Metrika manuscript No.
(will be inserted by the editor)

Convergence rates of density estimators for sums of powers of observations

Anton Schick · Wolfgang Wefelmeyer

Received: date / Accepted: date

Abstract Densities of functions of independent and identically distributed random observations can be estimated by a local U-statistic. It has been shown recently that, under an appropriate integrability condition, this estimator behaves asymptotically like an empirical estimator. In particular, it converges at the parametric rate. The integrability condition is rather restrictive. It fails for the sum of powers of two observations when the exponent is at least two. We have shown elsewhere that for exponent equal to two the rate of convergence slows down by a logarithmic factor on the support of the squared observation and is still parametric outside this support. For exponent greater than two, and on the support of the exponentiated observation, the estimator behaves like a classical density estimator: The bias is not negligible and the rate depends on the bandwidth. Outside the support, the rate is again parametric.

Keywords Local U-statistic · density estimator · convergence rate · Hoeffding decomposition

Mathematics Subject Classification (2000) 62G07 · 62G20

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

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

Wolfgang Wefelmeyer
Mathematisches Institut
Universität zu Köln
Weyertal 86-90
50931 Köln
Germany
E-mail: wefelm@math.uni-koeln.de

observations. Frees (1994) proposed as an estimator of $p(z)$ the local U-statistic

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

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

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

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 $\hat{f}$. As pointed out in Schick and Wefelmeyer (2008), if $\hat{f}(x) = (1/n)\sum_{j=1}^n K_b(x - X_j)$, then the estimator $\hat{p}_{SC}(z)$ is asymptotically equivalent to $\hat{p}_F(z)$ with $m = 2$, $q(X_1, X_2) = X_1 + aX_2$, and $k(y) = \int K(y - ax)K(x)\,dx$.

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 $q(X_1, \ldots, X_m) = u_1(X_1) + \cdots + u_m(X_m)$ 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 $q(X_1, \ldots, X_m)$ 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 the estimation of derivatives of convolutions of densities.

These results are less generally valid than appears at first sight. Consider the case

$$q(X_1, X_2) = |X_1|^\nu + |X_2|^\nu$$

for some positive ν. Then the Frees estimator is

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

For $\sqrt{n}$-consistency the above authors require the density of $|X_1|^\nu$ to be square-integrable. Let h and g denote the densities of $|X_1|$ and $|X_1|^\nu$, respectively. Then

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

and, with $\beta = 1/\nu$,

$$g(y) = \beta y^{\beta - 1} h(y^\beta).$$

If h is bounded, then the density g of $|X_1|^\nu$ is square-integrable if $\nu < 2$. However, if $\nu \ge 2$ and $\liminf_{y \to 0+} h(y) > 0$, then it is not. Indeed, the Frees estimator behaves differently for $\nu < 2$, $\nu = 2$ and $\nu > 2$. In the following we describe its asymptotic behavior at a fixed positive z. Throughout this paper we always impose the following condition on the kernel k.

(K) *The kernel k is a continuously differentiable function that integrates to one and has support* $[-1, 1]$.

In the ensuing discussion we restrict ourselves to the case when h is of bounded variation and the right-hand limit $h(0+)$ of h at 0 is positive. This rules out square-integrability of g. The requirement of bounded variation implies some smoothness of p, which is used to control the bias.

For $\nu < 2$, we can apply the arguments of Schick and Wefelmeyer (2004, 2008) and Giné and Mason (2007a) and obtain the following result.

Theorem 1 *Let $\nu < 2$. Suppose the density h is of bounded variation and $h(0+)$ is positive. Let $b \sim \sqrt{\log n}/n$. Then*

$$\sqrt{n}(\hat{p}_b(z) - p(z)) \xrightarrow{d} N\big(0, 4\operatorname{Var}(g(z - |X_1|^\nu))\big).$$

If $\nu = 2$, one typically obtains a rate of convergence of order $\sqrt{n/\log n}$. More precisely, Schick and Wefelmeyer (2008) derived the following result.

Theorem 2 *Let $\nu = 2$. Suppose h is of bounded variation and $h(0+)$ and $g(z-)$ are positive. Let $b \sim \sqrt{\log n}/n$. Then*

$$\sqrt{\frac{n}{\log n}}\,\Big(\hat{p}_b(z) - p(z)\Big) \xrightarrow{d} N(0, h^2(0+)g(z-)).$$

If $\nu > 2$, the results of this paper imply the following rate. Faster rates are possible under additional smoothness assumptions on p at z.

Theorem 3 *Let $\nu > 2$. Suppose h is of bounded variation and $h(0+)$ and $g(z-)$ are positive. Let $b \sim 1/n$. Then*

$$\hat{p}_b(z) - p(z) = O_P(n^{-\beta}).$$

Even in the case $\nu \geq 2$, the Frees estimator can be $\sqrt{n}$-consistent. But this requires that $g(z-) = 0$. More precisely, we have the following result. The case $\nu = 2$ was already obtained in Schick and Wefelmeyer (2008).

Theorem 4 *Let $\nu \geq 2$. Suppose h is of bounded variation with $h(0+)$ positive and g vanishes in a neighborhood of z. Let $b \sim \sqrt{\log n}/n$. Then*

$$\sqrt{n}(\hat{p}_b(z) - p(z)) \xrightarrow{d} N\big(0, 4\operatorname{Var}(g(z - |X_1|^\nu))\big).$$

If g has compact support and z is not in the support of g, then g vanishes in a neighborhood of z, and the Frees estimator is $\sqrt{n}$-consistent for all ν.

Our paper is organized as follows. In Section 2 we describe our results. The analysis is based on the Hoeffding decomposition of our estimator and on various propositions which treat the terms in this decomposition in different scenarios. A synthesis of results is given in a series of theorems. The resulting convergence results are uniform in the bandwidth and allow us to replace in the above theorems $\hat{p}_b(z)$ by $\hat{p}_{\hat{s}b}(z)$ with a positive random variables $\hat{s} = \hat{s}_n$ satisfying $\hat{s}_n + 1/\hat{s}_n = O_P(1)$. Proofs are given in Section 3.

2 Results

The density p of $|X_1|^\nu + |X_2|^\nu$ is the convolution $g * g$ of the density g of $|X_1|^\nu$ with itself. Thus we have the following representation for p,

$$p(z) = \int_0^z \beta(z-y)^{\beta-1} h((z-y)^\beta)\beta y^{\beta-1} h(y^\beta)\, dy,$$

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

$$p(z) = \beta^2 z^{2\beta-1} \int_0^1 h(z^\beta(1-s)^\beta) h(z^\beta s^\beta)(1-s)^{\beta-1} s^{\beta-1}\, ds.$$

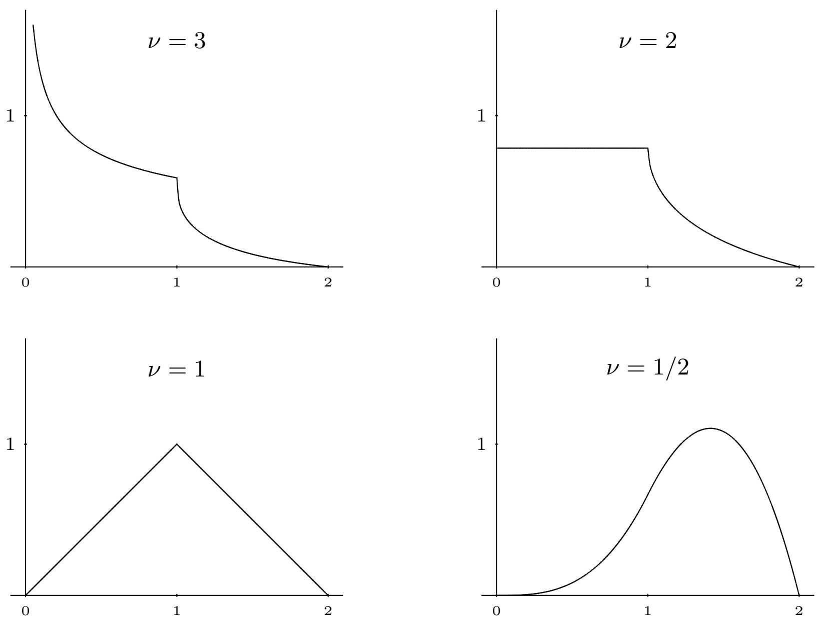
Of course, $p(z) = 0$ for negative z. Since the integrand is symmetric about $1/2$, we have

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

The representation shows that, for $\beta < 1/2$, the density p has a pole at 0 if h has a positive right-hand limit $h(0+)$ at 0. Indeed, then we have

$$\lim_{z \downarrow 0} z^{1-2\beta} p(z) = \beta^2 h^2(0+) \frac{\Gamma^2(\beta)}{\Gamma(2\beta)}.$$

Below are graphs of p for various values of ν when h is the uniform density on $(0, 1)$.



We study the behavior of the estimator $\hat{p}_b(z)$ at a fixed positive point z. Since $\hat{p}_b(z)$ is a U-statistic, we have the Hoeffding decomposition

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

where we write $A(\psi)$ for the centered average

$$A(\psi) = \frac{1}{n}\sum_{j=1}^{n}\big(\psi(z-|X_j|^\nu) - E[\psi(z-|X_j|^\nu)]\big)$$

and $U(\psi)$ for the degenerate U-statistic

$$U(\psi) = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\bar{\psi}(X_i, X_j)$$

with

$$\bar{\psi}(x,y) = \psi(z-|x|^\nu - |y|^\nu) - \psi * g(z-|x|^\nu) - \psi * g(z-|y|^\nu) + \psi * p(z).$$

Thus we need to treat the bias $p*k_b(z)-p(z)$, the average $A(g*k_b)$ and the degenerate U-statistic $U(k_b)$. To allow for random bandwidth we treat these expressions with b replaced by sb and with s running through a compact subinterval I of $(0,\infty)$. Let $C(I)$ denote the space of continuous functions on I endowed with the supremum norm. We first treat the degenerate U-statistic. A version of this result for the case $\nu = 2$ is already contained in Schick and Wefelmeyer (2008).

Proposition 1 *If p is bounded in a neighborhood of z, then the degenerate U-statistic process $\{n\sqrt{b}\,U(k_{sb}) : s \in I\}$ is tight in $C(I)$.*

The next three propositions deal with the average $A(g*k_{sb})$ under various assumptions. The case $\nu < 2$ is essentially known, see Schick and Wefelmeyer (2004), Giné and Mason (2007a) and Du and Schick (2008).

Proposition 2 *Let h have bounded variation. If $\nu < 2$ or g vanishes in a neighborhood of z, then*

$$\sup_{s\in I}\sqrt{n}|A(g*k_{sb}) - A(g)| = o_P(1)$$

and

$$\sqrt{n}A(g) \xrightarrow{d} N\big(0, \mathrm{Var}(g(z-|X_1|^\nu))\big).$$

The next result was proved in Schick and Wefelmeyer (2008).

Proposition 3 *Let $\nu = 2$. Let h have bounded variation and let $h(0+)$ and $g(z-)$ be positive. Suppose $\log(1/b)/\log n \to \gamma$ for some positive γ. Then*

$$\sup_{s\in I}\sqrt{\frac{n}{\log n}}\,\Big|A(g*k_{sb}) - A(g\mathbf{1}_{(b\log n,\infty)})\Big| = o_P(1)$$

and

$$\sqrt{\frac{n}{\log n}}\,A(g\mathbf{1}_{(b\log n,\infty)}) \xrightarrow{d} N(0, \gamma h^2(0+)g(z-)/4).$$

The following proposition treats $\nu > 2$ and is new.

Proposition 4 *Let $\nu > 2$. Let h have bounded variation and let $h(0+)$ and $g(z-)$ be positive. Then*

$$\sup_{s \in I} \sqrt{nb^{1-2\beta}} \, |A(g * k_{sb})| = O_P(1).$$

Finally we treat the bias term. For a positive α we let $\lfloor \alpha \rfloor$ denote the largest integer less than α. Recall that the function p is *α-smooth* at z if p is $\lfloor \alpha \rfloor$ times differentiable around z and if its $\lfloor \alpha \rfloor$-th derivative is Hölder at z with exponent $\alpha - \lfloor \alpha \rfloor$ in the sense that

$$\left| p^{(\lfloor \alpha \rfloor)}(z+t) - p^{(\lfloor \alpha \rfloor)}(z) \right| \le L|t|^{\alpha - \lfloor \alpha \rfloor}, \quad |t| < \delta,$$

for some $\delta > 0$ and some constant L. A Taylor expansion then gives the following well-known result.

Proposition 5 *If p is α-smooth at z for some positive α and the kernel also satisfies $\int x^i k(x)\,dx = 0$ for $i = 1, \ldots, \lfloor \alpha \rfloor$, then*

$$\sup_{s \in I} |p * k_{sb}(z) - p(z)| = O(b^\alpha).$$

The next lemma shows that p is α-smooth at z with exponent $\alpha = \min(1, \beta)$ if the density h is of bounded variation. The special case $\beta = 1/2$ was already obtained in Schick and Wefelmeyer (2008).

Lemma 1 *Suppose h is of bounded variation. Then p is Hölder at z with exponent $\alpha = \min(1, \beta)$.*

Higher Hölder exponents can be guaranteed under stronger assumptions on h.

Lemma 2 *Suppose h is of bounded variation and g vanishes in a neighborhood of z. Then p is Hölder at z with exponent 1.*

Lemma 3 *Suppose h is Lipschitz on $(0, \infty)$ in the sense that*

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

for some constant Λ. Then p is Hölder at z with exponent 1.

The above propositions and lemmas imply various asymptotic results for the Frees estimator $\hat{p}_{\hat{s}b}(z)$ with a random bandwidth $\hat{s}b$, where $\hat{s}$ is a positive random variable that changes with the sample size in such a way that

$$\hat{s} + \frac{1}{\hat{s}} = O_P(1).$$

We first consider the case when g vanishes in a neighborhood of z. This happens if z is outside the support of g. From Propositions 1, 2, 5 and Lemma 2 we obtain $\sqrt{n}$-consistency and asymptotic normality for all ν.

Theorem 5 *Suppose h is of bounded variation and g vanishes in a neighborhood of z. Let $nb^2 \to 0$ and $nb \to \infty$. Then*

$$\sqrt{n}\,(\hat{p}_{\hat{s}b}(z) - p(z)) \xrightarrow{d} N\big(0, 4\,\mathrm{Var}(g(z - |X_1|^\nu))\big).$$

Now we consider the case when both $h(0+)$ and $g(z-)$ are positive. The results for $\nu < 2$, $\nu = 2$ and $\nu > 2$ are then quite different. Under the minimal smoothness assumptions of Lemma 1 we have the following results.

Theorem 6 *Suppose h is of bounded variation. Let $h(0+)$ and $g(z-)$ be positive.*

1. *Let $\nu < 2$. Take b such that $nb \to \infty$ and $nb^{2\alpha} \to 0$, where $\alpha = \min(1, \beta)$, say $b \sim (n \log n)^{-1/(2\alpha)}$. Then*

$$\sqrt{n}\,(\hat{p}_{\hat{s}b}(z) - p(z)) \xrightarrow{d} N\big(0, 4\,\mathrm{Var}(g(z - |X_1|^\nu))\big).$$

2. *Let $\nu = 2$. Take $b \sim (\log n)^\tau / n$ for some $\tau < 1$. Then*

$$\sqrt{\frac{\log n}{n}}\,\Big(\hat{p}_{\hat{s}b}(z) - p(z)\Big) \xrightarrow{d} N(0, h^2(0+)g(z-)).$$

3. *Let $\nu > 2$. Take $b \sim 1/n$. Then*

$$n^\beta(\hat{p}_{\hat{s}b}(z) - p(z)) = O_P(1).$$

Under additional smoothness assumptions on p, better results are possible.

Theorem 7 *Suppose h is of bounded variation. Let $h(0+)$ and $g(z-)$ be positive. Suppose p is α-smooth at z for some $\alpha \geq \min(1, \beta)$, and the kernel k also fulfills $\int x^i k(x)\,dx = 0$ for $i = 1, \dots, \lfloor \alpha \rfloor$.*

1. *Let $\nu < 2$. Take b such that $nb \to \infty$ and $nb^{2\alpha} \to 0$, say $b \sim (n \log n)^{-1/(2\alpha)}$. Then*

$$\sqrt{n}\,(\hat{p}_{\hat{s}b}(z) - p(z)) \xrightarrow{d} N\big(0, 4\,\mathrm{Var}(g(z - |X_1|^\nu))\big).$$

2. *Let $\nu = 2$. Take $b \sim n^{-\gamma}$ with $1/(2\alpha) \leq \gamma \leq 1$. Then*

$$\sqrt{\frac{\log n}{n}}\,\Big(\hat{p}_{\hat{s}b}(z) - p(z)\Big) \xrightarrow{d} N(0, \gamma h^2(0+)g(z-)).$$

3. *Let $\nu > 2$. Take $b \sim n^{-\gamma}$ with $0 < \gamma \leq 1$. Then*

$$\hat{p}_{\hat{s}b}(z) - p(z) = O_P(b^\alpha) + O_P\Big(\frac{1}{\sqrt{nb^{1-2\beta}}}\Big).$$

   In particular, if $\gamma = 1/(1 + 2(\alpha - \beta))$ we obtain

$$\hat{p}_{\hat{s}b}(z) - p(z) = O_P(n^{-\alpha/(1+2(\alpha-\beta))}).$$

In the last two theorems, the conclusions for the case $\nu < 2$ remain valid without the requirement that $h(0+)$ and $g(z-)$ are positive.

For $\nu < 2$ the convergence rate of $\hat{p}_{\hat{s}b}(z)$ is $\sqrt{n}$ and is not affected by the choice of bandwidth within the admissible range. We recommend choosing a large bandwidth such as $(n \log n)^{-1/(2\alpha)}$.

For $\nu = 2$ the convergence rate of $\hat{p}_{\hat{s}b}(z)$ is $\sqrt{n/\log n}$ and does not depend on the bandwidth $b \sim n^{-\gamma}$. However, the asymptotic variance depends on the exponent γ and is minimized by $\gamma = 1/(2\alpha)$. Thus smoothness of p does not improve the rate of convergence, but allows for smaller asymptotic variances.

For $\nu > 2$ the estimator $\hat{p}_{\hat{s}b}(z)$ behaves like classical kernel density estimators: The bias is not negligible, and the rate is determined by the rates of the bias and variance terms. The optimal bandwidth is proportional to $n^{-1/(1+2(\alpha-\beta))}$. For the smallest admissible Hölder exponent $\alpha = \beta$, the optimal bandwidth is proportional to $1/n$ and one has

$$\hat{p}_{\hat{s}b}(z) - p(z) = O_P(n^{-\beta}).$$

Suppose we have independent observations $Z_1, \dots, Z_n$ from the density p itself. Then the kernel estimator

$$\tilde{p}_b(z) = \frac{1}{n} \sum_{j=1}^{n} k_b(z - Z_j)$$

based on these direct observations satisfies

$$\tilde{p}_b(z) - p(z) = O(b^\alpha) + O_P\Big(\frac{1}{\sqrt{nb}}\Big)$$

if p is α-smooth. The optimal bandwidth is $b \sim n^{-1/(1+2\alpha)}$, yielding

$$\tilde{p}_b(z) - p(z) = O_P(n^{-\alpha/(1+2\alpha)}).$$

Thus the estimator $\hat{p}_b(z)$ with optimal bandwidth has a faster rate of convergence than the kernel estimator $\tilde{p}_b(z)$ based on n observations from p.

3 Proofs

This section contains the proofs of Propositions 1, 2 and 4 and of Lemmas 1 to 3. We use $\|\psi\|$ and $\|\psi\|_q$ to denote the sup-norm and the L_q-norm, respectively, of a function ψ. By assumption (K) on the kernel k, there are constants M_1 and M_2 such that

$$\|k_{tb} - k_{sb}\|_1 = \|k_t - k_s\|_1 \le M_1|t - s|, \quad s, t \in I, \tag{3.1}$$

and

$$\sqrt{b}\|k_{tb} - k_{sb}\|_2 = \|k_t - k_s\|_2 \le M_2|t - s|, \quad s, t \in I. \tag{3.2}$$

For $c < d$ we set

$$\Gamma_z(c, d) = \sup_{c<y<d} g(z - y) \quad \text{and} \quad \Gamma(c, d) = \sup_{c<y<d} g(y).$$

Note that

$$\Gamma_z(c, d) = \Gamma(z - d, z - c).$$

In the proofs we shall use the fact that translation is continuous in L_2, i.e., for a square-integrable function $\bar{g}$ one has

$$\int (\bar{g}(x - v) - \bar{g}(x))^2 \, dx \to 0 \quad \text{as } v \to 0. \tag{3.3}$$

See Theorem 9.5 in Rudin (1987).

Proof of Proposition 1. Let $I = [l, r]$. For a bounded ψ, the summands in the degenerate U-statistic $U(\psi)$ are centered and uncorrelated. This follows from the fact that $E(\bar{\psi}(X_i, X_j)|X_i) = 0$ and $E(\bar{\psi}(X_i, X_j)|X_j) = 0$. Thus we have

$$E[U^2(\psi)] = \frac{2}{n(n-1)} E[\bar{\psi}^2(X_1, X_2)].$$

Straightforward calculations show that

$$E[\bar{\psi}^2(X_1, X_2)] \le E[\psi^2(z - |X_1|^\nu - |X_2|^\nu)] = \psi^2 * p(z).$$

If ψ has support in a compact set S, then

$$\psi^2 * p(z) \le \sup_{y \in S} p(z - y) \|\psi\|_2^2.$$

We will apply this inequality with $\psi = k_{sb}$ and $\psi = k_{tb} - k_{sb}$ for s and t in I. For such s and t, these functions have support contained in $S = [lb, rb]$. Note that the density p is bounded in a neighborhood of z and that $U(k_{tb}) - U(k_{sb})$ equals $U(k_{tb} - k_{sb})$. These properties and (3.2) yield

$$n^2 b E[U^2(k_{sb})] = O(1) \quad \text{and} \quad n^2 b E[(U(k_{tb}) - U(k_{sb}))^2] \le B|t - s|^2$$

for all s and t in I and some constant B. Thus tightness in $C(I)$ of the sequence $\{n\sqrt{b}\, U(k_{sb}) : s \in I\}$ follows from Theorem 12.3 in Billingsley (1968). □

Proof of Proposition 2. We shall show that

$$E[g^2(z - |X_1|^\nu)] = g^2 * g(z) = \int_0^z g(z - y) g^2(y)\, dy < \infty, \tag{3.4}$$

$$nE[(A(g * k_{sb}) - A(g))^2] \to 0, \tag{3.5}$$

$$nE[(A(g * k_{tb}) - A(g * k_{sb}))^2] \le B|t - s|^2, \tag{3.6}$$

for all s and t in I. The second conclusion of the proposition then follows from (3.4) and the central limit theorem. The first conclusion is a consequence of (3.5), (3.6) and Theorem 12.3 in Billingsley (1968).

Let us first consider the case when $\nu \ge 2$ and g vanishes in a neighborhood $(z - 2\delta, z + 2\delta)$ of z for some $\delta > 0$. Then the integral in (3.4) equals

$$\int_{2\delta}^z g(z - y) g^2(y)\, dy \le \Gamma(2\delta, z) \int_{2\delta}^z g(z - y) g(y)\, dy \le \beta \delta^{\beta - 1} \|h\| p(z).$$

This proves (3.4). Since $A(g * k_{sb}) - A(g) = A(g * k_{sb} - g)$ and

$$nE[A^2(\psi)] \le E[\psi^2(z - |X_1|^\nu)] = \psi^2 * g(z), \tag{3.7}$$

we can bound the left-hand side of (3.5) by

$$\begin{aligned}(g * k_{sb} - g)^2 * g(z) &= \int g(z - y) \Big(\int (g(y - sbu) - g(y)) k(u)\, du \Big)^2 dy \\ &\le \|k\|_1 \int g(z - y) \int (g(y - sbu) - g(y))^2 |k(u)|\, du\, dy.\end{aligned}$$

Thus (3.5) follows if we show that

$$\int g(z-y)(g(y-v)-g(y))^2\,dy \to 0 \quad \text{as } v \to 0. \tag{3.8}$$

Since g vanishes on the interval $(z-2\delta, z+2\delta)$, the range of integration can be restricted to $(2\delta, z-\delta)$ for $|v|<\delta$, and for such v the integral in (3.8) is bounded by

$$\begin{aligned} &\Gamma_z(2\delta, z-\delta)\int_{2\delta}^{z-\delta}(g(y-v)-g(y))^2\,dy \\ &\qquad \le \beta\delta^{\beta-1}\|h\|\int(\bar g(y-v)-\bar g(y))^2\,dy \end{aligned}$$

where $\bar g = g\mathbf{1}_{(\delta,z)}$. Since $\bar g$ is square-integrable, we obtain (3.8) from (3.3).

It follows from (3.4) and (3.8) that $g*g^2(z-v)\to g*g^2(z)$ as $v\to 0$. This implies that there is a constant M such that

$$g*g^2*|k_{tb}-k_{sb}|(z) \le M\|k_{sb}-k_{tb}\|_1, \quad s,t\in I.$$

We can bound the left-hand side of (3.6) by $(g*(k_{tb}-k_{sb}))^2*g(z)$. Using the inequality $(u*w)^2 \le \|w\|_1(u^2*|w|)$, which follows from the Cauchy–Schwarz inequality, we have

$$\begin{aligned} (g*(k_{tb}-k_{sb}))^2*g(z) &\le \|k_{tb}-k_{sb}\|_1 g*g^2*|k_{tb}-k_{sb}|(z) \\ &\le M\|k_{tb}-k_{sb}\|_1^2 \end{aligned}$$

for s and t in I. This bound and (3.1) imply (3.6).

Let us now consider the case $\nu<2$. Inspecting the above arguments we see that the desired (3.4) to (3.6) follow if we verify (3.4) and (3.8) in the present case. Note that now $\bar g = g\mathbf{1}_{(0,z)}$ is square-integrable. The integral in (3.4) is bounded by

$$\Gamma_z(0,z_*)\int_0^{z_*}g^2(y)\,dy + \Gamma(z_*,z)\int_{z_*}^{z}g(z-y)g(y)\,dy$$

with $z_* = z/2$. This proves (3.4). We have

$$\sup_{|v|<\delta}\ \sup_{z-\delta<y<z}|g(y-v)-g(y)| \le 2\Gamma(z-2\delta, z+\delta).$$

For $|v|<\delta$, we can bound the integral in (3.8) by the sum $S(v)+S$ of

$$\begin{aligned} S(v) &= \Gamma_z(-\delta, z-\delta)\int_{-\delta}^{z-\delta}(g(y-v)-g(y))^2\,dy \\ &\le \Gamma(\delta, z+\delta)\int(\bar g(y-v)-\bar g(y))^2\,dy \end{aligned}$$

and

$$\begin{aligned} S &= (2\Gamma(z-2\delta,z+\delta))^2\int_{z-\delta}^{z}g(z-y)\,dy \\ &\le (2\Gamma(z-2\delta,z+\delta))^2\|h\|\delta^\beta. \end{aligned}$$

Since $\bar g$ is square-integrable, we have $S(v)\to 0$ as $v\to 0$ by (3.3). This yields (3.8). □

Proof of Proposition 4. In view of Theorem 12.3 in Billingsley (1968) it suffices to show

$$nb^{1-2\beta}E[A^2(g*k_{sb})] = O(1) \tag{3.9}$$

and

$$nb^{1-2\beta}E[(A(g*k_{tb}) - A(g*k_{sb}))^2] \le B|t-s|^2 \tag{3.10}$$

for s and t in $I = [l, r]$. For this we show that there is a constant M such that

$$nE[A^2(g*\Delta_b)] \le b^{2\beta-1}M(\|\Delta\|_1 + \|\Delta\|_2)^2 \tag{3.11}$$

whenever $\Delta_b(x) = \Delta(x/b)/b$ for some bounded function Δ with support in $[-r, r]$. In view of (3.1) and (3.2), an application of this bound with $\Delta = k_s$ yields (3.9), and an application with $\Delta = k_t - k_s$ yields (3.10).

In view of (3.7), the left-hand side of (3.11) is bounded by

$$g*(g*\Delta_b)^2(z) = \int g(z-y)(g*\Delta_b(y))^2\,dy. \tag{3.12}$$

Since Δ_b has support contained in $[-rb, rb]$, the function $g*\Delta_b$ vanishes on $(-\infty, -rb]$. For $-rb < y < 2rb$ we use the bounds

$$\begin{aligned}(g*\Delta_b)^2(y) &= \Big(\int_{-rb}^{rb} g(y-x)\Delta_b(x)\,dx\Big)^2\\ &\le \int_{-rb}^{rb} g(y-x)\,dx \int_{-rb}^{rb} g(y-x)\Delta_b^2(x)\,dx\end{aligned}$$

and

$$\int_{-rb}^{rb} g(y-x)\,dx \le \int_0^{3rb} g(u)\,du \le \|h\|(3rb)^\beta$$

and obtain

$$\begin{aligned}J_1 &= \int_{-rb}^{2rb} g(z-y)(g*\Delta_b(y))^2\,dy\\ &\le \Gamma_z(-rb, 2rb)\|h\|(3rb)^\beta \int_{-rb}^{2rb}\int_{-rb}^{rb} g(y-x)\Delta_b^2(x)\,dx\,dy\\ &\le \Gamma_z(-rb, 2rb)\|h\|(3rb)^\beta \int_0^{3rb} g(u)\,du \int \Delta_b^2(x)\,dx\\ &\le \Gamma_z(-rb, 2rb)(\|h\|(3r)^\beta)^2 b^{2\beta-1}\|\Delta\|_2^2.\end{aligned}$$

For $y > 2rb$ we use the bound

$$(g*\Delta_b)^2(y) \le \|\Delta_b\|_1 \int g^2(y-x)|\Delta_b(x)|\,dx,$$

and for $|v| < r$ and $z_* = z/2$ we use the bounds

$$\begin{aligned}\int_{rb}^{z_*} g(z-u-bv)g^2(u)\,du &\le \Gamma_z(0, z_*+rb)\|h\|^2 \int_{rb}^{z_*} \beta^2 u^{2\beta-2}\,du\\ &\le \Gamma_z(0, z_*+rb)\|h\|^2\beta^2(rb)^{2\beta-1}/(1-2\beta)\end{aligned}$$

and

$$\int_{z_*}^{z+rb} g(z-u-bv)g^2(u)\,du \le \Gamma^2(z_*, z+rb)$$

and obtain

$$\begin{aligned} J_2 &= \int_{2rb}^{z} g(z-y)(g * \Delta_b(y))^2\,dy \\ &\le \|\Delta\|_1 \int_{2rb}^{z} g(z-y) \int g^2(y-bv)|\Delta(v)|\,dv \\ &\le \|\Delta\|_1 \int \int_{rb}^{z+rb} g(z-u-bv)g^2(u)\,du|\Delta(v)|\,dv \\ &\le \|\Delta\|_1^2 \Big(\Gamma^2(z_*, z+rb) + \Gamma_z(0, z_*+rb)\|h\|^2\beta^2(rb)^{2\beta-1}/(1-2\beta)\Big). \end{aligned}$$

The bounds for J_1 and J_2 yield the desired (3.11). □

In the proof of the lemmas we repeatedly use the inequalities

$$(s+t)^\gamma \le s^\gamma + t^\gamma \quad \text{and} \quad t^\gamma - s^\gamma \le (t-s)^\gamma, \tag{3.13}$$

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

$$w(s) = \beta^2(s(1-s))^{\beta-1}\mathbf{1}[0 < s < 1].$$

Recall that $\alpha = \min(1, \beta)$. It is easy to check that, for some $c > 0$,

$$\int_u^v w(s)\,ds \le c(v-u)^\alpha, \quad 0 < u < v < 1. \tag{3.14}$$

In view of the representation (2.1) we can write $p(z) = 2z^{2\beta-1}q(z)$ with

$$q(z) = \int_0^{1/2} h(z^\beta(1-s)^\beta)h(z^\beta s^\beta)w(s)\,ds.$$

Thus the Hölder properties of p follow from those of q.

Proof of Lemma 1. Since h is of bounded variation, we may assume that $h(y) = \int \mathbf{1}[0 \le t \le y]\varphi(dt)$, where φ is the difference $\varphi_1 - \varphi_2$ of two finite measures. Write $\mu = \varphi_1 + \varphi_2$, and for $0 \le u \le v < \infty$ set

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

Then

$$|h(v) - h(u)| \le \int_u^v \mu(dt). \tag{3.15}$$

We can write

$$\begin{aligned} |q(z_2) - q(z_1)| \le \|h\| &\int_0^{1/2} |h(z_1^\beta s^\beta) - h(z_2^\beta s^\beta)|w(s)\,ds \\ &+ \|h\| \int_0^{1/2} \Big|h(z_1^\beta(1-s)^\beta) - h(z_2^\beta(1-s)^\beta)\Big| w(s)\,ds. \end{aligned}$$

Using the substitution $u = 1 - s$ and the identity $w(s) = w(1 - s)$, we see that

$$|q(z_2) - q(z_1)| \le \|h\| \int_0^1 |h(z_1^\beta s^\beta) - h(z_2^\beta s^\beta)| w(s)\, ds \tag{3.16}$$

for $0 < z_1 < z_2$. Using (3.15) and (3.14), we can bound the integral on the right-hand side by

$$\begin{aligned} \int \int_{(z_1 s)^\beta}^{(z_2 s)^\beta} \mu(dt)\, w(s)\, ds &= \int_0^{z_2^\beta} \int_{t^\nu/z_2}^{t^\nu/z_1} w(s)\, ds\, \mu(dt) \\ &\le \int_0^{z_2^\beta} c\Big(\frac{t^\nu}{z_1} - \frac{t^\nu}{z_2}\Big)^\alpha \mu(dt) \end{aligned}$$

and thus obtain

$$|q(z_2) - q(z_1)| \le c\|h\| z_1^{-\alpha} (z_2 - z_1)^\alpha \mu[0, z_2^\beta]$$

for $0 < z_1 < z_2$. Since μ is a finite measure, this yields the desired result. □

Proof of Lemma 2. In view of Lemma 1 we may assume that $\beta < 1$. By assumption, there is a positive $\delta < z/4$ such that $h(t) = 0$ for all $|t - z^\beta| < 2\delta^\beta$. Let $|y^\beta - z^\beta| < \delta^\beta$ and $\eta = \delta/(z^\beta + \delta^\beta)^\nu$. For $0 < s < \eta$, we have $|y^\beta(1-s)^\beta - y^\beta| \le y^\beta s^\beta < \delta^\beta$ by (3.13). Thus the integrand in $q(y)$ is zero for such s and we can write

$$q(y) = \int_\eta^{1/2} h(y^\beta (1-s)^\beta) h(y^\beta s^\beta) w(s)\, ds.$$

Let $z^\beta - \delta^\beta < z_1^\beta < z_2^\beta < z^\beta + \delta^\beta$. From (3.16) we obtain

$$|q(z_1) - q(z_2)| \le \|h\| \beta^2 (\eta/2)^{\beta-1} \int_0^1 |h(z_1^\beta s^\beta) - h(z_2^\beta s^\beta)|\, ds.$$

With the help of (3.15) we find

$$\begin{aligned} \int_0^1 |h(z_1^\beta s^\beta) - h(z_2^\beta s^\beta)|\, ds &\le \int_0^1 \int_{z_1^\beta s^\beta}^{z_2^\beta s^\beta} \mu(dt)\, ds \\ &\le \int_0^{z_2^\beta} \Big(\frac{t^\nu}{z_1} - \frac{t^\nu}{z_2}\Big) \mu(dt) \\ &\le \frac{z_2 - z_1}{z_1} \mu(\mathbb{R}). \end{aligned}$$

This completes the proof. □

Proof of Lemma 3. From (3.16) we obtain

$$|q(z_2) - q(z_1)| \le \|h\| \Lambda |z_2^\beta - z_1^\beta| \int_0^1 s^\beta w(s)\, ds$$

for $0 < z_1 < z_2$. This is the desired result. □

References

1. Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.
2. 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.
3. Frees, E. W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.* **89**, 517–525.
4. Giné, E. and Mason, D. (2007a). On local U-statistic processes and the estimation of densities of functions of several variables. *Ann. Statist.* **35**, 1104–1145.
5. Giné, E. and Mason, D. (2007b). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.* **20**, 457–485.
6. Rudin, W. (1987). *Real and Complex Analysis*, 3rd ed. McGraw-Hill, New York.
7. Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canad. J. Statist.* **28**, 799–815.
8. Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.* **16**, 925–935.
9. 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.
10. Schick, A. and Wefelmeyer, W. (2008). Non-standard behavior of density estimators for sums of squared observations. Technical Report, Department of Mathematical Sciences, Binghamton University.