Convergence in distribution of Self-Normalized Sup-Norms of Kernel Density Estimators

Evarist Giné[1], Vladimir Koltchinskii[2], and Lyudmila Sakhanenko

Abstract. Let f_n denote a kernel density estimator of a density f on the real line, for a bounded, compactly supported probability kernel. Under relatively weak smoothness conditions on f and K it is proved, for every $0 < \beta < 1/2$, that the sequence

$$\hat{A}_n \left(\frac{\sqrt{nh_n}}{\|K\|_2 \|f_n\|_\infty^{1/2-\beta}} \sup_{t \in \hat{D}_{a_n}} \frac{|f_n(t) - f(t)|}{f_n^\beta(t)} - \hat{A}_n \right)$$

converges in distribution to the double exponential law. Here $\hat{A}_n$ is constructed from the sample, $a_n \to \infty$ as a power of n and $\hat{D}_{a_n} = \{t : f_n(t) \geq a_n^{-1}\}$. Thus, this result provides distribution free asymptotic confidence bands for densities on the real line.

1. Introduction

Let X, X_i, $i \in \mathbb{N}$, be i.i.d. real random variables with density f, let K be a kernel and $\{h_n\}$, the bandwidths, be a monotonically decreasing sequence such that $h_n \to 0$ and $nh_n \to \infty$, and let

$$f_n(t) := \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_n}\right), \quad n \in \mathbb{N}, \tag{1.1}$$

be the corresponding kernel density estimator. The object of this note is to obtain a distributional limit theorem for weighted supremum norms of $|f_n - f|$, suitably centered and normalized, so that the weigths, the centering and the normalization depend only on the data, the intervals over which the supremum is taken increases to the whole line, and the limit is distribution free. Obviously, such a result can be used for the construction of asymptotic confidence bands for f. However, speed of convergence, which would be nice to have, will not be considered.

1991 *Mathematics Subject Classification.* Primary 62G07; Secondary 62G20, 60F15.
Key words and phrases. kernel density estimates, convergence in distribution, self-normalized sup norms, (asymptotic) confidence bands for densities.
[1]Research partially supported by NSF Grant No. DMS-0070382. [2]Research partially supported by NSA Grant No. MDA904-02-1-0075.

This research builds upon recent previous work by the same authors in [2], where we obtained convergence in distribution to the double exponential distribution of statistics of the form

$$A_n\left(\|\Psi(t)(f_n(t)-Ef_n(t))\|_\infty - A_n\right), \tag{1.2}$$

thus extending to suprema over the whole line the classical result of Bickel and Rosenblatt ([1]). In that article we actually characterized such convergence by necessary and sufficient conditions in terms of the tail probabilities of $\Psi(X)$. Implicit in that work, and made explicit below, is the observation that if the n-th sup is taken over the set $\{t : f(t) \geq (nh_n)^{-\theta}\}$ for $\theta \in (0,1)$, then the tail probability condition becomes superfluous and the result holds for all densities satisfying some not too restrictive smoothness properties. However A_n still depends on Ψ which in turn must satisfy a condition related to f (namely, $\Psi(x)f(x)^\beta \leq c$ for some $0<\beta<1/2$). Our main result below shows that one can replace A_n by a function of the data, that the weights can also be made solely dependent on the data, and that the sup can be taken over the data-dependent set $\{t : f_n(t) \geq (nh_n)^{-\theta}\}$.

This article is far from being self-contained: it relies instead on results and proofs from [3] and [2], and the methods used originate as well with these papers, particularly the latter. What ultimately make these results possible are adaptations of classical work on the asymptotic distribution of weighted suprema of stationary Gaussian processes (Theorem 2 in [2]) in combination with the KMT approximation of the empirical process, and estimates of moments and tail probabilities of suprema of empirical processes over certain classes of functions (see [3] and [2]).

Next we describe (and so, we get out of the way) the general assumptions for the main theorem below. We begin with hypotheses on the density, that in our case will coincide with hypotheses on the weights as well because we will take our weights to be $f^{-\beta}$ and $f_n^{-\beta}$ for some $0<\beta<1/2$ in the main result. These hypotheses are:

(D) f is a non-vanishing bounded, continuous, piecewise monotone density on $\mathbb{R}$ satisfying the following two conditions: a) for every $\delta>0$ there exist $c\in(0,\infty)$ and $h_\delta>0$ such that for all $|\tau|\leq h_\delta$ and all $t\in\mathbb{R}$,

$$c^{-1}f^\delta(t) \leq \frac{f(t+\tau)}{f(t)} \leq cf^{-\delta}(t),$$

and b) for all $r>0$,

$$\sup_{f(t)\geq h^r, |\tau|\leq h}\left|\frac{f(t+\tau)}{f(t)}-1\right| = o\left(\frac{1}{|\log h|}\right) \quad \text{as} \quad h\to 0.$$

Note that symmetric exponential as well as normal densities satisfy this condition. For a discussion on this type of conditions see [3].

The hypotheses on the kernel are:

(K_α) K is a probability kernel of bounded variation, with support contained in $[-1/2, 1/2]$, and such that the function

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

satisfies that

$$r(t) = 1 - C|t|^\alpha + o\left(|t|^\alpha\right) \quad \text{as} \quad t \to 0,$$

for some $0 < \alpha \le 2$, and $\sup_{|t|>\varepsilon} |r(t)| < 1$ for all $\varepsilon > 0$. We set $\kappa := \|K\|_\infty$.

Finally, here is the hypothesis on the bandwidths h_n:

(H) h_t, $t \ge 1$, is regularly varying at infinity with exponent $-\tau$ for some $\tau \in (0,1)$, $h_t \searrow 0$ and $th_t \nearrow \infty$ as $t \nearrow \infty$.

Taking $\Psi = f^{-\beta}$ for some $0 < \beta < 1/2$, these conditions imply all the general assumptions (K, F, UH, D, W, WD and H) in [2].

The results in the present article complement Theorem 6 of [2] in substantial ways. It is thus worthwhile to conclude this introduction with a statement of this theorem. For simplicity we state it only in the case of non-vanishing densities (this is not a harmless simplification, but we make it anyway). That theorem deals with weighted suprema, and here are the assumptions on the weight function $\Psi : \mathbb{R} \mapsto \mathbb{R}_+$:

(W.a): Ψ is strictly positive, continuous and piecewise monotone.

(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 $x \in \mathbb{R}$,

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

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

$$\sup_{\Psi(x) \le h^{-r},\ |y| \le h} \left| \frac{\Psi(x+y)}{\Psi(x)} - 1 \right| = o\left(\frac{1}{|\log h|}\right) \quad \text{as } h \to 0.$$

$(\mathrm{W.D})_\beta$: For a fixed $\beta \in (0,2)$, $\|f^\beta\|_{\Psi,\infty} := \sup_{t\in\mathbb{R}} \Psi(t) f^\beta(t) < \infty$.

The following quantities are needed in order to define the norming and centering constants A_n: for $\alpha \in (0,2]$, the same α from hypothesis (K_α), we define the functions

$$\Psi_\alpha(u) := \sqrt{\frac{2}{\pi}} C^{1/\alpha} H_\alpha u^{2/\alpha - 1} e^{-u^2/2}, \quad u \ge 0, \tag{1.3}$$

where the constants H_α are absolute and are related to the asymptotic theory of stationary Gaussian processes (see [2]), with $H_2 = \sqrt{\pi}$, and

$$\Lambda_\alpha(u) := \int_{-\infty}^{\infty} \Psi_\alpha\left(\frac{u}{\Psi(y) f^{1/2}(y)}\right) dy, \quad u > 0. \tag{1.4}$$

With these definitions we also assume that the weight function Ψ and the density f satisfy (w.2): $\Lambda_\alpha(u_0) < \infty$ for some $0 < u_0 < \infty$ (hence for all $u \geq u_0$).

Finally, we define the norming and centering constants A_n: if $A_T > 0$ is a function of $T > 0$ satisfying the relation

$$\lim_{T\to\infty} T\Lambda_\alpha(A_T) = 1, \tag{1.5}$$

and h_n are the bandwidths, then

$$A_n := A_{h_n^{-1}}. \tag{1.6}$$

It can be proved (see [2]) that A_T is, up to multiplicative constants, of the order of $\sqrt{\log T}$ as $T \to \infty$. We then have:

Theorem 0. (Special case of Theorem 6 in [2].) *Assume the conditions (D), (K_α) for some $\alpha \in (0, 2]$, (H), (W.a)-(W.c), $(WD)_\beta$ for some $\beta \in (0, 1/2)$, and (w.2), and let Ψ be normalzied so that $\|\Psi f^{1/2}\|_\infty = 1$. Then, the condition*

$$\lim_{n\to\infty} n \Pr\Big\{\Psi(X) > \sqrt{nh_n|\log h_n|}\Big\} = 0,$$

is necessary and sufficient for

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

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

2. The limit theorem

We still need some extra definitions and hypotheses. For any $a > 0$ we set

$$D_a := \left\{t : f(t) \geq a^{-1}\right\},\ \hat{D}_a := \left\{t : f_n(t) \geq a^{-1}\right\}. \tag{2.1}$$

Taking

$$a_n = O\left((nh_n)^\theta\right), \tag{2.2}$$

for some $\theta \in (0, 1)$, we assume that for all $0 < \gamma < 1/2$

$$\sup_{t\in\mathbb{R}}\left(a_n^\gamma \wedge f^{-\gamma}(t)\right)\mathrm{Bias}_n(t) = o\left((nh_n|\log h_n|)^{-1/2}\right) \text{ as } n \to \infty \tag{2.3}$$

and also that for some $p < 1$

$$\int_{\mathbb{R}} f^p(t)dt < +\infty. \tag{2.4}$$

Here,

$$\mathrm{Bias}_n(t) := |\mathbb{E}f_n(t) - f(t)|.$$

It is well known and easy to check that if K is a probability kernel supported by $[-1/2, 1/2]$ and

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

is the modulus of continuity of f, then $\mathrm{Bias}_n(t) \leq \omega_f(t; h_n)$, and if moreover $\int_{-1/2}^{1/2} uK(u)du = 0$, $v = \int_{-1/2}^{1/2} u^2K(u)du$ and f is twice continuously differentiable then $\mathrm{Bias}_n(t) \le 2^{-1}v(f''(t) + \omega_{f''}(t; h_n))h_n^2$. This is the extent to which we will deal with the bias. Finally, we let $\beta \in (0, 1/2)$ and let $\hat{A}_n$ be defined as the solution of the equation

$$\int_{\hat{D}_{a_n}} \Psi_\alpha \left(\frac{\hat{A}_n \|f_n\|_\infty^{1/2-\beta}}{f_n^{1/2-\beta}(y)} \right) dy = h_n, \tag{2.5}$$

where Ψ_α is defined by (1.3) if the kernel K satisfies (K_α). The result we are going to prove is as follows:

Theorem 1. *Under assumptions (D), (K_α) for some $0 < \alpha \le 2$, (H), and (2.3) and (2.4) above, with $\beta \in (0, 1/2)$ and a_n, $\hat{D}_a$ and $\hat{A}_n$ as defined respectively in (2.2), (2.1) and (2.5), we have*

(2.6)

$$\lim_{n\to\infty} \Pr \left\{ \hat{A}_n \left(\frac{\sqrt{nh_n}}{\|K\|_2 \|f_n\|_\infty^{1/2-\beta}} \sup_{t \in \hat{D}_{a_n}} \frac{|f_n(t) - f(t)|}{f_n^\beta(t)} - \hat{A}_n \right) \le x \right\} = \exp\{-e^{-x}\}.$$

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

The proof will be based on Theorem 2 below and two lemmas.

Lemma 1. *Under the assumptions of Theorem 1, for any $\gamma \in (0, 1/2)$, the sequence*

$$\sqrt{\frac{nh_n}{|\log h_n|}} \sup_{t\in\mathbb{R}} \frac{|f_n(t) - f(t)|}{f^\gamma(t) \vee a_n^{-\gamma}}, \quad n \in \mathbb{N}, \tag{2.7}$$

is stochastically bounded.

Proof (Sketch). By the bias condition (2.3), we can eliminate the bias, that is, we can replace f by $\mathbb{E}f_n$ above, and then, the lemma follows along the steps in the proof of Theorem 2.1 in [3], with simplifications due to the particular form of the weights and to the fact that we are only proving tightness. The main difference with that theorem is that having a_n^γ in the denominator above allows us to obtain tightness without imposing any conditions on the tail of $\Psi(x) = f^{-\gamma}(x)$. The sup over $\mathbb{R}$ is decomposed into sups over sets A_n, B_n, C_n defined as follows:

$$A_n := \{t \in \mathbb{R} : f(t) < h_n^r\},$$

where r is chosen large enough, and, with $\varepsilon_n := 1/\log n$,

$$B_n := \left\{ t \in \mathbb{R} : h_n^r \le f(t) \le \varepsilon_n \left(\frac{|\log h_n|}{nh_n} \right)^{1/(2(1-\gamma))} \right\},$$

and

$$C_n := \left\{ t \in \mathbb{R} : f(t) > \varepsilon_n \left(\frac{|\log h_n|}{nh_n} \right)^{1/(2(1-\gamma))} \right\}.$$

Exactly as in [3], we can eliminate the centering for the sups over A_n and B_n. For the uncentered sup over A_n we proceed as follows. By (H) we can choose $\delta \in (0, 1)$ and $r > 0$ such that, with $p < 1$ as in (2.4), $nh_n^{r(1-\delta)(1-p)} \to 0$, and by (D), there exist $c<\infty$ and $n_0<\infty$ such that $t\in A_n$ and $|s-t|\le h_n/2$ imply that $f(s)\le cf^{1-\delta}(t)<ch_n^{r(1-\delta)}$ for all $n>n_0$. Therefore, setting $\lambda_n:=nh_n|\log h_n|^{1/2}$, we have that for these values of n, for p as in (2.4) and for some $C<\infty$,

$$\begin{aligned}\Pr\left\{\sup_{t\in A_n}\frac{f^{-\gamma}(t)\sum_{i=1}^n K((t-X_i)/h_n)}{\lambda_n}>\varepsilon\right\} &\le n\Pr\{f(X)\le ch_n^r\}\\ &= n\int_{\{t:f(t)\le ch_n^r\}} f(t)dt\\ &\le Cnh_n^{r(1-\delta)(1-p)}\int_{\mathbb{R}} f^p(t)dt\to 0.\end{aligned}$$

On the uncentered sups over B_n, following the proof of Theorem 2.1 in [3] (more concretely, the proof of (2.11) there), we can establish that for any $c>1$ and large enough n

$$\begin{aligned}\sup_{t\in B_n}\frac{(f^{-\gamma}(t)\wedge a_n^\gamma)\sum_{i=1}^n K((t-X_i)/h_n)}{\lambda_n} &\le \max_{1\le i\le n}\frac{c\kappa(f^{-\gamma}(X_i)\wedge a_n^\gamma)}{\lambda_n}\\ &+\max_{1\le j\le n}\frac{c\kappa(f^{-\gamma}(X_j)\wedge a_n^\gamma)I_j\sum_{1\le i\le n,i\ne j}I(|X_i-X_j|\le h_n)}{\lambda_n},\end{aligned}\tag{2.8}$$

where

$$I_j=I_{n,j}:=I\left(c^{-1}h_n^r\le f(t)\le c\varepsilon_n\left(\frac{|\log h_n|}{nh_n}\right)^{1/(2(1-\gamma))}\right).$$

Since $a_n^\gamma=o(\lambda_n)$, the first maximum at the right hand side of (2.8) tends to 0. To bound the second maximum, let P_j be conditional probability given X_j. Then, by a standard bound for binomial probabilities,

$$\begin{aligned}&P_j\left\{\frac{(f^{-\gamma}(X_i)\wedge a_n^\gamma)I_j\sum_{1\le i\le n,i\ne j}I(|X_i-X_j|\le h_n)}{\lambda_n}\ge\varepsilon\right\}\\ &\quad\le\left(\frac{(n-1)ep_j(f^{-\gamma}(X_j)\wedge a_n^\gamma)}{\lambda_n\varepsilon}\right)^{(\varepsilon\lambda_n/(f^{-\gamma}(X_j)\wedge a_n^\gamma))\vee 1},\end{aligned}\tag{2.9}$$

where, by condition (D),

$$2c^{-1}h_nf(X_j)\le p_j:=P_j\{|X-X_j|\le h_n\}\le 2ch_nf(X_j)\tag{2.10}$$

(provided that $I_j=1$). Using (2.10), we can bound (2.9) further by

$$\left(\frac{2ecnh_nf^{1-\gamma}(X_j)}{\lambda_n\varepsilon}\right)^{(\varepsilon\lambda_n/a_n^\gamma)\vee 1},$$

and if $I_j=1$ (otherwise the conditional probability in question is 0) by

$$\left(\frac{2ec^{2-\gamma}nh_n\varepsilon_n^{1-\gamma}|\log h_n|^{1/2}}{(nh_n)^{1/2}\lambda_n\varepsilon}\right)^{(\varepsilon\lambda_n/a_n^\gamma)\vee 1}\le\left(\frac{C_1\varepsilon_n^{1-\gamma}}{\varepsilon}\right)^{(\varepsilon\lambda_n/a_n^\gamma)\vee 1}$$

with some $C_1 < \infty$ (and all n large enough). Since $\varepsilon_n \to 0$ and $\lambda_n / a_n^\gamma \to \infty$ faster than a power of n, we can conclude that

$$\max_{1\le j\le n} P_j\Big\{\frac{(f^{-\gamma}(X_i)\wedge a_n^\gamma) I_j \sum_{1\le i\le n, i\ne j} I(|X_i - X_j| \le h_n)}{\lambda_n} \ge \varepsilon\Big\}$$
$$\le \left(\frac{C_1 \varepsilon_n^{1-\gamma}}{\varepsilon}\right)^{\varepsilon\lambda_n/a_n^\gamma} = O(n^{-A})$$

for any $A > 0$. This implies that

$$\Pr\Big\{\max_{1\le j\le n} \frac{(f^{-\gamma}(X_j)\wedge a_n^\gamma) I_j \sum_{1\le i\le n, i\ne j} I(|X_i - X_j| \le h_n)}{\lambda_n} \ge \varepsilon\Big\}$$
$$\le \sum_{j=1}^n \mathbb{E} I_j P_j\Big\{\frac{f^{-\gamma}(X_j)\wedge a_n^\gamma I_j \sum_{1\le i\le n, i\ne j} I(|X_i - X_j| \le h_n)}{\lambda_n} \ge \varepsilon\Big\}$$
$$\le nO(n^{-A}) \to 0$$

for $A > 1$.

The centered suprema over C_n can be handled as in the proof of Theorem 2.1 in [3], which, for this part, does not require the tail assumption of that theorem, and they are stochastically bounded even if we replace the weights $(f^\gamma \vee a_n^{-\gamma})^{-1}$ by $f^{-\gamma}$. □

The next result is implicit in [2] and constitutes a nice complement to Theorem 6 there (or Theorem 0 above) when no assumptions are made on the tail probabilities of X.

Theorem 2. *Assume the hypotheses of Theorem 1, and let $\Psi(t)$ be a piecewise monotone function that satisfies the hypotheses $(W.a) - (W.c)$ from [3] for $B_f = \mathbb{R}$, and, moreover, $\|\Psi f^\beta\|_\infty < \infty$ for some $0 < \beta < 1/2$, normalized so that $\|\Psi f^{1/2}\|_\infty = 1$. Let A_n be constants satisfying*

$$\int_{-\infty}^{\infty} \frac{1}{h_n} \Psi_\alpha \left(\frac{A_n}{\Psi(y) f^{1/2}(y)}\right) dy \to 1.$$

Then,

$$\lim_{n\to\infty} \Pr\left\{A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in D_{a_n}} |\Psi(t)(f_n(t) - f(t))| - A_n\right) \le x\right\} = e^{-e^{-x}} \tag{2.11}$$

for all $x \in \mathbb{R}$. If the bias condition (2.3) is not assumed, then $f(t)$ must be replaced by $\mathbb{E} f_n(t)$ in (2.11).

Proof. Recall the definition of the sets A_n, B_n and C_n in Lemma 1 for $\gamma = \beta$. (Although A_n is used for constants and for sets, it is clear what is meant by the context and there is no possibility of confusion.) Let

$$\tilde{D}_a := D_a \cap \{t : |t| \le a\},\ C_{n,a} := C_n \cap \tilde{D}_a^c.$$

Since $\Psi \le cf^{-\beta}$ and, on D_{a_n}, $f^{\beta}(t) \vee a_n^{-\beta} = f^{\beta}(t)$, it follows from the proof of Lemma 1 that

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{|\log h_n|}} \sup_{t\in A_n\cap D_{a_n}} |\Psi(t)(f_n(t) - \mathbb{E}f_n(t))| = 0 \text{ in pr.},$$

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{|\log h_n|}} \sup_{t\in B_n\cap D_{a_n}} |\Psi(t)(f_n(t) - \mathbb{E}f_n(t))| = 0 \text{ in pr.}$$

and it also follows from the proof of Theorem 2.1 in [3] that

$$\lim_{a\to\infty} \limsup_{n\to\infty} \mathbb{E}\sqrt{\frac{nh_n}{|\log h_n|}} \sup_{t\in C_{n,a}\cap D_{a_n}} |\Psi(t)(f_n(t) - \mathbb{E}f_n(t))| = 0$$

(see (2.20) there). By Lemma 5 in [2], for all $x \in \mathbb{R}$,

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

and therefore, replacing $\sqrt{\log h_n^{-1}}$ by $A_n + x/A_n$ in the previous limits, we get

$$\lim_{n\to\infty} \Pr\left\{A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in A_n\cap D_{a_n}} |\Psi(t)(f_n(t) - \mathbb{E}f_n(t))| - A_n\right) \ge x\right\} = 0, \tag{2.12}$$

$$\lim_{n\to\infty} \Pr\left\{A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in B_n\cap D_{a_n}} |\Psi(t)(f_n(t) - \mathbb{E}f_n(t))| - A_n\right) \ge x\right\} = 0 \tag{2.13}$$

and

$$\lim_{a\to\infty} \limsup_{n\to\infty} \Pr\left\{A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in C_{n,a}\cap D_{a_n}} |\Psi(t)(f_n(t) - \mathbb{E}f_n(t))| - A_n\right) \ge x\right\} = 0 \tag{2.14}$$

for all $x \in \mathbb{R}$. As in (4.12) in the proof of Theorem 6 in [3], we get for all a large enough and all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \Pr\left\{A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in \tilde{D}_a} |\Psi(t)(f_n(t) - \mathbb{E}f_n(t))| - A_n\right) \le x\right\} = e^{-e^{-x}}.$$

This together with (2.12)–(2.14) give

$$\lim_{n\to\infty} \Pr\left\{A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2} \sup_{t\in D_{a_n}} |\Psi(t)(f_n(t) - \mathbb{E}f_n(t))| - A_n\right) \le x\right\} = e^{-e^{-x}}.$$

The limit (2.11) is now a consequence of this limit and the bias condition (2.3). □

For simplicity, we have stated Theorem 2 for densities that do not vanish anywhere, but it is true as well (and with the same proof) for densities which may vanish outside an open set, as in [2].

We will use Theorem 2 for the weight function $\Psi(t) = \|f^{1/2-\beta}\|_\infty f^{-\beta}(t)$.

Lemma 2. *Under the assumptions of Theorem 1, with $0 < \delta < (1-\theta)/2$, and $\varepsilon_n := (nh_n)^{-\delta}$, we have*

$$\sup_{t \in D_{a_n} \cup \hat{D}_{a_n}} \left| \frac{f_n(t)}{f(t)} - 1 \right| = o_P(\varepsilon_n), \tag{2.15}$$

which also implies

$$\sup_{t \in D_{a_n} \cup \hat{D}_{a_n}} \left| \frac{f(t)}{f_n(t)} - 1 \right| = o_P(\varepsilon_n) \tag{2.16}$$

and

$$\sup_{t \in D_{a_n} \cup \hat{D}_{a_n}} \left| \left(\frac{f}{f_n}\right)^{\eta}(t) - 1 \right| = o_P(\varepsilon_n) \quad \text{and} \quad \sup_{t \in D_{a_n} \cup \hat{D}_{a_n}} \left| \left(\frac{f_n}{f}\right)^{\eta}(t) - 1 \right| = o_P(\varepsilon_n) \tag{2.17}$$

for all $\eta \in (0,1)$.

Proof. We take $\gamma > 0$ such that

$$\frac{1}{2} < 1 - \gamma < \frac{1}{2\theta} - \frac{\delta}{\theta}.$$

By Lemma 1,

$$\begin{aligned}
\sup_{t \in D_{a_n}} \left| \frac{f_n(t)}{f(t)} - 1 \right| &\le \sup_{t \in D_{a_n}} \frac{|f_n(t) - f(t)|}{f^{\gamma}(t)} \sup_{t \in D_{a_n}} f^{-(1-\gamma)}(t) \\
&\le \sup_{t \in \mathbb{R}} \frac{|f_n(t) - f(t)|}{f^{\gamma}(t) \vee a_n^{-\gamma}} \times O\left((nh_n)^{\theta(1-\gamma)}\right) \\
&= O_P\left(\sqrt{\frac{|\log h_n|}{nh_n}}\right) \times O\left((nh_n)^{\theta(1-\gamma)}\right) \\
&= O_P\left((nh_n)^{\theta(1-\gamma)-1/2} |\log h_n|^{1/2}\right) \\
&= o_P\left((nh_n)^{-\delta}\right) = o_p(\varepsilon_n). \qquad (2.18)
\end{aligned}$$

Lemma 1 and the same calculation that leads to (2.18) also give

$$\begin{aligned}
\sup_{t \in \hat{D}_{a_n}, f_n(t) \ge f(t)} \left| 1 - \frac{f(t)}{f_n(t)} \right| &= \sup_{t : f_n(t) \ge f(t) \vee a_n^{-1}} \frac{|f_n(t) - f(t)|}{f_n(t)} \\
&\le \sup_{t \in \mathbb{R}} \frac{|f_n(t) - f(t)|}{f^{\gamma}(t) \vee a_n^{-\gamma}} \sup_{t \in \hat{D}_{a_n}} f_n^{-(1-\gamma)}(t) \\
&\le \sup_{t \in \mathbb{R}} \frac{|f_n(t) - f(t)|}{f^{\gamma}(t) \vee a_n^{-\gamma}} \times O((nh_n)^{\theta(1-\gamma)}) \\
&\le o_P((nh_n)^{-\delta}),
\end{aligned}$$

in particular, with probability tending to 1,

$$\sup_{t\in\hat D_{a_n}, f_n(t)\ge f(t)} \frac{f_n(t)}{f(t)} \le (1-\varepsilon_n)^{-1}.$$

This and (2.18) show that

$$\sup_{t\in\hat D_{a_n}} \left|\frac{f_n(t)}{f(t)}-1\right| \le \max\Bigg(\sup_{t\in\hat D_{a_n}, f_n(t)\ge f(t)} \frac{|f_n(t)-f(t)|}{f_n(t)}\frac{f_n(t)}{f(t)}, \\ \sup_{t\in D_{a_n}} \frac{|f_n(t)-f(t)|}{f(t)}\Bigg)$$

$$= o_P\left(\frac{\varepsilon_n}{1-\varepsilon_n}\right) = o_P(\varepsilon_n), \tag{2.19}$$

and (2.15) follows from (2.18) and (2.19). Obviously, (2.16) is equivalent to (2.15). Finally, the inequality

$$|(1+x)^\eta - 1| \le \eta(1-|x|)^{\eta-1}|x|,\ |x|<1,$$

which holds for $\eta\in(0,1)$, and (2.15)-(2.16) imply

$$\sup_{t\in D_{a_n}\cup\hat D_{a_n}} \left|\left(\frac{f}{f_n}\right)^\eta(t)-1\right| = \sup_{t\in D_{a_n}\cup\hat D_{a_n}} \left|\left(1+\frac{f-f_n}{f_n}\right)^\eta(t)-1\right| = o_P(\varepsilon_n)$$

and similarly

$$\sup_{t\in D_{a_n}\cup\hat D_{a_n}} \left|\left(\frac{f_n}{f}\right)^\eta(t)-1\right| = o_P(\varepsilon_n).$$

□

Proof of Theorem 1. Let ε_n be as in Lemma 2. Note that if we define

$$E_n := \left\{\forall\ t\in D_{a_n}\ 1-\varepsilon_n < \frac{f_n(t)}{f(t)} < 1+\varepsilon_n \text{ and } 1-\varepsilon_n < \frac{f(t)}{f_n(t)} < 1+\varepsilon_n\right\}, \tag{2.20}$$

then Lemma 2 implies that

$$\Pr(E_n)\to 1 \ \text{ as } \ n\to\infty. \tag{2.21}$$

Moreover, on the event E_n, the following inclusions hold:

$$D_{a_n(1-\varepsilon_n)} \subset \hat D_{a_n} \subset D_{a_n(1+\varepsilon_n)}. \tag{2.22}$$

To prove convergence in distribution of

$$\hat A_n\left(\frac{\sqrt{nh_n}}{\|K\|_2\|f_n\|_\infty^{1/2-\beta}} \sup_{t\in\hat D_{a_n}} \frac{|f_n(t)-f(t)|}{f_n^\beta(t)} - \hat A_n\right)$$

to the double exponential law, where $\hat{A}_n$ is defined in (2.5), we begin with the decomposition

(2.23)
$$\frac{f_n(t)-f(t)}{\|f_n\|_\infty^{1/2-\beta} f_n^\beta(t)} = \frac{f_n(t)-f(t)}{\|f\|_\infty^{1/2-\beta} f^\beta(t)} + \frac{f_n(t)-f(t)}{\|f\|_\infty^{1/2-\beta} f^\beta(t)} \left[\frac{\|f\|_\infty^{1/2-\beta}}{\|f_n\|_\infty^{1/2-\beta}} \frac{f^\beta(t)}{f_n^\beta(t)} - 1\right].$$

Now,

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

for instance, by Lemma 1. Hence, proceeding as in the proof (2.17) in Lemma 2, we also have

$$\left|\frac{\|f\|_\infty^{1/2-\beta}}{\|f_n\|_\infty^{1/2-\beta}} - 1\right| = o_P(\varepsilon_n),$$

which, together with (2.17) for $\eta = \beta$ (note $|ab-1| \le |a-1||b| + |b-1|$) gives

$$\sup_{t \in D_{a_n} \cup \hat{D}_{a_n}} \left|\frac{\|f\|_\infty^{1/2-\beta}}{\|f_n\|_\infty^{1/2-\beta}} \frac{f^\beta(t)}{f_n^\beta(t)} - 1\right| = o_P(\varepsilon_n). \tag{2.24}$$

By the inclusions (2.22), we have that on E_n,

$$\sup_{t \in \hat{D}_{a_n}} \frac{|f_n(t)-f(t)|}{f^\beta(t)} \le \sup_{t \in D_{a_n(1+\varepsilon_n)}} \frac{|f_n(t)-f(t)|}{f^\beta(t)} \le (1+\varepsilon_n)^\beta \sup_{t \in \mathbb{R}} \frac{|f_n(t)-f(t)|}{f^\beta(t) \vee a_n^{-\beta}},$$

hence, by Lemma 1 and (2.21),

$$\sqrt{\frac{nh_n}{|\log h_n|}} \sup_{t \in \hat{D}_{a_n}} \frac{|f_n(t)-f(t)|}{f^\beta(t)} = O_P(1).$$

This and (2.24) then give

$$\sqrt{\frac{nh_n}{|\log h_n|}} \sup_{t \in \hat{D}_{a_n}} \frac{|f_n(t)-f(t)|}{f^\beta(t)} \left[\frac{\|f\|_\infty^{1/2-\beta}}{\|f_n\|_\infty^{1/2-\beta}} \frac{f^\beta(t)}{f_n^\beta(t)} - 1\right] = o_P(\varepsilon_n). \tag{2.25}$$

We will see below that $\hat{A}_n = O_P(\sqrt{\log n})$, so that, since ε_n tends to zero as a negative power of n, (2.25) shows that we can ignore the second term in the decomposition (2.23) and that it suffices to show that the random variables

$$\hat{A}_n\left(\frac{\sqrt{nh_n}}{\|K\|_2 \|f\|_\infty^{1/2-\beta}} \sup_{t \in \hat{D}_{a_n}} \frac{|f_n(t)-f(t)|}{f^\beta(t)} - \hat{A}_n\right)$$

converge in distribution to the double exponential law. That is, we must show that we can replace A_n by $\hat{A}_n$ in Theorem 2. To this end we argue as in the proof of Theorem 10 in [2]. First, by Theorem 2 above and Lemma 3 in [2], for any sequence A_n satisfying

$$h_n^{-1} \int_{D_{a_n}} \Psi_\alpha\Big(\frac{A_n \|f\|_\infty^{1/2-\beta}}{f^{1/2-\beta}(y)}\Big) dy \to 1 \text{ as } n \to \infty, \tag{2.26}$$

we have that
(2.27)
$$\lim_{n\to\infty} \Pr\left\{ A_n \left(\frac{\sqrt{nh_n}}{\|K\|_2 \|f\|_\infty^{1/2-\beta}} \sup_{t\in D_{a_n}} |f^{-\beta}(t)(f_n(t)-f(t))| - A_n \right) \le x \right\} = e^{-e^{-x}}.$$

for all $x \in \mathbb{R}$, and D_{a_n} in (2.26) can be replaced by $D_{a_n(1-\varepsilon_n)}$ or $D_{a_n(1+\varepsilon_n)}$. If A_n and $\bar{A}_n$ are two sequences satisfying (2.26), then they also satisfy the limit (2.27), and it follows that

$$\left| \frac{\bar{A}_n}{A_n} - 1 \right| = o(A_n^{-2}) \quad \text{as } n \to \infty. \tag{2.28}$$

On the other hand, if (2.28) holds, then A_n in (2.27) can be replaced by $\bar{A}_n$ (even if $\bar{A}_n$ is a sequence of random variables and (2.28) holds in probability, which is the case below).

Next, by Lemma 2 and arguing as in the proof of (2.24), we have

$$\sup_{t\in D_{a_n}\cup \hat{D}_{a_n}} \left| \left(\frac{\|f\|_\infty}{\|f_n\|_\infty} \frac{f_n(t)}{f(t)} \right)^{1/2-\beta} - 1 \right| = o_P(\varepsilon_n). \tag{2.29}$$

Since also the inclusion (2.22) holds on the event E_n, we have on an event $\tilde{E}_n$ whose probability tends to 1

$$\begin{aligned} \int_{D_{a_n(1-\varepsilon_n)}} \Psi_\alpha \left(\frac{\hat{A}_n \|f\|_\infty^{1/2-\beta}}{(1-\varepsilon_n) f^{1/2-\beta}(y)} \right) dy &\le \int_{\hat{D}_{a_n}} \Psi_\alpha \left(\frac{\hat{A}_n \|f_n\|_\infty^{1/2-\beta}}{f_n^{1/2-\beta}(y)} \right) dy = h_n \\ &\le \int_{D_{a_n(1+\varepsilon_n)}} \Psi_\alpha \left(\frac{\hat{A}_n \|f\|_\infty^{1/2-\beta}}{(1+\varepsilon_n) f^{1/2-\beta}(y)} \right) dy. \end{aligned} \tag{2.30}$$

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

$$\int_{D_{a_n(1-\varepsilon_n)}} \Psi_\alpha \left(\frac{A_n^- \|f\|_\infty^{1/2-\beta}}{f^{1/2-\beta}(y)} \right) dy = h_n$$

and

$$\int_{D_{a_n(1+\varepsilon_n)}} \Psi_\alpha \left(\frac{A_n^+ \|f\|_\infty^{1/2-\beta}}{f^{1/2-\beta}(y)} \right) dy = h_n$$

we derive from (2.30) (using the monotonicity of the functions involved) that on the event $\tilde{E}_n$

$$A_n^-(1-\varepsilon_n) \le \hat{A}_n \le A_n^+(1+\varepsilon_n). \tag{2.31}$$

Since we can replace A_n in (2.27) by A_n^+ or by A_n^-, we have

$$\left| \frac{A_n^+}{A_n} - 1 \right| = o(A_n^{-2}) \quad \text{and} \quad \left| \frac{A_n^-}{A_n} - 1 \right| = o(A_n^{-2}) \quad \text{as } n \to \infty.$$

By Lemma 5 in [2], A_n is of the order $|\log h_n|^{1/2}$ and also $\varepsilon_n = o(\frac{1}{|\log h_n|})$. It follows that

$$\left|\frac{\hat{A}_n}{A_n} - 1\right| = o_P(A_n^{-2}) \quad \text{as} \quad n \to \infty.$$

This implies that A_n can be replaced by $\hat{A}_n$ in (2.27), which, together with (2.25) and the fact that $\hat{A}_n = O_P\left(\sqrt{|\log h_n|}\right)$, completes the proof of the theorem. □

For a result analogous to Theorem 1 in the case $\beta = 0$, see Corollary 11 in [2]. In that corollary, the assumption on the modulus of continuity of f has the only effect of handling the bias, and can be replaced by a weaker hypothesis.

In the Bickel-Rosenblatt theorem ([1]), corresponding to the case $\beta = 1/2$, the constants A_n do not depend on f, and their theorem with $\mathbb{E}f_n$ replaced by f already provides asymptotic confidence bands for f, which moreover simplify by further replacing the weight $\sqrt{f}$ by $\sqrt{f_n}$, as these authors did; Lemma 2 above provides an explicit justification for this last replacement. Our Theorem 1 can be used to extend to the case $\beta \neq 1/2$ and to the whole line the corollary on confidence bands in the Introduction of [1], which corresponds to $\beta = 1/2$ and bounded intervals.

References

[1] P. J. Bickel and M. Rosenblatt, *On some global measures of the deviations of density function estimates.* Ann. Statist. **1** (1973), 86–118.

[2] E. Giné, V. I. Koltchinskii and L. Sakhanenko, *Kernel density estimators: convergence in distribution for weighted sup norms.* 2002, to appear. Can be downloaded from http://www.math.uconn.edu/˜gine/Publications/index.html.

[3] E. Giné, V. I. Koltchinskii and J. Zinn, *Weighted uniform consistency of kernel density estimators.* 2001, to appear. Can be downloaded from http://www.math.uconn.edu/˜gine/Publications/index.html.

Department of Mathematics, University of Connecticut, Storrs, CT 06269-3009, USA
E-mail address: `gine@uconnvm.uconn.edu`

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131-1141, USA
E-mail address: `vlad@math.unm.edu`

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