag{4.7}
$$

Furthermore, from assumptions (A) and (B) as well as the proofs of Theorems 1 and 2 of Boos (1986), it easily follows using (4.7) that for some $0 < C_i < \infty$, $i = 1, 2$, w.p. 1 for all large enough n uniformly in $a_n \leq h \leq b_n$,

$$
\|\widehat{F}_{n,h}-F_n\|_\infty \leq C_1\left[\frac{h\log n}{n}\right]^{1/2} + C_2 h^2. \tag{4.8}
$$

Since k vanishes outside $[-1,1]$, the integration in $I_{n2}(x)$ can be restricted to those y for which $|F(x)-F(y)| \leq h$. Therefore from (4.5)–(4.8) we now conclude that for some finite positive constant C_3, uniformly in $x \in \mathbb{R}$, we have w.p. 1 for all large enough n, uniformly in $a_n \le h \le b_n$,

$$\left|I_{n2}(x)\right| \le \left\{C_3\left[\frac{h\log n}{n}\right]^{1/2} + C_2h^2\right\}\hat{f}_{n,h}(x)$$
$$\le \left\{C_3\left[\frac{b_n\log n}{n}\right]^{1/2} + C_2b_n^2\right\}\hat{f}_{n,h}(x),$$

where

$$\hat{f}_{n,h}(x) := (nh)^{-1}\sum_{i=1}^{n} k\left(\frac{F(x)-F(X_i)}{h}\right)$$

is the kernel density estimator based on i.i.d. uniform (0, 1) random variables $F(X_i)$, $i = 1, \ldots, n$, evaluated at the point $F(x)$. Using assumptions (A), (B), (C) and Corollary 1 of Einmahl and Mason (2005), it readily follows that

$$\lim_{n\to\infty} \sup_{a_n\le h\le b_n} \frac{\sqrt{n}\|I_{n2}\|_\infty}{\sqrt{\log\log n}} = 0, \quad \text{a.s.} \tag{4.9}$$

Now consider $I_{n3}(x)$. To handle this term we shall need the following fact, which is a special case of Theorem 2 of Mason et al. (1983).

Fact *Let $X_1, X_2, \ldots$ be i.i.d. with continuous cumulative distribution F. For each $n \ge 1$ let F_n be the empirical distribution based on $X_1, \ldots, X_n$. For some constant $D > 0$, w.p. 1, as $n \to \infty$,*

$$\sup_{\{(x,y):|F(x)-F(y)|\ge n^{-1}\log n\}} \left\{\sqrt{\frac{n}{2\log n}}\frac{|F_n(x)-F_n(y)-(F(x)-F(y))|}{|F(x)-F(y)|^{1/2}}\right\} \to D. \tag{4.10}$$

Applying the above fact we see that w.p. 1 for all large n whenever

$$\left|F(x)-F(y)\right| \ge \frac{\log n}{n},$$

then

$$\left|F_n(x)-F_n(y)-\big(F(x)-F(y)\big)\right| \le 2D\left|F(x)-F(y)\right|^{1/2}\sqrt{\frac{\log n}{n}}.$$

Thus for any $C > 1$ w.p. 1 for all large n whenever $a_n \le h \le b_n$ and $|F(x)-F(y)| \ge Ch$ and keeping in mind that by (C.i),

$$na_n^2/\log n \to \infty,$$

we see that both $(F(x)-F(y))/h$ and $(F_n(x)-F_n(y))/h$ are of the same sign and lie outside of the interval $[-1, 1]$. Applying (4.8), we readily check that the same statement holds when $F_n(x)-F_n(y)$ is replaced by $\widehat{F}_{n,h}(x)-\widehat{F}_{n,h}(y)$. Since k' vanishes outside $[-1, 1]$, the integration in $I_{n3}(x)$ can be restricted to those y for which

$|\Delta| \leq 1$. Therefore we conclude from our discussion that for all n sufficiently large for each x we can restrict the integration to those y for which

$$\left|F(x) - F(y)\right| \leq Ch,$$

where $a_n \leq h \leq b_n$ and C is a constant greater than 1. Hence, (4.5)–(4.8) are once again applicable. Using assumptions (A), (B), (C) and the fact that k' is bounded, it easily follows that for some constant D_0, uniformly in $a_n \leq h \leq b_n$,

$$\|I_{n3}\|_\infty \leq \frac{D_0}{h^2}\left\{\frac{h \log n}{n} + h^4\right\},$$

which gives

$$\sup_{a_n \leq h \leq b_n} \|I_{n3}\|_\infty = O\left(\frac{\log n}{n a_n} + b_n^2\right), \quad \text{a.s.}$$

Therefore by the assumptions in (C),

$$\lim_{n\to\infty} \sup_{a_n \leq h \leq b_n} \frac{\sqrt{n}\|I_{n3}\|_\infty}{\sqrt{\log\log n}} = 0, \quad \text{a.s.} \tag{4.11}$$

Hence, from (4.9), (4.11), and the above Taylor series expansion we conclude that

$$\lim_{n\to\infty} \sup_{a_n \leq h \leq b_n} \frac{\sqrt{n}\|\widetilde{F}_{n,h} - I_{n1}\|_\infty}{\sqrt{\log\log n}} = 0, \quad \text{a.s.}$$

Finally, note that $I_{n1}(x)$ becomes the kernel estimator $\widehat{F}_{n,h}(\cdot)$ (for which (2.3)–(2.14) hold) by simply replacing in the latter x and X_i by $F(x)$ and $F(X_i)$, respectively. Keep in mind that we take the supremum over $a_n \leq h \leq b_n$ instead of $\frac{c \log n}{n} \leq h \leq b_n$.

Acknowledgements The authors acknowledge the constructive comments from the referees and the associate editor.

Appendix

A class of measurable real valued functions $\mathcal{G}$ defined on a measure space $(\mathcal{X}, \mathcal{A})$ is said to be a *pointwise measurable class* if there exists a countable subclass $\mathcal{G}_0$ of $\mathcal{G}$ so that we can find for any function g in $\mathcal{G}$ a sequence of functions $\{g_m\}$ in $\mathcal{G}_0$ for which $g_m(x) \to g(x)$, $x \in \mathcal{X}$. (See Example 2.3.4 in van der Vaart and Wellner 1996.)

Let $X, X_1, \ldots, X_n$ be i.i.d. from a probability space $(\mathcal{X}, \mathcal{A}, P)$ with common distribution μ. Further, let $\varepsilon_1, \ldots, \varepsilon_n$ be a sequence of independent Rademacher random variables independent of $X_1, \ldots, X_n$.

The following inequality is essentially due to Talagrand (1994) (see Einmahl and Mason 2000).

Inequality A.1 *Let $\mathcal{G}$ be a pointwise measurable class of functions satisfying for some $0 < M < \infty$, $\|g\|_\infty \le M$, $g \in \mathcal{G}$. Then for all $t > 0$ we have, for suitable finite constants $A_1, A_2 > 0$,*

$$\mathbb{P}\left\{\max_{1\le m\le n}\|\sqrt{m}\alpha_m\|_{\mathcal{G}} \ge A_1\left(\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i g(X_i)\right\|_{\mathcal{G}} + t\right)\right\}$$
$$\le 2\big(\exp\big(-A_2 t^2/n\sigma_{\mathcal{G}}^2\big) + \exp(-A_2 t/M)\big),$$

where $\sigma_{\mathcal{G}}^2 = \sup_{g\in\mathcal{G}} \mathrm{Var}(g(X))$.

It enables us to reduce many problems on almost sure convergence to investigating the moment quantity $\mu_n := \mathbb{E}\|\sum_{i=1}^n \varepsilon_i g(X_i)\|_{\mathcal{G}}$.

The following proposition proved in Einmahl and Mason (2000) is very helpful for obtaining bounds on this quantity, when the class $\mathcal{G}$ has a polynomial covering number. For a similar inequality, see Giné and Guillou (2001). Assume that there exists a finite valued measurable function G, called an envelope function, which satisfies for all $x \in \mathcal{X}$, $G(x) \ge \sup_{g\in\mathcal{G}} |g(x)|$. We define for $\epsilon > 0$

$$N(\epsilon, \mathcal{G}) := \sup_Q N\left(\epsilon\sqrt{Q\big(G^2\big)}, \mathcal{G}, d_Q\right),$$

where the supremum is taken over all probability measures Q on $(\mathcal{X}, \mathcal{A})$ for which $0 < Q(G^2) := \int G^2(y)Q(dy) < \infty$ and d_Q is the $L_2(Q)$-metric. As usual, $N(\epsilon, \mathcal{G}, d_Q)$ is the minimal number of balls $\{g : d_Q(g, f) < \epsilon\}$ of d_Q-radius ϵ needed to cover $\mathcal{G}$.

Proposition A.1 *Let $\mathcal{G}$ be a pointwise measurable class of bounded functions such that for some constants $\beta, \nu, C > 1$, $\sigma \le 1/(8C)$ and function G as above, the following four conditions hold*:

$$\mathbb{E}\big[G^2(X)\big] \le \beta^2; \tag{A.1}$$

$$N(\epsilon, \mathcal{G}) \le C\epsilon^{-\nu}, \quad 0 < \epsilon < 1; \tag{A.2}$$

$$\sigma_0^2 := \sup_{g\in\mathcal{G}} \mathbb{E}\big[g^2(X)\big] \le \sigma^2; \tag{A.3}$$

$$\sup_{g\in\mathcal{G}} \|g\|_\infty \le \frac{1}{2\sqrt{\nu+1}}\sqrt{n\sigma^2/\log(\beta \vee 1/\sigma)}. \tag{A.4}$$

Then we have for a universal constant A:

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i g(X_i)\right\|_{\mathcal{G}} \le A\sqrt{\nu n\sigma^2 \log(\beta \vee 1/\sigma)}. \tag{A.5}$$

Proposition A.2 *Under the assumptions (3.4)–(3.7) we have, for some $C_1 > 0$,*

$$N(\epsilon, \Gamma_r) \le C_1\epsilon^{-\nu_0-1}, \quad 0 < \epsilon < 1. \tag{A.6}$$

Proof Choose $0<\epsilon<1$ and select $0=h_0<h_1<\cdots<h_m=1$, $m>1$, such that $h_{i+1}^r-h_i^r<\epsilon/2$ for $i=0,\dots,m-1$. Notice that we can take

$$m\le 3\bigl(\epsilon/(2r)\bigr)^{-1}. \tag{A.7}$$

For each $i=1,\dots,m$, let P_i denote the probability measure on $\mathbb{R}^r$ with density

$$p_i(y)=\frac{1}{h_i^r}\mathbb{E}k\left(\frac{y-X}{h_i}\right),\quad y\in\mathbb{R}^r.$$

Further, let Y_i be an $\mathbb{R}^r$ valued random variable with density p_i. For each $i=1,\dots,m$ choose $g_{i,1},\dots,g_{i,n_i}$, $n_i>1$, so that for all $g\in\mathcal{G}$,

$$\min_{1\le j\le n_i}\sqrt{\mathbb{E}\bigl(g(Y_i)-g_{i,j}(Y_i)\bigr)^2}<M\epsilon/2. \tag{A.8}$$

By (3.7) we can take

$$n_i\le C_0(\epsilon/2)^{-\nu_0}. \tag{A.9}$$

For $1\le j\le n_i$, $1\le i\le m$, set

$$\varphi_{i,j}(\cdot)=\int_{\mathbb{R}^r}g_{i,j}(y)k\left(\frac{y-\cdot}{h_i}\right)dy.$$

Select

$$\varphi(\cdot)=\int_{\mathbb{R}^r}g(y)k\left(\frac{y-\cdot}{h}\right)dy\in\varGamma_r.$$

For the h appearing in $\varphi(\cdot)$ find $1\le i\le m-1$ so that $h_{i-1}<h\le h_i$ and $g_{i,j}$, $1\le j\le n_i$, so that (A.8) holds. Clearly, with

$$\varphi_i(\cdot)=\int_{\mathbb{R}^r}g(y)k\left(\frac{y-\cdot}{h_i}\right)dy,$$

we get by the triangle inequality,

$$\begin{aligned}&\sqrt{\mathbb{E}\bigl(\varphi(X)-\varphi_{i,j}(X)\bigr)^2}\\&\quad\le\sqrt{\mathbb{E}\bigl(\varphi(X)-\varphi_i(X)\bigr)^2}+\sqrt{\mathbb{E}\bigl(\varphi_i(X)-\varphi_{i,j}(X)\bigr)^2}.\end{aligned} \tag{A.10}$$

Notice that

$$\begin{aligned}\mathbb{E}\bigl(\varphi(X)-\varphi_i(X)\bigr)^2&=\mathbb{E}\left(\int_{\mathbb{R}^r}g(y)\left(k\left(\frac{y-X}{h}\right)-k\left(\frac{y-X}{h_i}\right)\right)dy\right)^2\\&\le M^2\mathbb{E}\left(\int_{\mathbb{R}^r}\left|k\left(\frac{y-X}{h}\right)-k\left(\frac{y-X}{h_i}\right)\right|dy\right)^2,\end{aligned}$$

which since H is nonincreasing and $k(u) = H(|u|)$,

$$= M^2\left(\int_{\mathbb{R}^r}\left\{H\left(\frac{|y|}{h_i}\right) - H\left(\frac{|y|}{h}\right)\right\}dy\right)^2$$
$$= M^2\left(h_i^r - h^r\right)^2 \le M^2\left(h_i^r - h_{i-1}^r\right)^2 \le \left(\frac{M\epsilon}{2}\right)^2. \tag{A.11}$$

Also observe that

$$\mathbb{E}\left(\varphi_i(X) - \varphi_{i,j}(X)\right)^2 = \mathbb{E}\left(\int_{\mathbb{R}^r}\left(g(y) - g_{i,j}(y)\right)k\left(\frac{y-X}{h_i}\right)dy\right)^2$$
$$\le h_i^{2r}\int_{\mathbb{R}^r}\left(g(y) - g_{i,j}(y)\right)^2 h_i^{-r}\mathbb{E}\left(k\left(\frac{y-X}{h_i}\right)\right)dy$$
$$= h_i^{2r}\mathbb{E}\left(g(Y_i) - g_{i,j}(Y_i)\right)^2 < \left(\frac{M\epsilon}{2}\right)^2. \tag{A.12}$$

Thus by (A.10), (A.11) and (A.12), we have $\sqrt{\mathbb{E}(\varphi(X) - \varphi_{i,j}(X))^2} \le M\epsilon$. Obviously, from (A.7) and (A.9) we get

$$\#\{\varphi_{i,j} : 1 \le j \le n_i, 1 \le i \le m\} \le 3\left(\epsilon/(2r)\right)^{-1} C_0(\epsilon/2)^{-\nu_0} =: C_1\epsilon^{-\nu_0 - 1}.$$

Hence (A.6) holds. □

References

Altman N, Léger C (1995) Bandwidth selection for kernel distribution function estimation. J Stat Plan Inference 46:195–214

Boos DD (1986) Rates of convergence for the distance between distribution function estimators. Metrika 33:197–202

Bowman A, Hall P, Prvan T (1998) Bandwidth selection for the smoothing of distribution functions. Biometrika 85:799–808

Chung K-L (1949) An estimate concerning the Kolmogoroff limit distribution. Trans Am Math Soc 67:36–50

Dony J, Mason DM (2008) Uniform in bandwidth consistency of conditional U-statistics. Bernoulli 14:1108–1133

Dony J, Einmahl U, Mason DM (2006) Uniform in bandwidth consistency of local polynomial regression function estimators. Aust J Stat 35:105–120

Einmahl U, Mason DM (2000) An empirical process approach to the uniform consistency of kernel-type function estimators. J Theor Probab 13:1–37

Einmahl U, Mason DM (2005) Uniform in bandwidth consistency of kernel-type function estimators. Ann Stat 33:1380–1403

Finkelstein H (1971) The law of the iterated logarithm for empirical distributions. Ann Math Stat 42:607–615

Giné E, Guillou A (2001) On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals. Ann Inst H Poincaré Probab Stat 37:503–522

Giné E, Nickl R (2008) Uniform central limit theorems for kernel density estimators. Probab Theory Relat Fields 141:333–387

Giné E, Nickl R (2009) An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. Probab Theory Relat Fields 143:569–596

Ledoux M, Talagrand M (1989) Comparison theorems, random geometry and some limit theorems for empirical processes. Ann Probab 17:596–631

Mason DM, Shorack GR, Wellner JA (1983) Strong limit theorems for oscillation moduli of the uniform empirical process. Z Wahrsch Verw Geb 65:83–97

Nolan D, Marron JS (1989) Uniform consistency of automatic and location-adaptive delta-sequence estimators. Probab Theory Relat Fields 80:619–632

Nolan D, Pollard D (1987) U-processes: Rates of convergence. Ann Stat 15:780–799

Polansky AM, Baker ER (2000) Multistage plug-in bandwidth selection for kernel distribution function estimates. J Stat Comput Simul 65:63–80

Sarda P (1993) Smoothing parameter selection for smooth distribution functions. J Stat Plan Inference 35:65–75

Serfling RJ (1982) Properties and applications of metrics on nonparametric density estimators. In: Nonparametric statistical inference, vols I, II, Budapest, 1980. Colloq. Math. Soc. János Bolyai, vol 32. North-Holland, Amsterdam, pp 859–873

Stute W (1982) The oscillation behavior of empirical processes. Ann Probab 10:86–107

Stute W (1984) The oscillation behavior of empirical processes: the multivariate case. Ann Probab 12:361–379

Swanepoel JWH (1987). A note on a functional law of the iterated logarithm for kdfe's with applications. FABWI-N-WST

Swanepoel JWH, Van Graan FC (2005) A new kernel distribution function estimator based on a nonparametric transformation of the data. Scand J Stat 32:551–562

Talagrand M (1994) Sharper bounds for Gaussian and empirical processes. Ann Probab 22:28–76

van der Vaart AW (1994) Weak convergence of smoothed empirical processes. Scand J Stat 21:501–504

van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, New York

van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes with applications to statistics. Springer, New York

Yukich JE (1992) Weak convergence of smoothed empirical processes. Scand J Stat 19:271–279