Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: http://www.tandfonline.com/loi/gnst20

Taylor series approximations of transformation kernel density estimators

Ola HÖssjeri & David Ruppert

To cite this article: Ola HÖssjeri & David Ruppert (1994) Taylor series approximations of transformation kernel density estimators, Journal of Nonparametric Statistics, 4:2, 165-177, DOI: 10.1080/10485259408832608

To link to this article: http://dx.doi.org/10.1080/10485259408832608

Published online: 12 Apr 2007.

Nonparametric Statistics, Vol. 4, pp. 165–177
Reprints available directly from the publisher
Photocopying permitted by license only

TAYLOR SERIES APPROXIMATIONS OF TRANSFORMATION KERNEL DENSITY ESTIMATORS

OLA HÖSSJER[1]* and DAVID RUPPERT[2]†

[1]*Department of Mathematical Statistics, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden,* [2]*School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801, USA*

(Received: May 11, 1993; Revised: November 13, 1993; Accepted January 4, 1994)

We examine the behaviour of a certain kind of transformation based kernel density estimator (TKDE). The transformation is a Taylor series approximation to a smoothed empirical cumulative distribution function computed from a pilot estimate. In this way a whole class of estimators is introduced, with a different number of terms m in the Taylor series. The case $m = 1$ corresponds to a standard varying bandwidth estimator, while the case $m = \infty$ corresponds to a TKDE with the smoothed empirical c.d.f. as transformation. We give an asymptotic expansion for any number of m. When $m = 1, 2$, the rate of convergence is the same as for an ordinary kernel density estimator using a second order kernel, and when $m \geq 3$ the rate of a fourth order kernel is obtained.

KEYWORDS: Bias reduction, higher order kernels, smoothed empirical distribution, Taylor series approximations, transformation of data, variable bandwidths.

1. INTRODUCTION

Suppose that we have an independent sample $X_1, \ldots, X_n$ from a density f. A popular 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 one. Let $\mu_j(K) = \int u^j K(u)\, du$ and define k as the smallest positive integer j with $\mu_j(K) \neq 0$. Then k is referred to as the order of K. If f is k times continuously differentiable at x, the bias of $\hat{f}(x; h)$ is asymptotically equivalent to $\mu_k(K) f^{(k)}(x) h^k / k!$ (Parzen, 1962; Bartlett, 1963; Singh, 1977, 1979). One way of reducing the bias is to use a higher order kernel ($k > 2$). However, it is easy to see that this forces K to take on negative vaues. As a result, the density estimate $\hat{f}$ itself may take on negative values. Even though the problem of negative values may be corrected (Gajek, 1986), the higher order KDE:s have a tendency to have an oscillating appearance, which is undesirable from the practitioners point of view.

Another way of reducing the bias was introduced by Ruppert and Cline (1992)

* Supported by the Swedish Natural Science Research Council, contract F-DP 6689-300.
† Partially supported by NSF Grant DMS-9002791.

and further considered by Hössjer and Ruppert (1993a). Let g be a smooth and increasing function. Transform the data to $Y_i = g(X_i)$, $i = 1, \ldots, n$, and estimate the transformed density

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

by means of a KDE

$$\tilde{f}_Y(y; g, h) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h}\right) \tag{1.1}$$

with a second order kernel. Proceed by transforming back using change-of-variables for probability density functions (p.d.f:s):

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

Note that this automatically produces a *bona fide* density, that is

$$\tilde{f}(x; g, h) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \tilde{f}(x; g, h)\, \mathrm{d}x = 1. \tag{1.3}$$

(When using a higher order KDE, only the second relation in (1.3) is guaranteed.) If g is non-stochastic, the asymptotic behaviour of $\tilde{f}(x; g, h)$ can easily be determined using standard methods for KDE:s. However, in order for $\tilde{f}(x; g, h)$ to have a small bias, it is crucial that g be close to F, the cumulative distribution function (c.d.f.) of the data. (More precisely, the derivatives of g must be close to those of F.)

Suppose g is chosen as a KDE of F, that is

$$g(x) = \hat{F}_1(x) = \int_{-\infty}^{x} \hat{f}_1(t), \quad \text{where} \quad \hat{f}_1(x) = \hat{f}(x; h_1)$$

for some bandwidth h_1. Then the transformation depends on the data, and the asymptotic analysis is significantly complicated. It turns out that $\hat{f}_2(x) = \tilde{f}(x; \hat{F}_1, h_2)$ has the same rate of convergence as a 4th order KDE (Ruppert and Cline, 1992) when h_1 and h_2 are of the same order. A complete asymptotic expansion for $\tilde{f}(x; \hat{F}_1, h_2)$ is obtained by Hössjer and Ruppert (1993a). In both of these papers, further iterates of the TKDE are also considered, and it is shown that the optimal rate of convergence is improved for each transformation performed. However, in this paper we will confine ourselves to one transformation.

We will investigate what happens when the transformation is chosen as

$$g(x') = \hat{F}_{1m}(x') := \sum_{j=1}^{m} \hat{F}_1^{(j)}(x) \frac{(x' - x)^j}{j!}, \tag{1.4}$$

a Taylor series approximation of $\hat{F}_1(\cdot) - \hat{F}_1(x)$ with m terms. (Actually, g depends on x, but this will not be made explicit in the notation.) The motivation for this is twofold:

- To introduce a new class of density estimators (indexed by m) with bias reduction, that are hybrids between well known estimators.
- An exact asymptotic expansion for $m = \infty$ was given by Hössjer and Ruppert (1993a) by methods completely different than those used here. In this paper we use similar techniques as in Ruppert and Cline (1992). We believe that the expansions obtained here shed further light on the results in Hössjer and Ruppert (1993a), by letting $m \to \infty$ (cf. Remark 4).

If $m = 1$ we obtain the local bandwidth KDE (cf. Jones (1991))

$$\tilde{f}(x; \hat{F}_{11}, h_2) = \frac{\hat{f}_1(x)}{nh_2} \sum_{i=1}^{n} K\left(\frac{\hat{f}_1(x)(x - X_i)}{h_2}\right) = \hat{f}\left(x; \frac{h_2}{\hat{f}_1(x)}\right).$$

In fact, $\tilde{f}(x; \hat{F}_{11}, h_2)$ is essentially a nearest neighbour density estimator (cf. e.g. Silverman, 1986, Section 5.2). Even though the Taylor series in (1.4) is not necessarily convergent when $m \to \infty$, we may say formally that $m = \infty$ corresponds to the TKDE with $g = \hat{F}_1$. (The density estimate is unaffected by adding a constant to the transformation g, so we may choose $g(\cdot) = \hat{F}_1(\cdot)$ as well as $g(\cdot) = \hat{F}_1(\cdot) - \hat{F}_1(x)$, the formal limit of (1.4) as $m \to \infty$.) By letting $m = 2, 3, \ldots,$ we obtain a whole class of TKDE:s that can be viewed as hybrids between the cases $m = 1$ and $m = \infty$.

There is one minor problem, however, that needs adjustment. When $1 < m < \infty$, we have no guarantee that the transformation g is monotone. Assuming that K is supported on $[-1, 1]$, we define

$$\hat{f}(x; g, h) = \begin{cases} 0, & g'(x) \le 0 \\ g'(x)(nh)^{-1} \sum_{i=1}^{n} K\left(\frac{g(x) - g(X_i)}{h}\right) I\left(|x - X_i| \le \frac{2h}{g'(x)}\right), & g'(x) > 0 \end{cases} \tag{1.5}$$

as an estimate of $f(x)$.* In this way, if $g(X_i)$ is close to $g(x)$ for values of X_i far away from x, X_i will not contribute to $\hat{f}(x; g, h)$. Since K is supported on $[-1, 1]$, $\hat{f}(x; g, h) = \tilde{f}(x; g, h)$ when $g = \hat{F}_{11}$, whereas if $g = \hat{F}_1$, the two estimators of $f(x)$ are asymptotically equivalent.

In this paper we will give an asymptotic expansion of

$$\hat{f}_{2m}(x) := \hat{f}(x; \hat{F}_{1m}, h_2) \tag{1.6}$$

for $m = 1, 2, \ldots$. One might expect that the asymptotic behaviour of $\hat{f}_{2m}$ gradually approaches that of $\hat{f}_2(x)$ as m increases. This is indeed so (see Remark 4). Quite surprisingly, however, the leading terms in the asymptotic expansion are identical for $m = 1$ and $m = 2$, so nothing is gained by just appending one new term in (1.4). The rate of convergence is the same as for a KDE with a 2nd order kernel when $m = 1, 2$. For higher values of m the convergence rate changes abruptly to that of a 4th order kernel. After that, when increasing m beyond 3 the rate of convergence is unaffected. However, the exact form (but not the rate) of the leading bias terms change until $m = 5$, but from then on it is unchanged. The

* Actually we may replace the factor 2 inside the indicator function in (1.5) by any other constant exceeding 1. Likewise, if K is supported on $[-B, B]$, we may choose any constant greater than B.

exact form of the stochastic term changes gradually, and it differs from the leading stochastic term of $\hat{f}_2(x)$ for all finite values of m.

Notice that the transformation $\hat{F}_{1m}(\cdot)$ depends on x when $1 \le m < \infty$, as opposed to the TKDE transformation $\hat{F}_1$. One might argue that this is a computational drawback, since the transformation has to be recomputed at each x. However, only $\hat{F}_1^{(j)}(\cdot), j = 1, \dots, m$ have to be computed, so the computational burden is not large for small values of m.

The paper is organized as follows. The regularity conditions and some notation is introduced in Section 2, whereas the main result is given in Section 3, together with some remarks. Finally, the more technical parts of the results are collected in the appendix.

2. REGULARITY CONDITIONS AND SOME NOTATION

The following assumptions will be used in Sections 3–5:

(A1) $X_1, \dots, X_n$ is an i.i.d. sample with common density f, x is in the interior of the support of f, $f(x) > 0$, f is l times differentiable in a neighbourhood of x, $l = l(m)$ (with m as defined in (1.4)) and $l(1) = l(2) = 2$, $l(3) = l(4) = 4$ and $l(m) = m - 1$ for $m \ge 5$. If $l(m) = m - 1$, $f^{(m-1)}$ exists and is bounded in a neighbourhood of x.

(A2) The kernel function K is non-negative, symmetric, supported on $[-1, 1]$, integrates to one and is $m - 1 + \varepsilon$ times continuously differentiable, where ε is an arbitrarily small positive number.*

(A3) The bandwidths h_1 and h_2 depend on n, in such a way that $h_2 = O(h_1)$ (but not necessarily $h_1 = O(h_2)$), $h_1 \to 0$ and $nh_2 \to \infty$ as $n \to \infty$.

Let $I(x, r)$ denote the closed real interval $[x - r, x + r]$. Introduce then the zero mean stochastic processes

$$W_n(x; \tilde{K}, h) = (nh)^{-1/2} \sum_{i=1}^{n} \left(\tilde{K}\left(\frac{x - X_i}{h}\right) - E\tilde{K}\left(\frac{x - X}{h}\right)\right) \tag{2.1}$$

and

$$\begin{aligned}\tilde{W}_n(x; g, \tilde{K}, h) = (nh)^{-1/2} \sum_{i=1}^{n} \Bigg(&\tilde{K}\left(\frac{g(x) - g(X_i)}{h}\right) I(|X_i - x| \le 2g'(x)^{-1}h) \\ &- E\left(\tilde{K}\left(\frac{g(x) - g(X)}{h}\right) I(|X - x| \le 2g'(x)^{-1}h)\right)\Bigg)\end{aligned} \tag{2.2}$$

for any function $\tilde{K}$. The function K will always denote the actual kernel function used in the definition of the TKDE whereas $\tilde{K}$ may denote an arbitrary function. Let

$$\alpha = \alpha(x) = \frac{h_2}{f(x)h_1}. \tag{2.3}$$

* For non-integer $r = [r] + \delta$, $0 < \delta < 1$, we say that K is r times continuously differentiable provided $\lim_{u \to t, u \ne t} |K^{([r])}(u) - K^{([r])}(t)|/|t - u|^{\delta}$ is a continuous function of t.

The dependence of α on x (and possibly also on n) will not be made explicit in the notation. Let also

$$K_\eta(\cdot) = K(\cdot/\eta)/\eta \tag{2.4}$$

for any $\eta > 0$.

3. THE MAIN RESULT

The main result may now be stated as follows:

THEOREM 1. Let x be a fixed real number. Given the regularity conditions (A1)–(A3), the density estimators defined in (1.6) satisfy

$$\hat{f}_{2m}(x) = f(x) + b(x; h_1, h_2) + (nh_1)^{-1/2} W_n(x; \tilde{K}, h_1) + R(x), \tag{3.1}$$

with

$$b(x; h_1, h_2) = \frac{\mu_2(K)}{2}\frac{f^{(2)}(x)}{f(x)^2} h_2^2, \quad m = 1, 2 \tag{3.2a}$$

$$= \tilde{b}(x)h_1^2h_2^2 + \frac{\mu_4(K)}{24}\left(\frac{f^{(4)}(x)}{f(x)^4} - \frac{10f^{(3)}(x)f'(x)}{f(x)^5}\right)h_2^4, \quad m = 3 \tag{3.2b}$$

$$= \tilde{b}(x)h_1^2h_2^2 + \frac{\mu_4(K)}{24}\frac{f^{(4)}(x)}{f(x)^4} h_2^4, \quad m = 4 \tag{3.2c}$$

$$= \tilde{b}(x)h_1^2h_2^2, \quad m \geq 5, \tag{3.2d}$$

$$\tilde{b}(x) = \frac{\mu_2(K)^2}{4}\left(\frac{f^{(2)}(x)^2}{f(x)^3} + \frac{3f'(x)f^{(3)}(x)}{f(x)^3} - \frac{3f'(x)^2 f^{(2)}(x)}{f(x)^4} - \frac{f^{(4)}(x)}{f(x)^2}\right)$$
$$= [-f(x) \times (f \circ F^{-1})^{(3)}(u)|_{u=F(x)}] \cdot [\mu_2(K)/2]^2, \tag{3.3}$$

$$\tilde{K}(t) = K_\alpha(t) - \sum_{j=1}^{[(m-1)/2]} \frac{\alpha^{2j}}{(2j)!}\mu_{2j}(K)K^{(2j)}(t) \tag{3.4}$$

and

$$R(x) = o_p(h_2^2 + (nh_2)^{-1/2}), \quad m = 1, 2 \tag{3.5a}$$

$$= o_p(h_2^4 + (nh_2)^{-1/2}), \quad m \geq 3. \tag{3.5b}$$

Before proving the theorem, let us make some remarks:

Remark 1. For $m = 1, 2$ we may express (3.1) in the following alternative way,

$$\hat{f}_{2m}(x) = f(x) + \frac{\mu_2(K)}{2} f^{(2)}(x)\left(\frac{h_2}{f(x)}\right)^2 + \left(\frac{nh_2}{f(x)}\right)^{-1/2} W_n\left(x; K, \frac{h_2}{f(x)}\right) + R(x),$$

by using (2.3), (2.4) and (3.4). This corresponds to a KDE with bandwidth $h_2/f(x)$.

Remark 2. Since $h_2 = O(h_1)$ it follows that the leading bias term in (3.1) is $O(h_2^2)$

for $m = 1, 2$ and $O(h_1^2 h_2^2)$ for $m \geq 3$. At a first glance, it looks like the main stochastic term $(nh_1)^{-1/2} W_n(x; \tilde{K}, h_1)$ in (3.1) only depends on h_1. However, it also depends on h_2, since $\tilde{K}$ contains α in its definition. In fact, it follows from (2.3), (2.4) and (3.4) that the main stochastic term is $O_p((nh_2)^{-1/2})$ for all m. For instance, for $m = 1, 2$, the main stochastic term $(nh_1)^{-1/2} W_n(x; \tilde{K}, h_1)$ equals $(nh_2/f(x))^{-1/2} W_n(x; K, h_2/f(x))$.

Remark 3. Assume that $h_i = c_i h$ with $c_i > 0$, $i = 1, 2$ and $h = n^{-\beta}$ for some $0 < \beta < 1$. Then

$$\begin{aligned} \hat{f}_{2m}(x) - f(x) &= O_p(h^2 + (nh)^{-1/2}) \quad \text{for} \quad m = 1, 2 \\ &= O_p(h^4 + (nh)^{-1/2}) \quad \text{for} \quad m \geq 3. \end{aligned}$$

Hence, the optimal choice of β is

$$\begin{aligned} \beta = 1/5 \quad &\Rightarrow \quad \hat{f}_{2m}(x) - f(x) = O_p(n^{-2/5}), \quad m = 1, 2 \\ \beta = 1/9 \quad &\Rightarrow \quad \hat{f}_{2m}(x) - f(x) = O_p(n^{-4/9}), \quad m \geq 3. \end{aligned}$$

Remark 4. The case $m = \infty$ was treated by Hössjer and Ruppert (1993a). A special case of Theorem 3.1 in that paper is that the expansion (3.1) holds for $\hat{f}_2$, with

$$\begin{aligned} b(x; h_1, h_2) &= \tilde{b}(x) h_1^2 h_2^2, \\ \tilde{K}(t) &= K(t) + K_\alpha(t) - K * K_\alpha(t), \end{aligned}$$

and $*$ denoting the convolution operator. If we make a formal Taylor expansion of K around t we obtain

$$\begin{aligned} K * K_\alpha(t) &= \sum_{j=0}^{\infty} \int K_\alpha(t-s) \frac{(s-t)^j}{j!} K^{(j)}(t)\, ds = \sum_{j=0}^{\infty} K^{(j)}(t) \int K(u) \frac{(-\alpha u)^j}{j!}\, du \\ &= \sum_{j=0}^{\infty} K^{(2j)}(t) \alpha^{2j} \frac{\mu_{2j}(K)}{(2j)!}, \end{aligned}$$

which leads to

$$\tilde{K}(t) = K_\alpha(t) - \sum_{j=1}^{\infty} \frac{\mu_{2j}(K)}{(2j)!} \alpha^{2j} K^{(2j)}(t).$$

Hence, we see that the leading bias term for $\hat{f}_2$ is the same as for $\hat{f}_{2m}$ when $m \geq 5$ and that the kernel is the formal asymptotic limit of (3.4).

Proof of Theorem 1. To use the methods of Ruppert and Cline (1992), we define the following class of transformations:

$$\begin{aligned} \mathcal{G}_m = \{g;\ & g \text{ is an } m\text{th order polynomial with } g(x) = 0, g'(x) \geq 0 \text{ and} \\ & |g^{(k)}(x) - f^{(k-1)}(x)| \leq c_n + c_n^{-1} (nh_1)^{-1/2} h_1^{-(k-1)}, k = 1, 2, \ldots, m\}, \end{aligned} \quad (3.6)$$

where $c_n \to 0$ as $n \to \infty$, so slowly that

$$c_n \gg \max(\omega_l(x; h_1), h_1^{1/2}, (nh_2)^{-1/4}), \quad (3.7)$$

with

$$\omega_l(x;h) := \sup_{x' \in I(x;h)} |f^{(l)}(x') - f^{(l)}(x)|,$$

and $l = l(m)$ is defined in (A.1). The class $\mathscr{G}_m$ actually depends on x, but this will not be made explicit in the notation.

We will tacitly assume that any function g mentioned in the proof belongs to $\mathscr{G}_m$. Suppose that n is so large that (A.4) and (A.5) in the appendix hold. Then (A.4) implies that we can define a local inverse of each $g \in \mathscr{G}_m$. Let

$$f_Y(y;g) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}, \quad y \in [-h_2, h_2]. \tag{3.8}$$

Because of (A.6) we may also assume that $g'(x) > 0$ for each $g \in \mathscr{G}_m$. Then by the definition of $\tilde{W}_n$ in (2.2) and a change of variables $\eta = g(x')/h_2$ we have

$$\begin{aligned}\hat{f}(x;g,h_2) &= \frac{g'(x)}{h_2}\int K\Big(\frac{g(x')}{h_2}\Big) I(|x'-x| \le 2g'(x)^{-1}h_2) f(x')\,\mathrm{d}x' \\ &\quad + (nh_2)^{-1/2} g'(x)\tilde{W}_n(x;g,K,h_2) \\ &= g'(x)\int_{-1}^{1} K(\eta) f_Y(\eta h_2; g)\,\mathrm{d}\eta + (nh_2)^{-1/2} g'(x)\tilde{W}_n(x;g,K,h_2). \end{aligned}\tag{3.9}$$

From now on, we will consider different values of m separately.

$m = 1, 2$

We may expand (3.9) as follows:

$$\hat{f}(x;g,h_2) = f(x) + b(x;h_1,h_2) + (nh_1)^{-1/2} W_n(x;K_\alpha,h_1) + \sum_{k=1}^{3} R_k(x;g), \tag{3.10}$$

with b defined in (3.2)

$$R_1(x;g) = g'(x)\int_{-1}^{1} K(\eta)\Big(f_Y(\eta h_2;g) - \sum_{j=0}^{l(m)} f_Y^{(j)}(0;g)\frac{(\eta h_2)^j}{j!}\Big)\,\mathrm{d}\eta, \tag{3.11}$$

$$R_2(x;g) = g'(x)(nh_2)^{-1/2}\tilde{W}_n(x;g,K,h_2) - (nh_1)^{-1/2} W_n(x;K_\alpha,h_1), \tag{3.12}$$

and

$$R_3(x;g) = g'(x)\Big(f_Y(0;g) + \frac{\mu_2(K)}{2} f_Y^{(2)}(0;g) h_2^2\Big) - (f(x) + b(x;h_1,h_2)). \tag{3.13}$$

By using (3.8) and the identity

$$f_Y^{(2)}(0;g) = \frac{f^{(2)}(x)}{g'(x)^3} - \frac{3f'(x)g^{(2)}(x)}{g'(x)^4} - \frac{f(x)g^{(3)}(x)}{g'(x)^4} + \frac{3f(x)g^{(2)}(x)^2}{g'(x)^5}, \tag{3.14}$$

it is possible to simplify (3.13) to

$$\begin{aligned} R_3(x;g) &\overset{m=1}{=} \Big(\frac{f^{(2)}(x)}{g'(x)^2} - \frac{f^{(2)}(x)}{f(x)^2}\Big)\frac{\mu_2(K)}{2} h_2^2 \\ &\overset{m=2}{=} \Big(\Big(\frac{f^{(2)}(x)}{g'(x)^2} - \frac{f^{(2)}(x)}{f(x)^2}\Big) + \frac{3g^{(2)}(x)}{g'(x)^3}\Big(\frac{f(x)g^{(2)}(x)}{g'(x)} - f'(x)\Big)\Big)\frac{\mu_2(K)}{2} h_2^2. \end{aligned}$$

By standard results for KDEs (cf. the proof of Lemma A.1),

$$R_3(x;\hat{F}_{11}) \overset{m=1}{=} O_p(\hat{f}_1(x)-f(x))h_2^2 = o_p(h_2^2) \tag{3.15a}$$

$$\overset{m=2}{=} O_p(|\hat{f}_1'(x)-f(x)| + |\hat{f}_1'(x)|\,(|\hat{f}_1'(x)-f'(x)| + |\hat{f}_1'(x)|\,|\hat{f}_1(x)-f(x)|))h_2^2$$

$$= o_p(h_2^2+(nh_2)^{-1/2}). \tag{3.15b}$$

The theorem for $m=1,2$ now follows from (3.10), (3.15), Lemma A.1 and Lemma A.3–A.4.

$m=3,4,5$
It follows from (3.9) that

$$\hat{f}(x;g,h_2) = f(x) + b(x;h_1,h_2) + (nh_1)^{-1/2}W_n(x;\tilde{K},h_1) + \sum_{k=1}^{4} R_k(x;g), \tag{3.16}$$

with b, $\tilde{K}$, R_1 and R_2 defined in (3.2), (3.4), (3.11) and (3.12) respectively,

$$R_3(x;g) = g'(x)\Big(f_Y(0;g) + \frac{\mu_2(K)}{2} f_Y^{(2)}(0;g)h_2^2\Big)$$

$$-\,(f(x)+\tilde{b}(x)h_1^2h_2^2) - \frac{\mu_2(K)}{2}\alpha^2(nh_1)^{-1/2}W_n(x;K^{(2)},h_1) \tag{3.17}$$

and

$$R_4(x;g) \overset{m=3}{=} \frac{\mu_4(K)}{24}h_2^4\Big(g'(x)f_Y^{(4)}(0;g) - \Big(\frac{f^{(4)}(x)}{f(x)^4} - \frac{10f^{(3)}(x)f'(x)}{f(x)^5}\Big)\Big) \tag{3.18a}$$

$$\overset{m=4}{=} \frac{\mu_4(K)}{24}h_2^4\Big(g'(x)f_Y^{(4)}(0;g) - \frac{f^{(4)}(x)}{f(x)^4}\Big) \tag{3.18b}$$

$$\overset{m=5}{=} \frac{\mu_4(K)}{24}h_2^4\Big(g'(x)f_Y^{(4)}(0;g) - \frac{(nh_1)^{-1/2}h_1^{-4}W_n(x;K^{(4)},h_1)}{f(x)^4}\Big). \tag{3.18c}$$

Inserting (3.3) and (3.14) into (3.17) yields

$$R_3(x;g) = \frac{\mu_2(K)}{2}h_2^2\bigg(\Big(-\frac{g^{(3)}(x)-f^{(2)}(x)}{g'(x)^2} + \frac{\mu_2(K)}{2}\frac{f^{(4)}(x)}{f(x)^2}h_1^2$$

$$+\frac{(nh_1)^{-1/2}h_1^{-2}W_n(x;K^{(2)},h_1)}{f(x)^2}\Big)$$

$$+\Big(\frac{g^{(3)}(x)}{g'(x)^3}(g'(x)-f(x)) - \frac{\mu_2(K)}{2}\frac{f^{(2)}(x)^2}{f(x)^3}h_1^2\Big)$$

$$+\Big(\frac{3g^{(2)}(x)}{g'(x)^3}(g^{(2)}(x)-f'(x)) - \frac{\mu_2(K)}{2}\frac{3f'(x)f^{(3)}(x)}{f(x)^3}h_1^2\Big)$$

$$+\left(-\frac{3g^{(2)}(x)^2}{g'(x)^4}(g'(x)-f(x))+\frac{\mu_2(K)}{2}\frac{3f'(x)^2f^{(2)}(x)}{f(x)^4}h_1^2\right)\Bigg)$$

$$:=\sum_{k=1}^{4}R_{3k}(x;g).$$

Since f is 4 times continuously differentiable around x we have (cf. (A.2))

$$\hat{f}_1^{(j)}(x)=f^{(j)}(x)+\frac{\mu_2(K)}{2}f^{(j+2)}(x)h_1^2+(nh_1)^{-1/2}h_1^{-j}W_n(x;K^{(j)},h_1)+o(h_1^2),$$

$$j=0,1,2. \quad (3.19)$$

By using (3.19) one finds after some calculations that

$$|R_{3k}(x;\hat{F}_{1m})|=o_p(h_2^4+(nh_2)^{-1/2}),\quad k=1,2,3,4. \quad (3.20)$$

We now turn our attention to R_4. We will need the expansion

$$\begin{aligned}f_Y^{(4)}(0;g)=&\frac{f^{(4)}(x)}{g'(x)^5}-\frac{10f^{(3)}(x)g^{(2)}(x)}{g'(x)^6}-\frac{10f^{(2)}(x)g^{(3)}(x)}{g'(x)^6}\\&+\frac{45f^{(2)}(x)g^{(2)}(x)^2}{g'(x)^7}-\frac{5f'(x)g^{(4)}(x)}{g'(x)^6}+\frac{60f'(x)g^{(2)}(x)g^{(3)}(x)}{g'(x)^7}\\&-\frac{105f'(x)g^{(2)}(x)^3}{g'(x)^8}+\frac{10f(x)g^{(3)}(x)^2}{g'(x)^7}+\frac{15f(x)g^{(2)}(x)g^{(4)}(x)}{g'(x)^7}\\&-\frac{105f(x)g^{(2)}(x)^2g^{(3)}(x)}{g'(x)^8}+\frac{105f(x)g^{(2)}(x)^4}{g'(x)^9}-\frac{f(x)g^{(5)}(x)}{g'(x)^6}.\end{aligned} \quad (3.21)$$

Let

$$F_{1m}(x')=\sum_{j=1}^{m}F^{(j)}(x)\frac{(x'-x)^j}{j!}.$$

It follows then from (3.21) that

$$\begin{aligned}F'_{1m}(x)f_Y^{(4)}(0;F_{1m})&\overset{m=3}{=}\frac{f^{(4)}(x)}{f(x)^4}-\frac{10f^{(3)}(x)f'(x)}{f(x)^5}\\&\overset{m=4}{=}\frac{f^{(4)}(x)}{f(x)^4}\\&\overset{m=5}{=}0.\end{aligned}$$

In order to prove that

$$|\hat{R}_4(x;\hat{F}_{1m})|=o_p(h_2^4+(nh_2)^{-1/2}) \quad (3.22)$$

it thus suffices to establish

$$\begin{aligned}&\hat{F}'_{1m}(x)f_Y^{(4)}(0;\hat{F}_{1m})-F'_{1m}(x)f_Y^{(4)}(0;F_{1m})\\&\overset{m=3,4}{=}o_p(1+(nh_2)^{-1/2}h_2^{-4}) \quad (3.23a)\\&\overset{m=5}{=}f(x)^{-4}(nh_1)^{-1/2}h_1^{-4}W_n(x;K^{(4)},h_1)+o_p(1+(nh_2)^{-1/2}h_2^{-4}). \quad (3.23b)\end{aligned}$$

But formula (3.23) follows, after some computation, from (3.19), (3.21) and the identity

$$\hat{f}_1^{(j)}(x) = f^{(j)}(x) + (nh_1)^{-1/2} h_1^{-j} W_n(x; K^{(j)}, h_1) + o(1), \quad j = 3, 4.$$

Now the theorem follows for $m = 3, 4, 5$ from (3.14), Lemma A.1, Lemma A.3–A.4, (3.20) and (3.22).

$m \geq 6$
The theorem may be proved similarly for higher values of m. For instance, when $m = 6$ (3.16) still holds, with the same remainder terms. Notice however, since $l(6) = 5$, $R_1(x; g)$ in (3.11) is different compared to $m = 3, 4, 5$. When $m = 7$, we have to add

$$R_5(x; g) = \frac{\mu_6(K)}{720} h_2^6 (f_Y^{(6)}(0; g) - f(x)^{-6}(nh_1)^{-1/2} h_1^{-6} W_n(x; K^{(6)}, h_1)),$$

to the sum on the RHS of (3.16). □

APPENDIX

Throughout the appendix, x is a fixed number, whereas x' varies in a neighbourhood of x. The following lemma states that the data-based transformation stays within $\mathcal{G}_m$ with probability tending to 1.

LEMMA A.1. Let $\hat{F}_{1m}$ be the transformation defined in (1.4). Then,

$$P(\hat{F}_{1m} \in \mathcal{G}_m) \to 1 \quad \text{as } n \to \infty. \tag{A.1}$$

Proof. Since $l(m) \geq m - 1$, we have for $k = 1, \ldots, m$,

$$\hat{F}_{1m}^{(k)}(x) = \hat{f}_1^{(k-1)}(x) = \int_{-1}^{1} K(\eta) f^{(k-1)}(x + \eta h_1)\, d\eta + (nh_1)^{-1/2} h_1^{-(k-1)} W_n(x; K^{(k-1)}, h_1). \tag{A.2}$$

Hence, because of (3.7),

$$\begin{aligned} \hat{F}_{1m}^{(k)}(x) - F^{(k)}(x) &= O_p((nh_1)^{-1/2} h_1^{-(k-1)}) \\ &\quad + \int_{-1}^{1} K(\eta)(f^{(k-1)}(x + \eta h_1) - f^{(k-1)}(x))\, d\eta \\ &= o_p(c_n^{-1}(nh_1)^{-1/2} h_1^{-(k-1)}) + o(c_n). \quad \square \end{aligned} \tag{A.3}$$

LEMMA A.2. For large enough n, each $g \in \mathcal{G}_m$ is strictly monotone increasing in a local neighbourhood of x and has a local inverse on $[-h_2, h_2]$. More precisely:

g is strictly monotone on the interval
$$[x - 2\max(g'(x)^{-1}, f(x)^{-1})h_2, x + 2\max(g'(x)^{-1}, f(x)^{-1})h_2] \tag{A.4}$$

and

$$\begin{cases} g(x') > h_2 \text{ for } x + 2\min(g'(x)^{-1}, f(x)^{-1})h_2 \\ \quad < x' < x + 2\max(g'(x)^{-1}, f(x)^{-1})h_2 \\ g(x') < -h_2 \text{ for } x - 2\max(g'(x)^{-1}, f(x)^{-1})h_2 \\ \quad < x' < x - 2\min(g'(x)^{-1}, f(x)^{-1})h_2 \end{cases} \tag{A.5}$$

Proof. By definition of $\mathcal{G}_m$ and (3.7),

$$\lim_{n\to\infty} \sup_{g\in\mathcal{G}_m} |g'(x) - f(x)| = 0. \tag{A.6}$$

We may therefore assume that n is so large that

$$\tfrac{2}{3}f(x) < g'(x) < \tfrac{4}{3}f(x), \quad \forall g \in \mathcal{G}_m.$$

Hence it suffices to show that for all $g \in \mathcal{G}_m$ and n large enough

$$g \text{ is strictly monotone on } [x - 3f(x)^{-1}h_2, x + 3f(x)^{-1}h_2] := I_x \tag{A.7}$$

and

$$\begin{cases} g(x') > h_2 \text{ when } x + \tfrac{3}{2}f(x)^{-1}h_2 < x' < x + 3f(x)^{-1}h_2 \\ g(x') < -h_2 \text{ when } x - 3f(x)^{-1}h_2 < x' < x - \tfrac{3}{2}f(x)^{-1}h_2. \end{cases} \tag{A.8}$$

Suppose we can show that

$$\lim_{n\to\infty} \sup_{\substack{g\in\mathcal{G}_m \\ x'\in I_x}} |g'(x') - f(x)| = 0. \tag{A.9}$$

Then (A.7) follows from (A.9) since $f(x) > 0$. Even (A.8) follows from (A.9) since the latter equation implies that

$$\inf_{\substack{g\in\mathcal{G}_m \\ x'\in I_x}} g'(x') > \tfrac{2}{3}f(x),$$

provided n is large enough. Thus, it remains to prove (A.9). According to (A.1), there exists a neighbourhood U_x of x and number $\bar{f}_m < \infty$ such that $|f^{(m-1)}|$ is bounded by $\bar{f}_m$ on U_x. Therefore, if n is so large that I_x is contained in U_x, the following holds uniformly for any $x' \in I_x$ and $g \in \mathcal{G}_m$;

$$\begin{aligned} |g'(x') - f(x)| &\le |f(x') - f(x)| + \sum_{j=1}^{m} |g^{(j)}(x) - f^{(j-1)}(x)| \frac{|x' - x|^{(j-1)}}{(j-1)!} \\ &\quad + 2\bar{f}_m \frac{|x' - x|^{(m-1)}}{(m-1)!} \\ &\le o(1) + \sum_{j=1}^{m} (c_n + c_n^{-1}(nh_1)^{-1/2}h^{-(j-1)}) \frac{(3f(x)^{-1}h_2)^{j-1}}{(j-1)!} + o(1) \\ &= o(1) + O(c_n + c_n^{-1}(nh_1)^{-1/2}) = o(1), \end{aligned}$$

since $h_2 = O(h_1)$ for all m. □

LEMMA A.3. Let $R_1(x; g)$ be defined as in (3.11). Then

$$\sup_{g \in \mathscr{G}_m} |R_1(x; g)| = \begin{cases} o(h_2^2 + (nh_2)^{-1/2}), & m = 1, 2 \\ o(h_2^4 + (nh_2)^{-1/2}), & m \geq 3. \end{cases} \tag{A.10}$$

Proof. By the definition of $R_1(x; g)$, it suffices to show that

$$\sup_{\substack{g \in \mathscr{G}_m \\ |y| \leq h_2}} |f_Y^{(l(m))}(y; g) - f_Y^{(l(m))}(0; g)| = \begin{cases} o(1 + h_2^{-l(m)}(nh_2)^{-1/2}), & 1 \leq m \leq 5 \\ o(h_2^{4-l(m)} + h_2^{-l(m)}(nh_2)^{-1/2}), & m \geq 5. \end{cases} \tag{A.11}$$

We confine ourselves to $m = 5$. The lemma may be proved similarly for other values of m. Formula (A.11) then reduces to

$$\sup_{\substack{g \in \mathscr{G}_4 \\ |y| \leq h_2}} |f_Y^{(4)}(y; g) - f_Y^{(4)}(0; g)| = o(1 + h_2^{-4}(nh_2)^{-1/2}). \tag{A.12}$$

By (3.21) we may write

$$f^{(4)}(y; g) = \frac{\sum_{j=0}^{4} Q_j(g^{-1}(y); g) f^{(j)}(g^{-1}(y))}{g'(g^{-1}(y))^9}, \tag{A.13}$$

where Q_j is a polynomial in $g', g^{(2)}, \ldots, g^{(5-j)}$. Let $J_x = [x - 2f(x)^{-1}h_2, x + 2f(x)^{-1}h_2]$. From the definition of $\mathscr{G}_4$ we obtain by Taylor expanding g around x that

$$\sup_{\substack{x' \in J_x \\ g \in \mathscr{G}_4}} |g^{(k)}(x')| = O(1 + c_n^{-1}(nh_1)^{-1/2} h_1^{-(k-1)}), \quad k = 1, 2, 3, 4, \tag{A.14}$$

and

$$\sup_{\substack{x' \in J_x \\ g \in \mathscr{G}_4}} |g^{(k)}(x') - g^{(k)}(x)| = \begin{cases} O(h_2 + c_n^{-1}(nh_1)^{-1/2} h_1^{-(k-1)}), & k = 1, 2, 3 \\ 0, & k = 4. \end{cases} \tag{A.15}$$

We also have

$$\sup_{\substack{x' \in J_x \\ g \in \mathscr{G}_4}} |f^{(k)}(x')| = O(1), \quad k = 1, 2, 3, 4, \tag{A.16}$$

and

$$\sup_{\substack{x' \in J_x \\ g \in \mathscr{G}_4}} |f^{(k)}(x') - f^{(k)}(x)| = o(c_n), \quad k = 1, 2, 3, 4. \tag{A.17}$$

We see from (3.21) that $Q_j(x'; g)$ is a linear combination of terms of the form

$$P(x'; g) = \prod_{k=1}^{5-j} g^{(k)}(x')^{\beta_k}, \tag{A.18}$$

where

$$\sum_{k=1}^{5-j} (k-1)\beta_k = 4 - j, \quad j = 0, 1, 2, 3, 4. \tag{A.19}$$

By expanding $f_Y^{(4)}(y; g) - f_Y^{(4)}(0; g)$, making use of (A.13), one sees that formula (A.12) will follow from (A.5), (A.9) and (A.14)–(A.17), if we also show that each function P of the form (A.18) satisfies

$$\sup_{\substack{x' \in J_x \\ g \in \mathscr{G}_4}} |P(x')| = O(1 + c_n^{-1}(nh_2)^{-1/2}h_2^{-4}) \tag{A.20}$$

and

$$\sup_{\substack{x' \in J_x \\ g \in \mathscr{G}_4}} |P(x') - P(x)| = o(1 + h_2^{-4}(nh_2)^{-1/2}). \tag{A.21}$$

Hence, it remains to prove (A.20)–(A.21). Starting with (A.20), we see from (A.14) and (A.19) that

$$\begin{aligned}\sup_{\substack{x' \in J_x \\ g \in \mathscr{G}_4}} |P(x')| &= O\left(\prod_{k=1}^{5-j} (1 + c_n^{-1}(nh_1)^{-1/2}h_1^{-(k-1)})^{\beta_k}\right) \\ &= O(1 + c_n^{-1}(nh_1)^{-1/2}h_1^{-(4-j)}),\end{aligned}$$

since $c_n \gg (nh_1)^{-1/2}$. This and the relation $h_2 = O(h_1)$ implies (A.20) for $j = 0, 1, 2, 3, 4$. Formula (A.21) may be proved in a similar, but more complicated way. □

LEMMA A.4. Let $R_2(x; g)$ be defined as in (3.12). Then

$$\sup_{g \in \mathscr{G}_m} |R_2(x; g)| = o_p((nh_2)^{-1/2}). \tag{A.22}$$

Proof. The proof, which makes use of some multiparameter stochastic process results in Bickel and Wichura (1971), can be found in Hössjer and Ruppert (1993b). □

Acknowledgement

We want to thank M.C. Jones and a referee for valuable comments on the manuscript.

References

Bartlett, M. S. (1963). Statistical estimation of density functions. *Sankhya 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.
Gajek, L. (1986). On improving density estimators which are not bona fide functions. *Ann. Statist.* **14,** 1612–1618.
Hössjer, O. and Ruppert, D. (1993a). Asymptotics for the transformation kernel density estimator. Report 1993:10, Dept. of Mathematical Statistics, Lund Institute of Technology.
Hössjer, O. and Ruppert, D. (1993b). Taylor series approximations of transformation kernel density estimators. Report 1993:12, Dept. of Mathematical Statistics, Lund Institute of Technology.
Jones, M. C. (1991). Variable kernel density estimates and variable kernel density estimates. *Australian J. of Statist.* **32,** 361–371.
Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33,** 1065–1076.
Ruppert, D. and Cline, D. (1992). Bias reduction in kernel density estimation by smoothed empirical transformations. To appear in the *Annals of Statistics.*
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 and New York.
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.