Probab. Theory Relat. Fields 130, 167–198 (2004)
Digital Object Identifier (DOI) 10.1007/s00440-004-0339-x

Evarist Giné · Vladimir Koltchinskii · Lyudmila Sakhanenko

Kernel density estimators: convergence in distribution for weighted sup-norms

Received: 2 July 2002 / Revised version: 24 December 2003 /
Published online: 3 March 2004 –

Abstract. Let f_n denote a kernel density estimator of a bounded continuous density f in the real line. Let $\Psi(t)$ be a positive continuous function such that $\|\Psi f^\beta\|_\infty<\infty$. Under natural smoothness conditions, necessary and sufficient conditions for the sequence $\sqrt{\frac{nh_n}{2\log h_n^{-1}}}\sup_{t\in\mathbf{R}}\left|\Psi(t)(f_n(t)-Ef_n(t))\right|$ (properly centered and normalized) to converge in distribution to the double exponential law are obtained. The proof is based on Gaussian approximation and a (new) limit theorem for weighted sup-norms of a stationary Gaussian process. This extends well known results of Bickel and Rosenblatt to the case of weighted sup-norms, with the sup taken over the whole line. In addition, all other possible limit distributions of the above sequence are identified (subject to some regularity assumptions).

1. Introduction

We consider the kernel density estimator f_n of an unknown density f in the real line based on a sample $(X_1,\dots,X_n)$ of size n of i.i.d. observations with density f, kernel K and bandwidth h_n such that $h_n\to 0$ and $nh_n\to\infty$ as $n\to\infty$:

$$f_n(t)=\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{t-X_i}{h_n}\right). \tag{1.1}$$

Our main goal is to study the convergence in distribution of the (properly centered and normalized) sequence

E. Giné: Departments of Mathematics and Statistics, University of Connecticut, Storrs, CT 06269-3009, USA. e-mail: gine@uconnvm.uconn.edu

V. Koltchinskii: Department of Mathematics and Statistics, University of New Mexico Albuquerque, NM 87131-1141, USA. e-mail: vlad@math.unm.edu

L. Sakhanenko: Department of Statistics and Probability, Michigan State University East Lansing, MI 48824-1027, USA. e-mail: luda@stt.msu.edu

1. Research partially supported by NSF Grant No. DMS-0070382.

2. Research partially supported by NSA Grant No. MDA904-02-1-0075 and NSF Grant No. DMS-0304861.

Mathematics Subject Classification (2000): Primary 62G07; secondary 62G20, 60F15

Key words and phrases: Kernel density estimator – Convergence in distribution – Weighted sup-norm

$$\sqrt{\frac{nh_n}{2\log h_n^{-1}}}\left\|\Psi(\cdot)(f_n - Ef_n)(\cdot)\right\|_\infty \tag{1.2}$$

where Ψ is a positive weight function that might depend on the density f (for instance, it might be $f^{-\beta}$ with some $\beta > 0$). In particular, under some regularity assumptions, we prove the following main result. Let

$$r(t) := \frac{\int_{-\infty}^{\infty} K(u)K(u+t)du}{\|K\|_2}.$$

Suppose it satisfies the condition

$$r(t) = 1 - C|t|^\alpha + o(|t|^\alpha) \tag{1.3}$$

for some $0 < \alpha \le 2,\ C > 0$. Define

$$\Psi_\alpha(u) := \sqrt{\frac{2}{\pi}}C^{1/\alpha}H_\alpha u^{2/\alpha-1}e^{-u^2/2},\ u \ge 0$$

(H_α is a known constant), and set

$$\Lambda_\alpha(u) := \int_{-\infty}^{\infty}\Psi_\alpha\big(u/w(y)\big)dy, \tag{1.4}$$

where $w := \Psi f^{1/2}$ is normalized so that $\|w\|_\infty = 1$. Let A_n denote the solution of the equation

$$\Lambda_\alpha(A_n) = h_n.$$

Then (subject to the regularity assumptions outlined in Section 1.1) the condition

$$\lim_{t\to\infty} t\Pr\{\Psi(X) \ge (th_t|\log h_t|)^{1/2}\} = 0$$

is necessary and sufficient for the convergence

$$\Pr\left\{A_n\left(\sqrt{nh_n}\,\frac{\left\|\Psi(\cdot)(f_n - Ef_n)(\cdot)\right\|_\infty}{\|K\|_2} - A_n\right) \le x\right\} \to e^{-e^{-x}} \text{ as } n \to \infty$$

for all $x \in \mathbf{R}$. Under simple extra conditions on the bias of f_n, results of this type can be reformulated as statements about the weighted uniform deviation of the kernel estimator from the density itself, which can then be used in hypotheses testing and in constructing confidence bands for unknown densities.

This continues the line of research initiated by Smirnov in the 30s who proved convergence in distribution of the uniform deviations (on a compact interval) of the histogram from the underlying density to the double exponential law and, more recently, by Bickel and Rosenblatt (1973) (see also the follow up papers of Konakov and Piterbarg (1984) and Rio (1994), where the conditions were significantly improved) who did the same for kernel density estimators.

Our approach is based on approximation of the empirical processes related to the kernel density estimator by Gaussian processes and on reducing the problem to the study of the asymptotic behavior of weighted suprema

$$\sup_{t\in\mathbf{R}} w\left(\frac{t}{T}\right)\left|\xi(t)\right|$$

of a stationary Gaussian process ξ as $T \to \infty$. We show that the above suprema, properly centered and normalized, converge in distribution to the double exponential law, which generalizes well known results about the limit behavior of suprema of stationary Gaussian processes in the unweighted case, see, e.g., Leadbetter, Lindgren and Rootzén (1986) or Piterbarg (1996). More precisely, if $r(t)$ denotes the covariance function of the process ξ and it satisfies condition (1.3), then defining A_T as the solution of the equation

$$\Lambda_\alpha(A_T) = T^{-1}$$

(with Λ_α defined by (1.4)), we show (subject to some further regularity conditions) that

$$\Pr\left\{A_T\left(\sup_{t\in\mathbf{R}} w\left(\frac{t}{T}\right)\left|\xi(t)\right| - A_T\right) \le x\right\} \to e^{-e^{-x}} \text{ as } T \to \infty\ \forall x \in \mathbf{R}.$$

This result might be of independent interest since it is related to the problem of analyzing the asymptotics of the probabilities of crossing curves of a certain shape by a stationary Gaussian process. It is given in Theorem 2 below.

Also, based upon recent results of Giné, Koltchinskii and Zinn (2001) (who studied convergence rates in probability and a.s. of weighted sup-norms of the deviations of kernel density estimators from their expectations), we determine all other possible nondegenerate limit distributions of (1.2) (subject to some regularity conditions).

We introduce some general assumptions used throughout the paper in Section 1.1. In section 2, we develop the Gaussian approximations needed to reduce the problem to the Gaussian case. Section 3 studies the asymptotic behavior of weighted sup-norms of stationary Gaussian processes. Section 4 contains the main results and their proofs and also studies which other limit distributions are possible (in addition to the double exponential). Theorems 6, 7, 8 and 9, in Section 4, particularly Theorem 6, contain the distributional limit for the deviation of the kernel density estimator and are the main results of this article. In Section 5, we sketch an approach to a version of the main results that might be used in statistical applications.

1.1. General assumptions

We apologize to the reader for the technical character of what follows, but it seems reasonable, for easier readability of the paper, to state the conditions we will be using throughout, thus getting them out of the way. Here and throughout, K is a kernel, f is a density on $\mathbf{R}$, Ψ is a weight function and h_n are the window sizes. Here are the assumptions we will be using (different sets on different instances).

(K1) K is a non-negative function of bounded variation with support in $[-1/2, 1/2]$. We will also assume in Section 3 that

(K2) the function

$$r(t) = \frac{\int_{-\infty}^{\infty} K(u)K(u+t)du}{\|K\|_2^2}$$

satisfies that $r(t) = 1 - C|t|^\alpha + o(|t|^\alpha)$ as $t \to 0$ for some $0 < \alpha \le 2$, and $\sup_{|t|>\varepsilon} |r(t)| < 1$ for all $\varepsilon > 0$.

We will require in Section 3 that the density f and the weight Ψ satisfy

(F1) $B_f = \{f > 0\}$ consists of a finite union of non-trivial disjoint intervals (half lines not excluded), and f and Ψ are both piecewise monotone on B_f. Moreover, f is bounded and Hölder continuous of some order $\alpha > 0$ on B_f, in particular, $\lim_{a\to\infty} \sup_{|t|>a} f(t) = 0$; and

(F2) the function $w = \Psi f^{1/2}$ satisfies conditions (w1), (w2) and (w3) below, in Section 3.

The set of assumptions (K1), (K2), (F1) and (F2) will be referred to as 'the additional hypotheses', since we will also require a set (or subsets of) conditions that we already encountered in Giné, Koltchinskii and Zinn (2001), some of which, although looking somewhat unusual, seem necessary when dealing with weighted sup norms over the whole of **R**. We refer to this article for comments on these conditions and for a number of examples illustrating their necessity. These are

(UH) (D.a)-(D.c), (W.a)-(W.c), $(\text{WD.a})_\beta$ for some $0 < \beta \le 1$, (H_1), (H_2) from Giné, Koltchinskii and Zinn (2001).

We refer to the conditions (UH) as the 'usual hypotheses', and they are as follows.

$(D.a)$ f is a bounded density on **R** continuous on its positivity set $B_f := \{t \in \mathbf{R} : f(t) > 0\}$, which is assumed to be open, and $\lim_{a\to\infty} \sup_{|t|>a} f(t) = 0$. (This condition is in fact implied by (F1).)

$(D.b)$ For all $\delta > 0$ there exist $c \in (0, \infty)$ and $h_0 > 0$ such that, for all $|y| \le h_0$ and all $x \in B_f, x + y \in B_f$,

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

$(D.c)$ For all $r > 0$,

$$\lim_{h\to 0} \sup_{\substack{x,y: f(x)\ge h^r, \\ x+y\in B_f, |y|\le h}} \left| \frac{f(x+y)}{f(x)} - 1 \right| = 0.$$

$(W.a)$ $\Psi : B_f \mapsto \mathbf{R}_+$ is a positive continuous function on B_f.

$(W.b)$ For all $\delta > 0$ there exist $c \in (0, \infty)$ and $h_0 > 0$ such that, for all $|y| \le h_0$ and all $x \in B_f, x + y \in B_f$,

$$\frac{1}{c} \Psi^{-\delta}(x) \le \frac{\Psi(x+y)}{\Psi(x)} \le c \Psi^\delta(x).$$

$(W.c)$ For all $r > 0$,

$$\lim_{h\to 0} \sup_{\substack{x,y:\Psi(x)\le h^{-r},\\ x+y\in B_f,|y|\le h}} \left|\frac{\Psi(x+y)}{\Psi(x)} - 1\right| = 0.$$

$(WD.a)_\beta$ $\|f^\beta\|_{\Psi,\infty} := \sup_{t\in B_f} |\Psi(t) f^\beta(t)| < \infty$, where β is a positive number.

(H_1) h_t, $t \ge 1$, is monotonically decreasing to 0 and th_t is a strictly increasing function diverging to infinity as $t \to \infty$, and

(H_2) h_t is regularly varying at infinity with exponent $-\eta$ for some $\eta \in (0, 1)$; in particular there exist $0 < \eta_0 \le \eta_1 < 1$ such that

$$\limsup_{t\to\infty} t^{\eta_0} h_t = 0 \quad\text{and}\quad \liminf_{t\to\infty} t^{\eta_1} h_t = \infty.$$

2. The Gaussian approximation

Let $D_a = \{|t| \le a,\ f(t) \ge 1/a\} \subseteq B_f$. Note that for a large enough $D_a \neq \emptyset$ consists of a finite union of non trivial bounded intervals. Let $\xi(t)$, $t \in \mathbf{R}$, be a stationary centered Gaussian process with covariance function $r(t) = \int_{-\infty}^{+\infty} K(u)K(u+t)du$. As we see below, this process has a sample continuous version, and we will always take such a version.

In this section we show that, under appropriate conditions on $A_n \to \infty$, if either of the two sequences

$$\left\{A_n\left(\sqrt{nh_n}\|\Psi(t)(f_n(t) - Ef_n(t))\|_{D_a} - A_n\right)\right\}$$

and

$$\left\{A_n\left(\|\Psi(t)\sqrt{f(t)}\xi(t/h_n)\|_{D_a} - A_n\right)\right\}$$

converges in distribution, so does the other. The method will consist in adapting the Komlós-Major-Tusnády approximation to general empirical processes, basically as done in Koltchinskii (1994).

To this end, we begin with the KMT approximation (Komlós, Major and Tusnády (1975)). KMT asserts that there exists a probability space with a sequence $\{\xi_i\}$ of i.i.d. uniform on [0, 1] rv's and a sequence of Brownian motions B_n, such that if $\alpha_n(t) = n^{-1/2}\sum_{i=1}^n \left(I_{[0,t]}(\xi_i) - t\right)$ and $W_n(t) = B_n(t) - tB_n(1)$, then

$$\Pr\left\{\|\alpha_n - W_n\|_\infty > \frac{x + C\log n}{\sqrt{n}}\right\} \le \Lambda e^{-\theta x},\quad 0 \le x < \infty,\ n \in \mathbf{N}, \tag{2.1}$$

where C, Λ and θ are universal positive constants. Extend the definition of α_n and W_n to measurable sets and integrable functions as usual, and, for a class of functions $\mathcal{F}$, set $\|H(g)\|_{\mathcal{F}} := \sup_{g\in\mathcal{F}} |H(g)|$. Set

$$\mathcal{F}_n := \left\{\Psi(t)K\left(\frac{t-\cdot}{h_n}\right) : t \in D_a\right\}. \tag{2.2}$$

Let F be the cdf of the density f and further define

$$\tilde{\mathcal{F}}_n := \{g \circ F^{-1} : g \in \mathcal{F}_n\},$$

where, as usual, $F^{-1}(y) := \inf\{x : F(x) \geq y\}$, $0 < y \leq 1$, and $F^{-1}(0) = F^{-1}(0+)$. Let $\mathcal{V}$ denote the set of functions on [0, 1] that are 0 at 1 and have total variation bounded by 1. Then $\mathcal{V} \subset \overline{co}(\mathcal{I})$, where

$$\mathcal{I} := \{I_{[0,t]} : t \in [0, 1]\}$$

Therefore

$$\|\alpha_n - W_n\|_{\mathcal{V}} = \|\alpha_n - W_n\|_\infty.$$

Now, by (K1) and monotonicity of the maps $x \mapsto (t - x)/h_n$ and F^{-1}, we have

$$\tilde{\mathcal{F}}_n \subset \|\Psi\|_{D_a} \|K\|_V \mathcal{V},$$

where $\|K\|_V$ denotes the total variation of K, and therefore, by (2.1), also that

$$\Pr\left\{\|\alpha_n - W_n\|_{\tilde{\mathcal{F}}_n} > \frac{\|\Psi\|_{D_a}\|K\|_V(x + C\log n)}{\sqrt{n}}\right\} \leq \Lambda e^{-\theta x} \qquad (2.3)$$

for all $0 \leq x < \infty$ and $n \in \mathbf{N}$. Since $F^{-1}(\xi)$ has law P (P denoting the distribution of X) if ξ is uniform on [0, 1], the above yields:

$$\Pr\left\{\|\nu_n - G_n\|_{\mathcal{F}_n} > \frac{\|\Psi\|_{D_a}\|K\|_V(x + C\log n)}{\sqrt{n}}\right\} \leq \Lambda e^{-\theta x} \qquad (2.4)$$

for all $0 \leq x < \infty$ and $n \in \mathbf{N}$, where ν_n is the empirical process based on the i.i.d.(P) sequence $X_i = F^{-1}(\xi_i)$,

$$\nu_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\delta_{X_i} - P),$$

and $G_n(g) := W_n(g \circ F^{-1})$ is a version of the P-Brownian bridge G_P, a centered Gaussian process with the covariance of $\delta_{X_1} - P$.

Our object, in different notation than above, is to show weak convergence of the sequence

$$A_n\left(\frac{1}{\sqrt{h_n}}\|\nu_n\|_{\mathcal{F}_n} - A_n\right) \qquad (2.5)$$

where A_n is an increasing sequence to be determined below, typically, of the order of $\sqrt{\log n}$, and this section's object is to show that the empirical process can be replaced by the Gaussian process $\xi(t)$ in (2.5). We have from (2.4), that, for any $L > C$,

$$\Pr\left\{\frac{A_n}{\sqrt{h_n}}\|\nu_n - G_n\|_{\mathcal{F}_n} > \frac{LA_n\|\Psi\|_{D_a}\|K\|_V \log n}{\sqrt{nh_n}}\right\} \leq \frac{\Lambda}{n^{\theta(L-C)}},$$

which implies that we can replace ν_n by $G_P =_d G_n$ in (2.5) as long as

$$A_n = o\Big(\frac{\sqrt{nh_n}}{\log n}\Big). \tag{2.6}$$

This is already a Gaussian reduction. In fact, $G_P(g) = B_P(g - E_P g) = B_P(g) - (E_P g)B_P(1)$ where B_P is P-Brownian motion, that is, it is a Gaussian process on $\mathcal{F}_n$ such that $E(B_P(g)B_P(h)) = E_P(gh)$ for all $g, h \in \mathcal{F}_n$, for all n. Since for $g \in \mathcal{F}_n$,

$$\int g dP \le h_n \|K\|_1 \|f\|_\infty \|\Psi\|_{D_a}$$

by change of variables, it follows that

$$E\Big\| \frac{A_n}{\sqrt{h_n}}(G_n - B_n)\Big\|_{\mathcal{F}_n} \le A_n \|K\|_1 \|f\|_\infty \|\Psi\|_{D_a} \sqrt{h_n},$$

which implies that we can replace ν_n by B_P in (2.5) as long as

$$A_n = o\big(h_n^{-1/2}\big). \tag{2.7}$$

We can write B_P on $\mathcal{F}_n$ as

$$B_P\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big) = \Psi(t)\int_{-\infty}^{\infty} K\Big(\frac{t-s}{h_n}\Big)\sqrt{f(s)}dB(s),$$

where B is standard Brownian motion (which follows easily by checking covariances). The Brownian motion integral here is sample bounded and sample continuous by e.g. Dudley's entropy condition (e.g., de la Peña and Giné (1999), p. 219) as follows. K being of bounded variation on a compact interval and Ψ bounded, it follows as in Giné, Koltchinskii and Zinn (2001), pages 15 and 16, that the class of functions

$$\mathcal{K}_1 := \Big\{\Psi(t)K\Big(\frac{t-\cdot}{h}\Big) : h > 0, t \in \mathbf{R}\Big\}$$

is a measurable VC (Vapnik-Červonenkis) type class of functions, so that there exist A and v (characteristics of the class) such that the covering numbers of the class for the L_2 distance with respect to any probability measure Q satisfy the uniform bound

$$N\big(\mathcal{K}_1, L_2(Q), \varepsilon\big) \le \Big(\frac{A}{\varepsilon}\Big)^v, \quad 0 < \varepsilon < 1.$$

Then, sample boundedness and continuity of $\{B_P(g) : g \in \mathcal{K}_1\}$ follows from Dudley's bound because $E(B_P(g_1) - B_P(g_2))^2 = \|g_1 - g_2\|^2_{L_2(P)}$. Thus, we can take a separable version of the process above with bounded sample paths.

Finally, we show that we can replace $B_P/\sqrt{h_n}$ in (2.5) by the process Y_n defined as

$$Y_n\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big) = \frac{1}{\sqrt{h_n}}\Psi(t)\sqrt{f(t)}\int_{-\infty}^{\infty} K\Big(\frac{t-s}{h_n}\Big)dB(s). \tag{2.8}$$

Note that, changing variables,

$$Y_n\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big) = \Psi(t)\sqrt{f(t)}\int_{-\infty}^{\infty} K\Big(\frac{t}{h_n}-u\Big)dB(u) = \Psi(t)\sqrt{f(t)}\xi(t/h_n), \tag{2.9}$$

where $\xi(t)$ is a stationary centered Gaussian process with covariance function

$$r(t) = \int_{-\infty}^{\infty} K(u)K(u+t)du \tag{2.10}$$

as in Bickel and Rosenblatt (1973). This process has a separable version with bounded sample paths by the same argument given for B_P: for instance, for $t \in [0, 1]$, $\xi(t) = 2B_Q(K(t-\cdot))$ where Q is the uniform distribution on $[-1/2, 3/2]$, so that we can proceed as above.

To prove the last reduction, first we see that, if

$$D_{a,n} := \{t \in D_a : |t-s| \geq h_n^{-1} \text{ for all } s \in D_a^c\},$$

then, for any $u_n \to \infty$,

$$\Pr\Big\{\Big\|Y_n\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big)\Big\|_{D_a\setminus D_{a,n}} > u_n\Big\} \to 0 \tag{2.11}$$

and

$$\Pr\Big\{\Big\|\frac{1}{\sqrt{h_n}}B_P\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big)\Big\|_{D_a\setminus D_{a,n}} > u_n\Big\} \to 0. \tag{2.12}$$

If this holds, assuming $A_n \to \infty$, and taking $u_n = A_n + x/A_n \to \infty$, we have

$$\begin{aligned}&\Pr\Big\{A_n\Big(\Big\|\frac{1}{\sqrt{h_n}}B_P\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big)\Big\|_{D_{a,n}} - A_n\Big) \leq x\Big\}\\ &-\Pr\Big\{A_n\Big(\Big\|\frac{1}{\sqrt{h_n}}B_P\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big)\Big\|_{D_a} - A_n\Big) \leq x\Big\} \to 0, \end{aligned}\tag{2.13}$$

and likewise for the process Y_n, so that we will be able to restrict attention to suprema over $D_{a,n}$ instead of D_a. To prove (2.11) and (2.12), we notice first that, for n large enough, $D_a \setminus D_{a,n}$ is a finite union of at most $m < \infty$ intervals of length h_n. We only need to consider one such interval, say $[b, b+h_n] \subset D_a$. For (2.11), we note that

$$E\Big\|Y_n\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big)\Big\|_{[b,b+h_n]} \leq \|\Psi\|_\infty\|\sqrt{f}\|_\infty E\sup_{0\leq t\leq 1}|\xi(t)| := M < \infty$$

by (2.9) and stationarity and sample boundedness of the Gaussian process $\xi(t)$ on $[0, 1]$. Then, by the Borell-Tsirel'son-Sudakov inequality, or its simpler Maurey-Pisier form ((3.2) in Ledoux and Talagrand (1991), page 57),

$$\Pr\Big\{\Big\|Y_n\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big)\Big\|_{[b,b+h_n]} > u_n\Big\} \leq \exp\Big(-\frac{2(u_n-M)^2}{\pi^2\sigma^2\|\Psi\|_\infty^2\|f\|_\infty}\Big)$$

for $u_n > M$, where $\sigma^2 = \sup_t E\xi^2(t) = \|K\|_2^2 < \infty$. This proves the limit (2.11). For (2.12), $\mathcal{K}_1$ being VC type with characteristics A and v, and observing that we can always enlarge $\mathcal{K}_1$ to contain zero, which makes the $L_2(P)$-diameter of the class a little larger, Dudley's entropy bound gives

$$E\Big\|\frac{1}{\sqrt{n}}B_P\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big)\Big\|_{[b,b+h_n]} \le L\int_0^{d_n}\sqrt{v\log\frac{A}{\varepsilon}}d\varepsilon$$

for a universal constant L, and where

$$d_n^2 = 4\sup_{[b,b+h_n]} E\Big|\frac{1}{\sqrt{n}}B_P\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big)\Big|^2 \le 4\|\Psi\|_\infty^2\|f\|_\infty\|K\|_2^2 < \infty.$$

Thus, the expected values above are uniformly bounded and the Borel-Sudakov-Tsirel'son inequality then yields the limit (2.12) just as it yielded (2.11).

Finally, we will consider

$$\begin{aligned}
&\Big\|\Big(\frac{B_P}{\sqrt{h_n}} - Y_n\Big)\Big(\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big)\Big\|_{D_{a,n}} \\
&= \Big\|\frac{1}{\sqrt{h_n}}\int_{-\infty}^{\infty}\Psi(t)K\Big(\frac{t-s}{h_n}\Big)\Big(1-\frac{\sqrt{f(t)}}{\sqrt{f(s)}}\Big)\sqrt{f(s)}dB(s)\Big\|_{D_{a,n}} \\
&= \frac{1}{\sqrt{h_n}}\Big\|B_P\Big(\Psi(t)K\Big(\frac{t-s}{h_n}\Big)\Big(1-\frac{\sqrt{f(t)}}{\sqrt{f(s)}}\Big)\Big)\Big\|_{D_{a,n}}.
\end{aligned}$$

To bound the expected value of this norm, we look at the class of functions

$$\mathcal{G}_n := \Big\{\Psi(t)K\Big(\frac{t-\cdot}{h_n}\Big)\Big(1-\frac{\sqrt{f(t)}}{\sqrt{f(\cdot)}}\Big) : t \in D_{a,n}\Big\}$$

and observe that it is a uniformly bounded VC type class (polynomial or Euclidean are different terms for the same; see e.g., Nolan and Pollard (1987), who were first to observe that the family of translations and dilations of K is VC type if K is of bounded variation). Uniform boundedness of the class follows because, since K is supported by $[-1/2, 1/2]$, the functions in this class are zero on D_a^c and are bounded by $\tau\|\Psi\|_{D_a}\|K\|_\infty$ on D_a, with $\tau = \sup\big\{|1-\sqrt{f(t)/f(s)}| : s, t \in D_a, |s-t| < \varepsilon\big\}$ for an appropriate $\varepsilon > 0$, which is finite by condition (D.c). The class $\mathcal{G}_n$ is measurable because the functions in the class are jointly measurable in the two variables, and to see it is of VC type we just note that $\mathcal{G}_n \subset \{g_1 + g_2 : g_1 \in \mathcal{K}_1, g_2 \in \mathcal{K}_{2,n}\}$ where

$$\mathcal{K}_{2,n} := \Big\{\Psi(t)\frac{\sqrt{f(t)}}{\sqrt{f(\cdot)}}K\Big(\frac{t-\cdot}{h_n}\Big) : t \in D_{a,n}\Big\}.$$

The class $\big\{\Psi(t)\sqrt{f(t)}K\big(\frac{t-\cdot}{h}\big) : t \in \mathbf{R}, h > 0\big\}$ is VC for the same reasons $\mathcal{K}_1$ is, and let us denote its VC characteristics by A and v; from this it follows trivially that, for all n, $\mathcal{K}_{2,n}$ is VC with characteristics $\sqrt{a}A$ and v (independent of n) because for any probability measure Q and $s, t \in D_{a,n}$,

$$E_Q\left(\frac{\Psi(t)\sqrt{f(t)}}{\sqrt{f(x)}}K\left(\frac{t-x}{h_n}\right)-\frac{\Psi(s)\sqrt{f(s)}}{\sqrt{f(x)}}K\left(\frac{s-x}{h_n}\right)\right)^2$$
$$\leq aE_Q\left(\Psi(t)\sqrt{f(t)}K\left(\frac{t-x}{h_n}\right)-\Psi(s)\sqrt{f(s)}K\left(\frac{s-x}{h_n}\right)\right)^2.$$

Since the functions in $\mathcal{G}_n$ are each the sum of a function in $\mathcal{K}_1$ and a function in $\mathcal{K}_{2,n}$, a simple argument gives that $\mathcal{G}_n$ is VC type with characteristics A and v independent of n (see e.g. Lemma 5.3.4 in de la Peña and Giné (1999)). Now, Dudley's entropy bound gives, as above,

$$E\|B_P\|_{\mathcal{G}_n} \leq L\int_0^{d_n}\sqrt{v\log(A/\varepsilon)}d\varepsilon$$

for a universal constant L, where $d_n^2 := 4\|E_P f^2\|_{\mathcal{G}_n} \leq 16h_n\|K\|_2^2\|\Psi\|_{D_a}^2\omega_{\sqrt{f}}^2(h_n)$, with the last inequality following by change of variables in two integrations. Here, $\omega_{\sqrt{f}}(\cdot)$ is the modulus of continuity for $\sqrt{f}$. This gives

$$\frac{1}{\sqrt{h_n}}E\left\|B_P\left(\Psi(t)K\left(\frac{t-s}{h_n}\right)\left(1-\frac{\sqrt{f(t)}}{\sqrt{f(s)}}\right)\right)\right\|_{D_{a,n}} = E\left\|B_P/\sqrt{h_n}-Y_n\right\|_{\mathcal{G}_n}$$
$$\leq L(K,\|f\|_\infty,\|\Psi\|_{D_a})\,\omega_{\sqrt{f}}(h_n)\sqrt{\log\frac{1}{\sqrt{h_n}\,\omega_{\sqrt{f}}(h_n)}}$$

for all n large enough. If f is Hölder α, then $\sqrt{f}$ is Hölder $\alpha/2$, and it follows that this expected value is of the order of

$$h_n^{\alpha/2}\sqrt{\log h_n^{-1}}.$$

We conclude from the last inequality and (2.13) that we can replace $\|\nu_n\|_{\mathcal{F}_n}/\sqrt{h_n}$ in (2.5) by $\|\sqrt{f(t)}\xi(t/h_n)\|_{D_a}$, where ξ is defined in (2.9) and (2.10), if

$$A_n = o\left(\frac{1}{h_n^{\alpha/2}\sqrt{|\log h_n|}}\right). \tag{2.14}$$

To summarize, with the assumption

$$A_n \nearrow \infty,\ h_n \searrow 0 \text{ and } A_n = o\left(\frac{1}{h_n^{\alpha/2}\sqrt{\log h_n^{-1}}}\wedge\frac{\sqrt{nh_n}}{\log n}\wedge\frac{1}{h_n^{1/2}}\right). \tag{A1}$$

we have proved:

Theorem 1. *Assuming (D.a), (D.c), (F1) except that f and Ψ do not need to be piecewise monotone, (A1) and Ψ continuous and bounded on B_f, if for some $a > 0$ either of the two sequences*

$$\left\{A_n\left(\sqrt{nh_n}\|\Psi(t)(f_n(t) - Ef_n(t))\|_{D_a} - A_n\right)\right\} \tag{2.5}$$

and

$$\left\{A_n\left(\|\Psi(t)\sqrt{f(t)}\xi(t/h_n)\|_{D_a} - A_n\right)\right\} \tag{2.15}$$

converges weakly, so does the other, where $\xi(t)$ is a stationary centered Gaussian process with covariance function $r(t) = \int_{-\infty}^{+\infty} K(u)K(u+t)du$.

Remark. Theorem 1 is designed to be used below on theorems for weighted suprema of $f_n - Ef_n$ *over all of* $\mathbf{R}$. If we are only interested in sups over $[-a, a]$, $a < \infty$, with the natural weight $1/\sqrt{f}$, which is what Bickel and Rosenblatt (1973) consider, the following reformulation, which has the same proof as Theorem 1, might be useful: Suppose that f satisfies (D.a) and that it is bounded away from zero and Hölder continuous (of some order $\alpha > 0$) on $[-a, a]$, and assume K satisfies condition (K.1) and A_n and h_n, $n \in \mathbf{N}$, satisfy (A.1). Then, if either of the two sequences

$$\left\{A_n\left(\sqrt{nh_n}\|(f_n(t) - Ef_n(t))/\sqrt{f(t)}\|_{[-a,a]} - A_n\right)\right\}$$

and

$$\left\{A_n\left(\|\xi(t/h_n)\|_{[-a,a]} - A_n\right)\right\}$$

converges weakly, so does the other. Essentially, this is the content of Propositions 2.1 and 2.2 in Bickel and Rosenblatt (1973), but, except for the fact that we do not consider kernels with unbounded support, they assume far more regularity for both, K and f.

3. Asymptotic distribution of weighted sup-norms of stationary Gaussian processes

Our object here is to find the limiting distribution of the sequence (2.15) for appropriate normalizing and centering constants A_n. The process $\xi(t)$ in (2.15) is stationary with covariance $r(t) = \int_{-\infty}^{\infty} K(u)K(u+t)du$. For ease of notation and for generality's sake, we replace the factor $\Psi(t)f^{1/2}(t)$ in (2.15) by a factor $w(t)$. We will also assume in this section, for simplicity, that $\|K\|_2 = 1$ (otherwise, K should be renormalized). Note that if K is a twice continuously differentiable, symmetric, bounded kernel with support in $[-1/2, 1/2]$ and normalized so that $\|K\|_2 = 1$, then

$$r(t) = 1 - Ct^2 + o(t^2) \text{ as } t \to 0, \text{ with } C = -\frac{1}{2}\int K(u)K''(u)du > 0,$$

$$\text{and } r(t) = 0 \text{ for } |t| > 1;$$

in fact, many other kernels satisfy this (e.g., the Epanechnikov normalized so that $\|K\|_2 = 1$); conceivably, with weaker assumptions on K we may have, for some $0 < \alpha \le 2$ and $C > 0$,

$$r(t) = 0 \text{ for } |t| > 1, \ r(t) = 1 - C|t|^\alpha + o(|t|^\alpha), \text{ for some } 0 < \alpha \le 2, C > 0 \tag{3.1}$$

and

$$\sup_{|t|>\varepsilon} |r(t)| < 1 \text{ for all } \varepsilon > 0. \tag{3.2}$$

So, we make this assumption on the process $\xi(t)$. By e.g. Albin (1990), p. 117, if $\xi(t)$ is a separable stationary process with covariance r as in (3.1) and (3.2), there exists $H_\alpha < \infty$ ($H_\alpha = \sqrt{\pi}$ if $\alpha = 2$) such that, whenever h satisfies that $\sup_{\varepsilon<t\le h} r(t) < 1$ for all $0 < \varepsilon < h$, we have

$$\lim_{u\to\infty} u^{1-2/\alpha} e^{u^2/2} \Pr\Big\{\sup_{0\le t\le h} |\xi(t)| > u\Big\} = h\sqrt{\frac{2}{\pi}} C^{1/\alpha} H_\alpha. \tag{3.3}$$

Let $w(t)$ be a non-negative function such that

($w1$) The support W of w consists of a finite number of disjoint closed intervals or half-lines, w is positive and continuous on its support, piecewise monotone, with $\|w\|_\infty = 1$ and such that $w(t) \to 0$ as $|t| \to \infty$.
Let

$$\Psi_\alpha(u) := \sqrt{\frac{2}{\pi}} C^{1/\alpha} H_\alpha u^{2/\alpha-1} e^{-u^2/2}, \ u \ge 0, \tag{3.4}$$

with $\Psi_\alpha(\infty) := 0$, and set

$$\Lambda_\alpha(u) := \int_{-\infty}^{\infty} \Psi_\alpha\big(u/w(y)\big)dy. \tag{3.5}$$

We also assume

($w2$)

$$\Lambda_\alpha(u_0) < \infty \text{ for some } u_0 < \infty$$

(then $\Lambda_\alpha(u) < \infty$ for all $u \ge u_0$). This holds under mild assumptions on w, and if e.g. w has a unique mode where it is almost flat, then, typically, $\Lambda_2(u)$ will be of the order of a constant times $e^{-u^2/2\|w\|_\infty} = e^{-u^2/2}$. Let A_T be the solution of the equation

$$\Lambda_\alpha(A_T) = \frac{1}{T}, \tag{3.6}$$

which exists and is unique for all T large enough, and $A_T \to \infty$ as $T \to \infty$. We make an additional assumption on $w(t)$:

$(w3)$

$$\sup_{\substack{t\in W, t+\tau\in W\\ |\tau|\le 1/T}} \left|\frac{w(t+\tau)}{w(t)} - 1\right| = o(A_T^{-2}) \text{ as } T\to\infty.$$

(We remark here that, by Lemma 5 below, $o(A_T^{-2})$ is equivalent to $o(1/\log T)$.)

The object of this section is to prove:

Theorem 2. *Let $\xi(t)$, $t\in\mathbf{R}$, be a separable centered, stationary Gaussian process with covariance satisfying (3.1) and (3.2). Let $w(t)$ be a non-negative function satisfying (w1)-(w3), and let A_T be as defined by (3.6). Then,*

$$\lim_{T\to\infty} \Pr\left\{A_T\Big(\sup_t w(t/T)|\xi(t)| - A_T\Big) \le x\right\} = \exp\big(-e^{-x}\big) \tag{3.7}$$

for all $x\in\mathbf{R}$.

Proof. We will follow the scheme of proof of Theorem 12.3.5 in Leadbetter, Lindgren and Rootzén (1983) with changes due to the facts that we are multiplying by $w(t)$ and supping over the whole line, and that we are considering sup of the absolute value of the process instead of sup of the process.

Step 1. For x fixed, let $u_T = \frac{x+A_T^2}{A_T}$. Then,

$$T\Lambda_\alpha(u_T) \to e^{-x} \text{ as } T\to\infty. \tag{3.8}$$

Proof. By the definitions of A_T ((3.6)) and u_T, it clearly suffices to show that

$$\lim_{t\to\infty} \frac{\Lambda_\alpha(t+x/t)}{\Lambda_\alpha(t)} = e^{-x}. \tag{3.9}$$

Let us assume $x>0$ as the argument for $x<0$ is similar. Set $\beta = 1-2/\alpha$. We obviously have

$$\frac{\Lambda_\alpha(t+x/t)}{\Lambda_\alpha(t)} \asymp \frac{\int w(y)^\beta e^{-t^2/2w^2(y)} e^{-x/w^2(y)} e^{-x^2/2t^2w^2(y)} dy}{\int w(y)^\beta e^{-t^2/2w^2(y)} dy}, \tag{3.10}$$

so that, since $x/w^2(y)\ge x$,

$$\limsup_{t\to\infty} \frac{\Lambda_\alpha(t+x/t)}{\Lambda_\alpha(t)} \le e^{-x}.$$

For any $0<c<1$ and, given $\varepsilon>0$, for all t large enough, we have that the right hand side of (3.10) is minorized by

$$\frac{\int_{w(y)>c} w(y)^\beta e^{-t^2/2w^2(y)} dy}{\int w(y)^\beta e^{-t^2/2w^2(y)} dy} \times \exp(-x/c^2 - \varepsilon x^2/c^2).$$

Hence, (3.9) will follow if we show that the ratio of integrals at the left of the last expression tends to one for each $0 < c < 1$, or what is the same, if

$$\lim_{t\to\infty} \frac{\int_{w(y)\le c} w(y)^\beta e^{-t^2/2w^2(y)}dy}{\int w(y)^\beta e^{-t^2/2w^2(y)}dy} = 0.$$

For any $c' > c$ and $t > u_0$, this quotient is dominated by

$$\frac{\int_{w(y)\le c} w(y)^\beta e^{-t^2/2w^2(y)}dy}{\int_{w(y)>c'} w(y)^\beta e^{-t^2/2w^2(y)}dy} \le \frac{\int_{w(y)\le c} w(y)^\beta e^{-u_0^2/2w^2(y)}dy}{\int_{w(y)>c'} w(y)^\beta e^{-u_0^2/2w^2(y)}dy} \times \frac{e^{-(t^2-u_0^2)/2c^2}}{e^{-(t^2-u_0^2)/2c'^2}} \to 0$$

as $t \to \infty$ since the first quotient does not depend on t and the second tends to zero. This completes the proof of the limit (3.8).

Step 2. We can replace in (3.7), the functions $w(t/T)$ by

$$w_T(t) := w(k/T) \text{ for } k \le t < k+1,\ k \in \mathbf{Z},\ \left[\frac{k}{T}, \frac{k+1}{T}\right) \subset W \tag{3.11}$$

and $w_T(t) = 0$ otherwise.

Proof. Set $w'_T(t) = w_T(t)$ on the support of w_T and $w'_T(t) = w(t/T)$ for $t \in TW \setminus \operatorname{supp} w_T$. Then, it is obvious from (w3) that we can replace $w(t/T)$ by $w'_T(t)$ in (3.7). Now, we can further replace $w'_T(t)$ by $w_T(t)$ because, by (w1), $w'_T(t) \ne w_T(t)$ only on a finite number of intervals of length at most 1, and for any such interval, say $[aT, k(aT))$, where $k(aT)$ is the smallest integer larger than or equal to aT, we have, by (3.3), that

$$\Pr\left\{\sup_{t\in[aT,k(aT))} w'_T(t)|\xi(t)| > u_T\right\} \le \Pr\left\{\sup_{t\in[0,1)} |\xi(t)| > \frac{u_T}{w(k(aT)/T)}\right\}$$

$$\asymp c\Psi_\alpha\Big(\frac{u_T}{w(k(aT)/T)}\Big) \to 0$$

as $T \to \infty$ since, in this case, $u_T \to \infty$ and $w(k(aT)/T) \to w(a) > 0$. The constant c is 1 if h can be taken to be 1 in (3.3), and otherwise, it is the smallest integer larger than the inverse of the largest h for which (3.3) holds.

Set $\tau := e^{-x}$. The proof of the theorem will be completed if we show:

Step 3 (main step). With the previous notation and assuming (3.8),

$$\lim_{T\to\infty} \Pr\left\{\sup_t w(t/T)|\xi(t)| \le u_T\right\} = e^{-\tau}. \tag{3.12}$$

Proof. To complete this step we will follow closely the proof of Theorem 12.3.5 in Leadbetter et al. (1983), on the asymptotic distribution of $\sup_{0\le t\le T} \xi(t)$. We require versions of several results in their Section 12.2 for $\sup|\xi(t)|$ (absolute values are not considered there). For instance, we will use the limit (3.3) instead of their Theorem 12.2.9 (note that the effect of taking absolute values of the process instead of the process itself consists only in changing the limit by a factor of 2).

Combining the first part of the proof of Theorem 9 in Albin (1990) with the proof of Lemma 12.2.3 of Leadbetter et al. (1983) easily gives the following: with the assumptions and notation of that lemma, if $q = a/u^{2/\alpha}$,

$$\frac{1}{\phi(u)/u} \Pr\left\{\max_{0 \le j \le n} |\xi(jq)| > u\right\} \to 2C^{1/\alpha} H_\alpha(n, a)$$

where the factor of 2 in the limit as $u \to \infty$ is the sole effect of considering absolute values. Then, because of this and (3.3), in Lemma 12.2.11 in Leadbetter et al. (1983) we can replace ξ by $|\xi|$, and obtain, in their notation,

$$0 \le \Pr\left\{|\xi(jq)| \le u,\ jq \in I\right\} - \Pr\left\{\sup_{t \in I} |\xi(t)| \le u\right\} \le 2\mu h \rho_a + o(\mu), \quad (3.13)$$

where $\rho_a = 1 - H_\alpha(a)/H_\alpha \to 0$ as $a \to 0$. In our notation, $2\mu := 2\mu(u) := \Psi_\alpha(u)$. [The main part of the proof of their Lemma 12.3.3 consists in obtaining the limit of

$$\Pr\left\{\max_{0<j<n} u(\xi(jq) - u) - x > -x | \xi(0) = u + x/u\right\},$$

and the first part of the proof of Theorem 9 in Albin (1990) computes the limit of

$$\Pr\left\{\max_{0<j<n} u(|\xi(jq)| - u) - x > -x | |\xi(0)| = u + x/u\right\};$$

both limits coincide; then one uses this in the decomposition

$$\begin{aligned}\Pr\left\{\max_{0 \le j < n} |\xi(jq)| > u\right\} &= \Pr\left\{u(|\xi(0)| - u) > 0\right\} \\ &+ \Pr\left\{u(|\xi(0)| - u) \le 0,\ \max_{0<j<n} u(|\xi(jq)| - u) > 0\right\}\end{aligned}$$

just as in the proof of Lemma 12.3.3 in Leadbetter et al. (1983) to obtain the above limit for $\Pr\{\max_{0 \le j \le n} |\xi(jq)| > u\}$.]

Define

$$\Lambda_{\alpha,T}(u) := \frac{1}{T} \sum_k \Psi_\alpha\big(u/w(kT^{-1})\big). \quad (3.14)$$

Since both w and Ψ_α are piecewise monotone, by standard approximation of integrals of monotone functions there exists C such that

$$\left|\Lambda_\alpha(u) - \Lambda_{\alpha,T}(u)\right| \le \frac{C}{T} \sup_y \Psi_\alpha\big(u/w(y)\big),$$

and, since Ψ_α is eventually decreasing and $0 < w(y) \le 1$ on W , we get that for all u large enough,

$$\left|\Lambda_\alpha(u) - \Lambda_{\alpha,T}(u)\right| \le \frac{C}{T} \Psi_\alpha(u),$$

which gives

$$\lim_{T \to \infty} T \Lambda_{\alpha,T}(u_T) = \tau \quad (3.8')$$

by (3.8), since $u_T \to \infty$ and $\Psi_\alpha(u_T) \to 0$.

We will use the following notation: for any measurable set A,

$$M_w(A) := \sup_{t\in A} w_T(t)|\xi(t)|,\ M(A) = \sup_{t\in A} |\xi(t)|,$$

and for all $k \in \mathbf{Z}$ and $\varepsilon \in (0, 1)$,

$$I_k := I_{[k,k+1-\varepsilon)},\ I_k^* := I_{(k+1-\varepsilon,k+1)},\ J_k = I_k \cup I_k^*.$$

Step 3_1. $\lim_{T\to\infty} \left[\Pr\Big\{ M_w(\cup I_k) \le u_T \Big\} - \Pr\Big\{ M_w(\mathbf{R}) \le u_T \Big\} \right] \le 2\varepsilon\tau.$

Proof. The above difference, which is nonnegative, is dominated by

$$\sum_k \Pr\Big\{ M_w(I_k^*) > u_T \Big\} = \sum_k \Pr\Big\{ \sup_{t\in I_k^*} |\xi(t)| > u_T/w_T(k) \Big\},$$

which, by the limit theorem (3.3), is in turn dominated, for all T large enough (depending on ε), by

$$2\varepsilon \sum_k \Psi_\alpha\big(u_T/w(k/T)\big) = 2\varepsilon T\,\Lambda_{\alpha,T}(u_T).$$

But this last expression tends to $2\varepsilon\tau$ by (3.8').
Step 3_2. Let $q = q_T = a/u_T^{2/\alpha}$ for some $a > 0$. Then,

$$\lim_{T\to\infty} \left[\Pr\Big\{ w_T(jq_T)|\xi(jq_T)| \le u_T,\ jq_T \in \cup I_k \Big\} - \Pr\Big\{ M_w(\cup I_k) \le u_T \Big\} \right] \le \tau\rho_a,$$

where $\rho_a \to 0$ as $a \to 0$.

Proof. We will write u for u_T and q for q_T in proofs. The above difference is non-negative and is dominated by

$$\sum_k \left(\Pr\Big\{ w(k/T)|\xi(jq)| \le u,\ jq \in I_k \Big\} - \Pr\Big\{ M(I_k) \le u/w(k/T) \Big\} \right).$$

Now we can apply inequality (3.13) (the version for absolute values of Lemma 12.2.11 in Leadbetter et al. (1983)) to get the last expression dominated by

$$\sum_k \Psi_\alpha\big(u/w(k/T)\big)(1-\varepsilon)\rho_a + o\big(\Psi_\alpha(u)\big) \le T\Lambda_{\alpha,T}(u)\rho_a \to \tau\rho_a,$$

where the inequality is valid for all T large enough.
Step 3_3.

$$\Pr\big\{ \big|\xi(jq_T)\big| \le u_T/w_T(k),\ jq_T \in I_k,\ k \in \mathbf{Z} \big\} - \prod_k \Pr\big\{ \big|\xi(jq_T)\big| \le u_T/w_T(k),\ jq_T \in I_k \big\} \to 0$$

as $T \to \infty$. (Recall the convention $c/\infty = 0$.)

Proof. Using a slight modification of Lemma 11.1.2 in Leadbetter et al. (1983) along the lines of Theorem 4.2.1 (a Slepian inequality for tail probabilities), we obtain that the absolute value of the above difference of probabilities is dominated by

$$\sum_{k\neq\ell}\sum_{iq\in I_k}\sum_{jq\in I_\ell}\frac{|r((i-j)q)|}{\sqrt{(1-r^2((i-j)q)}}\exp\left[-\frac{\big(w_T^{-2}(k)+w_T^{-2}(\ell)\big)u^2}{2(1+|r((i-j)q)|)}\right],$$

where the terms in $k=\ell$ are zero because the two processes that we are comparing have the same covariance structure within each block I_k. Since $r(t)=0$ for $|t|>1$ (we are assuming K supported on $[-1/2,1/2]$), for each k there are at most two $\ell\neq k$ such that the covariance $r(s-t)$, $s\in I_k$, $t\in I_\ell$, is not zero, namely $\ell=k-1$ and $\ell=k+1$; so, for each k there will be only about $2/q$ nonzero summands with index $jq\in I_\ell$. Also, when k differs from ℓ by 1, by $(w3)$, $w_T(k)/w_T(\ell)=w(k/T)/w(\ell/T)$ differs from 1 in $o(A_T^{-2})$. These observations imply that the above sum is smaller than or equal to

$$\frac{3}{q}\sum_k\sum_{\varepsilon\le sq\le 1}\frac{|r(sq)|}{\sqrt{1-r^2(sq)}}\exp\left[-\frac{1-o(A_T^{-2})}{1+|r(sq)|}\frac{u^2}{w^2(k/T)}\right],$$

as I_k and I_ℓ are at ε units apart. Also, since $\sup_{t\ge\varepsilon}|r(t)|:=\delta(\varepsilon):=\delta<1$ the above is dominated by a quantity that we can compare with $T\Lambda_{\alpha,T}(u_T)$, namely, by

$$\frac{3\delta}{\sqrt{1-\delta^2}}\frac{1}{q^2}\sum_k\exp\left(-\frac{u^2}{(1+\delta)w^2(k/T)}\right)$$

$$\le L_\delta T\Lambda_{\alpha,T}(u)\frac{1}{q^2}u^{1-\alpha/2}\exp\left[\left(\frac{1}{2}-\frac{1}{1+\delta}\right)u^2\right]\asymp L\tau o(u)\to 0$$

for some $L_\delta<\infty$.

Step 3_4.
$\limsup_{T\to\infty}\Big|\prod_k \Pr\{w_T(jq_T)|\xi(jq_T)|\le u_T,\ jq_T\in I_k\}$
$-\prod_k \Pr\{M_w(J_k)\le u_T\}\Big|$
$\le\tau(\rho_a+2\varepsilon).$

Proof. Repeating Step 3_2 for blockwise independent processes gives

$$0\le\prod_k\Pr\{w_T(jq)|\xi(jq)|\le u,\ jq\in I_k\}-\prod_k\Pr\{M_w(I_k)\le u\}\le\tau\rho_a,$$

and Step 3_1 likewise gives

$$0\le\prod_k\Pr\{M_w(I_k)\le u\}-\prod_k\Pr\{M_w(J_k)\le u\}\le 2\tau\varepsilon$$

for all u large enough.

The four sub-steps put together give that

$$\limsup_{T\to\infty}\left|\Pr\{M_w(\mathbf{R})\le u\}-\prod_k \Pr\{M_w(J_k)\le u\}\right|\le 2\tau(\rho_a+2\varepsilon),$$

and letting ε and a tend to zero, that this limit is zero. That is, taking Step 2 into account,

$$\lim_{T\to\infty}\left[\Pr\{\sup_t w(t/T)|\xi(t)|\le u\}-\prod_k \Pr\{\sup_{0\le t<1}|\xi(t)|\le u/w(k/T)\}\right]=0. \tag{3.15}$$

By (3.3) and (w2) the series $\sum_k \Pr\{M[0,1)>u/w(k/T)\}$ converges for all $u>u_0$, and by (3.3) $\max_k \Pr\{M[0,1)>u/w(k/T)\}\to 0$ as $u\to\infty$ (recall $\|w\|_\infty=1$). Therefore,

$$\sum_k \log \Pr\{M[0,1)\le u/w(k/T)\}\asymp -\sum_k \Pr\{M[0,1)>u/w(k/T)\}.$$

Again by (3.3),

$$\Pr\{M[0,1)>u/w(k/T)\}=\Psi_\alpha(u/w(k/T))+o(\Psi_\alpha(u/w(k/T)))$$

uniformly in k, and therefore (3.14) and (3.8') give

$$\lim_{T\to\infty}\prod_k \Pr\{\sup_{0\le t<1}|\xi(t)|\le u/w(k/T)\}=e^{-\tau}.$$

By (3.15), this concludes the proof of the theorem. □

Remark. A_T in theorem 2 does not need to satisfy equation (3.6), but only the following:

$$\lim_{T\to\infty} T\Lambda_\alpha(A_T)=1, \tag{3.6$'$}$$

and this follows directly from the proof.

Remark. If we take $w(t)=I_{|t|\le a}$ in Theorem 2, we get

$$\lim_{T\to\infty}\Pr\left\{A_T\left(\sup_{|t|\le a}|\xi(tT)|-A_T\right)\le x\right\}=\exp(-e^{-x}),$$

where A_T is any function of T such that

$$\lim_{T\to\infty} T\Psi_\alpha(A_T)=\frac{1}{2a}. \tag{3.7}$$

If we put together this result and the version of Theorem 1 in the remark following it, we obtain a result which is equivalent to Bickel and Rosenblatt's for kernels with bounded support. It is easy to see directly that $A_T\asymp\sqrt{\log T}$ so that, taking $A_n:=A_{h_n^{-1}}$, condition (A.1) becomes $h_n\searrow 0$ and $nh_n/\big((\log n)^2\sqrt{\log h_n^{-1}}\big)\to\infty$.

The conditions on the kernel K under which both theorems hold are (K.1) and (K.2). And the conditions on f are those in the remark following Theorem 1. Under these assumptions, and with the above definition of A_n, we conclude:

$$\lim_{n\to\infty} \Pr\left\{A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2}\left\|\frac{f_n - Ef_n}{\sqrt{f}}\right\|_{[-a,a]} - A_n\right) \le x\right\} = \exp\bigl(-e^{-x}\bigr), \quad x \in \mathbf{R}.$$

4. Kernel density estimators: Convergence in distribution of the uniform deviations from their means

As we will see below, the results in Giné and Guillou (2002) and in Giné, Koltchinskii and Zinn (2001) allow us to deal with $\|f_n - Ef_n\|_{D_a^c}$ as $a \to \infty$ (we recall $D_a = \{|t| \le a, f(t) \ge a\}$). So, in order to control the supremum over D_a we must show first that we can replace the supremum over $\mathbf{R}$ in Theorem 2 by the sup of the same variables over D_a without changing the centering or the normalization. We set, for $a > 0$ and $u > 0$,

$$W_a = \{t : |t| \le a, w(t) \ge a^{-1}\}, \ \Lambda_\alpha^{(a)}(u) = \int_{W_a} \Psi_\alpha(u/w(y))dy, \tag{4.1}$$

and, in general, for any measurable set $D \subset \mathbf{R}$,

$$\Lambda_\alpha^{(D)} = \int_D \Psi_\alpha(u/w(y))dy. \tag{4.2}$$

Lemma 3. *Let w be a non-negative function continuous on $\{w > 0\} \subseteq B_f$, where B_f is an open set, such that $\|w\|_\infty = 1$, $w(t) \to 0$ as $|t| \to \infty$ and $\Lambda_\alpha(u) > 0$ for some $0 < u < \infty$. Then, there exists $a < \infty$ such that, for any measurable set D satisfying $W_a \subseteq D \subseteq \{w > 0\}$ (in particular for W_b for $b \ge a$),*

$$\lim_{u\to\infty} \frac{\Lambda_\alpha(u)}{\Lambda_\alpha^{(D)}(u)} = 1. \tag{4.3}$$

Proof. By monotonicity of $\Lambda_\alpha^{(D)}$ with respect to D, it suffices to show that the limit (4.3) holds for $D = W_a$ for some $a < \infty$. Since $w(y) \to 0$ as $|y| \to \infty$, $\|w\|_\infty = 1$, and w is continuous on the open set B_f, there exists an interval I in B_f such that $M := \inf_{t\in I} w(t) > 1/2$ and $I \subseteq W_{a_1}$ for some $a_1 < \infty$, and there exists a_2 such that $\sup_{|y|>a_2} w(y) \le 1/(2\sqrt{2})$. Taking $a = \max\{a_1, a_2, 2\}$, we have

$$I \subseteq W_a \text{ and } M = \inf_{t\in I} w(t) > 1/2 > \sqrt{2} \sup_{|y|>a} w(y). \tag{4.4}$$

We also have $\sup_{y\in W_a} w(y) = 1$. We show that the limit (4.3) holds for $D = W_a$. Since

$$\frac{\Lambda_\alpha(u)}{\Lambda_\alpha^{(a)}(u)} = 1 + \frac{\int_{W_a^c} \Psi_\alpha(u/w(y))dy}{\int_{W_a} \Psi_\alpha(u/w(y))dy},$$

the lemma will follow if we show that both

$$\frac{\int_{|y|>a} \Psi_\alpha(u/w(y))dy}{\int_{W_a} \Psi_\alpha(u/w(y))dy} \to 0 \text{ and } \frac{\int_{|y|\le a,\ w(y)<a^{-1}} \Psi_\alpha(u/w(y))dy}{\int_{W_a} \Psi_\alpha(u/w(y))dy} \to 0$$

as $u \to \infty$. Since Ψ_α is eventually decreasing, for all u large enough the second quotient is dominated by

$$\frac{2a\Psi_\alpha(u/w(y))}{|I|\Psi_\alpha(u/M)} \le \frac{2a^{2/\alpha}e^{-u^2a^2/2}}{|I|M^{1-2/\alpha}e^{-u^2/2M^2}},$$

which tends to zero as $u \to \infty$ by (4.4). As for the first quotient, we have, also for all u large enough,

$$\begin{aligned}
\frac{\int_{|y|>a} \Psi_\alpha(u/w(y))dy}{\int_{W_a} \Psi_\alpha(u/w(y))dy} &\le \frac{1}{|I|}\int_{|y|>a} \frac{\Psi_\alpha(u/w(y))}{\Psi_\alpha(u/M)}dy \\
&= \frac{M^{2/\alpha-1}}{|I|}\int_{|y|>a} \frac{1}{(w(y))^{2/\alpha-1}} \\
&\quad \exp\Big\{-\frac{u^2}{2}\big(w^{-2}(y) - M^{-2}\big)\Big\}dy \\
&\le \frac{M^{2/\alpha-1}}{|I|}\int_{|y|>a} \frac{u/\sqrt{2}}{(w(y))^{2/\alpha-1}} \exp\Big\{-\frac{(u/\sqrt{2})^2}{2w^2(y)}\Big\}dy \\
&\le \frac{M^{2/\alpha-1}}{|I|C^{1/\alpha}H_\alpha\sqrt{2/\pi}}\Lambda_\alpha(u/\sqrt{2}) \to 0
\end{aligned}$$

as $u \to \infty$ by (4.4). □

This lemma allows us to apply Theorem 2 (and the remark following its proof) to the restriction of w to D without changing the norming constants A_T obtained from w via (3.6) or (3.6'), with the added advantage that the full condition (w3) is not required, but only its restriction to D. The effect of this is that the end result will apply to the normal distribution and to many other distributions with thin tails.

Corollary 4. *Let $\xi(t)$ be the Gaussian process of Theorem 2. Let $w(t)$ be a non-negative function satisfying conditions (w.1) and (w.2) and such that $\{w > 0\}$ is an open set, and let $a > 0$ be as in Lemma 3. Let D be a set consisting of a finite union of closed intervals or half lines such that $W_a \subseteq D \subseteq \{w > 0\}$. Assume further that*

$$\sup_{\substack{t\in D, t+\tau\in D\\ |\tau|\le 1/T}} \left|\frac{w(t+\tau)}{w(t)} - 1\right| = o(A_T^{-2}) \text{ as } T \to \infty. \qquad (w3_D)$$

and let A_T be as prescribed by (3.6) or (3.6'). Then,

$$\lim_{T\to\infty} \Pr\Big\{A_T\Big(\sup_{t\in D} w(t/T)|\xi(t)| - A_T\Big) \le x\Big\} = e^{-e^{-x}}$$

for all $x \in \mathbf{R}$.

A_T is not independent of f or Ψ, but its order of magnitude is, and we will need this to complete the proofs of our main results.

Lemma 5. *Under the hypotheses of Lemma 3, if A_T is defined by (3.6) or (3.6'), then*

$$0 < \liminf_{T\to\infty} \frac{A_T}{\sqrt{\log T}} \le \limsup_{T\to\infty} \frac{A_T}{\sqrt{\log T}} < \infty. \tag{4.5}$$

Proof. Let I and M be as in the proof of the previous lemma. Then, since, as shown in the previous proof,

$$\frac{\int_{|y|>a} \Psi_\alpha(u/w(y))dy}{\Psi_\alpha(u/M)} \to 0 \text{ as } u \to \infty$$

it follows that for all n large enough,

$$\begin{aligned}\frac{1}{T} = \Lambda_\alpha(A_T) &= \int_{|y|\le a} \Psi_\alpha\Big(\frac{A_T}{w(y)}\Big)dy + \int_{|y|>a} \Psi_\alpha\Big(\frac{A_T}{w(y)}\Big)dy \\ &\le 2a\Psi_\alpha(A_T) + \frac{\int_{|y|>a} \Psi_\alpha(A_T/w(y))dy}{\Psi_\alpha(A_T/M)}\Psi_\alpha\Big(\frac{A_T}{M}\Big) \le 3a\Psi_\alpha(A_T),\end{aligned}$$

since $M < 1$ and Ψ_α is eventually decreasing. So, there is a constant C independent of T such that, for all T large enough,

$$CA_T^{2/\alpha-1}\exp\Big(-\frac{A_T^2}{2}\Big) \ge \frac{1}{T},$$

which implies that $\limsup_T\big[A_T/\sqrt{\log T}\big] < \infty$. The left side of inequality (4.5) follows trivially from the observation that

$$\frac{1}{T} = \Lambda_\alpha(A_T) \ge \int_I \Psi_\alpha\Big(\frac{A_T}{w(y)}\Big)dy \ge |I|\Psi_\alpha\Big(\frac{A_T}{M}\Big).$$

This completes the proof of the lemma if we take (3.6) as the definition of A_T, but it is clear that the same proof with two obvious formal changes gives the lemma if one only assumes that A_T satisfies (3.6'). □

We now prove the distributional limit result for the kernel density estimator. Given f, Ψ, K and $\{h_n\}$ satisfying both the 'usual hypotheses' from Giné, Koltchinskii and Zinn (2001) and the 'additional hypotheses' from the Introduction, all the conditions for the validity of Theorems 1 and 2, hence also of the lemmas from this section and Corollary 4, are satisfied. Define $w := \Psi f^{1/2}$, with Ψ normalized so that $\|w\|_\infty = 1$ and, for this w, set

$$A_n := A_{h_n^{-1}},$$

that is

$$\Lambda_\alpha(A_n) = h_n \text{ or just } \lim_{n\to\infty} \frac{\Lambda_\alpha(A_n)}{h_n} = 1, \tag{4.6}$$

where Λ_α is defined by equation (3.5) for this w. We also recall, from the just mentioned reference, the notation

$$\|g\|_{\Psi,\infty} := \sup_{t\in B_f} |\Psi(t)g(t)|.$$

We then have

Theorem 6. *Assume f, Ψ, K and $\{h_n\}$ satisfy both the 'usual hypotheses' for some $0 < \beta < 1/2$, the 'additional hypotheses', and moreover, that either $B_f = \mathbf{R}$ or $K(0) = \|K\|_\infty$. Assume Ψ is normalized so that $\|\Psi f^{1/2}\|_\infty = 1$. Let A_n be defined as in (4.6). Then the condition*

$$\lim_{t\to\infty} t \Pr\left\{\Psi(X) > \sqrt{th_t|\log h_t|}\right\} = 0, \tag{4.7}$$

is necessary and sufficient for

$$\lim_{n\to\infty} \Pr\left\{A_n\Big(\frac{\sqrt{nh_n}}{\|K\|_2}\|f_n - Ef_n\|_{\Psi,\infty} - A_n\Big) \le x\right\} = \exp\{-e^{-x}\}. \tag{4.8}$$

Proof. We refer to the proof of Theorem 2.1 in Giné, Koltchinskii and Zinn (2001), where B_f is partitioned into four subsets, D_a and $\mathcal{A}_n$, $\mathcal{B}_n$ and $\mathcal{C}_{n,a}$, and where it is proved, under the 'usual hypotheses', that

$$\lim_{n\to\infty}\sqrt{\frac{nh_n}{\log\frac{1}{h_n}}}\sup_{t\in\mathcal{A}_n}|\Psi(t)(f_n(t) - Ef_n(t))| = 0 \text{ in pr.},$$

$$\sqrt{\frac{nh_n}{\log\frac{1}{h_n}}}\sup_{t\in\mathcal{B}_n}|\Psi(t)(f_n(t) - Ef_n(t))| = \kappa \max_{1\le i\le n}\frac{\Psi(X_i)}{\sqrt{nh_n|\log h_n|}} + o_P(1)$$

and

$$\lim_{a\to\infty}\limsup_{n\to\infty} E\sqrt{\frac{nh_n}{\log\frac{1}{h_n}}}\sup_{t\in\mathcal{C}_{n,a}}|\Psi(t)(f_n(t) - Ef_n(t))| = 0,$$

where

$$\mathcal{A}_n := \left\{t \in B_f : \ \Psi(t) > c_n^\beta(nh_n|\log h_n|)^{1/2}\right\},$$

$$\mathcal{B}_n := \left\{t \in B_f : \ f(t)\Psi(t) \le \varepsilon_n^{1-\beta}\Big(\frac{|\log h_n|}{nh_n}\Big)^{1/2}, \ \Psi(t) \le c_n^\beta(nh_n|\log h_n|)^{1/2}\right\},$$

$$\mathcal{C}_{n,a} := \left\{t \in D_a^c \cap B_f : \ f(t)\Psi(t) \ge \varepsilon_n^{1-\beta}(|\log h_n|/nh_n)^{1/2}\right\}$$

(see (2.8), (2.10) and (2.20) in the aforementioned reference). Condition (4.7) implies that $\lim_{n\to\infty}\max_{1\le i\le n}\Psi(X_i)/\sqrt{nh_n|\log h_n|} = 0$ in probability, so that the above expression for the sup over $\mathcal{B}_n$ also tends to zero in probability. By Lemma 5, for all $x \in \mathbf{R}$,

$$\limsup\frac{A_n + x/A_n}{\sqrt{\log\frac{1}{h_n}}} < \infty,$$

and therefore we can replace $\sqrt{\log h_n^{-1}}$ by $A_n + x/A_n$ for all $x \in \mathbf{R}$ in the previous limits. This substitution gives:

$$\lim_{n\to\infty} \Pr\left\{ A_n\Big(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in\mathcal{A}_n} |\Psi(t)(f_n(t) - Ef_n(t))| - A_n\Big) \geq x \right\} = 0, \tag{4.9}$$

$$\lim_{n\to\infty} \Pr\left\{ A_n\Big(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in\mathcal{B}_n} |\Psi(t)(f_n(t) - Ef_n(t))| - A_n\Big) \geq x \right\} = 0 \tag{4.10}$$

and

$$\lim_{a\to\infty} \limsup_{n\to\infty} \Pr\left\{ A_n\Big(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in\mathcal{C}_{n,a}} |\Psi(t)(f_n(t) - Ef_n(t))| - A_n\Big) \geq x \right\} = 0 \tag{4.11}$$

for all $x \in \mathbf{R}$. Now, set $w := \Psi f^{1/2}$. Since $\|\Psi f^{\beta}\|_\infty \leq c < \infty$, we have

$$W_a = \{t : |t| \leq a, w(t) \geq 1/a\} \subset D_{a'} = \{t : |t| \leq a', f(t) \geq 1/a'\}$$

where $a' := (ac)^{\frac{1}{1/2-\beta}} \vee a$. Hence, there exists $a_0 < \infty$ such that Corollary 4 applies to $D = D_a$ for all $a \geq a_0$. This and Theorem 1 give that for all $a > a_0$ and all $x \in \mathbf{R}$,

$$\lim_{n\to\infty} \Pr\left\{ A_n\Big(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in D_a} |\Psi(t)(f_n(t) - Ef_n(t))| - A_n\Big) \leq x \right\} = e^{-e^{-x}}. \tag{4.12}$$

Now sufficiency of (4.7) follows from (4.9)-(4.12).

Next we prove the necessity of (4.7). Since the limit distribution $\exp\{-e^{-x}\}$ in (4.8) is continuous, the convergence in (4.8) is uniform in x. Then for any $t > 1$

$$\begin{aligned}
&\limsup_{n\to\infty} \Pr\left\{ \frac{\sqrt{nh_n}}{A_n\|K\|_2} \|f_n - Ef_n\|_{\Psi,\infty} > t \right\} \\
&\leq \lim_{n\to\infty} \Pr\left\{ \frac{\sqrt{nh_n}}{A_n\|K\|_2} \|f_n - Ef_n\|_{\Psi,\infty} > \frac{A_n + A_n^2}{A_n^2} \right\} \\
&= \lim_{n\to\infty} \Pr\left\{ A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2} \|f_n - Ef_n\|_{\Psi,\infty} - A_n\right) > A_n \right\} \\
&= \lim_{n\to\infty} (1 - \exp\{-e^{-A_n}\}) = 0.
\end{aligned}$$

By Montgomery-Smith (1993) maximal inequality (e.g., de la Peña and Giné (1999), p. 6), the previous limit implies that

$$\lim_{n\to\infty} Pr\left\{ \frac{\sqrt{nh_n}}{A_n\|K\|_2} \frac{1}{nh_n} \max_{1\leq i\leq n} \|K((X_i - \cdot)/h_n) - EK((X - \cdot)/h_n)\|_{\Psi,\infty} > \frac{t}{30} \right\} = 0.$$

Then, following the proof of the necessity in Theorem 2.1 in Giné, Koltchinskii and Zinn (2001), one has

$$\lim_{n\to\infty} Pr\left\{ \frac{1}{A_n\sqrt{nh_n}} \max_{1\leq i\leq n} \Psi(X_i) > \frac{t\|K\|_2}{30} \right\} = 0,$$

which implies

$$\lim_{n\to\infty} nPr\left\{\Psi(X) > t\|K\|_2 A_n\sqrt{nh_n}/30\right\} = 0,$$

and (4.7) follows (in view of Lemma 5 and the regular variation of $\lambda_n := (nh_n|\log h_n|)^{1/2}$). □

Specially interesting is the case $\Psi = \|f\|_\infty^{-1/2}$, which yields convergence in distribution for the unweighted sup norm of $\|f_n - Ef_n\|_{B_f}$. In this case it is more efficient to invoke Theorem 3.3 (actually Proposition 3.2) and Remark 3.5 in Giné and Guillou (2002), to obtain, under weaker hypotheses, that

$$\lim_{a\to\infty}\limsup_{n\to\infty} Pr\left\{A_n\left(\sqrt{nh_n}\sup_{t\in D_a^c\cap B_f}|f_n(t)-Ef_n(t)| - A_n\right)\ge x\right\} = 0$$

for all $x \in \mathbf{R}$. Combining the hypotheses under which this limit holds with the hypotheses for Theorems 1 and 2, we obtain that the following theorem will hold under the following conditions:

i). Conditions on $\{h_n\}$*:*

$$h_n \searrow 0,\ \ nh_n \nearrow \infty,\ \ \frac{nh_n}{|\log h_n|(\log n)^2} \to \infty,$$
$$\frac{\log h_n}{\log\log n} \to \infty.$$

ii). Conditions on f*:* f is bounded, Hölder continuous and piecewise monotone on B_f, which is an open set consisting of the union of a finite number of intervals (or half-lines), and, for every $a > 0$,

$$\sup_{\substack{s,t\in D_a\\ |s-t|<1/T}}\left|1-\frac{\sqrt{f(t)}}{\sqrt{f(s)}}\right| = o\left(\frac{1}{\log T}\right).$$

iii). Conditions on K*:* the same as in Theorem 7 except that $K(0)$ needs not equal $\|K\|_\infty$ when B_f is not all of $\mathbf{R}$.

Theorem 7. *Under the conditions immediately above,*

$$\lim_{n\to\infty}\Pr\left\{A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2\|f\|_\infty^{1/2}}\|f_n - Ef_n\|_{B_f} - A_n\right)\le x\right\} = \exp\{-e^{-x}\},\quad (4.13)$$

where A_n *is defined by equation (4.6) with* $w = (f/\|f\|_\infty)^{1/2}$.

The next two theorems follow directly from results in Giné, Koltchinskii and Zinn (2001) (see Theorem 2.1, Corollary 2.4 and Theorem 3.1 there).

Let

$$\lambda_t := \lambda(t) := \sqrt{th_t|\log h_t|}.$$

For $\gamma > 0$ and $L > 0$, let $Z_{\gamma,L}$ be a nonnegative random variable with distribution function

$$\Pr\{Z_{\gamma,L}\le t\} = \exp\{-Lt^{-1/\gamma}\},\ \ t > 0.$$

If $L = 0$, we set $Z_{\gamma,L} := 0$.

Theorem 8. *Assume f, Ψ, K and $\{h_n\}$ satisfy the 'usual hypotheses' for some $0 < \beta < 1/2$. and moreover, that either $B_f = \mathbf{R}$ or $K(0) = \|K\|_\infty$. Let $\gamma := (1-\eta)/2$ (where $-\eta$ is the exponent of regular variation of h_n). Then, the sequence of random variables*

$$\sqrt{\frac{nh_n}{2|\log h_n|}}\Big\|f_n - Ef_n\Big\|_{\Psi,\infty}, \quad n \in \mathbf{N}, \tag{4.14}$$

converges in distribution if and only if the limit

$$\lim_{t\to\infty} t\,\Pr\Big\{\Psi(X) > \lambda_t\Big\} =: L \in [0, +\infty) \tag{4.15}$$

exists. In this case, the limit of the sequence (4.14) is equal in distribution to

$$\Big(\frac{\|K\|_\infty Z_{\gamma,L}}{\sqrt{2}}\Big)\bigvee\big(\|K\|_2\|f^{1/2}\|_{\Psi,\infty}\big).$$

Proof. According to Theorem 2.1 in Giné, Koltchinskii and Zinn (2001), the sequence (4.14) is stochastically bounded iff

$$\limsup_{t\to\infty} t\,\Pr\Big\{\Psi(X) > \lambda_t\Big\} < +\infty, \tag{4.15$'$}$$

and in this case

$$\sqrt{\frac{nh_n}{2|\log h_n|}}\Big\|f_n - Ef_n\Big\|_{\Psi,\infty} = \bigg(\|K\|_\infty \frac{\max_{1\le i\le n}\Psi(X_i)}{\sqrt{2}\lambda_n}\bigg) \bigvee\bigg(\|K\|_2\|f^{1/2}\|_{\Psi,\infty}\bigg) + o_p(1). \tag{4.16}$$

Of course, this is true under condition (4.15), which also implies that

$$\Pr\{\Psi(X) > u\} = \frac{L + o(1)}{\lambda^{-1}(u)} \text{ as } u \to \infty.$$

Since λ^{-1} is a regularly varying function (with exponent $1/\gamma$), we obtain that for all $x > 0$

$$\Pr\{\Psi(X) > x\lambda_n\} = \frac{L + o(1)}{\lambda^{-1}(x\lambda_n)} = \frac{Lx^{-1/\gamma} + o(1)}{n} \text{ as } n \to \infty.$$

This, of course, implies by a standard computation (e.g., Theorem 1.5.1 in Leadbetter, Lindgren and Rootzén (1986)) that the sequence

$$\frac{\max_{1\le i\le n}\Psi(X_i)}{\lambda_n}$$

converges in distribution to $Z_{\gamma,L}$, and this together with (4.16) yields the limit of (4.14) in distribution.

On the other hand, if (4.14) does converge in distribution, then it is stochastically bounded and (4.15') holds. This implies the representation (4.16) (by Theorem 2.1 in Giné, Koltchinskii and Zinn (2001)), which in turn implies the convergence in distribution of the sequence

$$\frac{\max_{1\le i\le n}\Psi(X_i)}{\lambda_n}\bigvee B,$$

where

$$B := \sqrt{2}\|K\|_2\|f^{1/2}\|_{\Psi,\infty}/\|K\|_\infty.$$

If G denotes the limit distribution of the last sequence, then we have, for all $x > B$, x a continuity point of G,

$$\Pr\Big\{\max_{1\le i\le n}\Psi(X_i)\le x\lambda_n\Big\}\to G(x) \text{ as } n\to\infty.$$

For all $x > B$ (continuity points of G) such that $G(x) > 0$ this gives, again as in Theorem 1.5.1 in Leadbetter, Lindgren and Rootzén (1986), that

$$n\Pr\Big\{\Psi(X) > x\lambda_n\Big\}\to -\log G(x) =: g(x).$$

Hence, we have (using simple properties of regularly varying functions) that

$$\Pr\Big\{\Psi(X) > u\Big\} = \frac{g(x)+o(1)}{\lambda^{-1}(u/x)} = \frac{g(x)x^{1/\gamma}+o(1)}{\lambda^{-1}(u)} \text{ as } u\to\infty.$$

This implies that, for some constant L, $g(x) = Lx^{-1/\gamma}$ and we have

$$\Pr\Big\{\Psi(X) > \lambda_t\Big\} = \frac{L+o(1)}{t},$$

so condition (4.15) holds. □

Theorem 9. *Assume f, Ψ, K and $\{h_n\}$ satisfy the 'usual hypotheses' for some $0<\beta\le 1$ and, moreover, that either $B_f = \mathbf{R}$ or $K(0) = \|K\|_\infty$. Let d_t be a strictly increasing regularly varying function with exponent γ such that $d_t/\lambda_t\to\infty$ and $d_t\ge ct^\beta$ for some $c>0$. The sequence of random variables*

$$\frac{nh_n}{d_n}\Big\|f_n - Ef_n\Big\|_{\Psi,\infty},\quad n\in\mathbf{N}, \tag{4.14$'$}$$

converges in distribution if and only if the limit

$$\lim_{t\to\infty} t\Pr\Big\{\Psi(X) > d_t\Big\} =: L\in[0,+\infty) \tag{4.17}$$

exists. Moreover, the limit in (4.14') is $\|K\|_\infty Z_{\gamma,L}$.

The proof is similar to the previous one and is based on Theorem 3.1 in Giné, Koltchinskii and Zinn (2001).

Remark. If d_t is a strictly increasing regularly varying function such that either $d_t = \lambda_t$, or $d_t/\lambda_t \to \infty$, $d_t \geq ct^\beta$ for some $c > 0$, and if condition (4.17) holds, then it's easy to see that, up to a linear rescaling, the cases described in Theorems 6, 8 and 9 are *the only cases* when the norm $\|f_n - Ef_n\|_{\Psi,\infty}$ (properly centered and normalized) converges to a non-degenerate distribution. The proof follows from a well known result of Khinchin (see Theorem 1.2.3 in Leadbetter, Lindgren and Rootzén (1986)).

Remark. Theorem 9 covers the asymptotic behavior of the statistic $\|(f_n - Ef_n)/\sqrt{f}\|_\infty$, and shows how this statistic behaves in a markedly different way from $\|(f_n - Ef_n)/\sqrt{f}\|_{[-a,a]}$ (that is, from the case considered by Bickel and Rosenblatt (1973), see also the last remark in Section 3 above). For instance, if $f(x) = 1/x^2$, $|x| > 1$, then Theorem 9 shows that $n^{-1}h_n\|(f_n - Ef_n)/\sqrt{f}\|_{[-1,1]^c} \to_d \|K\|_2 Z_{2,1}$ and that, in general, its behavior strongly depends on f. On the contrary, in the Bickel and Rosenblatt situation, A_n is independent of f, and in the situation of Theorems 6 and 7, although A_n depends on f, its order of magnitude does not (Lemma 5).

5. Data-dependent normings

The normings A_n in Theorem 6 depend on the unknown density f, which makes it difficult to use this result directly to develop hypotheses tests or confidence bands for unknown densities. In this section, we sketch an approach to a (statistically) more practical version of this result which takes care of this difficulty as well as of the bias of the kernel density estimator. It's not our goal at this moment to develop these more statistical aspects of the problem to their full extent, but only to indicate possible ways to do it. (We pursue this subject in Giné, Koltchinskii and Sakhanenko (2003), a subsequent paper that has however appeared before the present one.)

First, we define

$$\omega_f(t;\delta) := \sup\Big\{|f(t_1) - f(t_2)| : t_1, t_2 \in (t-\delta, t+\delta)\Big\}$$

and

$$\omega_f^\Psi(\delta) := \sup_{t\in\mathbf{R}} \Psi(t)\omega_f(t;\delta),$$

and note that, under the standard assumption $\int_{\mathbf{R}} K(x)dx = 1$, we have the following straightforward bound on the bias of f_n :

$$\|Ef_n - f\|_{\Psi,\infty} \leq \omega_f^\Psi(h_n). \tag{5.1}$$

In what follows we denote

$$w := \Psi\sqrt{f},\ \hat{w} := \hat{w}_n := \Psi\sqrt{f_n}.$$

Given $a > 0$, denote

$$\hat{W}_a := \{y : |y| \leq a,\ \hat{w}(y) \geq a^{-1}\}.$$

Given a sequence $a_n \to \infty$, let $\hat{A}_n$ be defined as the solution of the equation

$$\int_{\hat{W}_{a_n}} \Psi_\alpha\Big(\frac{\hat{A}_n}{\hat{w}_n(y)}\Big)dy = h_n.$$

Theorem 10. *Suppose that all the conditions of Theorem 6 hold (including condition (4.7)), that $\int_{\mathbf{R}} K(x)dx = 1$ and that, in addition,*

$$\omega_f^\Psi(h_n) = o\Big((nh_n|\log h_n|)^{-1/2}\Big) \text{ as } n \to \infty. \tag{5.2}$$

If

$$a_n^2 \sup_{|t|\le a_n} \Psi(t) = o\Big(\sqrt{nh_n}|\log h_n|^{-3/2}\Big), \tag{5.3}$$

for some sequence $a_n \to \infty$, then

$$\lim_{n\to\infty} \Pr\Big\{\hat{A}_n\Big(\frac{\sqrt{nh_n}}{\|K\|_2}\|f_n - f\|_{\Psi,\infty} - \hat{A}_n\Big) \le x\Big\} = \exp\{-e^{-x}\} \tag{5.4}$$

for all $x \in \mathbf{R}$, where $\hat{A}_n$ are as defined above.

Proof. We start with several simple observations. By bound (5.1) on the bias, condition (5.2) and Theorem 6, we have

$$\lim_{n\to\infty} \Pr\Big\{A_n\Big(\frac{\sqrt{nh_n}}{\|K\|_2}\|f_n - f\|_{\Psi,\infty} - A_n\Big) \le x\Big\} = \exp\{-e^{-x}\}, \tag{5.4}$$

where A_n is any sequence such that

$$h_n^{-1}\int_{\mathbf{R}} \Psi_\alpha\Big(\frac{A_n}{w(y)}\Big)dy \to 1 \text{ as } n \to \infty. \tag{5.5}$$

In particular, by Lemma 3, A_n can be chosen as a sequence such that

$$h_n^{-1}\int_{W_{b_n}} \Psi_\alpha\Big(\frac{A_n}{w(y)}\Big)dy \to 1 \text{ as } n \to \infty \tag{5.6}$$

for any $b_n \to \infty$ (since then (5.5) also holds). Note that if A_n and $\bar{A}_n$ are two sequences satisfying (5.5) (and, hence, also (5.4)), then

$$\Big|\frac{\bar{A}_n}{A_n} - 1\Big| = o(A_n^{-2}) \text{ as } n \to \infty. \tag{5.7}$$

On the other hand, if (5.7) holds, then A_n in (5.4) can be replaced by $\bar{A}_n$ (even if $\bar{A}_n$ is a sequence of r.v. and (5.7) holds in probability, which will be the case later in the proof). All this follows easily from the fact that two sequences of r.v. η_n and $C_n\eta_n + D_n$ converge in distribution to the same continuous r.v. (e.g., double exponential) if and only if $C_n \to 1$ and $D_n \to 0$ as $n \to \infty$.

Next we show that under condition (5.3)

$$\sup_{y\in W_{a_n}\cup\hat W_{a_n}}\left|\frac{\hat w(y)}{w(y)}-1\right| = o_P\left(\frac{1}{|\log h_n|}\right). \tag{5.8}$$

Indeed, we have

$$\begin{aligned}
\sup_{y\in W_{a_n}}\left|\frac{\hat w(y)}{w(y)}-1\right| &= \sup_{y\in W_{a_n}}\frac{\left|\sqrt{f_n(y)}-\sqrt{f(y)}\right|}{\sqrt{f(y)}}\\
&= \sup_{y\in W_{a_n}}\frac{\left|f_n(y)-f(y)\right|}{\sqrt{f(y)}\left(\sqrt{f_n(y)}+\sqrt{f(y)}\right)}\\
&\le \sup_{y\in W_{a_n}}\frac{\Psi^2(y)\left|f_n(y)-f(y)\right|}{\Psi^2(y)f(y)}\\
&\le a_n^2\sup_{|y|\le a_n}\Psi(y)\left\|f_n-f\right\|_{\Psi,\infty} =: \Delta_n. \qquad (5.9)
\end{aligned}$$

Quite similarly, we get

$$\sup_{y\in\hat W_{a_n}}\left|\frac{w(y)}{\hat w(y)}-1\right| \le \Delta_n. \tag{5.10}$$

Theorem 6 and Lemma 5 imply that

$$\left\|f_n-f\right\|_{\Psi,\infty} = O_P\left(\sqrt{\frac{|\log h_n|}{nh_n}}\right)$$

and this together with condition (5.3) implies that

$$\Delta_n = o_P\left(\frac{1}{|\log h_n|}\right). \tag{5.11}$$

Now (5.8) follows from (5.9)–(5.11). Note that (5.8) implies the existence of a sequence $\varepsilon_n\to 0$ such that

$$\varepsilon_n = o\left(\frac{1}{|\log h_n|}\right) \tag{5.12}$$

and

$$\Pr\left\{\sup_{y\in W_{a_n}\cup\hat W_{a_n}}\left|\frac{\hat w(y)}{w(y)}-1\right|\ge\varepsilon_n\right\}\to 0 \text{ as } n\to\infty. \tag{5.13}$$

To complete the proof, we observe that, on the event

$$E := \left\{\sup_{y\in W_{a_n}\cup\hat W_{a_n}}\left|\frac{\hat w(y)}{w(y)}-1\right|<\varepsilon_n\right\},$$

we have

$$(1-\varepsilon_n)w(y) \le \hat{w}(y) \le (1+\varepsilon_n)w(y) \text{ for all } y \in W_{a_n} \cup \hat{W}_{a_n},$$

which, in turn, yields the inclusion

$$W_{a_n(1-\varepsilon_n)} \subset \hat{W}_{a_n} \subset W_{a_n(1+\varepsilon_n)}.$$

Thus, we have (on the same event E)

$$\int_{W_{a_n(1-\varepsilon_n)}} \Psi_\alpha\Big(\frac{\hat{A}_n}{(1-\varepsilon_n)w(y)}\Big)dy \le \int_{\hat{W}_{a_n}} \Psi_\alpha\Big(\frac{\hat{A}_n}{\hat{w}_n(y)}\Big)dy = h_n$$
$$\le \int_{W_{a_n(1+\varepsilon_n)}} \Psi_\alpha\Big(\frac{\hat{A}_n}{(1+\varepsilon_n)w(y)}\Big)dy. \quad (5.14)$$

Defining now A_n^- and A_n^+ respectively as the solutions of the equations

$$\int_{W_{a_n(1-\varepsilon_n)}} \Psi_\alpha\Big(\frac{A_n^-}{w(y)}\Big)dy = h_n$$

and

$$\int_{W_{a_n(1+\varepsilon_n)}} \Psi_\alpha\Big(\frac{A_n^+}{w(y)}\Big)dy = h_n,$$

we conclude from (5.14) (using the monotonicity of the corresponding functions) that on the event E

$$A_n^-(1-\varepsilon_n) \le \hat{A}_n \le A_n^+(1+\varepsilon_n). \quad (5.15)$$

According to the observations made at the beginning of the proof, (5.4) holds with A_n replaced by A_n^+ or by A_n^-, which implies that (by (5.7))

$$\Big|\frac{A_n^+}{A_n} - 1\Big| = o(A_n^{-2})$$

and

$$\Big|\frac{A_n^-}{A_n} - 1\Big| = o(A_n^{-2}) \text{ as } n \to \infty.$$

Since, by Lemma 5, A_n is of the order $|\log h_n|^{1/2}$, it follows from (5.12), (5.13) and (5.15) that

$$\Big|\frac{\hat{A}_n}{A_n} - 1\Big| = o_p(A_n^{-2}) \text{ as } n \to \infty.$$

This implies that A_n can be replaced by $\hat{A}_n$ in (5.4), which completes the proof. □

Theorem 10 is not completely satisfactory in the sense that the weight Ψ itself might depend on f, so that, in the end, f has only been partially replaced by f_n. However this theorem is part of the solution. To illustrate this point we will specify a distribution free result for the simplest case, namely, the case Ψ =constant, which corresponds to Theorem 7.

Corollary 11. *Suppose that the conditions of Theorem 7 hold, that $B_f = \mathbf{R}$ and that $\int_{\mathbf{R}} K(x)dx = 1$. Assume further that*

$$\omega_f(h_n) = o\Big((nh_n|\log h_n|)^{-1/2}\Big)$$

and let $a_n \to \infty$ be such that

$$a_n^2 = o\Big(\sqrt{nh_n}|\log h_n|^{-3/2}\Big).$$

Let

$$\bar{w} = \bar{w}_n = \sqrt{f_n/\|f_n\|_\infty}$$

and let $\bar{A}_n$ be the sequence of random variables defined by the relation

$$\int_{\bar{W}_{a_n}} \Psi_\alpha\Big(\frac{\bar{A}_n}{\bar{w}_n(y)}\Big)dy = h_n,$$

where $\bar{W}_{a_n} = \{y: \ |y| \le a_n, \ \bar{w}_n(y) \ge a_n^{-1}\}$. Then we have both,

$$\lim_{n\to\infty} \Pr\left\{\bar{A}_n\left(\frac{\sqrt{nh_n}}{\|K\|_2\|f_n\|_\infty^{1/2}}\|f_n - f\|_\infty - \bar{A}_n\right) \le x\right\} = \exp\{-e^{-x}\}$$

and

$$\lim_{n\to\infty} \Pr\left\{\bar{A}_n\left(\frac{\sqrt{nh_n}}{\|K\|_2\|f_n\|_\infty^{1/2}}\|f_n - Ef_n\|_\infty - \bar{A}_n\right) \le x\right\} = \exp\{-e^{-x}\}$$

for all $x \in \mathbf{R}$.

Proof. The proof is just like that of Theorem 10 once we show that

$$\sup_{y\in W_{a_n}\cup\bar{W}_{a_n}} \left|\frac{\bar{w}(y)}{w(y)} - 1\right| = o_P\left(\frac{1}{|\log h_n|}\right). \tag{5.8$'$}$$

In this case we have

$$\begin{aligned}\sup_{y\in W_{a_n}} \left|\frac{\bar{w}(y)}{w(y)} - 1\right| &= \sup_{y\in W_{a_n}} \left|\frac{\sqrt{\|f\|_\infty f_n(y)}}{\sqrt{\|f_n\|_\infty f(y)}} - 1\right| \\ &\le \sqrt{\frac{\|f\|_\infty}{\|f_n\|_\infty}} \sup_{y\in W_{a_n}} \left|\frac{\sqrt{f_n(y)}}{\sqrt{f(y)}} - 1\right| + \left|\frac{\sqrt{\|f\|_\infty}}{\sqrt{\|f_n\|_\infty}} - 1\right|.\end{aligned}$$

Using (5.4) for $\Psi = 1/\|f\|_\infty^{1/2}$ and proceeding as in the previous proof, we have that on W_{a_n},

$$\left|\frac{\sqrt{f_n(y)}}{\sqrt{f(y)}} - 1\right| = \frac{|f_n(y) - f(y)|}{\sqrt{f(y)}(\sqrt{f_n(y)} + \sqrt{f(y)})} = o_P\left(\frac{1}{|\log h_n|}\right).$$

Moreover, again by (5.4), we have both, that $\sqrt{\|f_n\|_\infty}/\sqrt{\|f\|_\infty} \to 1$ and that

$$\left|\frac{\sqrt{\|f\|_\infty}}{\sqrt{\|f_n\|_\infty}} - 1\right| \le \frac{\|f - f_n\|_\infty}{\|f_n\|_\infty} = O_P\left(\sqrt{\frac{|\log h_n|}{nh_n}}\right).$$

Hence,

$$\sup_{y \in W_{a_n}} \left|\frac{\bar{w}(y)}{w(y)} - 1\right| = o_P\left(\frac{1}{|\log h_n|}\right).$$

Likewise, we get that

$$\sup_{y \in \bar{W}_{a_n}} \left|\frac{w(y)}{\bar{w}(y)} - 1\right| = o_P\left(\frac{1}{|\log h_n|}\right).$$

□

References

Albin, J.M.P.: On Extremal Theory for Stationary Processes. Ann. Probab. **18**, 92–128 (1990)

Bickel, P.J., Rosenblatt, M.: On some global measures of the deviations of density function estimates. Ann. Statist. **1**, 1071–1095 (1973)

de la Peña, V., Giné, E.: Decoupling, from dependence to independence. Springer-Verlag, New York (1999)

Giné, E., Guillou, A.: Rates of strong consistency for multivariate kernel density estimators. Ann. Inst. Henri Poincaré **38**, 907–922 (2000)

Giné, E., Zinn, J.: Some limit theorems for empirical processes. Ann. Probab. **12**, 929–989 (1984)

Giné, E., Koltchinskii, V., Sakhanenko, L.: Convergence in distribution of self-normalized sup-norms of kernel density estimators. High Dimensional Probability III, Progress in Probability **55**, 241–254 Birkhäuser, Basel (2003)

Giné, E., Koltchinskii, V., Zinn, J.: Weighted uniform consistency of kernel density estimators. To appear in Ann. Probab. (2001)

Koltchinskii, V. I.: Komlós-Major-Tusnády approximation for the general empirical process and Haar expansions of classes of functions. J. Theoret. Probab. **7**, 73–118 (1994)

Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent rv's and the sample df. I. Z. Wahrsch. verw. Gebiete **32**, 111–131 (1975)

Konakov, V.D., Piterbarg, V.I.: On the convergence rate of maximal deviations distributions for kernel regression estimates. J. Multivariate Analysis **15**, 279–294 (1984)

Leadbetter, M.R., Lindgren, G., Rootzén, H.: Extremes and related properties of random sequences and processes. Springer, New York (1986)

Ledoux, M., Talagrand, M.: Probability in Banach spaces. Springer, New York (1991)

Montgomery-Smith, S.J.: Comparison of sums of independent identically distributed random vectors. Probab. Math. Statist. **14**, 281–285 (1993)

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

Piterbarg, V.I.: Asymptotic methods in the theory of Gaussian processes and fields. Translations of Mathematical Monographs, 148. American Mathematical Society, Providence, RI (1996)

Rio, E.: Local invariance principles and their applications to density estimation. Probab. Theory and Related Fields **98** 21–45 (1994)