IMS Collections
From Probability to Statistics and Back: High-Dimensional Models and Processes
Vol. 9 (2013) 241–253

DOI: 10.1214/12-IMSCOLL917

Uniform in bandwidth limit laws for kernel distribution function estimators

David M. Mason[1,*] **and Jan W. H. Swanepoel**[2,*]

University of Delaware and North-West University

Abstract: We use results from modern empirical process theory to establish a uniform in bandwidth central limit theorem, laws of the iterated logarithm and Glivenko–Cantelli theorem for kernel distribution function estimators.

1. Introduction and main results

Let X, $X_1, X_2, \ldots,$ be i.i.d. taking values in $\mathbb{R}^d$ having multivariate cumulative distribution function [c.d.f.] F and for each $n \geq 1$ let

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\}, \qquad x \in \mathbb{R}^d,$$

denote the empirical distribution function based on the first n of these random variables, where for x, $y \in \mathbb{R}^d$, $y \leq x$ means that each component of y is $\leq$ each component of x. Next let K be a multivariate c.d.f. on $\mathbb{R}^d$. In this paper we shall obtain a uniform in bandwidth central limit theorem [CLT], law of the iterated logarithm [LIL] and a Glivenko-Cantelli theorem for the *kernel distribution function estimator [kdfe]*, which using the notation of Mason and Swanepoel [14], is defined as

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

The *smoothed empirical distribution function* is a special case of $\widehat{F}_{n,h}(x)$. To see this, let k be a density function and consider the kernel density estimator

$$f_{n,h}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right), \qquad x \in \mathbb{R}^d, \tag{1.1}$$

where k is a density on $\mathbb{R}^d$. The smoothed empirical distribution function becomes with $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$\widehat{S}_{n,h}(x) := \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_{n,h}(y)\, dy = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \qquad x \in \mathbb{R}^d,$$

*Research partially an NSF Grant.
†Research partially supported by NRF of South Africa.
[1]Statistics Program, University of Delaware, 206 Townsend Hall, Newark, DE 19716, USA, e-mail: davidm@udel.edu
[2]North-West University, Potchefstroom Campus, Potchefstroom, South Africa, e-mail: Jan.Swanepoel@nwu.ac.za
Primary 62F15, 62G07; secondary 62G08
Keywords and phrases: Empirical processes, kernel estimation, distribution function and uniform in bandwidth

where $K(t) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} k(y)\,dy$ for $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$.

Essential to our approach will be a CLT and compact LIL for the empirical process applied to the following class of functions. Consider for any $0 < b \le 1$

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

where $\mathbf{0}$ denotes the function of $x \in \mathbb{R}^d$ constantly equal to zero. Let $\mathbb{P}$ denote the probability measure induced by F on $\mathbb{R}^d$. The class $\mathcal{K}_b$ is bounded and separable as defined in Section 2. It will be shown in Section 1.1 that for any $0 < b \le 1$, $\mathcal{K}_b$ is also a $\mathbb{P}$-Donsker class. This means that the empirical process indexed by $\varphi = K(\frac{x-\cdot}{h}) - 1\{\cdot \le x\} \in \mathcal{K}_b$

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

which can be written as

$$\sqrt{n}\big\{ \widehat{F}_{n,h}(x) - F_n(x) - \big(\mathbb{E}\widehat{F}_{n,h}(x) - F(x)\big) \big\}, \tag{1.2}$$

converges weakly in the sense of Hoffmann-Jørgensen to a mean zero Gaussian process with covariance function $cov(\varphi_1(X), \varphi_2(X))$ and is continuous in the semi-metric

$$\rho(\varphi_1, \varphi_2) = \mathbb{E}\big[\varphi_1(X) - \varphi_2(X) - \mathbb{E}\big(\varphi_1(X) - \varphi_2(X)\big)\big]^2, \qquad \varphi_1, \varphi_2 \in \mathcal{K}_b.$$

Since for any $0 < b \le 1$, the class $\mathcal{K}_b$ is $\mathbb{P}$-Donsker we have

$$\sup_{\varphi \in \mathcal{K}_b} |\alpha_n(\varphi)| =: \Delta_n(b) = O_P(1), \tag{1.3}$$

and since it also satisfies all the conditions for the compact LIL for the empirical process (see (2.2) below), we get

$$\limsup_{n \to \infty} \frac{\Delta_n(b)}{\sqrt{2 \log\log n}} = \sigma(b) < \infty, \quad \text{a.s.}, \tag{1.4}$$

where

$$\begin{aligned} \sigma^2(b) &= \sup_{\varphi \in \mathcal{K}_b} \mathrm{Var}\big(\varphi(X)\big) \\ &= \sup_{0 < h \le b,\ x \in \mathbb{R}^d} \mathrm{Var}\left(K\left(\frac{x - X}{h}\right) - 1\{X \le x\} \right). \end{aligned}$$

For future use, we set

$$\begin{aligned} \mu(b) &= \sup_{\varphi \in \mathcal{K}_b} \big|\mathbb{E}\varphi(X)\big| \\ &= \sup_{0 < h \le b,\ x \in \mathbb{R}^d} \left| \mathbb{E}K\left(\frac{x - X}{h}\right) - F(x) \right|. \end{aligned}$$

Observe that

$$\begin{aligned} \frac{\Delta_n(b)}{\sqrt{2\log\log n}} &= \frac{\sup_{0<h\le b} \sup_{x \in \mathbb{R}^d} \sqrt{n} |\widehat{F}_{n,h}(x) - F_n(x) - (\mathbb{E}\widehat{F}_{n,h}(x) - F(x))|}{\sqrt{2\log\log n}} \\ &= \frac{\sup_{0<h\le b} \sup_{x \in \mathbb{R}^d} |\sqrt{n}\{\widehat{F}_{n,h}(x) - \mathbb{E}\widehat{F}_{n,h}(x)\} - \widetilde{\alpha}_n(x)|}{\sqrt{2\log\log n}}, \end{aligned}$$

where $\widetilde{\alpha}_n(x)$, $x \in \mathbb{R}^d$, is the classical empirical process

$$\widetilde{\alpha}_n(x) = \sqrt{n}\big\{F_n(x) - F(x)\big\}, \qquad x \in \mathbb{R}^d. \tag{1.5}$$

(Note that $\widetilde{\alpha}_n(x) := \alpha_n(1\{\cdot \le x\})$.) Summarizing these observations we get the following general proposition.

Proposition 1.1. *For any $0 < b \le 1$, (1.3) and (1.4) hold.*

Notice that (1.3) and (1.4) readily imply the following result.

Corollary 1.2. *Whenever $\sigma^2(b) \to 0$ as $b \searrow 0$, for any sequence of positive constants $0 < b_n < 1$ satisfying $b_n \to 0$,*

$$\Delta_n(b_n) = o_P(1) \tag{1.6}$$

and w.p. 1 as $n \to \infty$,

$$\frac{\Delta_n(b_n)}{\sqrt{2\log\log n}} = o(1). \tag{1.7}$$

Proof. Assertion (1.6) follows from the fact that $\mathbb{P}$-Donsker implies (cf. Theorem 3.7.2 in Dudley [3]) that for every $\varepsilon > 0$ there exist a $\delta > 0$ such that

$$\mathbb{P}\big\{\sup\big\{\|\varphi_1 - \varphi_2\|_{\mathbb{R}^d} : \rho(\varphi_1, \varphi_2) < \delta,\ \varphi_1, \varphi_2 \in \mathcal{K}_b\big\} > \varepsilon\big\} < \varepsilon,$$

$\mathbf{0} \in \mathcal{K}_b$ and the assumption that $\sigma^2(b) \to 0$ as $b \searrow 0$. To see that assertion (1.7) holds notice that for any $b > 0$

$$\limsup_{n\to\infty} \Delta_n(b_n)/\sqrt{2\log\log n} \le \limsup_{n\to\infty} \Delta_n(b)/\sqrt{2\log\log n} = \sigma(b),$$

and $\sigma(b) \to 0$ as $b \searrow 0$. $\square$

For $d \ge 1$ let

$$\mathcal{F} = \big\{f_x(\cdot) := 1\{\cdot \le x\} : x \in \mathbb{R}^d\big\}. \tag{1.8}$$

This is also a $\mathbb{P}$-Donsker class. Moreover it is *translation invariant* in the sense that for every $f_x(\cdot) \in \mathcal{F}$ and $y \in \mathbb{R}^d$ the function $f_x(\cdot + y) \in \mathcal{F}$. It is well-known that the classical empirical process $\widetilde{\alpha}_n(x)$ converges weakly to the d–variate Brownian bridge $B(x)$, $x \in \mathbb{R}^d$, i.e. B is a mean zero Gaussian process indexed by $\mathcal{F}$ with covariance function $\mathbb{E}(B(x)B(y)) = F(x \wedge y) - F(x)F(y)$, $x, y \in \mathbb{R}^d$, where $x \wedge y$ denotes the vector $(x_1 \wedge y_1, \ldots, x_d \wedge y_d)$. Corollary 1.2 implies that whenever $\sigma^2(b) \to 0$ as $b \searrow 0$, the *centered kdfe process*

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

converges weakly uniformly in bandwidth to $B(\cdot)$ in the sense that (1.6) holds.

Note that trivially we get from (1.7):

Corollary 1.3. *Whenever $\sigma^2(b) \to 0$ as $b \searrow 0$ and b_n is a sequence of positive constants $0 < b_n < 1$ satisfying $b_n \to 0$, and*

$$\sup_{0<h\le b_n}\sup_{x\in\mathbb{R}^d} \sqrt{n}\big|\mathbb{E}\widehat{F}_{n,h}(x) - F(x)\big| = o(\sqrt{\log\log n}), \tag{1.9}$$

then, w.p. 1,

$$\sup_{0<h\le b_n}\sup_{x\in\mathbb{R}^d} \sqrt{n}\big|\widehat{F}_{n,h}(x) - F_n(x)\big| = o(\sqrt{\log\log n}). \tag{1.10}$$

We shall infer our uniform in bandwidth LIL and Glivenko-Cantelli theorem from Corollaries 1.2 and 1.3.

First we can easily derive from (1.7) a uniform in bandwidth Finkelstein-type functional law of the iterated logarithm [FLIL] for the sequence of classes of random functions

$$\mathcal{F}_n = \left\{ \frac{\eta_{n,h}(\cdot)}{\sqrt{2\log\log n}} : \ 0 < h \le b_n \right\}.$$

Towards formulating our uniform in bandwidth FLIL, let $\ell^\infty(\mathcal{F})$ denote the space of bounded functions φ on $\mathcal{F}$ equipped with supremum norm

$$\|\varphi\|_{\mathcal{F}} = \sup_{x\in\mathbb{R}^d} \left|\varphi\big(1\{\cdot \le x\}\big)\right|.$$

Further let H denote the subset of $\ell^\infty(\mathcal{F})$ consisting of functions of $\mathcal{F}$ of the form

$$1\{\cdot \le x\} \mapsto \int_{\mathbb{R}^d} \big(1\{y \le x\} - F(x)\big) h(y) F(dy),$$

where $\int_{\mathbb{R}^d} h^2(y)F(dy) \le 1$. For all $\varepsilon > 0$ set

$$H^\varepsilon = \left\{ \phi : \phi \in \ell^\infty(\mathcal{F}) \text{ and } \inf_{\varphi\in H} \|\varphi - \phi\|_{\mathcal{F}} \le \varepsilon \right\}$$

and for each $\varphi \in H$ let

$$B_\varepsilon(\varphi) = \big\{ \phi : \phi \in \ell^\infty(\mathcal{F}) \text{ and } \|\varphi - \phi\|_{\mathcal{F}} \le \varepsilon \big\}.$$

A special case of the compact LIL formulated in Section 2 gives the following $\mathbb{R}^d$ version of the Finkelstein FLIL for the empirical process $\widetilde{\alpha}_n(\cdot)$: *w.p.* 1, for all $\varepsilon > 0$ there exists an $N > 0$ such that for all $n \ge N$,

$$\frac{\widetilde{\alpha}_n(\cdot)}{\sqrt{2\log\log n}} \in H^\varepsilon$$

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

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

When $d = 1$ and F is continuous this gives the Finkelstein [6] FLIL. Clearly it is routine to combine (1.7) with the $\mathbb{R}^d$ version of the Finkelstein FLIL just stated to infer the following FLIL.

Corollary 1.4. *Whenever $\sigma^2(b) \to 0$ as $b \searrow 0$, for any sequence of positive constants $0 < b_n < 1$ satisfying $b_n \to 0$, w.p. 1, for all $\varepsilon > 0$ there exists an N such that $\mathcal{F}_n \subset H^\varepsilon$ for all $n \ge N$ and for every $\varphi \in H$ there exists a subsequence $\{n_k\}$ such that for all $k \ge 1$,*

$$\frac{\eta_{n_k,h}(\cdot)}{\sqrt{2\log\log n_k}} \in B_\varepsilon(\varphi), \qquad \text{uniformly in } 0 < h \le b_{n_k}. \tag{1.11}$$

Notice that, in particular, under the conditions of Corollary 1.4 we can infer the following uniform in bandwidth version of the Chung [2] LIL

$$\limsup_{n\to\infty} \sup_{0<h\le b_n} \sup_{x\in\mathbb{R}^d} \frac{\sqrt{n}|\widehat{F}_{n,h}(x) - \mathbb{E}\widehat{F}_{n,h}(x)|}{\sqrt{2\log\log n}} = \sup_{x\in\mathbb{R}^d} \sqrt{F(x)(1-F(x))}, \text{ a.s.} \tag{1.12}$$

Recall that the classical Glivenko-Cantelli theorem says that *w.p.* 1

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^d} \left|F_n(x) - F(x)\right| = 0. \tag{1.13}$$

Next we get the following uniform in bandwidth Glivenko-Cantelli theorem for $\widehat{F}_{n,h}$.

Proposition 1.5. *For any sequence of positive constants $0 < b_n < 1$ satisfying $b_n \to 0$, w.p. 1, we have*

$$\lim_{n\to\infty} \sup_{0<h\le b_n} \sup_{x\in\mathbb{R}^d} \left|\widehat{F}_{n,h}(x) - \mathbb{E}\widehat{F}_{n,h}(x)\right| = 0 \tag{1.14}$$

and whenever $\mu(b) \to 0$ as $b \searrow 0$

$$\lim_{n\to\infty} \sup_{0<h\le b_n} \sup_{x\in\mathbb{R}^d} \left|\widehat{F}_{n,h}(x) - F(x)\right| = 0. \tag{1.15}$$

Proof. Assertion (1.14) follows immediately from (1.4) and (1.13). Whenever $\mu(b) \to 0$ as $b \searrow 0$, assertion (1.15) is an immediate consequence of (1.14). □

These uniform in bandwidth limit theorems take on a statistical importance when h is replaced by an estimator $\widehat{h}_n$ based on the sample $X_1, \dots, X_n$ chosen by some optimality criterion. They imply that as long as $\widehat{h}_n \to 0$, *w.p.* 1, the CLT, LIL and Glivenko–Cantelli theorem hold for their plug-in versions based on $\widehat{F}_{n,\widehat{h}_n}$. For more uniform in bandwidth results for kernel-type nonparametric function estimators and discussions of their uses consult Einmahl and Mason [5], Mason and Swanepoel [14] and Mason [13], as well as the references therein. For work on bandwidth selection for kernel estimators of the c.d.f. refer to Janssen, Swanepoel and Veraverbeke [10], Swanepoel [16] and Swanepoel and van Graan [17].

Our results have largely followed from bookkeeping. To apply them we must clarify when $\mu(b) \to 0$ and $\sigma^2(b) \to 0$ as $b \searrow 0$. In the next subsection we discuss conditions for this to happen. We shall also show that $\mathcal{K}_b$ is a $\mathbb{P}$-Donsker class for any $0 < b \le 1$.

1.1. Sufficient conditions and an open problem

We shall first show that $\mathcal{K}_b$ is $\mathbb{P}-$Donsker for any $0 < b \le 1$. To do this we shall apply a result of Giné and Nickl [8]. As above we use the notation $\mathcal{F} = \{1\{\cdot \le x\} : x \in \mathbb{R}^d\}$. For any $0 < h \le 1$ let $P_h(\cdot)$ be the probability measure induced on $\mathbb{R}^d$ by the multivariate c.d.f $K(\cdot/h)$ and let μ_h denote the corresponding probability measure induced on $\mathbb{R}^d$ by defining for any Borel subset $A \subset \mathbb{R}^d$, $\mu_h(A) = P_h(-A)$. Further let $\mathcal{M} = \{\mu_h : 0 < h \le 1\}$. Notice that for any $\mu_h \in \mathcal{M}$ and $f_x(\cdot) := 1\{\cdot \le x\} \in \mathcal{F}$

$$f_x * \mu_h(\cdot) = K\left(\frac{x - \cdot}{h}\right). \tag{1.16}$$

From identity (1.16) and the fact that $\mathcal{F}$ is a translation invariant $\mathbb{P}$-Donsker class, we can apply Lemma 2 of Giné and Nickl [8] to conclude that

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

is a $\mathbb{P}$-Donsker class. This implies via Theorem 3.8.1, p. 121, and Exercise 6, p. 127, in Dudley [3] that $\mathcal{K}_b$ is $\mathbb{P}$-Donsker for any $0 < b \le 1$.

We shall next derive some sufficient conditions for $\mu(b) \to 0$ and $\sigma^2(b) \to 0$ as $b \searrow 0$. To ease some needed calculations, let X, U and V be independent, where

X has c.d.f. F, U has c.d.f. K and V has c.d.f. K^2. Notice that

$$\mathbb{E}K\left(\frac{x-X}{h}\right) = \int_{\mathbb{R}^d} K\left(\frac{x-t}{h}\right)F(dt)$$
$$= P\{hU + X \le x\} = \int_{\mathbb{R}^d} F(x-hz)K(dz).$$

Therefore

$$(1.17)\qquad \begin{aligned}\mu(x,h) :=& \mathbb{E}\left(K\left(\frac{x-X}{h}\right) - 1\{X \le x\}\right)\\ =& \int_{\mathbb{R}^d} \big(F(x-hz) - F(x)\big)\, dK(z).\end{aligned}$$

Now noting that

$$\mathbb{E}K^2\left(\frac{x-X}{h}\right) = \int_{\mathbb{R}^d} K^2\left(\frac{x-t}{h}\right)F(dt)$$
$$= P\{hV + X \le x\} = \int_{\mathbb{R}^d} F(x-hz)K^2(dz)$$

and

$$\mathbb{E}\left(1\{X \le x\}K\left(\frac{x-X}{h}\right)\right) = P\{hU + X \le x, X \le x\}$$
$$= \int_{\mathbb{R}^d} F\big((x-hz)\wedge x\big)\, dK(z),$$

we see that

$$(1.18)\qquad \begin{aligned}\sigma^2(x,h) :=& \operatorname{Var}\left(K\left(\frac{x-X}{h}\right) - 1\{X \le x\}\right)\\ =& \int_{\mathbb{R}^d} \big(F(x-hz) - F(x)\big)\, dK^2(z)\\ &- 2\int_{\mathbb{R}^d} \Big(F\big((x-hz)\wedge x\big) - F(x)\Big)\, dK(z)\\ &- \left(\int_{\mathbb{R}^d} \big(F(x-hz) - F(x)\big)\, dK(z)\right)^2.\end{aligned}$$

We note for future reference that when $d = 1$

$$(1.19)\qquad \begin{aligned}\sigma^2(x,h) :=& \int_{\mathbb{R}} \big(F(x-hz) - F(x)\big)\, dK^2(z)\\ &- 2\int_0^{\infty} \big(F(x-hz) - F(x)\big)\, dK(z)\\ &- \left(\int_{\mathbb{R}} \big(F(x-hz) - F(x)\big)\, dK(z)\right)^2.\end{aligned}$$

It can be readily verified using equation (1.17) that whenever F is continuous or $K(\cdot) = 1\{\cdot \ge 0\}$ then

$$(1.20)\qquad \lim_{b\searrow 0} \mu(b) = 0$$

and using equation (1.18) that whenever F is continuous or degenerate or $K(\cdot) = 1\{\cdot \ge 0\}$ then

$$(1.21)\qquad \lim_{b\searrow 0} \sigma^2(b) = 0.$$

An open problem is to characterize when $\mu(b) \to 0$ and $\sigma^2(b) \to 0$ as $b \searrow 0$. In this subsection we shall solve it in the case $d = 1$ and in the process obtain the following refinement to Proposition 1.5 when $d = 1$, namely, we can add that when $d = 1$, (1.15) holds if and only if F is continuous or $K(\cdot) = 1\{\cdot \geq 0\}$. In addition, if F is not continuous and $K(0) - K(0-) < 1$, then, *w.p.* 1,

$$\liminf_{n\to\infty} \sup_{0<h\leq b_n} \sup_{x\in\mathbb{R}} \left|\widehat{F}_{n,h}(x) - F(x)\right| > 0. \tag{1.22}$$

When $d = 1$, a version of this result for $\widehat{F}_{n,h_n}$ for a fixed sequence h_n converging to zero can be inferred from those in Chacón and Rodríguez-Casal [1]. They base their arguments on probability inequalities. Yamato [18] was the first to show that the Glivenko-Cantelli theorem holds for $\widehat{F}_{n,h_n}$ when $d = 1$, F is continuous and h_n goes to zero at a certain rate.

Our refinement to Proposition 1.5 is an immediate consequence of the following lemma. The proof of part of it is contained in the arguments in the proof of Theorem 1 of Chacón and Rodríguez-Casal [1].

Lemma 1.6. *When $d = 1$, F is continuous or $K(\cdot) = 1\{\cdot \geq 0\}$ if and only if*

$$\lim_{b\searrow 0} \mu(b) = 0. \tag{1.23}$$

Proof. It has already been pointed out that if F is continuous then (1.23) holds, and obviously, (1.23) holds if $K(\cdot) = 1\{\cdot \geq 0\}$, since in this case for all $h > 0$ and $x \in \mathbb{R}$, $K(\frac{x-\cdot}{h}) = 1\{\cdot \leq x\}$.

Next, assume that (1.23) holds and F has a discontinuity point at x. In this case as $h \searrow 0$,

$$\int_{(-\infty,0]} F(x - hz)\, dK(z) \to \int_{(-\infty,0]} F(x)\, dK(z) = F(x)K(0)$$

and

$$\int_{(0,\infty)} F(x - hz)\, dK(z) \to \int_{(0,\infty)} F(x-)\, dK(z) = F(x-)\big(1 - K(0)\big).$$

This says that as $h \searrow 0$

$$F_h(x) - F(x) \to \big(F(x) - F(x-)\big)\big(K(0) - 1\big).$$

Hence if $K(0) < 1$, (1.23) cannot hold. Thus we must have $K(0) = 1$.

Suppose now that x is a discontinuity point and $K(0) = 1$, but $K(0) - K(0-) < 1$. In this case there exists a $\gamma > 0$ such that $K(-\gamma) > 0$. We get for any $h > 0$

$$\begin{aligned}
&\int_{\mathbb{R}} \big(F(x - h\gamma - zh) - F(x - h\gamma)\big)\, dK(z) \\
&\quad = \int_{(-\infty,0]} \big(F(x - h\gamma - zh) - F(x - h\gamma)\big)\, dK(z) \\
&\quad = \int_{-\gamma}^{0-} \big(F(x - h\gamma - zh) - F(x - h\gamma)\big)\, dK(z) \\
&\qquad + \int_{-\infty}^{-\gamma} \big(F(x - h\gamma - zh) - F(x - h\gamma)\big)\, dK(z),
\end{aligned}$$

which converges as $h \searrow 0$ to

$$\begin{aligned}&\big(F(x-) - F(x-)\big)\big(K(0-) - K(-\gamma+)\big) + \big(F(x) - F(x-)\big)K(-\gamma)\\&\quad = \big(F(x) - F(x-)\big)K(-\gamma) > 0.\end{aligned}$$

This shows that (1.23) cannot be true unless $K(\cdot) = 1\{\cdot \geq 0\}$.

An examination of the arguments just given shows if F is not continuous and $K(0) - K(0-) < 1$, then (1.23) cannot hold. Thus in order for (1.23) to be valid F must be continuous or $K(\cdot) = 1\{\cdot \geq 0\}$. □

Remark 1.7. Notice that we have also shown, when $d = 1$, that

$$\lim_{b \searrow 0} \mu(b) > 0 \tag{1.24}$$

if and only if F is not continuous and $K(0) - K(0-) < 1$.

The next lemma specifies exactly when $\sigma^2(b) \to 0$ as $b \searrow 0$, when $d = 1$.

Lemma 1.8. *When $d = 1$, F is continuous or F is degenerate or $K(\cdot) = 1\{\cdot \geq 0\}$ if and only if*

$$\lim_{b \searrow 0} \sigma^2(b) = 0. \tag{1.25}$$

Proof. Again, as has already been pointed out, if F is continuous then (1.25) is valid, and it holds trivially if $K(\cdot) = 1\{\cdot \geq 0\}$, since for all $h > 0$ and $x \in \mathbb{R}$, $K(\frac{x-\cdot}{h}) = 1\{\cdot \leq x\}$.

If F is degenerate there is an x_0 such that $F(x_0) - F(x_0-) = 1$. Thus with probability 1, $K(\frac{x-X}{h}) - 1\{X \leq x\} = K(\frac{x-x_0}{h}) - 1\{x_0 \leq x\}$, so clearly $\sigma^2(x, h) = 0$.

Now assume that F is not degenerate but is not continuous and recall the expression for $\sigma^2(x, h)$ given in (1.19).

Case 1. If x is a discontinuity point of F and $K(0) < 1$ as $h \searrow 0$, we get by using expression (1.19) for $\sigma^2(x, h)$ that

$$\begin{aligned}\sigma^2(x, h) \to\ & 2\big(1 - K(0)\big)\big(F(x) - F(x-)\big)\\&- \big(1 - K^2(0)\big)\big(F(x) - F(x-)\big) - \big(1 - K(0)\big)^2\big(F(x) - F(x-)\big)^2.\end{aligned}$$

This last expression equals

$$\big(1 - K(0)\big)^2\big[\big(F(x) - F(x-)\big) - \big(F(x) - F(x-)\big)^2\big],$$

which is strictly positive since F is not degenerate. Thus (1.25) cannot hold.

Case 2. If x is a discontinuity point of F and $K(0) = 1$, but $K(0) - K(0-) < 1$. In this case there exists a $\gamma > 0$ such that $K(-\gamma) > 0$. We get as $h \searrow 0$,

$$\begin{aligned}\sigma^2(x - h\gamma, h) &= \int_{(-\infty,0]} \big(F(x - h\gamma - hz) - F(x - h\gamma)\big)\, dK^2(z)\\&\quad - \left(\int_{(-\infty,0]} \big(F(x - h\gamma - hz) - F(x - h\gamma)\big)\, dK(z)\right)^2\\&\to \Big(\big(F(x) - F(x-)\big) - \big(F(x) - F(x-)\big)^2\Big)K^2(-\gamma) > 0.\end{aligned}$$

Hence (1.25) is not valid.

Examining the above arguments we see that if F is not continuous, $K(0) - K(0-) < 1$ and F is not degenerate then (1.25) does not hold. □

Remark 1.9. Notice that when $d = 1$, F is not continuous, F is not degenerate and $K(0) - K(0-) < 1$ if and only if

$$\lim_{b \searrow 0} \sigma^2(b) > 0. \tag{1.26}$$

Remark 1.10. One may conjecture that Lemmas 1.6 and 1.8, as well as the above addition to Proposition 1.5 remain true in dimension $d \geq 2$. However this is not the case. Consider (X, Y), $(X_1, Y_1), (X_2, Y_2), \ldots$, i.i.d.,where X is independent of Y and X has continuous c.d.f. F_X and Y has discontinuous c.d.f. F_Y. Consider the kernel defined for $(s, t) \in \mathbb{R}^2$ by

$$H(s, t) = K(s)1\{t \geq 0\},$$

where K is a continuous c.d.f. and its corresponding class of functions

$$\mathcal{H}_b = \left\{K\left(\frac{x - \cdot}{h}\right)1\{\cdot \leq y\} - 1\big\{(\cdot, \cdot) \leq (x, y)\big\} : (x, y) \in \mathbb{R}^2,\ 0 < h \leq b\right\}$$

for $0 < b \leq 1$. Using Lemma 22 of Nolan and Pollard [15] one sees that the classes of functions $\mathcal{F}$ and $\{K(\frac{x-\cdot}{h}) : x \in \mathbb{R},\ 0 < h \leq 1\}$ have polynomial covering numbers, from which one readily infers that $\mathcal{H}_1$ has a polynomial covering number. (See, for instance, Lemma A.1 in Einmahl and Mason [4].) This implies that $\mathcal{H}_b$ is $\mathbb{P}$-Donsker for any $0 < b \leq 1$.

Now

$$\mathbb{E}H\left(\frac{x - X}{h}, \frac{y - Y}{h}\right) = \mathbb{E}K\left(\frac{x - X}{h}\right)F_Y(y),$$
$$\mathbb{E}H^2\left(\frac{x - X}{h}, \frac{y - Y}{h}\right) = \mathbb{E}K^2\left(\frac{x - X}{h}\right)F_Y(y)$$

and

$$\mathbb{E}H\left(\frac{x - X}{h}, \frac{y - Y}{h}\right)1\big\{(X, Y) \leq (x, y)\big\} = \mathbb{E}\left[K\left(\frac{x - X}{h}\right)1\{X \leq x\}\right]F_Y(y).$$

From these expressions one readily verifies that $\mu(b) \to 0$ and $\sigma^2(b) \to 0$ as $b \searrow 0$, so that Corollaries 1.2-1.4 apply, along with the second part of Proposition 1.5. However, clearly the joint c.d.f. of (X, Y) is not continuous, not degenerate and $H(x, y) \neq 1\{(x, y) \geq (0, 0)\}$.

1.2. A straightforward generalization of our results

Though in our note we have focused on the kernel distribution function estimator, our analysis can be readily extended to the following more general indexed by function setup. Let $\mathcal{G}$ be a bounded separable translation invariant class of measurable real valued $\mathbb{P}-$Donsker functions defined on $\mathbb{R}^d$ and for any $0 < h \leq 1$ let μ_h be as above. Consider the process indexed by $g \in \mathcal{G}$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big\{g * \mu_h(X_i) - g(X_i) - \mathbb{E}g * \mu_h(X) + \mathbb{E}g(X)\big\}. \tag{1.27}$$

Specializing to $\mathcal{G} = \mathcal{F}$ we get (1.2). It should be clear to the reader that a straightforward extension of the arguments just given show that the obvious analogs of Proposition 1.1 and Corollaries 1.2-1.4 hold for the process (1.27).

1.3. Comparison with results in Mason and Swanepoel [14]

In this subsection $d = 1$. We assume that F satisfies the Lipschitz condition

$$|F(x) - F(y)| \le C|x - y|, \quad \text{for all } x, y \in \mathbb{R}, \text{ some } 0 < C < \infty, \tag{1.28}$$

and that

$$\int_{-\infty}^{\infty} |z|\, dK(z) < \infty. \tag{1.29}$$

Using empirical process inequalities Mason and Swanepoel [MS] [14] obtained a general uniform in bandwidth theorem that, when specialized to the kernel distribution function estimator, yields the following proposition.

Proposition 1.11 (Proposition 1 MS [14]). *Assume that F satisfies (1.28) and K fulfills(1.29). Then for $c > 0$, $0 < h_0 < 1$, w.p. 1, for some constant $0 < A(c) < \infty$,*

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n} \le h \le h_0} \sup_{x\in\mathbb{R}} \frac{\sqrt{n}|\widehat{F}_{n,h}(x) - F_n(x) - (\mathbb{E}\widehat{F}_{n,h}(x) - F(x))|}{\sqrt{h(|\log h| \vee \log\log n)}} = A(c). \tag{1.30}$$

They then pointed out that (1.30) readily implies the following corollaries, which show that under more smoothness assumptions one can obtain rates in our Corollaries 1.2 and 1.3, however with the supremum taken over $\frac{c\log n}{n} \le h \le b_n$ instead of $0 \le h \le b_n$.

Corollary 1.12 (Corollary 1 MS [14]). *Under the assumptions of Proposition 1.11, for any sequence of positive constants $0 < b_n < 1$ satisfying $b_n \to 0$ and $b_n \ge c\log n/n$, w.p. 1*

$$\begin{aligned} &\frac{\sup_{\frac{c\log n}{n} \le h \le b_n} \sup_{x\in\mathbb{R}} |\sqrt{n}\{\widehat{F}_{n,h}(x) - \mathbb{E}\widehat{F}_{n,h}(x)\} - \widetilde{\alpha}_n(x)|}{\sqrt{\log\log n}} \\ &\quad = O\left(\sqrt{b_n\left(\frac{|\log b_n|}{\log\log n} \vee 1\right)}\right) = o(1). \end{aligned} \tag{1.31}$$

Corollary 1.13 (Corollary 2 MS [14]). *Under the assumptions of Proposition 1.11, if $b_n \to 0$, $b_n \ge c\log n/n$ and*

$$\sup_{\frac{c\log n}{n} \le h \le b_n} \sup_{x\in\mathbb{R}} \sqrt{n}|\mathbb{E}\widehat{F}_{n,h}(x) - F(x)| = O\big(\sqrt{b_n(\log\log n \vee |\log b_n|)}\big) \tag{1.32}$$

then

$$\sup_{\frac{c\log n}{n} \le h \le b_n} \sqrt{n}\|\widehat{F}_{n,h} - F_n\|_\infty = O\big(\sqrt{b_n(\log\log n \vee |\log b_n|)}\big) \ a.s. \tag{1.33}$$

Notice that these results show that with added smoothness assumptions one can obtain almost sure rates in Proposition 1.1 and Corollaries 1.2 and 1.3. Mason and Swanepoel [14] also derived the following uniform in bandwidth refinement to the Finkelstein FLIL, where here we use the same notation as in Corollary 1.4.

Corollary 1.14 (Corollary 3 MS [14])**.** *Under the assumptions of Proposition 1.11, for any sequence of positive constants b_n satisfying $b_n \to 0$ and $b_n \geq c \log n/n$, w.p. 1, for all $\varepsilon > 0$ there exists an N such that $\mathcal{F}_n \subset H^\varepsilon$ for all $n \geq N$ and for every $\varphi \in H$ there exists a subsequence $\{n_k\}$ such that for all $k \geq 1$,*

$$\frac{\eta_{n_k,h}(\cdot)}{\sqrt{2\log\log n_k}} \in B_\varepsilon(\varphi), \text{ uniformly in } \frac{c\log n_k}{n_k} \leq h \leq b_{n_k}. \tag{1.34}$$

Remark 1.15. Clearly our Corollary 1.4 is an improvement of Corollary 3 of MS [14], since we do not require F to be Lipschitz and we can replace $\frac{c\log n_k}{n_k} \leq h \leq b_{n_k}$ by $0 < h \leq b_{n_k}$. We also point out that Giné and Nickl [9] have obtained under smoothness conditions a FLIL for the special case of the *kdfe process* when $\widehat{F}_{n,h_n}(x) = \int_{-\infty}^x f_{n,h_n}(y)dy$, where f_{n,h_n} is the kernel density estimator defined in (1.1) and h_n goes to zero such that $h_n \geq \log n/n$ and $\sup_{n\geq 1} \sqrt{n} h_n^{1+t} < \infty$ for some $t > 0$ depending on the smoothness of F. When the kernel k used to define f_{n,h_n} is a bounded density, Corollary 1.4 implies 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 1.16. We take this opportunity to point out the following corrections in Mason and Swanepoel [14]:

1. Replace "$(\log n)^{-1}$" by "$c \log n/n$" in their Corollaries 1-3 and three lines above equation (2.16).
2. In equation (2.5) replace "$O(\sqrt{b_n})$" by "$= O(\sqrt{b_n(\frac{|\log b_n|}{\log\log n} \vee 1)}) = o(1)$".
3. In equations (2.6) and (2.7) and in the two lines below their Corollary 2, replace "$\sqrt{b_n \log\log n}$" by "$\sqrt{b_n(\log\log n \vee |\log b_n|)}$".
4. Three lines below equation (2.15) replace "$0 < b_n < 1$ satisfying $b_n \geq (\log n)^{-1}$ and $\sqrt{n} b_n/\sqrt{\log\log n} = o(1)$" by "$b_n$ satisfying $b_n \geq c \log n/n$ and $b_n \to 0$".

2. Compact LIL

In this section we clarify for the convenience of the reader the compact LIL used in (1.4) above. The presentation in this section is adapted from material in Giné and Mason [7]. Let $X, X_1, X_2, \ldots,$ be i.i.d. random variables from a probability space $(\Omega, \mathcal{A}, P)$ to a measure space $(S, \mathcal{S})$. Consider an empirical process indexed by a class $\mathcal{F}$ of measurable real valued functions on $(S, \mathcal{S})$ defined by

$$\sqrt{n}(P_n - P)\varphi = \frac{\sum_{i=1}^n \varphi(X_i) - n\mathbb{E}\varphi(X)}{\sqrt{n}}, \qquad \varphi \in \mathcal{F}.$$

Assume that the class $\mathcal{F}$ is separable for P (P-separable) in the following sense:

Definition 1. A class $\mathcal{F}$ is separable for P if, for each n, the process $(P_n - P)\varphi$, $\varphi \in \mathcal{F}$, is separable. This means that there exists a countable set $\mathcal{F}_0 \subseteq \mathcal{F}$ such that for each φ in $\mathcal{F}$,

$$(P_n - P)\varphi \in \overline{\{(P_n - P)g : g \in \mathcal{F}_0, \|\varphi - g\|_{L_2(P)} \leq \varepsilon\}},$$

for every $\varepsilon > 0$, where $\overline{A}$ denotes the closure of a set A and

$$\|\varphi - g\|_{L_2(P)}^2 = \mathbb{E}\big(\varphi(X) - g(X)\big)^2.$$

In the following definition $\ell^\infty(\mathcal{F})$ denotes the space of bounded functions γ on $\mathcal{F}$, equipped with supremum norm $\|\gamma\|_{\mathcal{F}} = \sup_{\varphi\in\mathcal{F}} |\gamma(\varphi)|$.

Definition 2. We say that a P-separable class of functions $\mathcal{F}$ satisfies the compact LIL for P, whenever the sequence

$$\left\{ \frac{\sqrt{n}(P_n - P)\varphi}{\sqrt{2\log\log n}} : \varphi \in \mathcal{F} \right\}_{n=1}^{\infty}$$

is almost surely relatively compact in $\ell^\infty(\mathcal{F})$ with set of limit points

$$\mathcal{H} = \left\{\gamma \mapsto \mathbb{E}\left[\big(\gamma(X) - P\gamma\big)h(X)\right] : \mathbb{E}h^2(X) \le 1\right\}. \tag{2.1}$$

Note that, in particular, if $\mathcal{F}$ satisfies the compact LIL for P, then

$$\limsup_{n\to\infty} \sup_{\varphi\in\mathcal{F}} \left| \frac{\sqrt{n}(P_n - P)\varphi}{\sqrt{2\log\log n}} \right| = \sup_{\varphi\in\mathcal{F}} \big(\mathrm{Var}\big(\varphi(X)\big)\big)^{1/2}, \text{ a.s.} \tag{2.2}$$

Let us recall a LIL for empirical processes proved by Ledoux and Talagrand [11] in separable Banach spaces and stated in the language of empirical processes in Theorem 9 on p. 609 of Ledoux and Talagrand [12]. Let $\mathcal{F}$ be a separable for P class of functions in the sense of Definition 1.

In this situation, a $P-$separable class $\mathcal{F} \subset L_2(P)$ such that $\sup_{\varphi\in\mathcal{F}} |P\varphi| < \infty$ satisfies the compact LIL for P if and only if

- (a) $\mathcal{F}$ is totally bounded in L_2,
- (b) $\mathbb{E}(H^2/\log\log H) < \infty$ where $H = \sup_{\varphi\in\mathcal{F}} |\varphi|$, and
- (c) $\sup_{\varphi\in\mathcal{F}} |\frac{\sqrt{n}(P_n-P)\varphi}{\sqrt{\log\log n}}| \to 0$ in probability.

In particular, assuming separability, if $EH^2 < \infty$ and $\mathcal{F}$ is P-Donsker then $\mathcal{F}$ satisfies the compact LIL (since, $\mathcal{F}$ being P-Donsker, the sequence $\sup_{\varphi\in\mathcal{F}} |(P_n - P)\varphi/\sqrt{n}|$ is stochastically bounded).

Acknowledgment

The first named author thanks Jon Wellner for a question that led to his investigation of uniform in bandwidth limit theorems.

References

[1] Chacón, J. E. and Rodríguez-Casal, A. (2010). A note on the universal consistency of the kernel distribution function estimator. *Statistics and Probability Letters* **80** 1414–1419.

[2] Chung, K.-L. (1949). An estimate concerning the Kolmogoroff limit distribution. *Tran. Amer. Math. Soc.* **67** 36–50.

[3] Dudley, R. M. (1999). *Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics* **63**. Cambridge University Press, New York.

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

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

[6] Finkelstein, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.* **42** 607–615.

[7] Giné, E. and Mason, D. M. (2007). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.* **20** 457–485.

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

[9] Giné, E. and 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.

[10] Janssen, P., Swanepoel J. W. H. and Veraverbeke, N. (2007). Modifying the kernel distribution function estimator towards reduced bias. *Statistics* **41**(2) 93–103.

[11] Ledoux, M. and Talagrand, M. (1988). Characterization of the law of the iterated logarithm in Banach spaces. *Ann. Probab.* **16** 1242–1264.

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

[13] Mason, D. M. (2011). Proving consistency of non–standard kernel estimators. *Statistical Inference for Stochastic Processes.* To appear.

[14] Mason, D. M. and Swanepoel, J. W. H. (2011). A general result on the uniform in bandwidth consistency of kernel-type function estimators. *Test* **20** 72–94.

[15] Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15** 780–799.

[16] Swanepoel, J. W. H. (1988). Mean integrated squared error properties and optimal kernels when estimating a distribution function. *Comm. Statist. Theory Methods.* **17**(11) 3785–3799.

[17] Swanepoel, J. W. H., van Graan, F. C. (2005). A new kernel distribution function estimator based on a nonparametric transformation of the data. *Scandinavian Journal of Statistics—Theory and Applications.* **32** 551–562.

[18] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.* **15** 69–78.