The Annals of Statistics
1995, Vol. 23, No. 4, 1198–1222

ASYMPTOTICS FOR THE TRANSFORMATION KERNEL DENSITY ESTIMATOR

BY OLA HÖSSJER[1] AND DAVID RUPPERT[2]

Lund Institute of Technology and Cornell University

An asymptotic expansion is provided for the transformation kernel density estimator introduced by Ruppert and Cline. Let h_k be the bandwidth used in the kth iteration, $k = 1, 2, \ldots, t$. If all bandwidths are of the same order, the leading bias term of the lth derivative of the tth iterate of the density estimator has the form $\bar{b}_t^{(l)}(x) \prod_{k=1}^{t} h_k^2$, where the bias factor $\bar{b}_t(x)$ depends on the second moment of the kernel K, as well as on all derivatives of the density f up to order $2t$. In particular, the leading bias term is of the same order as when using an ordinary kernel density estimator with a kernel of order $2t$. The leading stochastic term involves a kernel of order $2t$ that depends on K, h_1 and $h_k/f(x)$, $k = 2, \ldots, t$.

1. Introduction. Suppose that we have an independent sample $X_1, \ldots, X_n$ from a density f. An extensively studied method of estimating f is the kernel density estimator (KDE)

$$\hat{f}(x; h) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where K is a symmetric kernel function that integrates to 1. The bias at x has a formal asymptotic expansion

$$\sum_{j=1}^{\infty} \frac{f^{(2j)}(x) \mu_{2j}(K)}{(2j)!} h^{2j}, \tag{1.1}$$

where $\mu_j(K) = \int u^j K(u)\, du$. One way of reducing the bias is to choose a kth order kernel, that is a kernel with the property $\mu_1(K) = \cdots = \mu_{k-1}(K) = 0$ and $\mu_k(K) \neq 0$, where k is an even positive integer. In this way the first $(k-1)/2$ terms in the expansion (1.1) vanish [Parzen (1962); Bartlett (1963); Singh (1977, 1979)].

A second approach is to let the bandwidth h depend on X_i, using a pilot estimate of f [cf. Abramson (1982, 1984), Silverman (1986), Jones (1990), Hall and Marron (1988) and Hall (1990)].

A third approach was introduced recently by Ruppert and Cline (1994) (henceforth denoted RC). They introduced the transformation kernel density

Received April 1993; revised September 1994.
[1] Supported by the Swedish Natural Science Research Council, contract F-DP 6689-300.
[2] Partially supported by NSF Grant DMS-90-02791.
AMS 1991 *subject classifications.* Primary 62G07; secondary 62G20.
Key words and phrases. Bias reduction, higher order kernels, smoothed empirical distribution, transformation to uniform distribution, variable bandwidths.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

estimator (TKDE), defined in the following way: Let g be a smooth monotone function. Transform the data to $Y_i = g(X_i)$, $i = 1, \dots, n$. Estimate the transformed density

$$f_Y(y; g) = f(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}$$

with a KDE

$$\hat{f}_Y(y; g, h) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h}\right),$$

and transform back by a change of variables:

$$\hat{f}(x; g, h) = \hat{f}_Y(g(x); g, h) g'(x). \tag{1.2}$$

The estimate $\hat{f}(x; g, h)$ defined in this way is the TKDE. In order for $\hat{f}(x; g, h)$ to have a small bias, it is crucial that g is chosen so that $\hat{f}_Y(y; g, h)$ has a small bias as an estimate of $f_Y(y; g)$. Ideally, $f_Y(y; g)$ becomes the uniform density on $(0, 1)$ if we choose $g = F$, the cumulative distribution function of $\{X_i\}$. According to (1.1) the bias of $\hat{f}_Y(y; g, h)$ then vanishes. Since F is unknown, this is of course not possible, but we can instead choose $g(x) = \hat{F}(x; h_1)$, where $\hat{F}(x; h_1)$ is the indefinite integral of a KDE $\hat{f}_1(x) = \hat{f}(x; h_1)$, so that g is close to F. It was shown in RC that

$$\hat{f}\left(x; \hat{F}(\cdot; h_1), h_2\right) - f(x) := \hat{f}_2(x) - f(x) = O_p(n^{-4/9}), \tag{1.3}$$

if $h_i = c_i n^{-1/9}$ and $c_i > 0$, $i = 1, 2$.

We may also iterate this process as follows: Let

$$\hat{f}_t(x) = \hat{f}\left(x; \hat{F}_{t-1}(\cdot), h_t\right), \qquad t = 2, 3, \dots, \tag{1.4}$$

where $\hat{F}_{t-1}$ is the indefinite integral of $\hat{f}_{t-1}$. As a generalization of (1.3), under certain smoothness conditions on f, it was established in RC that the lth derivative $\hat{f}_t^{(l)}(x)$ of $\hat{f}_t(x)$ satisfies

$$\hat{f}_t^{(l)}(x) - f^{(l)}(x) = O_p(n^{-(2t-l)/(4t+1)}), \qquad l = 1, \dots, t, \tag{1.5}$$

when $h_i = c_i n^{-1/(4t+1)}$ and $c_i > 0$, $i = 1, \dots, t$. The convergence in (1.5) is the same as when using a KDE with a $2t$th order kernel; see Singh (1979).

The purpose of this paper is to provide exact expansions for the RHS of formula (1.5) for any values of t and l. It turns out (Theorem 3.1 and Remarks 4.2 and 4.3) that the leading terms in the RHS of (1.5) are the lth derivative of

$$\bar{b}_t(x) \prod_{k=1}^{t} h_k^2 + n^{-1} \sum_{i=1}^{n} \left(\check{K}_{tx}(x - X_i) - E\check{K}_{tx}(x - X)\right), \tag{1.6}$$

where $\check{K}_{tx}$ is a $2t$th order kernel, the form of which depends on K, h_1 and $h_k/f(x)$, $k = 2, \dots, t$. Since it depends on the unknown $f(x)$, it is not computable. If K is supported on $[-1, 1]$, $\check{K}_{tx}$ is supported on $[-h_1 - \Sigma_2^t h_k/f(x), h_1 + \Sigma_2^t h_k/f(x)]$. The bias factor $\bar{b}_t(x)$ depends on $\mu_2(K)$ and $f^{(j)}(x)$, $j = 0, \dots, 2t$.

As a comparison, consider the asymptotic expansion of an ordinary KDE $\hat{f}(x;h)$ with a $2t$th order kernel [cf., e.g., Prakasa Rao (1983), Section 2.1]. It has leading bias the tth term in (1.1), which only depends on the unknown density in terms of $f^{(2t)}(x)$. On the other hand, the bias term in (1.6) depends on *all* derivatives up to order $2t$. The main stochastic term of $\hat{f}(x;h)$ is found by replacing $\check{K}_{tx}(\cdot)$ with $K(\cdot/h)/h$ in (1.6).

The paper is organized as follows: The regularity conditions on f, K and the bandwidths are formulated in Section 2 together with some notation. In Section 3 we establish the basic result of this paper: the expansion (1.6). Some remarks concerning our results are given in Section 4 and, finally, the more technical parts of the proofs are collected in the Appendix.

2. Regularity conditions and some notation. Let $t > 0$ and $l \geq 0$ be integers. The following assumptions will be used throughout the paper:

ASSUMPTION 1. $X_1, \dots, X_n$ is an i.i.d. sample with common density f, where $f(x) > 0$ and $f^{(2t+l)}$ is continuous in a neighbourhood of x.

ASSUMPTION 2. The function K is a nonnegative and symmetric kernel that integrates to 1 and is supported on $[-1, 1]$, and $K^{(2t+l-1)}$ is continuous.

ASSUMPTION 3. The bandwidths $h_1, \dots, h_t$ do not depend on x, $h_1 \to 0$ and $nh_1 \to \infty$ as $n \to \infty$ and, finally,

$$0 < \inf_n \frac{h_k(n)}{h_1(n)} \leq \sup_n \frac{h_k(n)}{h_1(n)} < \infty, \qquad k = 2, \dots, t. \tag{2.1}$$

The dependence of $h_1, \dots, h_t$ on n will be suppressed in the notation. The case when the bandwidths are stochastic (via some plug-in rule) and depend on x are treated in Hössjer and Ruppert (1993b). Let $I(x, r)$ denote the closed real interval $[x - r, x + r]$. Introduce the zero mean stochastic processes

$$W_n(x; \tilde{K}, h) = (nh)^{-1/2} \sum_{i=1}^{n} \left(\tilde{K}\left(\frac{x - X_i}{h} \right) - E\tilde{K}\left(\frac{x - X}{h} \right) \right) \tag{2.2}$$

and

$$\begin{aligned} \tilde{W}_n(x; g, \tilde{K}, h) = (nh)^{-1/2} \sum_{i=1}^{n} \Bigg(& \tilde{K}\left(\frac{g(x) - g(X_i)}{h} \right) \\ & - E\tilde{K}\left(\frac{g(x) - g(X)}{h} \right) \Bigg), \end{aligned} \tag{2.3}$$

for any function $\tilde{K}$. Put $\alpha_1(x) \equiv 1$ and

$$\alpha_k(x) = \frac{h_k}{f(x)h_1}, \qquad k = 2, \dots, t. \tag{2.4}$$

Note that $\alpha_k(x)$ may depend on n, but this will not be made explicit in the notation. Let $K_h(\cdot) = K(\cdot/h)/h$. For any $\alpha > 0$, we introduce the operator A_α: $C_0(\mathbb{R}) \to C_0(\mathbb{R})$ by means of

$$A_\alpha(\tilde{K}) = K_\alpha + \tilde{K} - K_\alpha * \tilde{K}, \tag{2.5}$$

with $*$ denoting the convolution operator and $C_0(\mathbb{R})$ the space of continuous and compactly supported functions on the real line. Define $\overline{K}_{1x}, \dots, \overline{K}_{tx}$ recursively through $\overline{K}_{1x} = K$ and

$$\overline{K}_{kx} = A_{\alpha_k(x)}\big(\overline{K}_{k-1,x}\big), \qquad k = 2, \dots, t. \tag{2.6}$$

Note that $\overline{K}_{kx}$ depends on x through the unknown $f(x)$ for $k = 2, \dots, t$, and it is supported on $[-\Sigma_1^k \alpha_j(x), \Sigma_1^k \alpha_j(x)]$ (since K is supported on $[-1, 1]$). We will also make use of the linear differential operator defined by

$$B(\tilde{f}) = \tilde{f}^{(2)} - \frac{f^{(2)}\tilde{f}}{f} - \frac{3f'\tilde{f}'}{f} + \frac{3(f')^2\tilde{f}}{f^2}, \tag{2.7}$$

for any function $\tilde{f}$ that is twice continuously differentiable in a neighbourhood of x. Here f denotes the density of X_i, so $B(\tilde{f})$ is well defined because of Assumption 1. Finally, the L_p-norm, $1 \le p \le \infty$, of a stochastic variable X will be denoted $\|X\|_p$.

3. The main result. We are now ready to formulate the main result. We will give a pointwise asymptotic expansion of $\hat{f}_t^{(l)}$. An asymptotic result for the error process $\hat{f}_t(x) - f(x)$ as a function of x requires different techniques and will be investigated in another paper.

THEOREM 3.1. *The lth derivative of the tth iterate TKDE has the asymptotic expansion*

$$\begin{aligned} \hat{f}_t^{(l)}(x) &= f^{(l)}(x) + b_t^{(l)}(x)h_1^{2t} \\ &\quad + (nh_1)^{-1/2} \frac{d^l}{dx^l} W_n\big(x; \overline{K}_{tx}, h_1\big) + R_t^{(l)}(x). \end{aligned} \tag{3.1}$$

The bias factor $b_t(x)$ is defined through the recursive scheme $b_1(x) = \mu_2(K)f^{(2)}(x)/2$ and

$$b_{k+1}(x) = -\frac{\mu_2(K)}{2}\alpha_{k+1}(x)^2 B(b_k)(x), \qquad k = 1, 2, \dots, t-1. \tag{3.2}$$

The remainder term may be decomposed as $R_t(x) = R_{t1}(x) + R_{t2}(x)$. The first term $R_{t1}(x)$ is nonstochastic with

$$\sup_{x' \in I(x, C_1 h_1)} \big|R_{t1}^{(j)}(x')\big| = o\big(h_1^{2t}\big), \qquad j = 0, 1, \dots, l, \tag{3.3}$$

and the second term satisfies

$$\sup_{x' \in I(x, C_1 h_1)} \left| R_{t2}^{(j)}(x') \right| = o_p\left((nh_1)^{-1/2} h_1^{-j} \right), \qquad j = 0, 1, \ldots, l, \tag{3.4}$$

for any constant $C_1 > 0$.

In order to prove Theorem 3.1, we need the following lemma, which is the main technical result of this paper. Given an asymptotic expansion of an estimate $\hat{f}(x)$ of $f(x)$, it provides an asymptotic expansion of $\tilde{f}(x) = \hat{f}(x; \hat{F}, h)$, where $\hat{F}$ is the indefinite integral of $\hat{f}$. The proof of this lemma is given in the Appendix.

LEMMA 3.2. *Suppose that the density estimate* $\hat{f}$ *is nonnegative with asymptotic expansion*

$$\hat{f}(x) = f(x) + \hat{b}(x) h_1^{2k} + \hat{R}_1(x) + \hat{R}_2(x) + (nh_1)^{-1/2} W_n\left(x; \hat{K}_x, h_1\right),$$

where $\hat{K}_x$ *is a kernel (possibly depending on x) and* $\hat{b}$ *and* $\hat{R}_1$ *are nonstochastic functions of x. Let J be a fixed nonnegative integer* $(\le 2t + l - 4)$ *and* C_1 *an arbitrary positive constant. Suppose there exists an integer N such that*

$$\bigcup_{\substack{x' \in I(x, C_1 h_1) \\ n \ge N}} \operatorname{supp}\left(\hat{K}_{x'} \right) \quad \text{is bounded} \tag{3.5}$$

and a constant $C > 0$ *such that for any fixed* $C_2 > C$,

$$\sup_{\substack{x' \in I(x, C_1 h_1) \\ |y - x'| \le C_2 h_1}} \left| \frac{d^j}{dx'^j} \hat{K}_x\left(\frac{y - x'}{h_1} \right) \right| = O(h_1^{-j}), \qquad j = 0, 1, \ldots, J + 3, \tag{3.6}$$

and

$$\sup_{\substack{x' \in I(x, C_1 h_1) \\ |y - x'| \le C_2 h_1}} \left| \frac{d^j}{dx'^j} \left(\hat{K}_{x'}\left(\frac{y - x'}{h_1} \right) - \hat{K}_x\left(\frac{y - x'}{h_1} \right) \right) \right| = o\left(h_1^{-j} \right), \tag{3.7}$$

$$j = 0, 1, \ldots, J + 3.$$

Assume also that

$$\sup_{x' \in I(x, C_1 h_1)} \left| \hat{b}^{(j)}(x') \right| = O(1), \qquad j = 0, 1, \ldots, J + 2, \tag{3.8}$$

$$\sup_{x' \in I(x, C_1 h_1)} \left| \hat{R}_1^{(j)}(x') \right| = o\left(h_1^{2k} \right), \qquad j = 0, 1, \ldots, J + 2 \tag{3.9}$$

and

$$\sup_{x' \in I(x, C_1 h_1)} \left| \hat{R}_2^{(j)}(x') \right| = o_p((nh_1)^{-1/2} h_1^{-j}), \qquad j = 0, 1, \ldots, J + 2. \tag{3.10}$$

Define now $\tilde{f}(x) = \hat{f}(x; \hat{F}, h)$, *with*

$$0 < \inf_n \frac{h(n)}{h_1(n)} \le \sup_n \frac{h(n)}{h_1(n)} < \infty. \tag{3.11}$$

Then $\tilde{f}(x)$ *has an asymptotic expansion*

$$\begin{aligned} \tilde{f}(x) &= f(x) + \tilde{b}(x)h_1^{2k+2} + \tilde{R}_1(x) + \tilde{R}_2(x) \\ &\quad + (nh_1)^{-1/2} W_n\big(x; \tilde{K}_x, h_1\big), \end{aligned} \tag{3.12}$$

where $\tilde{b}(x) = -\mu_2(K)\alpha(x)^2 B(\hat{b})(x)/2$, $\tilde{K}_x = A_{\alpha(x)}(\hat{K}_x)$ *and* $\alpha(x) = h/(h_1 f(x))$. *Furthermore,* $\tilde{b}$ *and* $\tilde{R}_1$ *are nonstochastic with*

$$\sup_{x' \in I(x, C_1 h_1)} \big|\tilde{b}^{(j)}(x')\big| = O(1), \qquad j = 0, 1, \dots, J, \tag{3.13}$$

$$\sup_{x' \in I(x, C_1 h_1)} \big|\tilde{R}_1^{(j)}(x')\big| = o\big(h_1^{2k+2}\big), \qquad j = 0, 1, \dots, J \tag{3.14}$$

and

$$\sup_{x' \in I(x, C_1 h_1)} \big|\tilde{R}_2^{(j)}(x')\big| = o_p\big((nh_1)^{-1/2} h_1^{-j}\big), \qquad j = 0, 1, \dots, J. \tag{3.15}$$

Finally, $\tilde{K}_x$ *satisfies*

$$\bigcup_{\substack{x' \in I(x, C_1 h_1) \\ n \ge N}} \operatorname{supp}\big(\tilde{K}_{x'}\big) \quad \textit{is bounded}, \tag{3.16}$$

$$\sup_{\substack{x' \in I(x, C_1 h_1) \\ |y - x'| \le C_2 h_1}} \left| \frac{d^j}{dx'^j} \tilde{K}_x\left(\frac{y - x'}{h_1}\right)\right| = O(h_1^{-j}), \qquad j = 0, 1, \dots, J+1 \tag{3.17}$$

and

$$\sup_{\substack{x' \in I(x, C_1 h_1) \\ |y - x'| \le C_2 h_1}} \left| \frac{d^j}{dx'^j} \left(\tilde{K}_{x'}\left(\frac{y - x'}{h_1}\right) - \tilde{K}_x\left(\frac{y - x'}{h_1}\right)\right)\right| = o\big(h_1^{-j}\big), \tag{3.18}$$

$$j = 0, 1, \dots, J+1.$$

PROOF OF THEOREM 3.1. For $k = 1, \dots, t$, we will show (using induction w.r.t. k) that

$$\begin{aligned} \hat{f}_k(x) &= f(x) + b_k(x)h_1^{2k} + (nh_1)^{-1/2} W_n\big(x; \overline{K}_{kx}, h_1\big) \\ &\quad + R_{k1}(x) + R_{k2}(x), \end{aligned} \tag{3.19}$$

where b_k is defined recursively in (3.2) and for any constant $C_1 > 0$,

$$\sup_{x' \in I(x, C_1 h_1)} \big|b_k^{(j)}(x')\big| = O(1), \qquad j = 0, 1, \dots, 2(t-k) + l, \tag{3.20}$$

R_{k1} is nonstochastic with

$$\sup_{x' \in I(x, C_1 h_1)} \left| R_{k1}^{(j)}(x') \right| = o(h_1^{2k}), \qquad j = 0, 1, \dots, 2(t-k) + l, \tag{3.21}$$

$$\sup_{x' \in I(x, C_1 h_1)} \left| R_{k2}^{(j)}(x') \right| = o_p\big((nh_1)^{-1/2} h_1^{-j}\big), \qquad j = 0, 1, \dots, 2(t-k) + l, \tag{3.22}$$

for some integer $N > 0$,

$$\bigcup_{\substack{x' \in I(x, C_1 h_1) \\ n \geq N}} \operatorname{supp}\big(\overline{K}_{kx'}\big) \quad \text{is bounded,} \tag{3.23}$$

and there exists a constant $C > 0$ such that for any fixed $C_2 > C$,

$$\sup_{\substack{x' \in I(x, C_1 h_1) \\ |y - x'| \leq C_2 h_1}} \left| \frac{d^j}{dx'^j} \overline{K}_{kx}\left(\frac{y - x'}{h_1} \right) \right| = O(h_1^{-j}), \qquad j = 0, 1, \dots, 2(t-k) + l + 1 \tag{3.24}$$

and

$$\sup_{\substack{x' \in I(x, C_1 h_1) \\ |y - x'| \leq C_2 h_1}} \left| \frac{d^j}{dx'^j} \left(\overline{K}_{kx'}\left(\frac{y - x'}{h_1} \right) - \overline{K}_{kx}\left(\frac{y - x'}{h_1} \right) \right) \right| = o(h_1^{-j}), \qquad j = 0, 1, \dots, 2(t-k) + l + 1. \tag{3.25}$$

The theorem will then follow by putting $k = t$ in (3.19), (3.21) and (3.22).

For $k = 1$ we have

$$\hat{f}_1(x) = \int K(\eta) f(x + \eta h_1)\, d\eta + (nh_1)^{-1/2} W_n(x; K, h_1).$$

Since $\overline{K}_{1x} = K$ and $b_1(x) = \mu_2(K) f^{(2)}(x)/2$, it follows that (3.19)–(3.25) are satisfied with $R_{12}(x) = 0$ and

$$R_{11}(x) = \int K(\eta) f(x + \eta h_1)\, d\eta - \left(f(x) + \frac{\mu_2(K)}{2} f^{(2)}(x) h_1^2 \right).$$

[Since f is $2t + l$ times differentiable in a neighbourhood of x, (3.20)–(3.21) are satisfied when $k = 1$, and since K is $2t + l - 1$ times continuously differentiable with compact support, (3.22)–(3.25) are satisfied.]

Suppose now that we have shown (3.19)–(3.25) for a certain $k < t$. Then (3.19)–(3.25) also holds for $k + 1$ because of Lemma 3.2, with $\hat{f}(x) = \hat{f}_k(x)$, $\hat{b} = b_k$, $\hat{R}_j = R_{kj}$, $\hat{K}_x = \overline{K}_{kx}$, $J = 2(t-k) + l - 2$, $\tilde{f} = \hat{f}_{k+1}$, $h = h_{k+1}$, $\tilde{b} = b_{k+1}$, $\tilde{R}_j = R_{k+1,j}$, $\tilde{K}_x = \overline{K}_{k+1,x}$ and $\alpha(x) = \alpha_k(x)$. This completes the induction step. □

The following proposition shows that the kernel $\overline{K}_{tx}$ defined in (2.6) is indeed at $2t$th order kernel.

PROPOSITION 3.3. *The function $\overline{K}_{tx}$ defined in* (2.6) *is a* $2t$-*th order kernel with*

$$\mu_k(\overline{K}_{tx}) = \begin{cases} 1, & k = 0, \\ 0, & k = 1,\dots,2t-1, \\ (2t)!\left(\dfrac{\mu_2(K)}{2}\right)^t \left(\prod_{k=1}^{t} \alpha_k(x)\right)^2 (-1)^{t-1}, & k = 2t, \end{cases} \tag{3.26}$$

for $t = 1, 2, \dots$.

PROOF. We proceed by induction w.r.t. t. Since $\overline{K}_{1x} = K$, (3.26) holds when $t = 1$ because of Assumption 2. Suppose now $t \ge 2$ and that we have established the expansion (3.26) of $\mu_{k'}(\overline{K}_{t'x})$ for $t' = 1,\dots,t-1$ and $k' = 0,1,\dots,2t'$. Then, for $k \in \{0,1,\dots,2t\}$,

$$\begin{aligned}
&\mu_k(\overline{K}_{tx}) \\
&\quad = \mu_k\left(K_{\alpha_t(x)} + \overline{K}_{t-1,x} - K_{\alpha_t(x)} * \overline{K}_{t-1,x}\right) \\
&\quad = \alpha_t(x)^k \mu_k(K) + \mu_k(\overline{K}_{t-1,x}) \\
&\qquad - \sum_{j=0}^{k} \binom{k}{j} \alpha_t(x)^{k-j} \mu_{k-j}(K) \mu_j(\overline{K}_{t-1,x})
\end{aligned}$$

$$= \begin{cases}
\mu_0(K) + \mu_0(\overline{K}_{t-1,x}) - \mu_0(\overline{K}_{t-1,x}) = 1, & k = 0, \\
\alpha_t(x)^k \mu_k(K) + \mu_k(\overline{K}_{t-1,x}) - \alpha_t(x)^k \mu_k(K) = 0, & k = 1,\dots,2t-3, \\
\alpha_t(x)^{2t-2} \mu_{2t-2}(K) + \mu_{2t-2}(\overline{K}_{t-1,x}) & \\
\quad - \left(\alpha_t(x)^{2t-2} \mu_{2t-2}(K) + \mu_{2t-2}(\overline{K}_{t-1,x})\right) = 0, & k = 2t-2, \\
0, & k = 2t-1, \\
\alpha_t(x)^{2t} \mu_{2t}(K) + \mu_{2t}(\overline{K}_{t-1,x}) & \\
\quad - \left(\alpha_t(x)^{2t} \mu_{2t}(K) + \binom{2t}{2} \alpha_t(x)^{2} \mu_2(K) \mu_{2t-2}(\overline{K}_{t-1,x})\right. & \\
\quad \left. + \mu_{2t}(\overline{K}_{t-1,x})\right) & \\
\quad = -2t(2t-1)\dfrac{\mu_2(K)}{2} \alpha_t(x)^2 \mu_{2t-2}(\overline{K}_{t-1,x}), & k = 2t.
\end{cases}$$

We have used the fact that $\overline{K}_{tx}$ is an even function (this may also be established using induction w.r.t. t), which in particular implies that $\mu_k(\overline{K}_{tx}) = 0$ for k odd. □

4. Further remarks.

REMARK 4.1. We see from (3.1) that $\hat{f}_t^{(l)}(x) - f^{(l)}(x) = O_p(h_1^{2t} + (nh_1)^{-1/2}h_1^{-l})$, as for a $2t$th order KDE. Suppose $h_k = c_k n^{-\beta}$, $k = 1, \dots, t$, with $c_1, \dots, c_t > 0$ and $0 < \beta < 1$. Then the optimal choice of exponent is $\beta = \beta(t, l) = 1/(4t + 2l + 1)$.

REMARK 4.2. Let $S_k = \{(i_1, \dots, i_k);\ 1 \le i_1 < \cdots < i_k \le t\}$ denote all ordered subsets of $\{1, \dots, t\}$ of size k. It is then easy to show by induction w.r.t. t that $\overline{K}_{tx}$ has the following explicit expansion:

$$\overline{K}_{tx} = \sum_{k=1}^{t} (-1)^{k-1} \sum_{(i_1, \dots, i_k) \in S_k} K_{\alpha_{i_1}(x)} * \cdots * K_{\alpha_{i_k}(x)}. \tag{4.1}$$

As special cases we have [remember that $\alpha_1(x) = 1$]

$$\overline{K}_{2x} = K + K_{\alpha_2(x)} - K * K_{\alpha_2(x)}$$

and

$$\begin{aligned}\overline{K}_{3x} = K &+ K_{\alpha_2(x)} + K_{\alpha_3(x)} - K * K_{\alpha_2(x)} - K * K_{\alpha_3(x)} \\ &- K_{\alpha_2(x)} * K_{\alpha_3(x)} + K * K_{\alpha_2(x)} * K_{\alpha_3(x)}.\end{aligned}$$

Put $\tilde{h}_1(x) \equiv h_1$ and $\tilde{h}_k(x) = h_k/f(x)$, $k = 2, \dots, t$. It follows from (4.1), (2.2) and (2.4) that the main stochastic term in (3.1) may be written as

$$n^{-1} \sum_{i=1}^{n} \big(\check{K}_{tx}(x - X_i) - E\check{K}_{tx}(x - X)\big), \tag{4.2}$$

with

$$\check{K}_{tx} = \sum_{k=1}^{t} (-1)^{k-1} \sum_{(i_1, \dots, i_k) \in S_k} K_{\tilde{h}_{i_1}(x)} * \cdots * K_{\tilde{h}_{i_k}(x)}. \tag{4.3}$$

This formulation of the stochastic term is more symmetric with respect to the bandwidths.

REMARK 4.3. The bias term $b_t(x)h_1^{2t}$ may also be written as

$$\overline{b}_t(x) \prod_{k=1}^{t} h_k^2, \tag{4.4}$$

with $\overline{b}_1(x) = b_1(x) = \mu_2(K)f^{(2)}(x)/2$ and

$$\overline{b}_{k+1}(x) = -\frac{\mu_2(K)}{2} f(x)^{-2} B(\overline{b}_k)(x), \qquad k = 1, 2, \dots, t-1.$$

Introducing the operator $\tilde{f} \to \tilde{B}(\tilde{f}) = \tilde{f}^{-2}B(\tilde{f})$, we obtain the nonrecursive formulation

$$\bar{b}_t(x) = (-1)^{t-1}\left(\frac{\mu_2(K)}{2}\right)^t \tilde{B}^{t-1}(f^{(2)})(x), \tag{4.5}$$

with $\tilde{B}^{t-1} = \tilde{B} \circ \cdots \circ \tilde{B}$ iterated $t-1$ times and $\tilde{B}^0$ is the identity operator. The advantage of this formulation is that $\bar{b}_t(x)$ only depends on $\mu_2(K)$ and all derivatives of f up to order $2t$, not on the bandwidths.

REMARK 4.4. We have only treated the case when all bandwidths are of the same order [cf. (2.1)]. However, formulas (4.2)–(4.5) suggest that generally the bias term is $O(h_1^2 h_2^2 \cdots h_t^2)$ while the stochastic term is $O_p((n(\min_k h_k)^{2l+1})^{-1/2})$. If, for instance, $h_k = c_k n^{-1/(4t+1)}$ for $k = 1, \dots, t-1$ and $h_t/h_1 \to 0$ as $n \to \infty$, the stochastic term dominates (cf. Theorem 2.3 in RC).

REMARK 4.5. The bias term in (3.1) when $t = 2$ and $l = 0$ has the form (cf. Proposition 3.3)

$$\begin{aligned} &\frac{\mu_4(\bar{K}_{2x})}{24}\left(f^{(4)}(x) - \frac{(f^{(2)}(x))^2}{f(x)} - \frac{3f'(x)f^{(3)}(x)}{f(x)} + \frac{3(f'(x))^2 f^{(2)}(x)}{f(x)^2}\right)h_1^4 \\ &= -\frac{\mu_2(K)^2}{4}\left(\frac{f^{(4)}(x)}{f(x)^2} - \frac{(f^{(2)}(x))^2}{f(x)^3} - \frac{3f'(x)f^{(3)}(x)}{f(x)^3} + \frac{3(f'(x))^2 f^{(2)}(x)}{f(x)^4}\right)h_1^2 h_2^2. \end{aligned} \tag{4.6}$$

Since h_1 and h_2 are fixed constants (not depending on x) it is more instructive to consider the RHS of (4.6), since there is a hidden dependence on x in the factor $\mu_4(\bar{K}_{2x})$ of the LHS. We see that $f(x)$ enters with powers 2, 3 and 4 in the denominators of the terms in the bias. This suggests that the (second iterate) TKDE estimates spikes of a density well [$f(x)$ is large] while tails are estimated worse [$f(x)$ is small]. This agrees with the empirical studies in RC. One way of improving the estimate of $f(x)$ in the tails is to let h_2 depend on x. This idea has been explored in Hössjer and Ruppert (1993b).

REMARK 4.6. The bias term for the second iterate TKDE (4.6) vanishes identically on intervals for densities having the form $f(x) = ag(bx + c)$ on that same interval, where a, b and c are constants and $g(x) = 1$, x, x^{-3} or e^{-x}. This should be compared with the ordinary KDE, using a fourth order kernel. Then the factor $f^{(4)}(x)$ appearing in the bias term vanishes identically for $g(x) = 1$, x, x^2 and x^3 (and linear combinations of these functions).

REMARK 4.7. As mentioned in Section 2, $\operatorname{supp}(\overline{K}_{tx}) = [-\sum_{k=1}^t \alpha_k(x), \sum_{k=1}^t \alpha_k(x)]$. Hence, the main stochastic term of the asymptotic expansion in Theorem 3.1 is influenced by data satisfying

$$(4.7) \quad |X_i - x| \le h_1 \sum_{k=1}^{t} \alpha_k(x) = h_1 + f(x)^{-1} \sum_{k=2}^{t} h_k, \qquad i = 1, \dots, n.$$

If all h_k are of the same order, (4.7) suggests that $\hat{f}_t(x)$ cannot converge as $t \to \infty$, since more and more data are involved for computing the higher order iterates. The optimal number of iterations in terms of asymptotic mean squared error depends in a complicated way on K, n and f as well as on the bandwidths. However, the simulations in RC indicate that $\hat{f}_t$ changes substantially during the first three—perhaps four—iterations, but only slowly after that. After four iterations, bias does not decrease much further, but variance does increase. We recommend three or four iterations in practice.

REMARK 4.8. Why does $\hat{f}_{k+1}(x)$ have a smaller bias than $\hat{f}_k(x)$, $k = 1, 2, \dots, t-1$? Note that

$$(4.8) \qquad \hat{f}_{k+1}(x) = \hat{f}_k(x) \hat{f}_Y\big(\hat{F}_k(x); \hat{F}_k, h_{k+1}\big),$$

where $\hat{f}_k(x)$ is an estimate of $f(x)$ and $\hat{f}_Y(\hat{F}_k(x); \hat{F}_k, h_{k+1})$ is an estimate of 1. The bias terms are $O(h_1^{2k})$ for both of these factors, while the stochastic terms are $O_p((nh_1)^{-1/2})$. However, the bias term of $\hat{f}_k(x)$ is $-f(x)$ times the bias term of $\hat{f}_Y(\hat{F}_k(x); \hat{F}_k, h_{k+1})$ plus terms of order h_1^{2k+2}. Hence, the two bias terms cancel (up to higher order terms) in (4.8), and the bias of $\hat{f}_Y(\hat{F}_k(x); \hat{F}_k, h_{k+1})$ corrects that of $\hat{f}_k(x)$. This follows from the expansion (A.8) for $\hat{f}_Y$ and the one in Lemma 3.2 for $\hat{f}(x) = \hat{f}_k(x)$.

APPENDIX

We start by proving a technical lemma needed in the proof of Lemma 3.2. Given a function $\tilde{K}: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, define $Z_n(x) = (nh_1)^{-1} \sum_{i=1}^n (\tilde{K}(x; X_i) - E\tilde{K}(x; X))$, a zero mean stochastic process. Actually, $\tilde{K}$ depends on n, but this will not be made explicit in the notation. With C_1 and C_2 two positive constants, introduce the following two subsets of $\mathbb{R}^2$:

$$(A.1) \qquad I_1(C_1, C_2) = \{(x', y); |x' - x| \le C_1 h_1, |x' - y| > C_2 h_1\}$$

and

$$(A.2) \qquad I_2(C_1, C_2) = \{(x', y); |x' - x| \le C_1 h_1, |x' - y| \le C_2 h_1\}.$$

(The dependence of I_1 and I_2 on x is suppressed in this notation.) Denote the partial derivatives w.r.t. the first argument of $\tilde{K}$ as $\tilde{K}^{(j)}(x, y) = \partial^j \tilde{K}(x, y)/\partial x^j$ and put

$$(A.3) \qquad \|\tilde{K}^{(j)}\|_{C_1, C_2} = \sup_{(x', y) \in I_2(C_1, C_2)} \big|\tilde{K}^{(j)}(x', y)\big|.$$

It will be tacitly understood in Lemma A.1 that the number J is the same as in Lemma 3.2.

LEMMA A.1. *Suppose there exists an integer N such that for all* $n \geq N$,

$$\tilde{K}(x', y) = 0, \qquad \forall (x', y) \in I_1(C_1, C_2), \tag{A.4}$$

and that $\tilde{K}$ *is* $J + 1$ *times continuously differentiable w.r.t. the first argument with*

$$\|\tilde{K}^{(j)}\|_{C_1, C_2} = o(h_1^{-j}), \qquad j = 0, 1, \dots, J + 1. \tag{A.5}$$

Then

$$\sup_{x' \in I(x, C_1 h_1)} \left| Z_n^{(j)}(x') \right| = o_p\left((nh_1)^{-1/2} h_1^{-j} \right), \qquad j = 0, 1, \dots, J. \tag{A.6}$$

If in (A.5), *o is replaced by O, then* (A.6) *still holds with* O_p *instead of* o_p.

PROOF. The proof [when we have $o(h_1^{-j})$ in (A.5)] consists of first proving that

$$\begin{aligned} &E\left(Z_n^{(j)}(x + t_2 h_1) - Z_n^{(j)}(x + t_1 h_1) \right)^2 \\ &\qquad = o(1)(nh_1)^{-1} h_1^{-2j} (t_2 - t_1)^2, \end{aligned} \tag{A.7}$$

uniformly for $t_1, t_2 \in [-C_1, C_1]$ [cf. Hössjer and Ruppert (1993a) for details]. The lemma then follows from Bickel and Wichura [(1971), (1) and Theorem 1 with $\gamma = \beta = 2$]. When o is replaced by O in (A.6), the proof is the same; we just replace $o(1)$ by $O(1)$ in (A.7). □

PROOF OF LEMMA 3.2. We will first (Steps 1 and 2) give a proof of (3.12)–(3.15) when $\hat{K}_{x'}$ is replaced by $\hat{K}_x$ for all $x' \in I(x, C_1 h_1)$, and $\tilde{K}_{x'}$ is replaced by $A_{\alpha(x')}(\hat{K}_x)$. Then the extension to a kernel that depends on x' is given in Step 3. Finally, we prove that $\tilde{K}_{x'}$ satisfies (3.16)–(3.18) in Step 4. Without ambiguity, we will put $\hat{K} = \hat{K}_x$ in Steps 1 and 2.

Step 1. *Asymptotic expansion of* $\hat{f}_Y$. In this step, we will show that $\hat{f}_Y(x; \hat{F}, h)$ has the asymptotic expansion

$$\begin{aligned} &\hat{f}_Y\left(\hat{F}(x); \hat{F}, h \right) \\ &\quad = 1 + f(x)^{-1} \Big((nh_1)^{-1/2} W_n(x; K_{\alpha(x)}, h_1) \\ &\qquad\qquad - (nh_1)^{-1/2} W_n\left(x; K_{\alpha(x)} * \hat{K}, h_1 \right) \\ &\qquad\qquad - \tfrac{1}{2} \mu_2(K) \alpha(x)^2 B(\hat{b})(x) h_1^{2k+2} - \hat{b}(x) h_1^{2k} \\ &\qquad\qquad\qquad + f(x)^{-1} \hat{b}(x)^2 h_1^{4k} - \hat{R}_1(x) \Big) \\ &\qquad + \overline{R}_1(x) + \overline{R}_2(x), \end{aligned} \tag{A.8}$$

with $\overline{R}_1$ nonstochastic satisfying

$$\sup_{x' \in I(x, C_1 h_1)} \left| \overline{R}_1^{(j)}(x') \right| = o\left(h_1^{2k+2} \right), \qquad j = 0, 1, \dots, J, \tag{A.9}$$

and

$$\sup_{x' \in I(x, C_1 h_1)} \left| \overline{R}_2^{(j)}(x') \right| = o_p\left((nh_1)^{-1/2} h_1^{-j} \right), \qquad j = 0, 1, \dots, J. \tag{A.10}$$

By the definition of $\hat{f}_Y$ in Section 1,

$$\hat{f}_Y\left(\hat{F}(x); \hat{F}, h\right) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{\hat{F}(X_i) - \hat{F}(x)}{h} \right). \tag{A.11}$$

We will make a Taylor expansion of each term in (A.11). For this we need some preliminaries. Let $\bar{f}(x) = f(x) + \hat{b}(x) h_1^{2k} + \hat{R}_1(x)$ be the nonstochastic part of the density $\hat{f}(x)$. Put $\tilde{F}(x) = \int_{-\infty}^{x} \bar{f}(t)\,dt$ and

$$\overline{F}(y) = \begin{cases} \tilde{F}(y), & y \in I(x, (C_1 + 2C_2)h_1), \\ \tilde{F}(x + (C_1 + 2C_2)h_1), & y > x + (C_1 + 2C_2)h_1, \\ \tilde{F}(x - (C_1 + 2C_2)h_1), & y < x - (C_1 + 2C_2)h_1. \end{cases}$$

Assume $C_2 > \max_n (h(n)/h_1(n))/f(x)$. Then, it follows from (3.8) and (3.9) [cf. (Assumption 1)] that

$$\left| \overline{F}(y) - \overline{F}(x') \right| > h \quad \text{whenever } (x', y) \in I_1(C_1, C_2), \tag{A.12}$$

for large enough n. Introduce $u(x; y) = (F(y) - F(x))/h$, $\overline{u}(x; y) = (\overline{F}(y) - \overline{F}(x))/h$ and $\hat{u}(x; y) = (\hat{F}(y) - \hat{F}(x))/h$. By the Taylor expansion,

$$\begin{aligned} &K(\hat{u}(x; y)) \\ &\quad = K(u(x; y)) + K'(u(x; y))(\hat{u}(x; y) - u(x; y)) \\ &\qquad + \tilde{K}_1(x; y) + \tilde{K}_2(x; y) + \tilde{K}_3(x; y), \end{aligned} \tag{A.13}$$

with

$$\tilde{K}_1(x; y) = \int_{u(x; y)}^{\overline{u}(x; y)} K^{(2)}(t)(\overline{u}(x; y) - t)\,dt,$$

$$\tilde{K}_2(x; y) = (K'(\overline{u}(x; y)) - K'(u(x; y)))(\hat{u}(x; y) - \overline{u}(x; y))$$

and

$$\tilde{K}_3(x; y) = \int_{\overline{u}(x; y)}^{\hat{u}(x; y)} K^{(2)}(t)(\hat{u}(x; y) - t)\,dt.$$

Insertion of (A.13) into (A.11) gives

$$\begin{aligned} &\hat{f}_Y\left(\hat{F}(x); \hat{F}, h\right) \\ &\quad = 1 + (nh)^{-1/2} \tilde{W}_n(x; F, K, h) \\ &\qquad + (nh)^{-1} \sum_{i=1}^{n} K'(u(x; X_i))(\hat{u}(x; X_i) - u(x; X_i)) \\ &\qquad + (nh)^{-1} \sum_{i=1}^{n} \tilde{K}_1(x; X_i) \\ &\qquad + (nh)^{-1} \sum_{i=1}^{n} \tilde{K}_2(x; X_i) + (nh)^{-1} \sum_{i=1}^{n} \tilde{K}_3(x; X_i). \end{aligned} \tag{A.14}$$

Let $\hat{\Delta}(t) = \int_{-\infty}^{t} \hat{K}(s)\,ds$ and $\check{f}(x) = E\hat{K}((X-x)/h_1)/h_1$. Then

$$\hat{u}(x;y) - u(x;y)$$

$$\text{(A.15)}\qquad = (nh)^{-1}\sum_{j=1}^{n}\left(\hat{\Delta}\left(\frac{y-X_j}{h_1}\right) - \hat{\Delta}\left(\frac{x-X_j}{h_1}\right)\right) + h^{-1}\int_x^y \left(\hat{b}(t)h_1^{2k} - \check{f}(t) + \hat{R}_1(t) + \hat{R}_2(t)\right)dt.$$

Put also

$$\tilde{K}_4(x;y) = h^{-1}h_1^{2k}K'(u(x;y))\int_x^y \hat{b}(t)\,dt,$$

$$\hat{K}_5(x;y) = h^{-1}K'(u(x;y))\int_x^y \hat{R}_1(t)\,dt$$

and

$$\tilde{K}_6(x;y) = h^{-1}K'(u(x;y))\int_x^y \hat{R}_2(t)\,dt.$$

Hence, insertion of (A.15) into the third term in the RHS of (A.14) yields

$$(nh)^{-1}\sum_{i=1}^{n} K'(u(x;X_i))(\hat{u}(x;X_i) - u(x;X_i))$$

$$\text{(A.16)}\qquad \begin{aligned} &= (nh)^{-2}\sum\sum_{1\le i,j\le n} H(x;X_i,X_j) \\ &\quad -(nh^2)^{-1}\sum_{i=1}^{n} K'(u(x;X_i))\int_x^{X_i}\check{f}(t)\,dt \\ &\quad +(nh)^{-1}\sum_{i=1}^{n}\tilde{K}_4(x;X_i) + (nh)^{-1}\sum_{i=1}^{n}\tilde{K}_5(x;X_i) \\ &\quad +(nh)^{-1}\sum_{i=1}^{n}\tilde{K}_6(x;X_i), \end{aligned}$$

where $H(x;y,z) = K'(u(x;y))(\hat{\Delta}((y-z)/h_1) - \hat{\Delta}((x-z)/h_1))$. It follows from (A.14) and (A.16) that (A.8) holds with

$$\overline{R}_1(x) = \sum_{k=1}^{3}\overline{R}_{1k}(x),$$

$$\overline{R}_2(x) = \sum_{k=1}^{8}\overline{R}_{2k}(x),$$

$$\overline{R}_{11}(x) := h^{-1}E\tilde{K}_1(x;K) - f(x)^{-2}\hat{b}(x)^2h_1^{4k},$$

$$\begin{aligned}\overline{R}_{12}(x) := h^{-1}E\tilde{K}_4(x;X) &+ f(x)^{-1}\hat{b}(x)h_1^{2k} \\ &+ \tfrac{1}{2}f(x)^{-1}\mu_2(K)\alpha(x)^2B(\hat{b})(x)h_1^{2k+2},\end{aligned}$$

$$\overline{R}_{13}(x) := h^{-1}E\tilde{K}_5(x;X) + f(x)^{-1}\hat{R}_1(x),$$

$$\overline{R}_{21}(x) := (nh)^{-1/2}\tilde{X}_n(x;F,K,h) - f(x)^{-1}(nh_1)^{-1/2}W_n(x;K_{\alpha(x)},h_1),$$

$$\overline{R}_{22}(x) := (nh)^{-1}\sum_{i=1}^{n}\big(\tilde{K}_1(x;X_i) - E\tilde{K}_1(x;X)\big),$$

$$\overline{R}_{23}(x) := (nh)^{-1}\sum_{i=1}^{n}\tilde{K}_2(x;X_i),$$

$$\overline{R}_{24}(x) := (nh)^{-1}\sum_{i=1}^{n}\tilde{K}_3(x;X_i),$$

$$\begin{aligned}\overline{R}_{25}(x) := {}& (nh)^{-2}\sum_{1\le i,j\le n}\sum H(x;X_i,X_j)\\ &-(nh^2)^{-1}\sum_{i=1}^{n}K'(u(x;X_i))\int_x^{X_i}\check{f}(t)\,dt\\ &+f(x)^{-1}(nh_1)^{-1/2}W_n\Big(x;K_{\alpha(x)}*\hat{K},h_1\Big),\end{aligned}$$

$$\overline{R}_{26}(x) := (nh)^{-1}\sum_{i=1}^{n}\big(\tilde{K}_4(x;X_i) - E\tilde{K}_4(x;X)\big),$$

$$\overline{R}_{27}(x) := (nh)^{-1}\sum_{i=1}^{n}\big(\tilde{K}_5(x;X_i) - E\tilde{K}_5(x;X)\big)$$

and

$$\overline{R}_{28}(x) := (nh)^{-1}\sum_{i=1}^{n}\tilde{K}_6(x;X_i).$$

It remains to prove (A.9) and (A.10) by showing that each $\overline{R}_{1k}$ satisfies (A.9) and each $\overline{R}_{2k}$ satisfies (A.10). For the nonstochastic terms $\overline{R}_{11} - \overline{R}_{13}$, we will make use of the following: Let

$$\begin{aligned}\xi(x',\eta) &= F^{-1}(F(x')+\eta h)\\ &= x' + \frac{\eta h}{f(x')} - \frac{1}{2}\frac{f'(x')}{f(x')^3}(\eta h)^2\\ &\quad+\frac{1}{6}\left(\frac{3f'(x')^2}{f(x')^5} - \frac{f^{(2)}(x')}{f(x')^4}\right)(\eta h)^3 + \rho(x';\eta).\end{aligned}\tag{A.17}$$

Since f is $J+2$ times continuously differentiable around x with $f(x) > 0$, it follows by making a Taylor expansion of F^{-1} around $F(x')$, that for any $C > 0$,

$$\sup_{\substack{x'\in I(x,C_1h_1)\\ |\eta|\le C}} \big|\rho^{(j)}(x';\eta)\big| = o\big(h_1^3\big), \qquad j = 0,1,\dots,J.\tag{A.18}$$

$\overline{R}_{11}$. Let $\hat{c}(x) = \hat{b}(x)h_1^{2k} + \hat{R}_1(x)$. Then

$$\begin{aligned}\overline{R}_{11}(x) &= \left(\tfrac{1}{2}h^{-3}\int K^{(2)}(u(x;y))(y-x)^2 f(y)\,dy\,\hat{c}(x)^2 - f(x)^{-2}\hat{b}(x)^2 h_1^{4k}\right)\\ &\quad+\left(\tfrac{1}{2}h^{-3}\int K^{(2)}(u(x;y))\left(\left(\int_x^y \hat{c}(t)\,dt\right)^2 - ((y-x)\hat{c}(x))^2\right)f(y)\,dy\right)\\ &\quad+\left(h^{-1}\int\int_{u(x;y)}^{\overline{u}(x;y)}(K^{(2)}(s)-K^{(2)}(u(x;y)))(\overline{u}(x;y)-s)\,ds\,f(y)\,dy\right)\\ &:= \overline{R}_{111}(x)+\overline{R}_{112}(x)+\overline{R}_{113}(x).\end{aligned}$$

By change of variables,

$$\overline{R}_{111}(x') = \tfrac{1}{2}h^{-2}\int_{-1}^{1}K^{(2)}(\eta)(\xi(x',\eta)-x')^2\,d\eta\,\hat{c}(x')^2 - f(x')^{-2}\hat{b}(x')^2 h_1^{4k},$$

and from (3.8), (3.9) and (A.18) it follows that $\sup_{x'\in I(x,C_1h_1)}|\overline{R}_{111}^{(j)}(x')| = o(h_1^{4k}) = o(h_1^{2k+2})$ for $j=0,1,\dots,J$. Similarly, by making the same change of variables and using (3.8), (3.9) and (A.18), it is possible to prove

$$\sup_{x'\in I(x,C_1h_1)}|\overline{R}_{112}^{(j)}(x')| = O(h_1^{4k+1}) = o(h_1^{2k+2})$$

and

$$\sup_{x'\in I(x,C_1h_1)}|\overline{R}_{113}^{(j)}(x')| = O(h_1^{6k}) = o(h_1^{2k+2})$$

for $j=0,1,\dots,J$. For details, see Hössjer and Ruppert (1993a).

$\overline{R}_{12}$. By the definition of $\tilde{K}_4$ and the change of variables $\eta = (F(y)-F(x))/h$,

$$h^{-1}E\tilde{K}_4(x;X) = h^{-1}h_1^{2k}\int_{-1}^{1}K'(\eta)\int_x^{\xi(x;\eta)}\hat{b}(t)\,dt\,d\eta. \tag{A.19}$$

We now perform a Taylor expansion of the factor $\int_x^{\xi(x;\eta)}\hat{b}(t)\,dt$ in (A.19) and use the definition of $B(\hat{b})$. This results in

$$\begin{aligned}\overline{R}_{12}(x) &= h^{-1}E\tilde{K}_4(x;X) + f(x)^{-1}\frac{\mu_2(K)\alpha(x)^2}{2}B(\hat{b})(x)h_1^{2k+2}\\ &\quad+f(x)^{-1}\hat{b}(x)h_1^{2k}\\ &= \Bigg(h^{-1}\int_{-1}^{1}K'(\eta)(\xi(x;\eta)-x)\,d\eta + f(x)^{-1}\\ &\qquad+\frac{1}{2}\mu_2(K)\left(\frac{3f'(x)^2}{f(x)^5}-\frac{f^{(2)}(x)}{f(x)^4}\right)h^2\Bigg)\hat{b}(x)h_1^{2k}\end{aligned}$$

$$+\left(\frac{1}{2}h^{-1}\int_{-1}^{1}K'(\eta)(\xi(x;\eta)-x)^2\,d\eta - \frac{3}{2}\mu_2(K)\frac{f'(x)}{f(x)^4}h^2\right)\hat{b}'(x)h_1^{2k}$$

$$+\left(\frac{1}{6}h^{-1}\int_{-1}^{1}K'(\eta)(\xi(x;\eta)-x)^3\,d\eta + \frac{1}{2}\frac{\mu_2(K)}{f(x)^3}h^2\right)\hat{b}^{(2)}(x)h_1^{2k}$$

$$+\left(\frac{1}{2}h^{-1}\int_{-1}^{1}K'(\eta)\int_{x}^{\xi(x;\eta)}\left(\hat{b}^{(2)}(t)-\hat{b}^{(2)}(x)\right)(\xi(x;\eta)-t)^2\,dt\,d\eta\right)h_1^{2k}$$

$$:= \overline{R}_{121}(x)+\overline{R}_{122}(x)+\overline{R}_{123}(x)+\overline{R}_{124}(x).$$

Observe next that $\mu_k(K')=\int_{-1}^{1}K'(\eta)\eta^k\,d\eta$ equals 0 for k even, -1 for $k=1$ and $-3\mu_2(K)$ for $k=3$. In conjunction with (A.17) and (A.18) this implies

$$\sup_{x'\in I(x,C_1h_1)}\left|\overline{R}_{12i}^{(j)}(x')\right| = o(h_1^{2k+2}),\qquad i=1,2,3,\ j=0,1,\dots,J. \tag{A.20}$$

Formula (A.20) for $i=4$ follows from (3.8) and (A.17) and (A.18). For details, see Hössjer and Ruppert (1993a).

$\overline{R}_{13}$. This term is handled in the same way as $\overline{R}_{12}$, with $\hat{R}_1$ instead of $h_1^{2k}\hat{b}$.

Now for the stochastic terms $\overline{R}_{21}-\overline{R}_{28}$, we first need some estimates that we will make frequent use of later. Put $v(x;y)=f(x)(y-x)/h$ and remember the notation in (A.3). Then it is easily seen that

$$\begin{aligned}\|u^{(j)}\|_{C_1,C_2} &= \sup_{(x',y)\in I_2(C_1,C_2)}\left|\frac{\partial^j u(x',y)}{\partial x'^j}\right|\\ &\le \begin{cases} C, & \text{when } j=0,\\ Ch_1^{-1}, & \text{when } j=1,\dots,J+1,\end{cases}\end{aligned} \tag{A.21}$$

$$\|v^{(j)}\|_{C_1,C_2} \le \begin{cases} C, & \text{when } j=0,\\ Ch_1^{-1}, & \text{when } j=1,\dots,J+1,\end{cases} \tag{A.22}$$

$$\|\overline{u}^{(j)}\|_{C_1,C_2} \le \begin{cases} C, & \text{when } j=0,\\ Ch_1^{-1}, & \text{when } j=1,\dots,J+1,\end{cases} \tag{A.23}$$

and

$$\|\hat{u}^{(j)}\|_{C_1,C_2} = O_p\left(h_1^{-j}\right),\quad \text{when } j=0,1,\dots,J+1. \tag{A.24}$$

The last identity holds since $\hat{u}^{(j)}(x',y)=-h^{-1}\hat{f}^{(j-1)}(x')$ when $j\ge 1$, and according to (3.8)–(3.10) the RHS of (A.24) is $O_p(h_1^{-1}(1+(nh_1)^{-1/2}h_1^{-j+1}))=$

$O_p(h_1^{-j})$, since $h_1 \gg n^{-1}$. We also need the following results:

$$\|u^{(j)} - v^{(j)}\|_{C_1,C_2} \le \begin{cases} Ch_1, & \text{when } j = 0, \\ C, & \text{when } j = 1, \\ Ch_1^{-1}, & \text{when } j = 2, \dots, J+1, \end{cases} \tag{A.25}$$

$$\|\bar{u}^{(j)} - u^{(j)}\|_{C_1,C_2} \le \begin{cases} Ch_1^{2k}, & \text{when } j = 0, \\ Ch_1^{2k-1}, & \text{when } j = 1, \dots, J+1, \end{cases} \tag{A.26}$$

and

$$\|\hat{u}^{(j)} - \bar{u}^{(j)}\|_{C_1,C_2} = O_p((nh_1)^{-1/2} h_1^{-j}), \quad \text{when } j = 0, \dots, J+1. \tag{A.27}$$

For instance, (A.25) follows since

$$u(x',y) - v(x',y) = h^{-1} \int_{x'}^{y} (f(t) - f(x'))\,dt$$

and

$$u'(x',y) - v'(x',y) = -f'(x')(y - x')/h.$$

We start with $\bar{R}_{21}$, $\bar{R}_{22}$, $\bar{R}_{26}$ and $\bar{R}_{27}$, for which we will make use of Lemma A.1.

$\bar{R}_{21}$. We may write $\bar{R}_{21}(x) = (nh_1)^{-1} \sum_{i=1}^{n} (\tilde{K}(x; X_i) - E\tilde{K}(x; X))$, where $\tilde{K}(x; y) = (K((F(y) - F(x))/h) - K(f(x)(y - x)/h))h_1/h$. It suffices to establish that (A.4) and (A.5) of Lemma A.1 are satisfied. By the choice of C_2 before (A.12), it is clear that (A.4) holds. On the other hand, (A.5) follows from (A.21), (A.22) and (A.25). For instance, $\|\tilde{K}\|_{C_1,C_2} = O(\|u - v\|_{C_1,C_2}) = O(h_1)$, and for the first derivative, $\tilde{K}'(x; y) = (u'(x; y)K'(u(x; y)) - v'(x; y) \times K'(v(x; y)))h_1/h$, which implies $\|\tilde{K}'\|_{C_1,C_2} = O(\|u' - v'\|_{C_1,C_2} + \|v'\|_{C_1,C_2} \|u - v\|_{C_1,C_2}) = O(1)$. Continuing in this way, one shows that $\|\tilde{K}^{(j)}\|_{C_1,C_2} = O(h_1^{1-j})$, $j = 0, 1, \dots, J+1$, which implies (A.5).

$\bar{R}_{22}$. Again, it suffices to show that $\tilde{K}_1$ satisfies (A.4) and (A.5) of Lemma A.1. By the choice of C_2 and construction of $\bar{F}$, it follows that

$$\begin{aligned} &|x' - x| \le C_1 h_1, \quad y > x' + C_2 h_1 \;\Rightarrow\; u(x',y), \bar{u}(x';y) > 1, \\ &|x' - x| \le C_1 h_1, \quad y < x' - C_2 h_1 \;\Rightarrow\; u(x',y), \bar{u}(x';y) < -1, \end{aligned} \tag{A.28}$$

for large enough n. Hence, by the definition of $\tilde{K}_1$, $(x',y) \in I_1(C_1, C_2) \Rightarrow \tilde{K}_1(x',y) = 0$, which proves (A.4). Formula (A.5) is a consequence of (A.21), (A.23) and (A.26). Using the expansion $\tilde{K}_1(x'; y) = K(\bar{u}(x'; y)) - K(u(x'; y)) - (\bar{u}(x',y) - u(x',y))K'(u(x',y))$, it follows that $\|\tilde{K}_1^{(j)}\|_{C_1,C_2} = O(h_1^{2k-j}) = o(h_1^{-j})$, $j = 0, 1, \dots, J+1$, and (A.5) is proved.

$\bar{R}_{26}$ *and* $\bar{R}_{27}$. It suffices to establish that $\tilde{K}_4$ and $\tilde{K}_5$ satisfy (A.4) and (A.5) of Lemma A.1; we omit the details.

Next we consider $\bar{R}_{23}$, $\bar{R}_{24}$ and $\bar{R}_{28}$, which are treated in a similar manner.

$\overline{R}_{23}$. It follows from (A.28) that for all n that are large enough, $(x', y) \in I_1(C_1, C_2) \Rightarrow \tilde{K}_2(x'; y) = 0$. Hence, for all $x' \in I(x, C_1 h_1)$,

$$\overline{R}_{23}^{(j)}(x') = (nh)^{-1} \sum_{i=1}^{n} \tilde{K}_2^{(j)}(x'; X_i) I(|X_i - x| \le (C_1 + C_2)h_1).$$

Hence

$$\sup_{|x'-x| \le C_1 h_1} \left|\overline{R}_{23}^{(j)}(x')\right| = O_p\left((nh_1)^{-1} \|\tilde{K}_2^{(j)}\|_{C_1, C_2} nh_1\right) = O_p\left(\|\tilde{K}_2^{(j)}\|_{C_1, C_2}\right).$$

It therefore suffices to show that

$$\|\tilde{K}_2^{(j)}\|_{C_1, C_2} = O_p\left(h_1^{2k-j}(nh_1)^{-1/2}\right) = o_p\left(h_1^{-j}(nh_1)^{-1/2}\right), \qquad j = 0, 1, \dots, J.$$

This is actually a consequence of (A.21), (A.23), (A.24), (A.26) and (A.27). For instance,

$$\|\tilde{K}_2\|_{C_1, C_2} = O\left(\|\overline{u} - u\|_{C_1, C_2} \|\hat{u} - \overline{u}\|_{C_1, C_2}\right) = O_p\left(h_1^{2k}(nh_1)^{-1/2}\right)$$

and $\tilde{K}_2' = (\overline{u}' K'(\overline{u}) - u' K'(u))(\hat{u} - \overline{u}) + (K(\overline{u}) - K(u))(\hat{u}' - \overline{u}')$, from which it follows that

$$\begin{aligned} \|\tilde{K}_2'\|_{C_1, C_2} &= O\Big(\big(\|\overline{u}' - u'\|_{C_1, C_2} + \|u'\|_{C_1, C_2} \|\overline{u} - u\|_{C_1, C_2}\big) \|\hat{u} - \overline{u}\|_{C_1, C_2} \\ &\qquad + \|\overline{u} - u\|_{C_1, C_2} \|\hat{u}' - \overline{u}'\|_{C_1, C_2}\Big) \\ &= O_p\left(h_1^{2k-1}(nh_1)^{-1/2}\right). \end{aligned}$$

$\overline{R}_{24}$. Since $\hat{F}$ is nondecreasing, formulas (3.8)–(3.10) and the lower bound on C_2 given before (A.12) imply $P(|\hat{F}(y) - \hat{F}(x')| > h \; \forall (x', y) \in I_1(C_1, C_2)) \to 1$ as $n \to \infty$. This formula and (A.12) give $P(\tilde{K}_3(x', y) = 0 \; \forall (x', y) \in I_1(C_1, C_2)) \to 1$ as $n \to \infty$, which implies, in the same way as for $\overline{R}_{23}$, $\sup_{|x'-x| \le C_1 h_1} \left|\overline{R}_{24}^{(j)}(x')\right| = O_p(\|\tilde{K}_2^{(j)}\|_{C_1, C_2})$, so it suffices to show that

$$\text{(A.29)} \qquad \|\tilde{K}_3^{(j)}\|_{C_1, C_2} = o_p\left(h_1^{-j}(nh_1)^{-1/2}\right), \qquad j = 0, 1, \dots, J.$$

For instance, $\|\tilde{K}_3\|_{C_1, C_2} = O(\|\hat{u} - \overline{u}\|_{C_1, C_2}^2) = O_p((nh_1)^{-1}) = o_p((nh_1)^{-1/2})$ and $\tilde{K}' = \int_{\overline{u}}^{\hat{u}} (K^{(2)}(t)\hat{u}' - K^{(2)}(\overline{u})\overline{u}')\, dt$, from which it is possible to deduce (A.29) when $j = 1$. Higher values of j are handled similarly.

$\overline{R}_{28}$. This term is handled in the same way as $\overline{R}_{23}$ and $\overline{R}_{24}$.

$\overline{R}_{25}$. The proof for this term is based on U-statistics theory and given in Lemma A.2.

Step 2. Asymptotic expansion of $\tilde{f}$. By assumption, $\tilde{f}(x) = \hat{f}(x)\hat{f}_Y(\hat{F}(x); \hat{F}, h)$, with $\hat{f}(x) = f(x) + \hat{b}(x)h_1^{2k} + \hat{R}_1(x) + \hat{R}_2(x) + (nh_1)^{-1/2} W_n(x; \hat{K}, h_1)$. It follows by multiplying this identity with (A.8) and collecting terms that

$$\begin{aligned} \tilde{f}(x) &= f(x) + \tilde{b}(x)h_1^{2k+2} + (nh_1)^{-1/2} \\ &\qquad \times W_n\left(x; \hat{K} + K_{\alpha(x)} - \hat{K} * K_{\alpha(x)}, h_1\right) + \tilde{R}_1(x) + \tilde{R}_2(x), \end{aligned}$$

with

$$\begin{aligned}\tilde{R}_1(x) &= f(x)\overline{R}_1(x) + \left(\hat{b}(x)h_1^{2k} + \hat{R}_1(x)\right)\\ &\quad\times\Big(-\tfrac{1}{2}f(x)^{-1}\mu_2(K)\alpha(x)^2 B(\hat{b})(x)h_1^{2k+2}\\ &\qquad + f(x)^{-2}\hat{b}(x)^2 h_1^{4k} + \overline{R}_1(x)\Big)\\ &\quad - f(x)^{-2}\hat{R}_1(x)^2\end{aligned}$$

and

$$\begin{aligned}\tilde{R}_2(x) &= f(x)\overline{R}_2(x) + \left(\hat{b}(x)h_1^{2k} + \hat{R}_1(x)\right)\\ &\quad\times\left(f(x)^{-1}(nh_1)^{-1/2}W_n\left(x; K_{\alpha(x)} - K_{\alpha(x)} * \hat{K}, h_1\right) + \overline{R}_2(x)\right)\\ &\quad + (nh_1)^{-1/2}W_n\left(x; \hat{K}, h_1\right)\left(\hat{f}_Y(x) - 1\right) + \hat{R}_2(x)\hat{f}_Y(x).\end{aligned} \tag{A.30}$$

Now (3.13)–(3.15) follows from (3.6), (3.8)–(3.10), (A.9) and (A.10). [The supremum of the W_n-processes involved in (A.30) are taken care of using Assumption 2, (3.6) and Lemma A.1.]

Step 3. *The dependence of $\hat{K}_{x'}$ on x'.* We will show in this step that $\hat{K}_x$ may be replaced by $\hat{K}_{x'}$ and $\tilde{K}_{x'}$ by $A_{\alpha(x')}(\hat{K}_x)$ for all $x' \in I(x, C_1 h_1)$, without affecting the conclusion of Lemma 3.2. If we can establish that

$$\begin{aligned}\sup_{x' \in I(x, C_1 h_1)} &\left|(nh_1)^{-1/2}\frac{d^j}{dx'^j}\Big(W_n\left(x'; \hat{K}_{x'}, h_1\right)\right.\\ &\qquad\left.- W_n\left(x'; \hat{K}_x, h_1\right)\Big)\right|\\ &= o_p\left((nh_1)^{-1/2}h_1^{-j}\right), \qquad j = 0, 1, \dots, J+2,\end{aligned} \tag{A.31}$$

and

$$\begin{aligned}\sup_{x' \in I(x, C_1 h_1)} &\left|(nh_1)^{-1/2}\frac{d^j}{dx'^j}\Big(W_n\left(x'; A_{\alpha(x')}\left(\hat{K}_{x'}\right), h_1\right)\right.\\ &\qquad\left.- W_n\left(x'; A_{\alpha(x')}\left(\hat{K}_x\right), h_1\right)\Big)\right|\\ &= o_p\left((nh_1)^{-1/2}h_1^{-j}\right), \qquad j = 0, 1, \dots, J,\end{aligned} \tag{A.32}$$

we may incorporate the difference between the stochastic processes in (A.31) into $\hat{R}_2(x')$ and the difference in (A.32) into $\tilde{R}_2(x')$. An appeal to what we have already proved in Steps 1 and 2 then proves the whole lemma. Observe that

$$\begin{aligned}(nh_1)^{-1/2}&\left(W_n\left(x'; \hat{K}_{x'}, h_1\right) - W_n\left(x'; \hat{K}_x, h_1\right)\right)\\ &= (nh_1)^{-1}\sum_{i=1}^{n}\left(\tilde{K}_1(x'; X_i) - E\tilde{K}_1(x'; X)\right),\end{aligned}$$

with $\tilde{K}_1(x'; y) = \hat{K}_{x'}((y - x')/h_1) - \hat{K}_x((y - x')/h_1)$. However, (A.31) then follows from Lemma A.1 (with $J + 2$ instead of J), (3.5) and (3.7). Formula (A.32), finally, is derived from Lemma A.1, with $\tilde{K}_2(x'; y) = A_{\alpha(x')}(\hat{K}_{x'}) - A_{\alpha(x')}(\hat{K}_x)$. Since $\tilde{K}_1$ satisfies (A.4) and (A.5), we may verify that $\tilde{K}_2$ satisfies these equations as well by using the methods in Step 4.

Step 4. *Verification of* (3.16)–(3.18). By definition,

$$\tilde{K}_x = \hat{K}_x + K_{\alpha(x)} - \hat{K}_x * K_{\alpha(x)}, \tag{A.33}$$

so (3.16) follows from (3.5) and the fact that $\alpha(\cdot)$ is bounded in a neighbourhood of x. Formula (3.17) will follow if we can show that $d^j\tilde{K}_x(\eta)/d\eta^j$ is bounded in η for $j = 0, 1, \dots, J + 1$. However, (3.6) implies that $d^j\hat{K}_x(\eta)/d\eta^j$ is bounded. Thus, (3.17) follows from (A.33), (3.6), Assumption 2 and the fact that $J + 1 < 2t + l - 1$. It remains to verify (3.18). Hence, we have to show that

$$\sup_{\substack{x' \in I(x, C_1 h_1) \\ y \in I(x', C_2 h_1)}} \left|\tilde{K}^{(j)}(x'; y)\right| = o\left(h_1^{-j}\right), \qquad j = 0, 1, \dots, J + 1, \tag{A.34}$$

where

$$\begin{aligned}
\tilde{K}(x'; y) &= \tilde{K}_{x'}\left(\frac{y - x'}{h_1}\right) - \tilde{K}_x\left(\frac{y - x'}{h_1}\right) \\
&= \left(\hat{K}_{x'}\left(\frac{y - x'}{h_1}\right) - \hat{K}_x\left(\frac{y - x'}{h_1}\right)\right) \\
&\quad + \left(K_{\alpha(x')}\left(\frac{y - x'}{h_1}\right) - K_{\alpha(x)}\left(\frac{y - x'}{h_1}\right)\right) \\
&\quad - \left(\hat{K}_{x'} * K_{\alpha(x')}\left(\frac{y - x'}{h_1}\right) - \hat{K}_x * K_{\alpha(x)}\left(\frac{y - x'}{h_1}\right)\right) \\
&:= \tilde{K}_1(x'; y) + \tilde{K}_2(x'; y) - \tilde{K}_3(x'; y).
\end{aligned}$$

Obviously, it suffices to prove (A.34) with $\tilde{K}_m$ instead of $\tilde{K}$, $m = 1, 2, 3$. By the induction hypothesis, the case $m = 1$ is already done. For $\tilde{K}_2$, (A.34) follows from the facts that

$$\sup_{x' \in I(x, C_1 h_1)} |\alpha(x') - \alpha(x)| = o(1) \tag{A.35}$$

and (since $J + 1 < 2t + l$)

$$\sup_{x' \in I(x, C_1 h_1)} \left| \alpha^{(j)}(x') \right| = O(1), \qquad j = 0, 1, \ldots, J + 1, \tag{A.36}$$

which are both consequences of Assumption 1. For instance,

$$\begin{aligned} &\tilde{K}_2'(x'; y) \\ &\quad = -\frac{\alpha'(x')}{\alpha(x')^2} K\left(\frac{y - x'}{\alpha(x')h_1}\right) - \frac{\alpha'(x')}{\alpha(x')^3} K'\left(\frac{y - x'}{\alpha(x')h_1}\right)\frac{y - x'}{h_1} \\ &\qquad - \left(\frac{1}{\alpha(x')^2 h_1} K'\left(\frac{y - x'}{\alpha(x')h_1}\right) - \frac{1}{\alpha(x)^2 h_1} K'\left(\frac{y - x'}{\alpha(x)h_1}\right)\right). \end{aligned} \tag{A.37}$$

The first two terms in the RHS of (A.37) are clearly $O(1)$ because of (A.36), and the difference of the last two terms is $o(h^{-1})$ as a consequence of (A.35). It remains to consider $\tilde{K}_3$. Note that

$$\begin{aligned} &\tilde{K}_3^{(j)}(x'; y) \\ &\quad = \frac{1}{h_1} \int \sum_{m=0}^{j} \binom{j}{m} \\ &\qquad \times \left(\frac{d^m \hat{K}_{x'}((z - x')/h_1)}{dx'^m} \frac{d^{(j-m)} K_{\alpha(x')}((y - z - x')/h_1)}{dx'^{(j-m)}} \right. \\ &\qquad \left. - \frac{d^m \hat{K}_{x}((z - x')/h_1)}{dx'^m} \frac{d^{(j-m)} K_{\alpha(x)}((y - z - x')/h_1)}{dx'^{(j-m)}} \right) dz. \end{aligned}$$

Now (A.34), with $\tilde{K}_3$ in place of $\tilde{K}$, follows from the fact that we have already established (A.34) for $\tilde{K}_1$ and $\tilde{K}_2$, and from the relation

$$\sup_{\substack{x' \in I(x, C_1 h_1) \\ \xi \in I(x', C_2 h_1)}} \left(\left| \frac{d^j \hat{K}_x((\xi - x')/h_1)}{dx'^j} \right| + \left| \frac{d^j K_{\alpha(x)}((\xi - x')/h_1)}{dx'^j} \right| \right) = O(h_1^{-j}),$$

$$j = 0, 1, \ldots, J + 1. \quad \square$$

LEMMA A.2. *Let $\overline{R}_{25}$ be the remainder term defined in the proof of Lemma* 3.2, *Step* 1. *Then*

$$\sup_{x' \in I(x, C_1 h_1)} \left| \overline{R}_{25}^{(j)}(x') \right| = o_p((nh_1)^{-1/2} h_1^{-j}), \qquad j = 0, 1, \ldots, J. \tag{A.38}$$

PROOF. The first term in the definition of $\overline{R}_{25}$ is a U-statistic of order 2 [cf. Serfling (1980), Chapter 5], which we decompose in the traditional way:

Put $H_1(x; y) = EH(x; y, X)$, $H_2(x; z) = EH(x; X, z)$, $H_0(x) = EH(x; X_1, X_2) = EH_1(x; X) = EH_2(x; X)$ and finally $\tilde{H}(x; y, z) = H(x; y, z) - H_1(x; y) - H_2(x; z) + H_0(x)$. Then

$$
\begin{aligned}
&(nh)^{-2} \sum_{1 \le i, j \le n} \sum H(x; X_i, X_j) \\
&\quad = \frac{n-1}{(nh)^2} \sum_{i=1}^{n} (H_2(x; X_i) - H_0(x)) + \frac{n-1}{(nh)^2} \sum_{i=1}^{n} H_1(x; X_i) \\
&\qquad + (nh)^{-2} \sum_{i \ne j} \sum \tilde{H}(x; X_i, X_j) + (nh)^{-2} \sum_{i=1}^{n} H(x; X_i, X_i).
\end{aligned}
$$

Since $H_1(x; y) = K'(u(x; y)) \int_x^y \check{f}(t)\, dt$, we may write $\overline{R}_{25} = \overline{R}_{251} + \overline{R}_{252} + \overline{R}_{253} + \overline{R}_{254}$, with

$$
\begin{aligned}
\overline{R}_{251}(x) &:= \frac{n-1}{(nh)^2} \sum_{i=1}^{n} H_1(x; X_i) - (nh^2)^{-1} \sum_{i=1}^{n} K'(u(x; X_i)) \int_x^{X_i} \check{f}(t)\, dt \\
&= -(nh)^{-2} \sum_{i=1}^{n} K'(u(x; X_i)) \int_x^{X_i} \check{f}(t)\, dt, \\
\overline{R}_{252}(x) &:= (nh)^{-2} \sum_{i=1}^{n} H(x; X_i, X_i), \\
\overline{R}_{253}(x) &:= \frac{n-1}{(nh)^2} \sum_{i=1}^{n} (H_2(x; X_i) - EH_2(x; X)) \\
&\quad + f(x)^{-1} (nh_1)^{-1/2} W_n\left(x; K_{\alpha(x)} * \hat{K}, h_1\right)
\end{aligned}
$$

and

$$
\overline{R}_{254}(x) := (nh)^{-2} \sum_{i \ne j} \sum \tilde{H}(x; X_i, X_j),
$$

It suffices to prove (A.38) for each of these four terms.

$\overline{R}_{251}$ *and* $\overline{R}_{252}$. These terms are handled in the same way as $\overline{R}_{23}$.

$\overline{R}_{253}$. Integration by parts yields

$$
H_2(x; z) = -h \int K\left(\frac{F(y) - F(x)}{h}\right) \hat{K}\left(\frac{y - z}{h_1}\right) dy / h_1,
$$

so that

$$
\begin{aligned}
\overline{R}_{253}(x) = {} & \frac{n-1}{n}(nh_1)^{-1}\sum_{i=1}^{n}\left(\tilde{K}(x;X_i) - E\tilde{K}(x;X)\right)\\
& + f(x)^{-1}(n^2h_1)^{-1}\sum_{i=1}^{n}\left(K_{\alpha(x)} * \hat{K}\left(\frac{X_i - x}{h_1}\right)\right.\\
& \left. - EK_{\alpha(x)} * \hat{K}\left(\frac{X-x}{h_1}\right)\right)
\end{aligned}
\tag{A.39}
$$

with

$$
\begin{aligned}
\tilde{K}(x;z) &= \frac{h_1}{h^2}H_2(x;z) + f(x)^{-1}K_{\alpha(x)} * \hat{K}\left(\frac{x-z}{h_1}\right)\\
&= -h^{-1}\int\left(K\left(\frac{F(y)-F(x)}{h}\right) - K\left(\frac{f(x)(y-x)}{h}\right)\right)\hat{K}\left(\frac{z-y}{h_1}\right)dy.
\end{aligned}
$$

The first term in (A.39) is treated as $\overline{R}_{21}$, and the second one as $\overline{R}_{23}$.

$\overline{R}_{254}$. This term is a degenerate U-statistic of order 2. Based on moment estimates for such statistics [cf. Serfling (1980), page 183], it is possible to prove

$$
E\left(R_{254}^{(j)}(x + t_2h_1) - R_{n54}^{(j)}(x + t_1h_1)\right)^2 = o(1)(nh_1)^{-1}h_1^{-2j}(t_2 - t_1)^2,
$$

uniformly for $t_1, t_2 \in [-C_1, C_1]$; see Hössjer ad Ruppert (1993a). The relation (A.38) for $\overline{R}_{254}$ then follows from Bickel and Wichura [(1971), (1) and Theorem 1 with $\gamma = \beta = 2$]. □

Acknowledgments. We want to thank an Associate Editor and two referees for valuable comments on an earlier version of this paper.

REFERENCES

ABRAMSON, I. S. (1982). On bandwidth variation in kernel estimation—a square root law. *Ann. Statist.* **10** 1217–1223.

ABRAMSON, I. S. (1984). Adaptive density flattening—a metric distortion principle for combating bias in nearest neighbour methods. *Ann. Statist.* **12** 880–886.

BARTLETT, M. S. (1963). Statistical estimation of density functions. *Sankhyā Ser. A* **25** 245–254.

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

HALL, P. (1990). On the bias of variable bandwidth curve estimators. *Biometrika* **77** 529–535.

HALL, P. and MARRON, J. S. (1988). Variable window width kernel estimates of probability densities. *Probab. Theory Related Fields* **80** 37–49.

HÖSSJER, O. and RUPPERT, D. (1993a). Asymptotics for the transformation kernel density estimator. Report 10, Dept. Mathematical Statistics, Lund Institute of Technology.

HÖSSJER, O. and RUPPERT, D. (1993b). Transformation kernel density estimators using stochastic bandwidths. Report 12, Dept. Mathematical Statistics, Lund Institute of Technology.

JONES, M. C. (1990). Variable kernel density estimates and variable kernel density estimates. *Austral. J. Statist.* **32** 361–371.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic Press, Orlando, FL.

RUPPERT, D. and CLINE, D. (1994). Bias reduction in kernel density estimation by smoothed empirical transformations. *Ann. Statist.* **22** 185–210.

SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SILVERMAN, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

SINGH, R. S. (1977). Estimation of derivatives of a density. *Ann. Statist.* **5** 400–404.

SINGH, R. S. (1979). Mean squared errors of estimates of a density and its derivatives. *Biometrika* **66** 177–180.

DEPARTMENT OF MATHEMATICAL STATISTICS
LUND INSTITUTE OF TECHNOLOGY
BOX 118
S-221 00 LUND
SWEDEN

COLLEGE OF ENGINEERING
E & TC BUILDING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-3801