Weak convergence of the supremum distance for supersmooth kernel deconvolution

Bert van Es
Korteweg-de Vries Institute for Mathematics
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
The Netherlands
vanes@science.uva.nl

Shota Gugushvili
Eurandom
Technische Universiteit Eindhoven
P.O. Box 513
5600 MB Eindhoven
The Netherlands
gugushvili@eurandom.tue.nl

March 5, 2008

Abstract

Let $f_{nh}(x)$ denote the deconvolution kernel density estimator. In this paper we establish the asymptotic distribution of the supremum distance $\sup_{x\in[0,1]} |f_{nh}(x) - \mathrm{E}\,[f_{nh}(x)]|$, which provides a global measure of performance of the deconvolution kernel density estimator. We consider the supersmooth deconvolution problem, in particular deconvolution for error distributions with characteristic functions that have an exponential tail like the characteristic function of a normal density. It turns out that the asymptotics are essentially different from corresponding results in ordinary smooth deconvolution. We also briefly discuss the method of construction of the uniform confidence intervals for the target density f.

Keywords: Deconvolution, kernel density estimator, Rayleigh distribution, supremum distance.
AMS subject classification: 62G07

1 Introduction and results

Consider the classical deconvolution problem: let $X_1, \dots, X_n$ be i.i.d. observations, where $X_i = Y_i + Z_i$ and Y_i and Z_i are independent. Assume that the unobservable Y_i have distribution function F and density f, and that the random variables Z_i have a known density k. Note that the density g of X_i is equal to the convolution of f and k. The nonparametric deconvolution problem is the problem of estimating f or F from the observations X_i. Thus we want to recover the distribution of Y_i using the contaminated measurements X_i. Additional information on measurement error models and many practical examples can be found in Carroll et al. (2006).

A popular density estimator for this problem is the deconvolution kernel density estimator introduced in Carroll & Hall (1988) and Stefanski & Carroll (1990). This estimator is defined as

$$f_{nh}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\phi_w(ht)\phi_{emp}(t)}{\phi_k(t)}\,dt = \frac{1}{nh}\sum_{j=1}^{n} v_h\Big(\frac{x - X_j}{h}\Big), \tag{1}$$

with

$$v_h(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\phi_w(s)}{\phi_k(s/h)}\, e^{-isu} ds.$$

Here w denotes a *kernel function*, $h > 0$ is a *bandwidth*, ϕ_{emp} is the empirical characteristic function of the sample defined by $\phi_{emp}(t) = (1/n)\sum_{j=1}^{n} e^{itX_j}$, and ϕ_w and ϕ_k denote the characteristic functions of w and k, respectively. Note that (1) is not a standard kernel density estimator, because the kernel function v_h depends on the bandwidth h. For an introduction to the estimator (1) see e.g. Wand & Jones (1995).

The rate of decay to zero at minus and plus infinity of the modulus of the characteristic function ϕ_k, and consequently the smoothness of k, is crucial to the asymptotic behaviour of (1). Two cases have been distinguished, the *ordinary smooth case*, where $|\phi_k|$ decays algebraically to zero, and the *supersmooth case*, where it decreases exponentially. The asymptotics in the ordinary smooth case are essentially the same as for a kernel estimator of a higher order derivative of a density, see e.g. Fan (1991), Fan & Liu (1997) and van Es & Kok (1998). The asymptotics in the supersmooth case have been studied e.g. in Fan (1991) and van Es & Uh (2004, 2005).

Notice that the above papers study local properties of the estimator (1), i.e. its pointwise behaviour. We, on the other hand, will focus on the asymptotic behaviour of the supremum distance of the estimator to its expectation, which provides a global measure of its performance. Accordingly, define

$$M_n = \sup_{0\le x\le 1} |f_{nh}(x) - \mathrm{E}\,[f_{nh}(x)]|. \tag{2}$$

The fact that the supermum is taken over $[0, 1]$ is not a restriction of generality and is for convenience only. One could have considered any interval

$[a, b]$. An alternative here is to consider the integrated squared error of the estimator f_{nh}, which is defined by

$$\mathrm{ISE}[f_{nh}] = \int_{-\infty}^{\infty} (f_{nh}(x) - \mathrm{E}\,[f_{nh}(x)])^2 dx,$$

since it also provides a global measure of performance of f_{nh}, and study its asymptotic distribution. This was done in Holzmann & Boysen (2006).

The asymptotic distribution of the supremum distance similar to (2), namely

$$\sup_{x\in[0,1]} \frac{1}{\sqrt{g(x)}} |g_{nh}(x) - \mathrm{E}\,[g_{nh}(x)]|,$$

for an ordinary kernel density estimator g_{nh} in the direct density estimation setting (i.e. in the error-free case) was derived in Bickel & Rosenblatt (1973). Owing in a certain sense to the similarity of the asymptotics in the ordinary smooth deconvolution problem to that in the direct density estimation problem, qualitatively similar results were obtained in Bissantz et al. (2007) in the ordinary smooth deconvolution problem for the supremum distance $\sup_{x\in[0,1]}(g(x))^{-1/2}|f_{nh}(x) - \mathrm{E}\,[f_{nh}(x)]|$. Normalisation with $\sqrt{g(x)}$ is explainable by the fact that the expression for the asymptotic variance in the asymptotic normality theorem for the estimator $f_{nh}(x)$ in the ordinary smooth deconvolution problem involves $g(x)$, see Fan (1991). No direct extension of the methods used in Bickel & Rosenblatt (1973) to the supersmooth deconvolution problem is possible and derivation of the asymptotic distribution of (2) requires a different approach. This is precisely the task of the present paper. Notice that in (2) we do not have to normalise with $\sqrt{g(x)}$, because the asymptotic variance in the asymptotic normality theorem for this case does not depend on g, but only on the error density k (in some global way), see van Es & Uh (2005).

We now state the conditions on the density k and kernel w, which will be used throughout the paper. The condition on k which defines supersmooth deconvolution is given in Condition 1.

Condition 1. *Assume that*

$$\phi_k(t) = C|t|^{\lambda_0} \exp\left[-|t|^{\lambda}/\mu\right] (1 + o(|t|^{-1})) \tag{3}$$

as $|t| \to \infty$, *for a constant* $0 < \lambda \leq 2$ *and some constants* $\mu > 0, \lambda_0 \in \mathbb{R}$ *and* $C \in \mathbb{R}$. *Furthermore, let* $\phi_k(t) \neq 0$ *for all* $t \in \mathbb{R}$.

Condition 1 is stronger than the usual condition on k in supersmooth deconvolution given e.g. in van Es & Uh (2005), where the term $o(|t|^{-1})$ is not present and one just has the asymptotic equivalence.

Condition 2. *Let* ϕ_w *be real-valued, symmetric and have support* $[-1, 1]$. *Let* $\phi_w(0) = 1$, *and assume* $\phi_w(1 - t) = At^{\alpha} + o(t^{\alpha})$ *as* $t \downarrow 0$ *for some constants* A *and* $\alpha \geq 0$.

Examples of kernel functions and their characteristic functions satisfying Condition 2 are the sinc kernel

$$w(x) = \sin x/(\pi x),$$
$$\phi_w(t) = I_{[-1,1]}(t),$$

where $\alpha = 0$ and $A = 1$, and the kernel used for simulations in Fan (1992),

$$w(x) = \frac{48x(x^2-15)\cos x - 144(2x^2-5)\sin x}{\pi x^7},$$
$$\phi_w(t) = (1-t^2)^3 I_{[-1,1]}(t),$$

where $\alpha = 3$ and $A = 8$.

Our main theorem establishes the asymptotic distribution of M_n. Since it will appear repeatedly in the paper, we will write $\zeta(h)$ for $\exp(1/(\mu h^\lambda))$.

Theorem 1. *Assume Condition 1 for $\lambda = 2$ and Condition 2 and let $\mathrm{E}\,[X_j^2] < \infty$. Let V denote a positive random variable with a Rayleigh distribution with density $f_V(x) = x\exp[-x^2/2]I_{[x\geq 0]}$. Then, as $n \to \infty$ and $h \to 0$,*

$$\frac{\sqrt{n}}{h^{\lambda(1+\alpha)+\lambda_0-1}\zeta(h)}\, M_n \xrightarrow{\mathcal{D}} \frac{1}{2}\sqrt{2}\,\frac{A}{\pi C}\left(\frac{\mu}{\lambda}\right)^{1+\alpha}\Gamma(\alpha+1)\,V, \tag{4}$$

where Γ denotes the gamma function.

By assuming $\lambda = 2$ we restrict ourselves to deconvolution problems for error distributions with characteristic functions that have an exponential tail like the characteristic function of a normal density. The most important case covered by this condition is standard normal deconvolution, where $\lambda = 2, \lambda_0 = 0, \mu = 2$ and $C = 1$. The condition $\lambda = 2$ seems to be essential in the proof of Lemma 3, specifically in (15), where we prove a condition for tightness of the remainder process $R_n^{(1)}$. Whether it can be relaxed by other approaches, avoiding tightness, remains open.

The rate of convergence in Theorem 1 once again reflects the difficulty of the supersmooth deconvolution problem compared to the ordinary smooth deconvolution. Furthermore, unlike in ordinary smooth deconvolution, see Bissantz et al. (2007), in order to obtain the asymptotic distribution of M_n, we do not have to subtract a drift term. This also has a parallel when considering the asymptotics of the ISE$[f_{nh}]$ in the supersmooth deconvolution, see Holzmann & Boysen (2006) for additional details. Notice also that unlike the direct density estimation or the ordinary smooth deconvolution, see Bickel & Rosenblatt (1973) and Bissantz et al. (2007), the limit distribution in (4) is not Gumbel, which confirms the conjecture in Bissantz et al. (2007) for the case $\lambda = 2$.

One application of Theorem 1 is in construction of uniform confidence intervals for f. Noting that the value of

$$\mathrm{P}\left(\frac{\sqrt{n}}{h^{\lambda(1+\alpha)+\lambda_0-1}\zeta(h)}\, M_n \leq x\right) \tag{5}$$

is approximately given by the value of the distribution function of the random variable on the right-hand side of (4) at the point x, one can invert (5) in the usual way to obtain the uniform confidence band for $\mathrm{E}\left[f_{nh}(x)\right]$ on $[0,1]$. However, in reality we are interested in the confidence band for f. It is well-known that $\mathrm{E}\left[f_{nh}(x)\right] = f * w_h(x)$, where the function $w_h(y) = (1/h)w(y/h)$ and $*$ denotes the convolution operator, see e.g. Wand & Jones (1995). Hence $\mathrm{E}\left[f_{nh}(x)\right]$ is a smoothed version of $f(x)$. Using the identity

$$\mathrm{E}\left[f_{nh}(x)\right] = f(x) + (f * w_h(x) - f(x)),$$

it turns out that we have to deal with the bias of the estimator $f_{nh}(x)$, which is given by $f * w_h(x) - f(x)$. Note that this expression coincides with the bias of an ordinary kernel density estimator based on a sample from f. A possible way to reduce it is to undersmooth the estimator $f_{nh}(x)$, i.e. to take h relatively small, cf. Bissantz et al. (2007). We do not pursue the question of uniform confidence bands any further, since it requires a thorough simulation study, which lies outside the scope of the present paper.

When $\lambda_0 = 0$, by Lemma 5 of van Es & Uh (2005) in (4) one can equivalently normalize with $\sqrt{n}\int_0^1 \phi_w(s)\exp[s^\lambda/(\mu h^\lambda)]ds/h$. The latter normalisation should be preferred for smaller sample sizes (and consequently larger h) for reasons explained in van Es & Gugushvili (2008).

2 Proof of Theorem 1

The proof of Theorem 1 is based on a decomposition of $f_{nh}(x)$ in van Es & Uh (2005), which is the basis of the proof of their asymptotic normality theorem. We have

$$\begin{aligned} f_{nh}(x) = {} & \frac{1}{\pi C}\, h^{\lambda_0-1}\int_\epsilon^1 \phi_w(s)s^{-\lambda_0}\exp(s^\lambda/(\mu h^\lambda))ds\frac{1}{n}\sum_{j=1}^n \cos\left(\frac{X_j - x}{h}\right) \\ & + R_n^{(1)}(x) + R_n^{(2)}(x) + R_n^{(3)}(x), \end{aligned} \tag{6}$$

where $R_n^{(l)}(x) = (1/n)\sum_{j=1}^n R_{n,j}^{(1)}(x), l = 1,2,3$, and

$$\begin{aligned}
R_{n,j}^{(1)}(x) &= \frac{1}{C}\frac{1}{\pi}h^{\lambda_0-1}\int_\epsilon^1 \Big(\cos\Big(s\Big(\frac{X_j-x}{h}\Big)\Big) - \cos\Big(\frac{X_j-x}{h}\Big)\Big) \\
&\quad \times \phi_w(s)s^{-\lambda_0}\exp(s^\lambda/(\mu h^\lambda))ds \\
R_{n,j}^{(2)}(x) &= \frac{1}{2\pi h}\int_{-\epsilon}^{\epsilon}\exp\Big(is\Big(\frac{X_j-x}{h}\Big)\Big)\phi_w(s)\frac{1}{\phi_k(s/h)}ds \\
R_{n,j}^{(3)}(x) &= \frac{1}{2\pi h}\Big(\int_{-1}^{-\epsilon}+\int_\epsilon^1\Big)\exp\Big(is\Big(\frac{X_j-x}{h}\Big)\Big)\phi_w(s) \\
&\quad \times\Big(\frac{1}{\phi_k(s/h)}-\frac{1}{C}\Big(\frac{|s|}{h}\Big)^{-\lambda_0}\exp(|s|^\lambda/(\mu h^\lambda))\Big)ds.
\end{aligned}$$

We will write $R_n^{(l)}, l = 1,2,3$ for the stochastic processes $R_n^{(l)} = (R_n^{(l)}(x))_{x\in[0,1]}$. Notice that these processes belong to the space $C[0,1]$.

Now the rough idea is to derive the asymptotic distribution of the supremum of the first summand in (6) minus its expectation and to show that the remainder terms are negligible. Define the process U_n as

$$U_n(x) = \frac{1}{\sqrt{n}}\sum_{j=1}^n U_{n,j}(x), \tag{7}$$

where

$$U_{n,j}(x) = \cos\Big(\frac{X_j-x}{h}\Big) - \mathrm{E}\Big[\cos\Big(\frac{X_j-x}{h}\Big)\Big] \tag{8}$$

Note that this is a process with expectation equal to zero at every x. Write

$$S_n = \sup_{0\le x\le 1}|U_n(x)|. \tag{9}$$

Lemma 1. *Under the conditions of Theorem 1 we have, as $n\to\infty$ and $h\to 0$,*

$$S_n \xrightarrow{\mathcal{D}} \sup_{0\le x\le 2\pi}|W(x)|,$$

where W is a zero mean Gaussian process on $[0,2\pi]$ with covariance function $\mathrm{Cov}(W(x_1),W(x_2)) = (1/2)\cos(x_1-x_2)$.

Proof. Replacing x by yh, by the periodicity of the cosine function we have for $h \leq (2\pi)^{-1}$ that

$$\begin{aligned}
S_n &= \sup_{0\leq x\leq 1} |U_n(x)| \\
&= \sup_{0\leq x\leq 1} \Big|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big(\cos\Big(\frac{X_j - x}{h}\Big) - \mathrm{E}\Big[\cos\Big(\frac{X_j - x}{h}\Big)\Big]\Big)\Big| \\
&= \sup_{0\leq y\leq 1/h} \Big|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big(\cos\Big(\frac{X_j - yh}{h}\Big) - \mathrm{E}\Big[\cos\Big(\frac{X_j - yh}{h}\Big)\Big]\Big)\Big| \\
&= \sup_{0\leq y\leq 1/h} \Big|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\cos(Y_j - y) - \mathrm{E}\left[\cos(Y_j - y)\right])\Big| \\
&= \sup_{0\leq y\leq 2\pi} \Big|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\cos(Y_j - y) - \mathrm{E}\left[\cos(Y_j - y)\right])\Big|, \\
&= \sup_{0\leq y\leq 2\pi} |W_n(y)|,
\end{aligned}$$

where

$$Y_j = \frac{X_j}{h} \mod 2\pi, \tag{10}$$

and the process W_n on $[0, 2\pi]$ is given by

$$W_n(y) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (W_{n,j}(y) - \mathrm{E}\left[W_{n,j}(y)\right]) \tag{11}$$

with $W_{n,j}(y) = \cos(Y_j - y)$.

By Lemma 6 of van Es & Uh (2005) we know that $Y_j \xrightarrow{\mathcal{D}} \mathrm{Un}(0, 2\pi)$ as $h \to 0$ for each j, where $\mathrm{Un}(0, 2\pi)$ denotes the uniform distribution on $[0, 2\pi]$. Hence by the dominated convergence theorem we get that

$$\begin{aligned}
&\mathrm{Cov}\Big(\cos\Big(Y_j - y_1\Big), \cos\Big(Y_j - y_2\Big)\Big) \\
&\to \frac{1}{2\pi}\int_0^{2\pi} \cos(u - y_1)\cos(u - y_2)du = \frac{1}{2}\cos(y_1 - y_2).
\end{aligned}$$

It follows that we have to study the convergence of the process $W_n(x) - \mathrm{E}\left[W_n(x)\right]$ which belongs to $C[0, 2\pi]$. According to Prohorov's theorem and in particular Theorem 8.1 of Billingsley (1968), it suffices to show weak convergence of the finite dimensional distributions and tightness of the sequence. By the multivariate central limit theorem in the triangular array scheme or Cramer-Wold device, see Theorem 7.7 in Billingsley (1968), the finite dimensional distributions of the process W_n converge to multivariate normal distributions with covariances given by $\mathrm{Cov}(W(y_1), W(y_2)) = (1/2)\cos(y_1 - y_2)$. To prove tightness, we will verify conditions of Theorem 12.3 of Billingsley (1968). First of all, notice that the sequence $W_n(0)$ is tight, because the asymptotic normality of $W_n(0)$ follows by a univariate Lyapunov central limit theorem in a trinagular array scheme, see Theorem 7.3 in Billingsley (1968). Furthermore, for an arbitrary positive η,

$$\begin{aligned}
&P\Big(|W_n(y_2) - \mathrm{E}\,[W_n(y_2)] - (W_n(y_1) - \mathrm{E}\,[W_n(y_1)])| \geq \eta\Big) \\
&\leq \frac{1}{\eta^2}\,\mathrm{Var}[W_{n,j}(y_2) - W_{n,j}(y_1)] \\
&\leq \frac{1}{\eta^2}\,\mathrm{E}\,[(W_{n,j}(y_2) - W_{n,j}(y_1))^2] \\
&\leq \frac{1}{\eta^2}\,(y_2 - y_1)^2,
\end{aligned}$$

which follows from the fact that

$$\begin{aligned}
|\cos(Y_j - y_2) - \cos(Y_j - y_1)| &= \left|2\sin\left(\frac{2Y_j - y_2 - y_1}{2}\right)\sin\left(\frac{y_1 - y_2}{2}\right)\right| \\
&\leq |y_1 - y_2|.
\end{aligned}$$

Here we used the inequality $|\sin x| \leq |x|$. Therefore W_n converges weakly to a zero mean Gaussian process W on $[0, 2\pi]$ with covariance function $\mathrm{Cov}(W(y_1), W(y_2)) = (1/2)\cos(y_1 - y_2)$. By the continuous mapping theorem, see Theorem 5.1 in Billingsley (1968), the supremum of $|W_n|$ then converges weakly to the supremum of the absolute value of the limit process, which proves the lemma. □

Lemma 2. *With V as in Theorem 1, we have*

$$\sup_{0\leq x\leq 2\pi} |W(x)| \stackrel{\mathcal{D}}{=} \frac{1}{2}\sqrt{2}\,V. \tag{12}$$

Proof. Let N_1 and N_2 denote two independent standard normal random variables and let us define the process $\tilde{W}$ by $\tilde{W} = (\tilde{W}(x))_{x\in[0,2\pi]}$, where

$$\tilde{W}(x) \stackrel{\mathcal{D}}{=} \frac{1}{2}\sqrt{2}(N_1\cos x + N_2 \sin x). \tag{13}$$

Since the covariance function $\mathrm{Cov}(W(x_1), W(x_2))$ of the process W, given by $(1/2)\cos(x_1 - x_2)$, equals $\mathrm{Cov}(\tilde{W}(x_1), \tilde{W}(x_2))$ by

$$\begin{aligned}
&\mathrm{Cov}\left(\frac{1}{2}\sqrt{2}(N_1\cos x_1 + N_2\sin x_1), \frac{1}{2}\sqrt{2}(N_1\cos x_2 + N_2\sin x_2)\right) \\
&\qquad = \frac{1}{2}\,(\cos x_1\cos x_2 + \sin x_1 \sin x_2) = \frac{1}{2}\,\cos(x_1 - x_2),
\end{aligned}$$

it follows that $W \stackrel{\mathcal{D}}{=} \tilde{W}$.

Next write

$$\begin{aligned}
&\frac{1}{2}\sqrt{2}(N_1\cos x + N_2\sin x)\\
&= \frac{1}{2}\sqrt{2}\sqrt{N_1^2+N_2^2}\Big(\frac{N_1}{\sqrt{N_1^2+N_2^2}}\cos x + \frac{N_2}{\sqrt{N_1^2+N_2^2}}\sin x\Big)\\
&= \frac{1}{2}\sqrt{2}\sqrt{N_1^2+N_2^2}(\cos\xi\cos x + \sin\xi\sin x)\\
&= \frac{1}{2}\sqrt{2}\sqrt{N_1^2+N_2^2}\cos(x-\xi), \qquad (14)
\end{aligned}$$

for a ξ such that $\cos\xi = N_1/\sqrt{N_1^2+N_2^2}$ and $\sin\xi = N_2/\sqrt{N_1^2+N_2^2}$. The supremum of the absolute value of (14) is equal to $(1/2)\sqrt{2}\sqrt{N_1^2+N_2^2} = (1/2)\sqrt{2}V$, where V has a Rayleigh distribution. This entails (12). □

Lemma 3. *Let* $a_n = \sqrt{n}h^{-\lambda(1+\alpha)-\lambda_0+1}(\zeta(h))^{-1}$ *denote the normalising sequence in Theorem 1. For* $l = 1, 2, 3$ *we have*

$$a_n(R_n^{(l)} - \mathrm{E}\,[R_n^{(l)}]) \xrightarrow{\mathrm{P}} \mathbf{0}$$

as $n \to \infty$ *and* $h \to 0$. *Here* $\mathbf{0}$ *denotes the zero process on* $[0,1]$.

Proof. To prove the lemma, we will apply Prohorov's theorem, and in particular Theorem 8.1 of Billingsley (1968). Firstly, notice that for a fixed x the remainder terms $a_n(R_n^{(l)}(x) - \mathrm{E}\,[R_n^{(l)}(x)])$ vanish in probability, which was proved in van Es & Uh (2005). This implies that the finite dimensional vectors of the processes $a_n(R_n^{(l)} - \mathrm{E}\,[R_n^{(l)}])$ also converge in probability to null vectors. To establish tightness, we will again verify conditions of Theorem 12.3 of Billingsley (1968). Notice that when $x = 0$, the sequence $a_n(R_n^{(l)}(0) - \mathrm{E}\,[R_n^{(l)}(0)])$ is tight, since it converges to zero in probability.

Furthermore, for an arbitrary positive η we have

$$\begin{aligned}
P\Big(a_n|R_n^{(1)}(x_2) - \mathrm{E}\,[R_n^{(1)}(x_2)] - (R_n^{(1)}(x_1) - \mathrm{E}\,[R_n^{(1)}(x_1)])| \geq \eta\Big)& \\
&\leq \frac{a_n^2}{\eta^2}\,\mathrm{Var}[R_n^{(1)}(x_2) - R_n^{(1)}(x_1)] \\
&= \frac{a_n^2}{\eta^2}\,\frac{1}{n}\,\mathrm{Var}[R_{n,1}^{(1)}(x_2) - R_{n,1}^{(1)}(x_1)] \\
&\leq \frac{a_n^2}{\eta^2}\,\frac{1}{n}\,\mathrm{E}\,[(R_{n,1}^{(1)}(x_2) - R_{n,1}^{(1)}(x_1))^2] \\
&\leq \frac{a_n^2}{\eta^2}\,\frac{1}{C^2}\frac{1}{\pi^2}\frac{1}{n}\,h^{2(\lambda_0-1)}(x_2-x_1)^2 \\
&\qquad \times K^2\Big(\int_\epsilon^1 \Big(\frac{1-s}{h^2}\Big)\phi_w(s)s^{-\lambda_0}\exp(s^\lambda/(\mu h^\lambda))ds\Big)^2 \\
&= K^2\frac{a_n^2}{\eta^2}\,\frac{1}{C^2}\frac{1}{\pi^2}\frac{1}{n}\,h^{2(\lambda_0-2)-2}(x_2-x_1)^2 \\
&\qquad \times \Big(\int_\epsilon^1 \Big(1-s)\phi_w(s)s^{-\lambda_0}\exp(s^\lambda/(\mu h^\lambda))ds\Big)^2 \\
&= O\Big(\frac{1}{n}\,h^{2(\lambda_0-2)-2+2(2+\alpha)\lambda}(\zeta(h))^2 a_n^2\Big)\frac{1}{\eta^2}(x_2-x_1)^2 \\
&= O\Big(h^{2(\lambda-2)}\Big)\frac{1}{\eta^2}\,(x_2-x_1)^2 \qquad (15) \\
&= O(1)\frac{1}{\eta^2}\,(x_2-x_1)^2.
\end{aligned}$$

where K is some constant. Here we used Lemma 5 of van Es & Uh (2005), which states that

$$\begin{aligned}
&\int_\epsilon^1 s^{-\lambda_0}(1-s)^\beta\phi_w(s)\exp(s^\lambda/(\mu h^\lambda))ds \\
&\quad \sim \quad A\Big(\frac{\mu}{\lambda}h^\lambda\Big)^{1+\alpha+\beta}\zeta(h)\Gamma(\alpha+\beta+1), \qquad (16)
\end{aligned}$$

and the fact that for $0 \leq s \leq 1$ and $0 \leq x_1 < x_2 \leq 1$ we have

$$
\begin{aligned}
&\Big|\cos\Big(s\Big(\frac{X_j-x_2}{h}\Big)\Big)-\cos\Big(\frac{X_j-x_2}{h}\Big)-\cos\Big(s\Big(\frac{X_j-x_1}{h}\Big)\Big)+\cos\Big(\Big(\frac{X_j-x_1}{h}\Big)\Big)\Big|\\
&\quad=\Big|\int_{x_1}^{x_2}\int_s^1\frac{\partial^2}{\partial u\partial v}\Big\{\cos\Big(v\Big(\frac{X_j-u}{h}\Big)\Big)-\cos\Big(\frac{X_j-u}{h}\Big)\Big\}dudv\Big|\\
&\quad=\Big|\int_{x_1}^{x_2}\int_s^1\Big\{\frac{1}{h}\sin\Big(v\Big(\frac{X_j-u}{h}\Big)\Big)+v\Big(\frac{X_j-u}{h^2}\Big)\cos\Big(\frac{X_j-u}{h}\Big)\Big\}dudv\Big|\\
&\quad\le\int_{x_1}^{x_2}\int_s^1\frac{1}{h^2}\left(|X_j|+1+h\right)dudv\\
&\quad\le\frac{1}{h^2}\left(|X_j|+1+h\right)(1-s)|x_1-x_2|.
\end{aligned}
$$

Hence the process $a_n(R_n^{(1)}-\mathrm{E}\,[R_n^{(1)}])$ is tight.

In order to prove tightness of the process $a_n(R_n^{(2)}-\mathrm{E}\,[R_n^{(2)}])$, note that, as above, for positive η

$$
\begin{aligned}
&P\Big(a_n|R_n^{(2)}(x_2)-\mathrm{E}\,[R_n^{(2)}(x_2)]-(R_n^{(2)}(x_1)-\mathrm{E}\,[R_n^{(2)}(x_1)])|\ge\eta\Big)\\
&\qquad\le\frac{a_n^2}{\eta^2}\,\frac{1}{n}\,\mathrm{E}\,[(R_{n,1}^{(2)}(x_2)-R_{n,1}^{(2)}(x_1))^2]\\
&\qquad\le 4\frac{a_n^2}{\eta^2}\,\frac{1}{4\pi^2h^2}\,\frac{1}{n}\Big(\int_{-\epsilon}^{\epsilon}\frac{s}{h}\phi_w(s)\frac{1}{\phi_k(s/h)}\,ds\Big)^2(x_2-x_1)^2\\
&\qquad=4\frac{a_n^2}{\eta^2}\,\frac{1}{4\pi^2h^4}\,\frac{1}{n}\Big(\int_{-\epsilon}^{\epsilon}s\phi_w(s)\frac{1}{\phi_k(s/h)}\,ds\Big)^2(x_2-x_1)^2\\
&\qquad\le 4a_n^2\frac{1}{4\pi^2h^4}\,\frac{1}{n}\,(2\epsilon)^2\epsilon^2\Big(\int_{-\epsilon}^{\epsilon}\frac{1}{\phi_k(s/h)}\,ds\Big)^2\frac{1}{\eta^2}(x_2-x_1)^2\\
&\qquad\le 4a_n^2\frac{1}{\pi^2h^4}\,\frac{1}{n}\,\epsilon^4\Big(\sup_{-\epsilon\le s\le\epsilon}\frac{1}{\phi_k(s/h)}\,ds\Big)^2\frac{1}{\eta^2}(x_2-x_1)^2\\
&\qquad\le 4a_n^2\frac{2}{C^2\pi^2}\,\frac{1}{n}\,(\epsilon/h)^{4-2\lambda_0}\exp(2(\epsilon/h)^{\lambda}/\mu)\frac{1}{\eta^2}(x_2-x_1)^2\\
&\qquad=o(1)\frac{1}{\eta^2}(x_2-x_1)^2,
\end{aligned}
$$

where K is some constant and where we used the fact that for $0\le s\le 1$,

$$
\begin{aligned}
&\Big|\exp\Big(is\Big(\frac{X_j-x_2}{h}\Big)\Big)-\exp\Big(is\Big(\frac{X_j-x_1}{h}\Big)\Big)\Big|\\
&\qquad\qquad\le\Big|\cos\Big(s\Big(\frac{X_j-x_2}{h}\Big)\Big)-\cos\Big(s\Big(\frac{X_j-x_1}{h}\Big)\Big)\Big|\\
&\qquad+\Big|\sin\Big(s\Big(\frac{X_j-x_2}{h}\Big)\Big)-\sin\Big(s\Big(\frac{X_j-x_1}{h}\Big)\Big)\Big|\le\frac{2s}{h}|x_1-x_2|,
\end{aligned}
\tag{17}
$$

which follows by converting the differences of sines and cosines into products and using the fact that $|\sin x|\le|x|$. Consequently, the process $a_n(R_n^{(2)}-\mathrm{E}\,[R_n^{(2)}])$ is tight.

To prove tightness of the process $a_n(R_n^{(3)} - \mathrm{E}\,[R_n^{(3)}])$, we first introduce the function u, given by

$$u(y) = \frac{C|y|^{\lambda_0}\exp(-|y|^\lambda/\mu)}{\phi_k(y)} - 1. \tag{18}$$

By Condition 1 this function is bounded on $\mathbb{R}\backslash(-\delta,\delta)$, where δ is an arbitrary positive number. Moreover, by (3) the function $xu(x)$ is also bounded and both functions vanish at plus and minus infinity. It follows that $(s/h)u(s/h)$ is bounded and tends to zero for all fixed s with $|s| \geq \epsilon$ as $h \to 0$.

Using the function u, rewrite $R_{n,j}^{(3)}(x)$ as follows

$$\begin{aligned} R_{n,j}^{(3)}(x) &= \frac{1}{2\pi h}\Big(\int_{-1}^{-\epsilon} + \int_{\epsilon}^{1}\Big)\exp\Big(is\Big(\frac{X_j - x}{h}\Big)\Big)\phi_w(s) \\ &\quad\times\Big(\frac{1}{\phi_k(s/h)} - \frac{1}{C}\Big(\frac{|s|}{h}\Big)^{-\lambda_0}\exp(|s|^\lambda/(\mu h^\lambda))\Big)ds \\ &= \frac{1}{2\pi h}\Big(\int_{-1}^{-\epsilon} + \int_{\epsilon}^{1}\Big)\exp\Big(is\Big(\frac{X_j - x}{h}\Big)\Big)\phi_w(s) \\ &\quad\times\frac{1}{C}\Big(\frac{|s|}{h}\Big)^{-\lambda_0}\exp(|s|^\lambda/(\mu h^\lambda))u(s/h)ds. \end{aligned}$$

Next note that, as above, for positive η we have by (17) that

$$\begin{aligned} &P\Big(a_n|R_n^{(3)}(x_2) - \mathrm{E}\,[R_n^{(3)}(x_2)] - (R_n^{(3)}(x_1) - \mathrm{E}\,[R_n^{(3)}(x_1)])| \geq \eta\Big) \\ &\quad\leq \frac{a_n^2}{\eta^2}\frac{1}{n}\mathrm{E}\,[(R_{n,1}^{(3)}(x_2) - R_{n,1}^{(3)}(x_1))^2] \\ &\quad\leq \frac{a_n^2}{\eta^2}\frac{1}{4\pi^2h^2}\frac{1}{n}\Big(\Big(\int_{-1}^{-\epsilon} + \int_{\epsilon}^{1}\Big)\phi_w(s) \\ &\qquad\times\frac{1}{C}\Big(\frac{|s|}{h}\Big)^{-\lambda_0}\exp(|s|^\lambda/(\mu h^\lambda))\frac{s}{h}\,u\Big(\frac{s}{h}\Big)ds\Big)^2(x_2 - x_1)^2 \\ &\quad= o(1)\frac{1}{\eta^2}(x_2 - x_1)^2 \end{aligned}$$

and hence $a_n(R_n^{(3)} - \mathrm{E}\,[R_n^{(3)}])$ is tight. By Prohorov's theorem each of the three processes now converges weakly to the zero process. Since the convergence in distribution to a constant entails convergence to the same constant in probability, this concludes the proof of the lemma. □

Finally, we combine the obtained results to prove Theorem 1.

Proof of Theorem 1. The proof is immediate from Lemmas 1–3 just proved, the fact that by (16)

$$\begin{aligned} &a_n\frac{1}{\pi C}h^{\lambda_0 - 1}\int_{\epsilon}^{1}\phi_w(s)s^{-\lambda_0}\exp(s^\lambda/(\mu h^\lambda))ds\frac{1}{\sqrt{n}} \\ &\quad\sim\frac{A}{\pi C}\Big(\frac{\mu}{\lambda}\Big)^{1+\alpha}\Gamma(\alpha + 1), \end{aligned}$$

and Theorems 4.1 and 5.1 of Billingsley (1968). □

References

Bissantz, N., Dümbgen, L., Holzmann H. & Munk, A. (2007). Nonparametric confidence bands in deconvolution density estimation. *J. Roy. Statist. Soc. Ser. B* **69**, 483–506.

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

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Carroll, R. J. & Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Stat. Assoc.* **83**, 1184–1186.

Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective.* 2nd edn. Chapman & Hall/CRC Press, Boca Raton.

van Es, A. J. & Kok, A. R. (1998). Simple kernel estimators for certain nonparametric deconvolution problems. *Statist. Probab. Lett.* **39**, 151–160.

van Es, A. J. & and Uh, H.-W. (2004). Asymptotic normality of nonparametric kernel-type deconvolution density estimators: crossing the Cauchy boundary. *J. Nonparametr. Stat.* **16**, 261–277.

van Es, A. J. & and Uh, H.-W. (2005). Asymptotic normality of kernel type deconvolution estimators. *Scand. J. Statist.* **32**, 467–483.

van Es, B. & Gugushvili, S. (2008). Some thoughts on the asymptotics of the deconvolution kernel density estimator. Available online at arXiv:0801.2600[stat.ME].

Fan, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **53**, 97–110.

Fan, J. (1992). Deconvolution for supersmooth distributions. *Canad. J. Statist.* **20**, 155–169.

Fan, J. & Liu, Y. (1997). A note on asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **59**, 138–141.

Holzmann, H. & Boysen, L. (2006). Integrated square error asymptotics for supersmooth deconvolution. *Scand. J. Statist.* **33**, 849–860.

Stefanski, L. A. & Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **2**, 169–184.

Wand, M. P. & Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall, London.