NON-STANDARD BEHAVIOR OF DENSITY ESTIMATORS FOR SUMS OF SQUARED OBSERVATIONS

By Anton Schick
and Wolfgang Wefelmeyer

Binghamton University and Universität zu Köln

It has been shown recently that, under an appropriate integrability condition, densities of functions of independent and identically distributed random variables can be estimated at the parametric rate by a local U-statistic, and a functional central limit theorem holds. For the sum of two squared random variables, the integrability condition typically fails. We show that then the estimator behaves differently for different arguments.

At points in the support of the squared random variable, the rate of the estimator slows down by a logarithmic factor and is independent of the bandwidth, but the asymptotic variance depends on the rate of the bandwidth, and otherwise only on the density of the squared random variable at this point and at zero. A functional central limit theorem cannot hold.

Of course, for bounded random variables, the sum of squares is more spread out than a single square. At points outside the support of the squared random variable, the estimator behaves classically. Now the rate is again parametric, the asymptotic variance has a different form and does not depend on the bandwidth, and a functional central limit theorem holds.

1. Introduction. Suppose that $X_1, \dots, X_n$ are independent observations with density f. It is sometimes of interest to estimate the density p of a transformation $q(X_1, \dots, X_m)$ of m of these observations, with $m \geq 2$. Frees (1994) proposed as estimator of $p(z)$ the local U-statistic

$$\hat{p}_F(z) = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} k_b(z - q(X_{i_1}, \dots, X_{i_m}))$$

with $k_b(x) = k(x/b)/b$ for a kernel k and a bandwidth b. He showed that this estimator can be pointwise $\sqrt{n}$-consistent under some assumptions on f and q. Saavedra and Cao (2000) consider the function $q(X_1, X_2) = X_1 + aX_2$. They obtain pointwise $\sqrt{n}$-consistency for their convolution estimator

$$\hat{p}_{SC}(z) = \int \hat{f}(z - ax)\hat{f}(x)\, dx$$

AMS 2000 subject classifications: Primary 62G07; secondary 62G20.

Keywords and phrases: convolution density estimator, smoothness of convolutions, asymptotically linear estimator.

with a kernel estimator $\hat{f}$ of f. This is a plug-in estimator which replaces the unknown density f in the representation of $p(z)$ by $\hat{f}$. The estimator $\hat{p}_{SC}(z)$ is asymptotically equivalent to $\hat{p}_F(z)$ with $m = 2$ and $q(X_1, X_2) = X_1 + aX_2$, and with k replaced by the kernel K defined by $K(y) = \int k(y - ax)k(x)\, dx$.

It is even possible to obtain $\sqrt{n}$-consistency in various norms, together with functional central limit theorems in the corresponding spaces. Schick and Wefelmeyer (2004, 2007) prove such results for transformations of the form $q(X_1, \dots, X_m) = u_1(X_1) + \dots + u_m(X_m)$ and $q(X_1, X_2) = X_1 + X_2$ in the sup-norm and in L_1-norms. Giné and Mason (2007a) consider general transformations $q(X_1, \dots, X_m)$ and obtain such results in the L_p-norms. Their results hold locally uniformly in the bandwidth. Giné and Mason (2007b) prove a law of the iterated logarithm for the estimator. Du and Schick (2007) generalize some of these results to derivatives of convolutions of densities. More general results applicable to the estimation of densities of sums of independent random variables are Nickl (2007) and (2009).

We want to show that the above results are less generally valid than appears at first sight. Consider the case $q(X_1, X_2) = u(X_1) + u(X_2)$. In order to prove $\sqrt{n}$-consistency of the estimator, the above authors require the density of $u(X_1)$ to be square-integrable. But this assumption is typically already violated if u has a derivative that vanishes at a single point, for example if $u(X_1) = X_1^2$ and the density of X_1 is bounded away from zero in a neighborhood of zero. How critical is the assumption of square-integrability? In particular, is the Frees estimator $\hat{p}(z)$ for the density of $q(X_1, X_2) = X_1^2 + X_2^2$ at the point z still $\sqrt{n}$-consistent if f is bounded away from zero in a neighborhood of zero? We show that the answer to this question depends on z.

For $q(X_1, X_2) = X_1^2 + X_2^2$, the Frees estimator is

$$\hat{p}(z) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} k_b(z - X_i^2 - X_j^2). \tag{1.1}$$

Let h and g denote the densities of $|X_1|$ and X_1^2, respectively. Then

$$h(y) = (f(y) + f(-y))\mathbf{1}[y > 0]$$

and

$$g(y) = \frac{1}{2\sqrt{y}} h(\sqrt{y}).$$

Hence p is the convolution $g * g$ of g with itself.

We assume that h has bounded variation. Then h has finite left- and right-hand limits for all positive arguments. Assume that the right-hand limit $h(0+)$ of h at 0 is positive. We get different results for $\hat{p}(z)$ depending on whether the left-hand limit $g(z-)$ of g at z is positive or not.

Our arguments differ on the positive and negative parts of the support of the kernel k. This is why we state our results separately for kernels with support $[0,1]$ and $[-1,0]$. We also assume that the kernels are bounded densities.

Let us first consider the case when $g(z-)$ is positive. For kernels with support $[0,1]$ and $[-1,0]$, we show in Theorem 1 that $\sqrt{n/\log n}(\hat{p}(z)-p(z))$ is asymptotically normal with variance $h^2(0+)g(z-)$ if the bandwidth is proportional to $\sqrt{\log n}/n$. This non-standard choice of bandwidth is used to control the bias. Under the present assumptions on h, the density p is guaranteed to be Hölder with exponent $1/2$ only, so that the bias is of order $b^{1/2}$.

We can choose larger bandwidths if p is known to be smoother at z. Specifically, if p is Hölder at z with exponent α for $1/2 < \alpha \leq 1$, we can choose a bandwidth of order $n^{-1/(2\alpha)}$ and obtain that $\sqrt{n/\log n}(\hat{p}(z)-p(z))$ is asymptotically normal with variance $h^2(0+)g(z-)/(2\alpha)$. Thus, under additional smoothness assumptions on p, a smaller asymptotic variance can be achieved by choice of bandwidth, but the rate of convergence cannot be improved.

The asymptotic behavior of $\hat{p}(z)$ is governed by observations X_j with X_j^2 close to z. This implies that $\Delta(z_1)$ and $\Delta(z_2)$ are asymptotically independent for different z_1 and z_2, where $\Delta(z) = \sqrt{n/\log n}(\hat{p}(z) - p(z))$. In particular, functional central limit theorems for Δ are not possible. This is analogous to known results for classical kernel estimators.

Now we consider the case when $g(z-)$ is zero. Then the above asymptotic variances reduce to zero, indicating that better rates for $\hat{p}(z)$ are possible. Suppose that g is left Hölder at z, say $h(\sqrt{z-s}) = O(s^\beta)$ as $s \downarrow 0$, where β is positive. For a kernel with support $[0,1]$ we show in Theorem 2 that then $\hat{p}(z)$ behaves quite differently. If the bandwidth is again proportional to $\sqrt{\log n}/n$, we obtain that $\sqrt{n}(\hat{p}(z)-p(z))$ is asymptotically normal with variance given by the variance of $2g(z-X_1^2)$. This is analogous to the recent results on $\sqrt{n}$-consistent density estimation of transformations. The same result holds for a kernel with support $[-1,0]$ under a two-sided Hölder condition on g, say $h(\sqrt{z-s}) = O(|s|^\beta)$ as $s \to 0$, where β is again positive.

If the observations are bounded, then the support of p is larger than the support of g. Outside the support of g we have $g(z-) = 0$ and $h(\sqrt{z-s}) = 0$ for small positive s. Hence $\hat{p}(z)$ has the parametric rate for z outside the support of g. Functional central limit theorems in the space $C(I)$ of continuous functions on a compact interval I are possible as long as I is a subinterval of the complement of the support of g.

In Theorem 3 we combine the above results on one-sided kernels to obtain better rates under additional smoothness properties on p using higher-order kernels. Specifically, if k is a bounded symmetric density on $[-1,1]$, the bandwidth is proportional to $(n \log n)^{-1/4}$, and p has a second derivative at z, then $\sqrt{n/\log n}(\hat{p}(z) - p(z))$ is asymptotically normal with variance $h^2(0+)g(z-)/4$. If $g(z-) = 0$ and g is Hölder at z, then $\sqrt{n}(\hat{p}(z) - p(z))$ is asymptotically normal with variance given by the variance of $2g(z - X_1^2)$.

In Proposition 3 we show that our results remain valid if we replace the fixed bandwidth b by a random bandwidth $\hat{s}b$ with $\hat{s}$ a positive random variable such that $\hat{s} + 1/\hat{s} = O_p(1)$ provided we choose a smooth kernel.

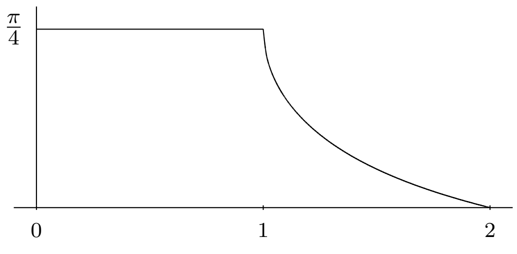
Let us illustrate the above with two special cases.

Example 1. Suppose that f is the standard normal density. Then p is the exponential density with mean 2 which is Hölder with exponent 1 at each positive z. We find $h(0+) = (2/\pi)^{1/2}$ and $g(z-) = g(z) = (2\pi z)^{-1/2} \exp(-z/2) > 0$ for all positive z. Thus $\sqrt{n/\log n}(\hat{p}(z) - p(z))$ is asymptotically normal with variance $\pi^{-3/2}(z/2)^{-1/2} \exp(-z/2)/(2\alpha)$ for a bandwidth of order $n^{-1/(2\alpha)}$ with $1/2 \le \alpha \le 1$. This holds for bounded kernels with compact support. Since p has a second derivative for positive z, the smaller variance $\pi^{-3/2}(z/2)^{-1/2} \exp(-z/2)/4$ can be achieved by using a symmetric kernel and a bandwidth proportional to $(n \log n)^{-1/4}$.

Example 2. Suppose that f is the uniform density on $(0,1)$. Using (2.1) one calculates

$$p(z) = \frac{\pi}{4}\,\mathbf{1}[0 < z \le 1] + \left(\frac{\pi}{4} - \arcsin\sqrt{\frac{z-1}{z}}\right)\mathbf{1}[1 < z < 2].$$

A graph of this density is given next.



We have $h(0+) = 1$ and $g(z-) = 1/(2\sqrt{z})\mathbf{1}[0 < z \le 1]$. If $0 < z < 1$, then $g(z-)$ is positive and $\sqrt{n/\log n}(\hat{p}(z) - p(z))$ is asymptotically normal with variance $1/(4\alpha\sqrt{z})$ for a bandwidth of order $n^{-1/(2\alpha)}$ with $1/2 \le \alpha$. This result holds also at $z = 1$ if the kernel has support $[0, 1]$. Since p is right Hölder with exponent $1/2$ at $z = 1$, $\sqrt{n/\log n}(\hat{p}(z) - p(z))$ is asymptotically normal for $z = 1$ with variance $1/2$ if the kernel has support $[-1, 0]$ and the bandwidth is proportional to $\sqrt{\log n}/n$. For $z > 1$ we have $g(z-) = 0$, and $\sqrt{n}(\hat{p}(z) - p(z))$ is asymptotically normal with variance

$$\frac{1}{2}\int_{l(z)}^{1} \frac{1}{z-y}\frac{1}{\sqrt{y}}\,dy - \frac{1}{4}\Big(\int_{l(z)}^{1} \frac{1}{\sqrt{z-y}}\frac{1}{\sqrt{y}}\,dy\Big)^2,$$

where $l(z) = \min\{1, z-1\}$. This variance is zero for $z \geq 2$ so that $\sqrt{n}(\hat{p}(z) - p(z)) = o_p(1)$ for such z.

It is not very surprising that densities of transformations $q(X_1, \dots, X_m)$ can often be estimated at the rate $\sqrt{n}$. Such densities at a point are represented as functionals of other densities, and the plug-in principle can be invoked. The slightly worse rate in our case $q(X_1, X_2) = X_1^2 + X_2^2$ is explained by the fact that the influence function for the corresponding density is just barely not square-integrable. For sum of more than two squares, and for sums $|X_1|^\nu + |X_2|^\nu$ with $\nu < 2$, $\sqrt{n}$-consistency can again be achieved. For sums $|X_1|^\nu + |X_2|^\nu$ with $\nu > 2$ the Frees estimator behaves more like an ordinary density estimator; it has the slower rate $n^{-1/\nu}$ when the density of $|X_1|^\nu$ is positive at $z-$. See Schick and Wefelmeyer (2009).

If $g(z-)$ is zero, then by Proposition 2 the error of $\hat{p}(z)$ is approximated up to $o_p(1/\sqrt{n})$ by the average $2A(z, g)$ defined before Proposition 1. Our model is locally asymptotically normal, and one can show that $p(z)$ is a Hellinger differentiable functional of the underlying density f. This implies that $\hat{p}(z)$ is asymptotically efficient in the sense of a nonparametric version of the convolution theorem or a local asymptotic minimax theorem.

If $g(z-)$ is positive, then by Proposition 1 the error of $\hat{p}(z)$ is approximated up to $o_p(\sqrt{\log n/n})$ by the average $2A(z, g\mathbf{1}_{(rb,\infty)})$ with $r \sim \log n$. We will show elsewhere that an extended version of local asymptotic normality still holds with a normalizing rate of order $\sqrt{n/\log n}$. With the help of such a result we can show that the rate $\sqrt{n/\log n}$ is optimal for estimating $p(z)$ in general (and not just by local U-statistics) and address asymptotic efficiency of the estimator $\hat{p}(z)$. We get a non-standard rate because of the non-standard assumption on the density: it is a convolution of two densities that are not square-integrable.

The literature contains several other density estimation problems under non-standard assumptions. For discontinuous densities, it is of interest to estimate location and size of a jump. We refer to Liebscher (1990) and Chu and Cheng (1996). For densities f with support bounded (to the right, say) by an unknown a and falling off as $f(a-z) \sim z^{\alpha-1}$ with $\alpha > 0$, the maximal observation has rate $n^{-1/\alpha}$. For these classical results see e.g. Embrechts, Klüppelberg and Mikosch (1997) or de Haan and Ferreira (2006). A more general problem is to estimate the boundary of the support of a multivariate density that falls off steeply at the boundary, or *frontier estimation*. Recent references are Korostelev, Simar and Tsybakov (1995) and Hall and Park (2004).

Our results are stated in Section 2; the proofs are in Section 3.

2. Results. We have the following representation for the density p,

$$p(z) = \int_0^z \frac{h(\sqrt{z-y})}{2\sqrt{z-y}} \frac{h(\sqrt{y})}{2\sqrt{y}}\, dy,$$

valid for $z > 0$. Using the substitution $y = zs$ we find for such z that

$$p(z) = \int_0^1 \frac{h(\sqrt{z(1-s)})h(\sqrt{zs})}{4\sqrt{(1-s)s}}\, ds.$$

Of course, $p(z) = 0$ for negative z. The representation shows that p is bounded if h is. Since the integrand is symmetric about $1/2$, we have

$$p(z) = 2\int_0^{1/2} \frac{h(\sqrt{z(1-s)})h(\sqrt{zs})}{4\sqrt{(1-s)s}}\, ds. \tag{2.1}$$

We study the behavior of the estimator $\hat{p}(z)$ at a fixed positive point z. We shall do so under the following condition on the bandwidth and under two alternative conditions on the kernel.

(B) *The bandwidth* b *satisfies* $b \to 0$ *and* $nb \to \infty$.

(K+) *The kernel* k *is a bounded density with support* $[0, 1]$.

(K–) *The kernel* k *is a bounded density with support* $[-1, 0]$.

Let k^+ and k^- be kernels as specified in (K+) and (K–), respectively. Write $\hat{p}_b^+$ and $\hat{p}_b^-$ for the corresponding Frees estimators (1.1), stressing the dependence on the kernel and the bandwidth b. Then we can define a new kernel k by $k(y) = \lambda k^-(y/c)/c + (1-\lambda)k^+(y/d)/d$ for $0 \le \lambda \le 1$ and positive c and d. This kernel is a bounded density with support $[-c, d]$. The Frees estimator $\hat{p}$ corresponding to this kernel k is of the form $\hat{p} = \lambda\hat{p}_{cb}^- + (1-\lambda)\hat{p}_{db}^+$. Thus the properties of the estimator corresponding to k are easily derived from those of the estimators corresponding to k^+ and k^-.

When working with condition (K+) we will control the bias by assuming that the density p is *left Hölder* at z with exponent α at least 1/2. This means that

$$|p(z-s)-p(z)| \leq Cs^{\alpha}, \quad 0 < s < \delta,$$

for some small positive δ and some constant C. When working with condition (K–) we will control the bias by assuming that the density p is *right Hölder* at z with exponent α at least 1/2. In both cases the bias $E[\hat{p}(z)-p(z)] = p * k_b(z) - p(z)$ is of order b^{α}.

The following lemma shows that p is (left and right) Hölder at z with exponent 1/2 if the density h is of bounded variation.

LEMMA 1. *Suppose h is of bounded variation. Then p is Hölder with exponent* 1/2 *at each positive argument.*

If h falls off to 0 from the left of $\sqrt{z}$ sufficiently fast, then p is left Hölder at z with an exponent larger that 1/2. To make this precise, we use the following assumption.

(H–) *There are positive constants B and β such that for some positive δ*

$$h(\sqrt{z-s}) \leq Bs^{\beta}, \quad 0 < s < \delta.$$

LEMMA 2. *Suppose h is of bounded variation and (H–) holds. Then p is left Hölder at z with exponent greater than* 1/2.

When working with (K–) we require a two-sided version of (H–).

(H) *There are positive constants B and β such that for some positive δ*

$$h(\sqrt{z-s}) \leq B|s|^{\beta}, \quad |s| < \delta.$$

LEMMA 3. *Suppose h is of bounded variation and (H) holds. Then p is Hölder at z with exponent greater than* 1/2.

Higher Hölder exponents can be guaranteed under stronger assumptions on h. For example, if h is *uniformly Lipschitz* on $(0,\infty)$,

$$|h(v)-h(u)| \leq \Lambda(v-u), \quad 0 < u < v,$$

for some constant Λ, then p is Hölder with exponent 1 at each positive argument. Indeed, we see from (3.4) below that now

$$\begin{aligned}|p(z_2)-p(z_1)| &\le 2\|h\|\Lambda \int_0^1 (\sqrt{z_2 s}-\sqrt{z_1 s})\frac{1}{4\sqrt{s(1-s)}}\,ds \\ &\le \|h\|\Lambda(\sqrt{z_2}-\sqrt{z_1})\int_0^1 \frac{1}{2\sqrt{1-s}}\,ds \\ &\le \|h\|\Lambda\frac{z_2-z_1}{\sqrt{z_2}+\sqrt{z_1}}, \qquad 0<z_1<z_2.\end{aligned}$$

Here $\|h\|$ denotes the sup-norm of h.

We first study our estimator under condition (K+). The target density is zero for negative arguments. A kernel with support on the positive axis guarantees that the estimator $\hat{p}$ vanishes for negative arguments.

If $h(0+)$ and $h(\sqrt{z}-)$ are positive, then $g(z-X_1^2)$ is not square-integrable. This is the reason why we lose the parametric rate of $\hat{p}(z)$, as shown next. For this we compare $\hat{p}(z)$ with its expected value $E[\hat{p}(z)] = p * k_b(z)$. To state our results, for a function ψ on the real line we introduce the notation

$$\begin{aligned}A(z,\psi) &= \frac{1}{n}\sum_{j=1}^n \big(\psi(z-X_j^2) - E[\psi(z-X_j^2)]\big) \\ &= \frac{1}{n}\sum_{j=1}^n \big(\psi(z-X_j^2) - \psi * g(z)\big).\end{aligned}$$

PROPOSITION 1. *Let h be of bounded variation with $h(0+)$ positive. Let (K+) or (K–) hold. Suppose the bandwidth b satisfies (B) and*

$$\log(1/b)/\log n \to \gamma \tag{2.2}$$

for some γ with $0<\gamma\le 1$. Then for $r\sim \log n$ we have

$$\hat{p}(z) - p * k_b(z) = 2A(z, g\mathbf{1}_{(rb,\infty)}) + o_p\Big(\sqrt{\log n/n}\Big) \tag{2.3}$$

and $\sqrt{n/\log n}A(z, g\mathbf{1}_{(rb,\infty)})$ is asymptotically normal with mean 0 and variance $h^2(0+)g(z-)\gamma$. Hence $\sqrt{n/\log n}(\hat{p}(z)-p*k_b(z))$ is asymptotically normal with mean 0 and variance $h^2(0+)g(z-)\gamma$.

The requirement (2.2) is met by the choice $b\sim n^{-\gamma}\log^\tau n$ for some real τ which needs to be positive if $\gamma=1$ in order to fulfill (B). If p is left Hölder with exponent α with $1/2\le\alpha\le 1$, then the bias $p*k_b(z)-p(z)$ is of order b^α. Thus $\sqrt{n/\log n}(\hat{p}(z)-p(z))$ is asymptotically normal with mean 0 and variance $h^2(0+)g(z-)\gamma$ if $\gamma > 1/(2\alpha)$ or if $\gamma = 1/(2\alpha)$ and $\tau < 1/(2\alpha)$. The smallest asymptotic variance corresponds to the choice $\gamma = 1/(2\alpha)$. If we know only that the density h is of bounded variation, then α may be as small as $1/2$, and $\gamma = 1$.

THEOREM 1. *Let h be of bounded variation with $h(0+)$ positive. Let (K+) or (K–) hold and suppose that $b \sim \sqrt{\log n}/n$. Then $\sqrt{n/\log n}(\hat{p}(z) - p(z))$ is asymptotically normal with mean 0 and variance $h^2(0+)g(z-)$.*

Proposition 1 shows that the estimator $\hat{p}(z)$ cannot have a faster rate than $\sqrt{\log n/n}$ if $g(z-)$ is positive.

We now address the case when $g(z-) = 0$.

PROPOSITION 2. *Let h be of bounded variation. Let (B) hold. Assume that either (H–) and (K+) or (H) and (K–) hold. Then*

$$\hat{p}(z) - p * k_b(z) = 2A(z, g) + o_p(1/\sqrt{n}), \tag{2.4}$$

and $\sqrt{n}(\hat{p}(z) - p * k_b(z))$ is asymptotically normal with mean 0 and variance $4\operatorname{Var}(g(z - X_1^2))$.

Since the Hölder exponent α is usually unknown, we should choose b as in Theorem 1 and obtain the following result.

THEOREM 2. *Let h be of bounded variation. Assume that either (H–) and (K+) or (H) and (K–) hold. Let $b \sim \sqrt{\log n}/n$. Then $\sqrt{n}(\hat{p}(z) - p(z))$ is asymptotically normal with mean 0 and variance $4\operatorname{Var}(g(z - X_1^2))$.*

The choice $b \sim \sqrt{\log n}/n$ works in both theorems and gives optimal convergence rate $\sqrt{\log n/n}$ if $g(z-)$ is positive, and the rate $1/\sqrt{n}$ if $g(z-)$ is zero and the assumptions of Theorem 2 hold.

If we are sure that h is Lipschitz, then we can choose $b \sim 1/\sqrt{n \log n}$ and obtain that $\sqrt{n/\log n}(\hat{p}(z) - p(z))$ is asymptotically normal with mean 0 and variance $h^2(0+)g(z-)/2$ if $g(z-)$ is positive, and $\sqrt{n}(\hat{p}(z) - p(z))$ is asymptotically normal with mean 0 and variance $4\operatorname{Var}(g(z - X_1^2))$ under the assumptions of Theorem 2.

If p has a second derivative at z and the kernel k has mean zero and compact support, then the bias is of order $O(b^2)$. Thus we have the following result.

THEOREM 3. *Assume that h is of bounded variation and that p is twice differentiable at z. Suppose that k is a bounded symmetric density with support $[-1, 1]$ and $b \sim (n \log n)^{-1/4}$. Then the following hold.*

(i) If $h(0+)$ is positive, then $\sqrt{n/\log n}(\hat{p}(z)-p(z))$ is asymptotically normal with mean 0 and variance $h^2(0+)g(z-)/4$.
(ii) If (H) holds, then $\sqrt{n}(\hat{p}(z)-p(z))$ is asymptotically normal with mean 0 and variance $4\,\mathrm{Var}(g(z-X_1^2))$.

We conclude this section by showing that we can replace in the above results the constant bandwidth b by a random bandwidth of the form $\hat{s}b$ for some positive random variable $\hat{s}$ such that $\hat{s}+1/\hat{s}=O_p(1)$. This requires that we also use a continuously differentiable kernel.

It should be clear that the treatment of the bias is unaffected by this choice of random bandwidth. To treat the variance term, write $\hat{p}_{sb}(z)$ for the Frees estimator with bandwidth sb and $\bar{p}_{sb}(z)=p*k_{sb}(z)$ for its mean, where s is positive. In order to use a random bandwidth $\hat{s}b$ as above, we need to show that for each compact interval I contained in $(0,\infty)$ one has

$$\sup_{s\in I}|\hat{p}_{sb}(z)-\bar{p}_{sb}(z)-2A(z,g\mathbf{1}_{(b\log n,\infty)})|=o_p\Big(\sqrt{\log n/n}\Big) \tag{2.5}$$

or, respectively,

$$\sup_{z\in I}|\hat{p}_{sb}(z)-\bar{p}_{sb}(z)-2A(z,g)|=o_p(1/\sqrt{n}). \tag{2.6}$$

Note that (2.5) is a uniform version of (2.3) with b replaced by sb and $sr=\log n$, and (2.6) is a uniform version of (2.4) with b replaced by sb. Thus (2.5) holds if the sequence $\{\sqrt{n/\log n}(\hat{p}_{sb}(z)-\bar{p}_{sb}(z)),s\in I\}$ of processes is tight in $C(I)$, while (2.6) holds if the sequence $\{\sqrt{n}(\hat{p}_{sb}(z)-\bar{p}_{sb}(z)),s\in I\}$ of processes is tight in $C(I)$. Conditions for tightness are given next.

PROPOSITION 3. *Suppose the kernel k is a continuously differentiable density with support contained in $[-1,1]$. Let h have bounded variation and let I be a compact interval in $(0,\infty)$. Then the sequence $\{\sqrt{n/\log n}(\hat{p}_{sb}(z)-\bar{p}_{sb}(z)),s\in I\}$ of processes is tight in $C(I)$. If also (H–) holds, then the sequence $\{\sqrt{n}(\hat{p}_{sb}(z)-\bar{p}_{sb}(z)),s\in I\}$ of processes is tight in $C(I)$.*

3. Proofs.

This section contains the proofs of Lemmas 1 to 3 and of Propositions 1 to 3. In what follows we repeatedly use the inequalities

$$(s+t)^\gamma\le s^\gamma+t^\gamma\quad\text{and}\quad t^\gamma-s^\gamma\le(t-s)^\gamma \tag{3.1}$$

valid for $0\le s<t$ and $0<\gamma\le 1$.

PROOF OF LEMMA 1. Since h is of bounded variation, we may assume that $h(y)=\int\mathbf{1}[0\le t\le y]\nu(dt)$, where ν is the difference $\nu_1-\nu_2$ of two finite measures. Write $\mu = \nu_1 + \nu_2$, and set

$$\int_u^v r(t)\mu(dt) = \int r(t)\mathbf{1}[u < t \le v]\mu(dt)$$

for $0 \le u \le v < \infty$. Then

$$|h(v) - h(u)| \le \int_u^v \mu(dt). \tag{3.2}$$

Let

$$w(s) = \frac{\mathbf{1}[0 < s < 1]}{2\sqrt{s} \cdot 2\sqrt{1-s}}.$$

It is easy to check that

$$\int_u^v w(s)\,ds \le \sqrt{v-u}, \quad 0 < u < v < 1. \tag{3.3}$$

We can write

$$\begin{aligned}|p(z_2) - p(z_1)| \le \|h\| \int |h(\sqrt{z_1 s}) - h(\sqrt{z_2 s})| w(s)\,ds \\ + \|h\| \int \Big| h\Big(\sqrt{z_1(1-s)}\Big) - h\Big(\sqrt{z_2(1-s)}\Big)\Big| w(s)\,ds.\end{aligned}$$

Using the substitution $u = 1 - s$ and the identity $w(s) = w(1-s)$, we see that the two integrals on the right-hand side are the same. Thus we have

$$|p(z_2) - p(z_1)| \le 2\|h\| \int |h(\sqrt{z_1 s}) - h(\sqrt{z_2 s})| w(s)\,ds \tag{3.4}$$

for $0 < z_1 < z_2$. Using (3.2) and (3.3), we can bound the integral on the right-hand side by

$$\begin{aligned}\int \int_{\sqrt{z_1 s}}^{\sqrt{z_2 s}} \mu(dt)\, w(s)\,ds &= \int_0^{\sqrt{z_2}} \int_{t^2/z_2}^{t^2/z_1} w(s)\,ds\,\mu(dt) \\ &\le \sqrt{\frac{1}{z_1} - \frac{1}{z_2}} \int_0^{\sqrt{z_2}} t\,\mu(dt)\end{aligned}$$

and thus obtain

$$|p(z_2) - p(z_1)| \le 2\|h\| \sqrt{\frac{z_2 - z_1}{z_1}}\, \mu[0, \sqrt{z_2}], \quad 0 < z_1 < z_2.$$

Since μ is a finite measure, this yields the desired result. $\square$

Proof of Lemma 2. It suffices to show that

$$|p(uz) - p(z)| \le D(1-u)^{\alpha}, \quad 1/2 < u < 1, \tag{3.5}$$

for some positive D and some $\alpha > 1/2$. Since h is of bounded variation, it follows that h is bounded, and thus we may assume that (H–) holds with $\delta = z$. Thus

$$h(\sqrt{zv}) \le Bz^{\beta}(1-v)^{\beta}, \quad 0 \le v \le 1. \tag{3.6}$$

From now on we assume w.l.g. that $\beta \le 1/2$. In view of the representation (2.1), we can bound the left-hand side of (3.5) by $2(I_1 + I_2)$, where

$$\begin{aligned}
I_1 &= \int_0^{1/2} \frac{|h(\sqrt{uz(1-s)}) - h(\sqrt{z(1-s)})| h(\sqrt{uzs})}{4\sqrt{s(1-s)}}\, ds \\
&\le \|h\| \int_0^{1/2} \frac{|h(\sqrt{uz(1-s)}) - h(\sqrt{z(1-s)})|}{2\sqrt{s}}\, ds, \\
I_2 &= \int_0^{1/2} \frac{h(\sqrt{z(1-s)}) |h(\sqrt{uzs}) - h(\sqrt{zs})|}{4\sqrt{s(1-s)}}\, ds \\
&\le Bz^{\beta} \int_0^{1/2} |h(\sqrt{uzs}) - h(\sqrt{zs})| s^{\beta - 1/2}\, ds.
\end{aligned}$$

Using (3.2) and (3.1) with $\gamma = \beta + 1/2$, we can write

$$\begin{aligned}
\int_0^{1/2} |h(\sqrt{uzs}) - h(\sqrt{zs})| s^{\beta-1/2}\, ds &= \int_0^{1/2} \int_{\sqrt{uzs}}^{\sqrt{zs}} \mu(dt) s^{\beta - 1/2}\, ds \\
&= \int_0^{\sqrt{z/2}} \int_{t^2/z}^{t^2/(uz)} s^{\beta-1/2}\, ds\, \mu(dt) \\
&\le \int_0^{\sqrt{z/2}} \frac{\left(\frac{t^2}{uz} - \frac{t^2}{z}\right)^{\beta+1/2}}{\beta + 1/2} \mu(dt) \\
&\le \frac{(1-u)^{\beta+1/2}}{\beta+1/2} \mu(0,\infty).
\end{aligned}$$

Let $0 < \eta < 1/2$. Using (3.6) with $v = 1-s$ and $v = u(1-s)$, we obtain

$$\begin{aligned}
I_{11} &= \int_0^{\eta} \left| h\Big(\sqrt{uz(1-s)}\Big) - h\Big(\sqrt{z(1-s)}\Big) \right| \frac{ds}{2\sqrt{s}} \\
&\le \int_0^{\eta} Bz^{\beta} \big((1-u(1-s))^{\beta} + s^{\beta}\big) \frac{ds}{2\sqrt{s}}.
\end{aligned}$$

By (3.1) we have $(1-u(1-s))^\beta = (s+(1-s)(1-u))^\beta \le s^\beta + (1-s)^\beta(1-u)^\beta$ and therefore

$$\begin{aligned} I_{11} &\le Bz^\beta \int_0^\eta (2s^\beta + (1-u)^\beta)\frac{ds}{2\sqrt{s}} \\ &= Bz^\beta\Big(\frac{\eta^{\beta+1/2}}{\beta+1/2} + (1-u)^\beta\sqrt{\eta}\Big). \end{aligned}$$

Using (3.2),

$$\begin{aligned} I_{12} &= \int_\eta^{1/2} \Big|h\Big(\sqrt{uz(1-s)}\Big) - h\Big(\sqrt{z(1-s)}\Big)\Big|\frac{ds}{2\sqrt{s}} \\ &\le \frac{1}{2\sqrt{\eta}}\int_\eta^{1/2}\int_{\sqrt{uz(1-s)}}^{\sqrt{z(1-s)}} \mu(dt)\, ds \\ &= \frac{1}{2\sqrt{\eta}}\int_{\sqrt{uz/2}}^{\sqrt{z(1-\eta)}}\int_{1-t^2/(uz)}^{1-t^2/z} ds\, \mu(dt) \\ &\le \frac{1}{2\sqrt{\eta}}\frac{1-u}{u}\mu(0,\infty). \end{aligned}$$

The above shows that there is a constant C such that

$$|p(uz) - p(z)| \le C\Big(\eta^{\beta+1/2} + (1-u)^\beta\sqrt{\eta} + \frac{1-u}{\sqrt{\eta}} + (1-u)^{\beta+1/2}\Big)$$

for all η and u such that $0 < \eta \le 1/2 < u < 1$. Taking η to be the smaller of $1/2$ and $(1-u)^{1/(\beta+1)}$, we see that (3.5) holds with $\alpha = (2\beta+1)/(2\beta+2)$. $\square$

Proof of Lemma 3. In view of Lemma 2 we only need to show that

$$|p(uz) - p(z)| \le D(u-1)^\alpha, \quad 1 < u < 3/2, \tag{3.7}$$

for some positive D and some $\alpha > 1/2$. In view of (H) and since h is bounded we may assume that

$$h(\sqrt{zv}) \le Bz^\beta|1-v|^\beta, \quad 0 \le v \le 3/2,$$

and $0 < \beta \le 1/2$. We then calculate as in the proof of Lemma 2 that

$$|p(uz) - p(z)| \le C\Big(\eta^{\beta+1/2} + (u-1)^\beta\sqrt{\eta} + \frac{u-1}{\sqrt{\eta}} + (u-1)^{\beta+1/2}\Big)$$

for some constant C and for all η and u such that $0 < \eta \le 1/2$ and $1 < u < 3/2$. Taking η to be the smaller of $1/2$ and $(u-1)^{1/(\beta+1)}$, we see that (3.7) holds with $\alpha = (2\beta+1)/(2\beta+2)$. $\square$

PROOF OF PROPOSITION 1. For $c < d < z$, we have

$$\Gamma(c,d) = \sup_{c<s<d} g(z-s) \le \frac{\|h\|}{2\sqrt{z-d}}.$$

It follows from the assumptions on h that $\Gamma(0,t) \to g(z-)$ as $t \downarrow 0$. Since $2\sqrt{t}g(t) \le \|h\|$, we obtain for $0 < u < v$,

$$\int_u^v g(t)\,dt \le \|h\|(\sqrt{v}-\sqrt{u}) \le \|h\|\sqrt{v-u}. \tag{3.8}$$

Note that for bounded ψ,

$$nE[A^2(z,\psi)] = \psi^2 * g(z) - (\psi * g(z))^2 \le \psi^2 * g(z). \tag{3.9}$$

The Hoeffding decomposition of the U-statistic $\hat{p}(z)$ is

$$\hat{p}(z) = p * k_b(z) + 2A(z, g * k_b) + U(z)$$

where

$$U(z) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \Big(k_b(z - X_i^2 - X_j^2) \\ - g * k_b(z - X_i^2) - g * k_b(z - X_j^2) + p * k_b(z)\Big).$$

We have

$$n(n-1)E[U^2(z)] \le 2E[k_b^2(z - X_1^2 - X_2^2)] = 2k_b^2 * p(z) \tag{3.10}$$

and

$$k_b^2 * p(z) = \int p(z-bu)k^2(u)\,du/b \le \|p\| \int k^2(u)\,du/b. \tag{3.11}$$

Since p and k are bounded, we obtain

$$U(z) = O_p(1/(nb^{1/2})) = o_p(1/\sqrt{n}) \tag{3.12}$$

by (B).

Let $[-a,c]$ denote the support of k_b. We have $[-a,c] = [0,b]$ under (K+) and $[-a,c] = [-b,0]$ under (K–). Set $g_b = g * k_b$. We will now show that

$$T_1 = A(z, g_b \mathbf{1}_{(-\infty,rb)}) = O_p\Big(\sqrt{r\Gamma(-a,rb)/n}\Big) \tag{3.13}$$

and

$$T_2 = A(z, g_b \mathbf{1}_{[rb,\infty)}) - A(z, g\mathbf{1}_{[rb,\infty)}) = o_p(1/\sqrt{n}). \tag{3.14}$$

Note that $g_b(y) = 0$ for $y \le -a$. For $y > -a$, applications of the Cauchy–Schwarz inequality and (3.8) yield

$$g_b^2(y) = \Big(\int_{-a}^{c} g(y-u) k_b(u)\, du \Big)^2 \le \int_{-a}^{b} g^2(y-u) k_b(u)\, du \tag{3.15}$$

and

$$g_b^2(y) \le \int_{-a}^{c} g(y-u)\, du \int_{-a}^{c} g(y-u) k_b^2(u)\, du \le \sqrt{b}\|h\|\|k_b\| g_b(y). \tag{3.16}$$

Now we use (3.9) and (3.16) to derive the bound

$$nE[T_1^2] \le \sqrt{b}\|h\|\|k_b\| \int_{-a}^{rb} g(z-y) g_b(y)\, dy.$$

In view of (3.8) we obtain

$$\int_{-a}^{rb} g(z-y) g_b(y)\, dy \le \Gamma(-a, rb) \int_{-a}^{rb} g_b(y)\, dy \le \Gamma(-a, rb) \|h\| \sqrt{rb + a}.$$

Thus we have $nE[T_1^2] \le \|h\|^2 \|k\| \Gamma(-a, rb) \sqrt{r+1}$ which yields (3.13).

Note that

$$|g_b - g| \mathbf{1}_{[s,\infty)} \le \frac{\|h\|}{\sqrt{s-b}}, \quad s > b. \tag{3.17}$$

Next we show that there is a constant C such that

$$\int_s^\infty (g_b(t) - g(t))^2\, dt \le \frac{Cb}{s-b}, \quad s > b. \tag{3.18}$$

For $b < s \le t$ and $|u| < b$ we have

$$\begin{aligned} |g(t-u) - g(t)| &\le \frac{|h(\sqrt{t-u}) - h(\sqrt{t})|}{2\sqrt{t-u}} + h(\sqrt{t}) \Big| \frac{1}{2\sqrt{t-u}} - \frac{1}{2\sqrt{t}} \Big| \\ &\le \frac{|h(\sqrt{t-u}) - h(\sqrt{t})|}{2\sqrt{s-b}} + \frac{b\|h\|}{4(t-b)^{3/2}}. \end{aligned} \tag{3.19}$$

Since h is of bounded variation and the map φ defined by $\varphi(t) = \sqrt{\max(0,t)}$ is nondecreasing, $h \circ \varphi$ is of bounded variation. Thus, as shown in the proof of Lemma 8 of Schick and Wefelmeyer (2007), there is a constant L such that

$$\int |h \circ \varphi(t-u) - h \circ \varphi(t)|\, dt \le L|u|, \quad u \in \mathbb{R}.$$

In particular, we have

$$\int_s^\infty |h(\sqrt{t-u}) - h(\sqrt{t})|\,dt \le L|u|, \quad |u| \le s.$$

Using this and (3.19) we obtain

$$\int_s^\infty |g_b(t) - g(t)|\,dt \le \sup_{0<u<b} \int_s^\infty |g(t-u) - g(t)|\,dt \le \frac{(L + \|h\|)b}{2\sqrt{s-b}}$$

for $s > b$. From this and (3.17) we derive (3.18) with $C = \|h\|(L + \|h\|)/2$.

To prove (3.14), we write $T_2 = A(z, (g_b - g)\mathbf{1}_{[rb,\infty)})$. Using (3.8), (3.9), (3.17) and (3.18) and setting $z_r = z - 1/r$, we obtain the bound

$$\begin{aligned} nE[T_2^2] &\le \int_{rb}^z g(z-t)(g_b(t) - g(t))^2\,dy \\ &\le \Gamma(rb, z_r) \int_{rb}^{z_r} (g_b(t) - g(t))^2\,dt + \frac{\|h\|^2}{z_r - b} \int_{z_r}^z g(z-t)\,dt \\ &\le \frac{\|h\|\sqrt{r}Cb}{rb - b} + \frac{\|h\|^3}{(z_r - b)\sqrt{r}} = O(1/\sqrt{r}), \end{aligned}$$

which implies (3.14).

The Hoeffding decomposition and relations (3.12)–(3.14) imply

$$\sqrt{n/\log n}(\hat{p}(z) - p * k_b(z)) = 2T + o_p(1),$$

where

$$T = \sqrt{n/\log n}A(z, g\mathbf{1}_{[rb,\infty)}) = \sum_{j=1}^n (Z_{nj} - E[Z_{nj}])$$

with

$$Z_{nj} = \frac{1}{\sqrt{n \log n}} (g\mathbf{1}_{[rb,\infty)})(z - X_j^2).$$

We have

$$\sqrt{n \log n}E[Z_{nj}] = g\mathbf{1}_{[rb,\infty)} * g(z) \le g * g(z) = p(z).$$

With $e = rb + 1/r$ we derive

$$(g\mathbf{1}_{[rb,e]})^2 * g(z) = \int_{rb}^e \frac{h^2(\sqrt{t})g(z-t)}{4t}\,dt = \Big(\frac{1}{4}h^2(0+)g(z-) + o(1)\Big) \log\Big(\frac{e}{rb}\Big),$$

$$(g\mathbf{1}_{(e,z/2]})^2 * g(z) = \int_e^{z/2} \frac{h^2(\sqrt{t})g(z-t)}{4t}\,dt \le -\log\Big(\frac{2e}{z}\Big)\|h\|^2\Gamma(e, z/2),$$

$$(g\mathbf{1}_{(z/2,\infty)})^2 * g(z) = \int_{z/2}^{z} \frac{h^2(\sqrt{t})g(z-t)}{4t}\,dt \le \frac{\|h\|^2}{2z}.$$

Since $-\log e = o(\log n)$ and $\log(1/b)/\log n \to \gamma$, we obtain

$$nE[Z_{nj}^2] = \frac{(g\mathbf{1}_{[rb,\infty)})^2 * g(z)}{\log n} = \frac{1}{4}h^2(0+)g(z-)\gamma + o(1)$$

and thus

$$\sum_{j=1}^{n} \operatorname{Var} Z_{nj} = \frac{1}{4}h^2(0+)g(z-)\gamma + o(1).$$

Also,

$$0 \le Z_{nj} \le \frac{\|h\|}{\sqrt{rbn\log n}} = o(1).$$

This implies the Lindeberg condition. Thus, by the Lindeberg–Feller central limit theorem, T is asymptotically normal with mean 0 and variance $h^2(0+)g(z-)\gamma/4$. □

Proof of Proposition 2. We use the notation of the proof of Proposition 1. The stronger requirements on b are not needed for the proofs of (3.12), (3.13) and (3.14). By our assumptions on h,

$$g^2(z-t)g(t) \le B^3(z-t)^{\beta-1}t^{2\beta-1/2}\mathbf{1}[0<t<z].$$

Thus $g(z-X_1^2)$ has a finite second moment. Therefore $\sqrt{n}A(z,g)$ is asymptotically normal with mean 0 and variance $\operatorname{Var} g(z-X_1^2)$, and

$$A(z, g\mathbf{1}_{(-\infty,rb)}) = o_p(1/\sqrt{n}).$$

This and the results of the proof of Proposition 1 yield

$$\hat{p}(z) - p * k_b(z) = 2A(z,g) + 2T_1 + o_p(1/\sqrt{n}).$$

Under (K+) and (H–) we have $r\Gamma(-a,rb) = r\Gamma(0,rb) \le Br^{1+\beta}b^{\beta}$, and under (K–) and (H) we have $r\Gamma(-a,rb) \le Br^{1+\beta}b^{\beta}$. Thus in each case we have $T_1 = o_p(1/\sqrt{n})$ in view of (3.13). □

Proof of Proposition 3. Write $I = [L,R]$. The assumptions on the kernel imply that there is a constant K such that

$$\|k_t - k_s\|_1 + \|k_t - k_s\|_2 \le K|t-s|, \quad s,t \in I.$$

The Hoeffding decomposition yields

$$\hat{p}_{sb}(z) - \bar{p}_{sb}(z) = 2A_s + U_s$$

where $A_s = A(z, g * k_{sb})$ and

$$U_s = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \Big(k_{sb}(z - Y_i - Y_j) - g * k_{sb}(z - Y_i) - g * k_{sb}(z - Y_j) + \bar{p}_{sb}(z) \Big).$$

Now fix s and t in I and set $\Delta = k_t - k_s$ and $\Delta_b = k_{tb} - k_{sb}$. Then $\Delta_b(x) = \Delta(x/b)/b$. Using (3.10) and (3.11) with k replaced by Δ we derive the inequality

$$n(n-1)bE[(U_t - U_s)^2] \le 2\|p\| \|\Delta\|_2^2.$$

We have the identity $A_t - A_s = A(z, g * k_{tb} - g * k_{sb}) = A(z, g * \Delta_b)$. This and (3.9) yield the bound

$$nE[(A_t - A_s)^2] \le g * (g * \Delta_b)^2(z) = \int g(z-y)(g * \Delta_b)^2(y)\, dy.$$

Since Δ_b has support contained in $[-Rb, Rb]$, an application of the Cauchy–Schwarz inequality and the use of (3.8) yield

$$\begin{aligned}(g * \Delta_b)^2(y) &\le \int g(y-u)\Delta_b^2(u)\, du \int_{-Rb}^{Rb} g(y-u)\, du \\ &\le g * \Delta_b^2(y)\sqrt{2Rb}\|h\|\end{aligned}$$

and

$$\int_{-\infty}^{2Rb} g * \Delta_b^2(y)\, dy \le \|\Delta_b\|_2^2 \int_0^{3Rb} g(v)\, dv \le \|\Delta_b\|_2^2 \sqrt{3bR}\|h\|.$$

These inequalities and the identity $b\|\Delta_b\|_2^2 = \|\Delta\|_2^2$ give the bound

$$I_1 = \int_{-\infty}^{2Rb} g(z-y)(g * \Delta_b)^2(y)\, dy \le \sqrt{6}R\Gamma(-\infty, 2Rb)\|h\|^2 \|\Delta\|_2^2.$$

An application of the Cauchy–Schwarz inequality yields

$$(g * \Delta_b)^2(y) \le \int g^2(y-u)|\Delta_b(u)|\, du \|\Delta_b\|_1, \quad y > 2Rb.$$

Since $\|\Delta_b\|_1 = \|\Delta\|_1$, we obtain with $\delta > 0$ that

$$\begin{aligned}I_2 &= \int_{2Rb}^{2Rb+\delta} g(z-y)(g * \Delta_b)^2(y)\, dy \\ &\le \|\Delta\|_1 \iint_{2Rb < y + bu < 2Rb + \delta} g(z - y - bu) g^2(y) |\Delta(u)|\, dy\, du \\ &\le \|\Delta\|_1^2 \Gamma(2Rb, 2Rb + \delta)\|h\|^2 \int_{Rb}^{3Rb+\delta} \frac{1}{4x}\, dx\end{aligned}$$

and

$$I_3 = \int_{2Rb+\delta}^{z} g(z-y)(g*\Delta_b)^2(y)\,dy \le \|\Delta\|_1^2\|h\|\frac{1}{2\sqrt{\delta}}\|p\|.$$

The above show that there is a constant M such that

$$n^2bE[(U_t - U_s)^2] \le M|t-s|^2, \quad s,t \in I, \tag{3.20}$$

and

$$nE[(A_t - A_s)^2] \le M|t-s|^2 \log n, \quad s,t \in I. \tag{3.21}$$

The latter can be improved if (H–) holds. Then we have

$$g(z-y-bu) \le B(y+bu)^\beta \le B(2y)^\beta, \quad 2Rb < y, \quad |u| \le R,$$

and obtain

$$\begin{aligned} I_2 &\le \|\Delta\|_1 \iint_{2Rb<y+bu<2Rb+\delta} 2^\beta B y^\beta g^2(y)|\Delta(u)|\,dy\,du \\ &\le \|\Delta\|_1^2\|h\|^2 2^\beta B \int_{Rb}^{3Rb+\delta} y^{\beta-1}\,dy. \end{aligned}$$

Thus if (H–) holds we obtain

$$nE[(A_t - A_s)^2] \le M|t-s|^2, \quad s,t \in I. \tag{3.22}$$

The result now follows from Theorem 12.3 in Billingsley (1968). □

Acknowledgments. A first version of this paper was finished while the authors stayed at the Oberwolfach Institute, June 3–16, 2007, under the program "Research in Pairs". Anton Schick was supported by NSF Grant DMS0405791.

REFERENCES

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

[2] Chu, C. K. and Cheng, P. E. (1996). Estimation of jump points and jump values of a density function. *Statist. Sinica* **6**, 79–95.

[3] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory. An Introduction.* Springer Series in Operations Research and Financial Engineering. Springer, New York.

[4] Du, J. and Schick, A. (2007). Root-n consistency and functional central limit theorems for estimators of derivatives of convolutions of densities. *Internat. J. Statist. Management Syst.* **2**, 67–87.

[5] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events. For Insurance and Finance.* Applications of Mathematics 33. Springer, Berlin.

[6] Frees, E. W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.* **89**, 517–525.

[7] Giné, E. and Mason, D. (2007a). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.* **35**, 1105–1145.
[8] Giné, E. and Mason, D. (2007b). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.* **20**, 457–485.
[9] Hall, P. and Park, B. U. (2004). Bandwidth choice for local polynomial estimation of smooth boundaries. *J. Multivariate Anal.* **91**, 240–261.
[10] Korostelev, A. P., Simar, L. and Tsybakov, A. B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* **23**, 476–489.
[11] Liebscher, E. (1990). Kernel estimators for probability densities with discontinuities. *Statistics* **21**, 185–196.
[12] Nickl, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields* **138**, 411–449. Erratum: **141** (2008) 331–332.
[13] Nickl, R. (2009). On convergence and convolutions of random signed measures. *J. Theoret. Probab.* **22**, 38–56.
[14] Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canad. J. Statist.* **28**, 799–815.
[15] Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.* **16**, 925–935.
[16] Schick, A. and Wefelmeyer, W. (2007). Root-n consistent density estimators of convolutions in weighted L_1-norms. *J. Statist. Plann. Inference* **37**, 1765–1774.
[17] Schick, A. and Wefelmeyer, W. (2009). Convergence rates of density estimators for sums of powers of observations. *Metrika* **69**, 249–264.

Anton Schick
Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902-6000
USA
e-mail: anton@math.binghamton.edu
url: math.binghamton.edu/anton/

Wolfgang Wefelmeyer
Mathematisches Institut
Universität zu Köln
Weyertal 86-90
50931 Köln
Germany
e-mail: wefelm@math.uni-koeln.de
url: www.mi.uni-koeln.de/∼wefelm/