Probab. Theory Relat. Fields 105, 159 – 192 (1996)

Probability Theory and Related Fields

Asymptotic bias and variance for a general class of varying bandwidth density estimators

Ola Hössjer

Lund University, Department of Mathematical Statistics, Box 118, S-221 00 Lund, Sweden

Received: 21 September 1994/In revised form: 7 November 1995

Summary. We consider a general class of varying bandwidth estimators of a probability density function. The class includes the Abramson estimator, transformation kernel density estimator (TKDE), Jones transformation kernel density estimator (JTKDE), nearest neighbour type estimator (NN), Jones–Linton–Nielsen estimator (JLN), Taylor series approximations of TKDE (TTKDE) and Simpson's formula approximations of TKDE (STKDE). Each of these estimators needs a pilot estimator. Starting with an ordinary kernel estimator $\hat{f}_1$, it is possible to iterate and compute a sequence of estimates $\hat{f}_2, \ldots, \hat{f}_t$, using each estimate as a pilot estimator in the next step. The first main result is a formula for the bias order. If the bandwidths used in different steps have a common order $h = h(n)$, the bias of $\hat{f}_k$ is of order $h^{2k \wedge m}$, $k = 1, \ldots, t$. Here h^m is the bias order of the ideal estimator (defined by using the unknown f as pilot). The second main result is a recursive formula for the leading bias and stochastic terms in an asymptotic expansion of the density estimates. If $m < \infty$, it is possible to make $\hat{f}_t$ asymptotically equivalent to the ideal estimator.

Mathematics Subject Classifications (1991): 62G07, 62G20

1 Introduction

Given independent and identically distributed real valued random variables $X_1, \ldots, X_n$ with common distribution F, a well known estimator of the probability density function $f = F'$ at x is the kernel estimator (KDE)

$$\hat{f}_1(x; h_1) = \frac{1}{nh_1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_1}\right),$$

with K a non-negative, symmetric kernel function that integrates to one and h_1 the bandwidth. A disadvantage of $\hat{f}_1$ is that the bandwidth h_1 does not adjust for location. For instance, it is advisable to use a smaller bandwidth at locations where f has a spike, and a larger one at the tails of f. This can be accomplished by letting the bandwidth depend on x and/or the data. A general class of varying bandwidth estimators has the form

$$(1.1)\qquad \hat{f}_2(x;\mathbf{h}_2) = \frac{1}{nh_2}\sum_{i=1}^{n}\hat{\beta}_2(x,X_i;h_1)K\left(\frac{(x-X_i)\hat{\alpha}_2(x,X_i;h_1)}{h_2}\right),$$

with $\mathbf{h}_2 = (h_1,h_2), \hat{\alpha}_2(x,z;h_1) = P_2(x,z;\hat{f}_1(\cdot;h_1))$ and $\hat{\beta}_2(x,z;h_1) = Q_2(x,z;\hat{f}_1(\cdot;h_1))$. Here P_2 and Q_2 are functionals $\mathbb{R}\times\mathbb{R}\times\mathscr{M}\to\mathbb{R}$, with $\mathscr{M}$ an appropriate class of real valued functions on the real line. The effective bandwidth of $\hat{f}_2$ is $h_2/\hat{\alpha}_2(x,X_i;h_1)$ for values of x close to X_i. The quantity $\hat{\alpha}_2$ thus measures how the bandwidth varies with location. The other quantity $\hat{\beta}_2$ is usually close to $\hat{\alpha}_2$, but it can also incorporate a multiplicative correction factor. Examples of estimators within this class are certain versions of nearest neighbour estimators (NN) (originally proposed by Loftsgaarden and Quesenberry 1965) and the transformation kernel density estimator (TKDE) (Ruppert and Cline 1994). These two estimators are usually not formulated as in (1.1). We explain this point a little more in Appendix A. Other examples are the Abramson estimator (Abramson 1982), M.C. Jones' proposed variation of the TKDE (JTKDE) (Hössjer and Ruppert 1993), Taylor series approximations of TKDE (TTKDE) (Hössjer and Ruppert 1994), a Simpson's formula approximation of the TKDE (STKDE), and the Jones–Linton–Nielsen estimator (JLN) (Jones et al. 1995). See Table 1 for details. Strictly speaking, the JTKDE and JLN estimators are based on multiplicative bias reduction methods with effectively constant bandwidths, but they can nevertheless be put into the general framework (1.1). The estimator of Breiman et al. (1977) also belongs to this class. See also Jones (1990) for a comparison of different types of varying bandwidth estimators.

Continuing as in (1.1), we may recursively compute estimates $\hat{f}_2,\ldots,\hat{f}_t$ according to

(1.2)
$$\hat{f}_k(x;\mathbf{h}_k) = \frac{1}{nh_k}\sum_{i=1}^{n}\hat{\beta}_k(x,X_i;\mathbf{h}_{k-1})K\left(\frac{(x-X_i)\hat{\alpha}_k(x,X_i;\mathbf{h}_{k-1})}{h_k}\right),\quad k=2,\ldots,t,$$

with $\mathbf{h}_k = (h_1,h_2,\ldots,h_k)$, $\hat{\alpha}_k(x,z;\mathbf{h}_{k-1}) = P_k(x,z;\hat{f}_{k-1})$ and $\hat{\beta}_k(x,z;\mathbf{h}_{k-1}) = Q_k(x,z;\hat{f}_{k-1})$. (Here $\hat{f}_{k-1}$ means $\hat{f}_{k-1}(\cdot,\mathbf{h}_{k-1})$.) Notice that we allow different functionals P_k and Q_k at each iteration, and $\hat{f}_1$ corresponds to $P_1=1$ and $Q_1=1$. (For technical reasons, the exact definitions of $\hat{f}_k, \hat{\alpha}_k$ and $\hat{\beta}_k$ will be changed slightly in Sect. 5.)

All the functionals considered in this paper have the form

$$(1.3)\qquad \begin{aligned} P_k(x,z;g) &= \sum_{l=0}^{q_k} P_{kl}(x,z;g)(z-x)^l \\ Q_k(x,z;g) &= \sum_{l=0}^{q_k} Q_{kl}(x,z;g)(z-x)^l, \end{aligned}$$

where $P_{kl}(x,z;g)$ and $Q_{kl}(x,z;g)$ depend on $g, g^{(1)}, \ldots, g^{(l)}$. Hence, P_k and Q_k depend on the first q_k derivatives of g. Observe that $q_k = 0$ for all functionals in Table 1 except the TTKDE.

Table 1. Examples of varying density functionals

Estimator	$P(x,z;g)$	$Q(x,z;g)$	$s(k)$
KDE	1	1	2
NN-type	$g(x)$	$g(x)$	2
Abramson	$g(z)^{1/2}$	$g(z)^{1/2}$	$2k \wedge 4$
TKDE	$\frac{1}{z-x}\int_x^z g(v)\,dv$	$g(x)$	$2k$
TTKDE	$\sum_{j=0}^{q} \frac{(z-x)^j}{(j+1)!} g^{(j)}(x)$	$g(x)$	$2k \wedge (2[q/2]+2)$
STKDE	$\frac{1}{6}g(x) + \frac{2}{3}g((x+z)/2) + \frac{1}{6}g(z)$	$g(x)$	$2k \wedge 4$
JTKDE	$\frac{1}{g(x)(z-x)}\int_x^z g(v)\,dv$	1	$2k$
JLN	1	$g(x)/g(z)$	$2k$

The main result of this paper (Theorem 5.1) is an asymptotic expansion

$$\hat{f}_k(x;\mathbf{h}_k) = f(x) + b_k(x;\mathbf{h}_k) + W_k(x;\mathbf{h}_k) + \text{remainders}\,, \tag{1.4}$$

where b_k is the main bias term for the kth step and

$$W_k(x;\mathbf{h}_k) = \frac{1}{n}\sum_{i=1}^{n}(L_k(x,X_i;\mathbf{h}_k) - EL_k(x,X;\mathbf{h}_k)) \tag{1.5}$$

the main stochastic term, and the remainders are asymptotically negligible.

Even though b_k and L_k have been derived in various special cases (see the references in Sect. 5), we give a general formula for computing these quantities. Previous results in the literature also require (P_k, Q_k) to be the same for all k, whereas we allow them to vary with k. The remainder term estimates are derived in L^p-norm uniformly over compact intervals. For the TKDE and JTKDE functionals for instance, this generalizes pointwise results obtained in Hössjer and Ruppert (1993, 1995).

We will refer to L_k as the effective kernel of $\hat{f}_k$, since the stochastic part of $\hat{f}_k$ is essentially the same as for a kernel estimator with kernel L_k. Notice however that L_k may depend on f, so the corresponding kernel estimator may be ideal. Assuming that the bandwidths $h_1, \ldots, h_t$ used in the different steps are of the same order $h = h(n)$ and that f is sufficiently smooth, one consequence of Theorem 5.1 is that $b_k = O(h^{s(k)})$, where the numbers $s(1), \ldots, s(t)$ will be defined in Sect. 2 (see also Table 1 for examples) in terms of the ideal estimators corresponding to $\hat{f}_2, \ldots, \hat{f}_t$. (The ideal estimator $\hat{f}_k^{\text{id}}$ is defined by replacing $\hat{f}_{k-1}$ by f in the definitions of $\hat{\alpha}_k$ and $\hat{\beta}_k$). This means that the bias and variance of $\hat{f}_k$ have the same order of magnitude as for a KDE with a kernel of order $s(k)$. In particular the choice $h(n) = O(n^{-1/(2s(t)+1)})$ implies that the bias and stochastic parts have comparable size at the last iteration. As a consequence, $\hat{f}_t - f = O_p(n^{-s(t)/(2s(t)+1)})$.

The local variation of f around x is crucial for determining $b_k(x;\mathbf{h}_k)$ (as it is for ordinary kernel estimates). On the other hand, f can be considered constant around x when we derive $L_k(x,u;\mathbf{h}_k)$. This simplifies the form of L_k a lot.

Within the framework of our theory, it is possible to prove that if $\hat{f}_t^{\text{id}}$ has nonzero bias, then $\hat{f}_t$ is asymptotically equivalent to $\hat{f}_t^{\text{id}}$, provided t is chosen large enough and that h_t is of smaller order than $h_1,\ldots,h_{t-1}$. This is applicable for the Abramson, TTKDE and STKDE functionals. The resulting estimators have high rates of convergence and simple asymptotic mean squared error (AMSE) formulas. For the Abramson functional, this answers affirmatively an open problem; whether or not it is possible to construct an adaptive estimator that is asymptotically equivalent to the ideal one.

We hasten to add that all results in this paper are asymptotic in nature. Indeed, the work by Marron and Wand (1992) indicates that larger sample sizes are needed for higher order methods before the asymptotic expansions are valid. The finite sample behaviour of many estimators considered in this paper (as well as many others) are investigated by Jones and Signorini (1996).

There is a technical problem with varying bandwidth estimators when $\hat{\alpha}_k$ depends on X_i and becomes small in the tails of f. As a result, many terms in (1.2) will contribute to $\hat{f}_k(x)$, even when X_i is far away from x. This can be overcome by clipping or truncating $\hat{\alpha}_k$ from below away from zero. In this paper, the truncation is taken care of through Conditions (vii) and (viii) in Sect. 5. In fact, we also truncate $\hat{\beta}_k$ from below in the same way as $\hat{\alpha}_k$, to assure that $\hat{f}_k$ has a small bias. A more detailed analysis of clipping is given by Terrell and Scott (1992), Hall et al. (1995) and McKay (1995). Notice that positivity of the estimators is guaranteed even without this truncation for all the functionals in Table 1, as long as K is non-negative.

In Sect. 2 we will define the ideal estimators and bias exponents $s(k)$. The recursive formulas for b_k are defined in Sect. 3, and the ones for L_k in Sect. 4. Regularity conditions and the main result are given in Sect. 5. In Sect. 6 we derive the form b_k and L_k for the examples listed in Table 1. The case of different bandwidth orders and the asymptotic equivalence between $\hat{f}_t$ and $\hat{f}_t^{\text{id}}$ are discussed in Sect. 7. Finally, the proofs are gathered in the appendices.

Throughout the paper C and ε will denote positive numbers whose value may change from line to line. On the other hand, numbered constants like $C_0, C_1, \bar{C}_1, \varepsilon_1, \delta_0$ are considered fixed. We denote the L_p-norm $(E|X|^p)^{1/p}$ by $\|X\|_{L_p}$, and the natural numbers as $\mathbb{N} = \{0,1,2,\ldots\}$. Let g be a real-valued function defined on a subset of $\mathbb{R}^p$, and $\mathbf{j} = (j_1,\ldots,j_{p'}) \in \mathbb{N}^{p'}$, $p' \leqq p$, is a multi-index. Partial derivatives of g are written as $g^{(\mathbf{j})}(\mathbf{y}) := \partial^{|\mathbf{j}|}g(\mathbf{y})/(\partial\mathbf{y}^{\mathbf{j}})$, where $\mathbf{y} = (y_1,\ldots,y_p)$, $|\mathbf{j}| = j_1 + \cdots + j_{p'}$ and $\mathbf{y}^{\mathbf{j}} = y_1^{j_1}\cdots y_{p'}^{j_{p'}}$. For $\Upsilon \subset \mathbb{R}^p$ we put $\|g\|_\Upsilon = \sup_{\mathbf{y}\in\Upsilon}|g(\mathbf{y})|$.

2 Ideal estimator and bias order

In this section we assume that the density $f \in C^{\infty}(\mathbb{R})$ is bounded away from zero in a neighbourhood of x. Given $k \in \{2,\dots,t\}$ and functionals P_k and Q_k, the ideal estimator corresponding to $\hat{f}_k$ is

$$\text{(2.1)} \qquad \hat{f}_k^{\text{id}}(x;h_k) = \frac{1}{nh_k}\sum_{i=1}^{n} \beta_k(x,X_i)K\left(\frac{(x-X_i)\alpha_k(x,X_i)}{h_k}\right), \quad k=2,\dots,t\,,$$

with

$$\text{(2.2)} \qquad \alpha_k(x,z) = P_k(x,z;f)$$

$$\beta_k(x,z) = Q_k(x,z;f)\,.$$

Standard arguments give[1]

$$\text{(2.3)} \qquad \hat{f}_k^{\text{id}}(x;h_k) = f_{bk}^{\text{id}}(x;h_k) + W_k^{\text{id}}(x;h_k) + o_p((nh_k)^{-1/2})\,,$$

with

$$\text{(2.4)} \qquad f_{bk}^{\text{id}}(x;h_k) = \int \beta_k(x,z)K\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right) f(z)\,dz\,,$$

the non-stochastic (or biased) part of $\hat{f}_k^{\text{id}}$,

$$\text{(2.5)} \qquad W_k^{\text{id}}(x;h_k) = \frac{1}{n}\sum_{i=1}^{n} (L_k^{\text{id}}(x,X_i;h_k) - EL_k^{\text{id}}(x,X;h_k))$$

the main stochastic term, and

$$\text{(2.6)} \qquad L_k^{\text{id}}(x,u;h_k) = \frac{\beta_k(x,x)}{h_k}K\left(\frac{(x-u)\alpha_k(x,x)}{h_k}\right)$$

the effective kernel. Using a result of Hall (1990), f_{bk}^{id} has the formal Taylor series expansion

$$\text{(2.7)} \qquad f_{bk}^{\text{id}}(x;h_k) = \sum_{j=0}^{\infty} \gamma_{kj}(x)h_k^j\,,$$

with

$$\text{(2.8)} \qquad \gamma_{kj}(x) = (-1)^j \frac{\mu_j(K)}{j!}\left[\frac{\beta_k(x,z)f(z)}{\alpha_k(x,z)^{j+1}}\right]_{z=x}^{(0,j)}\,,$$

and $\mu_j(K) = \int u^j K(u)\,du$. Assuming that K is an even function, symmetry implies $\gamma_{kj}(x) = 0$ for j odd. Consistency as $h_k \to 0$ requires

$$\text{(2.9)} \qquad f(x) = \gamma_{k0}(x) \iff \alpha_k(x,x) = \beta_k(x,x)\,.$$

[1] Actually, this requires that the clipping problem described in Sect. 1 is taken care of

We define the order of (P_k, Q_k) as

$$(2.10) \qquad m(k) = \max\{j > 0;\ j \text{ even},\ \gamma_{kj}(x) = 0 \text{ for any } f\} + 2\,,$$

with $m(k) = \infty$ if $\gamma_{kj}(x) = 0$ for all even and positive j and $m(k) = 2$ if the set in (2.10) is empty. Notice that $m(k)$ does not depend on x or f, since we vary f over all $C^\infty(\mathbb{R})$-functions with $f(x) > 0$ in (2.10). If we choose another x we may translate the functions f correspondingly.

Even though $m(k)$ was defined in terms of the ideal estimator $\hat{f}_k^{\text{id}}$, it has importance for the bias b_k of $\hat{f}_k$. We will prove in Theorem 5.1 (or, more specifically, in Lemma B.6) that $b_k = O(h^{s(k)})$, where h is the common order of $h_1, \ldots, h_t$ and $\{s(k)\}$ are defined through

$$(2.11) \qquad s(1) = 2 \quad \text{and} \quad s(k) = m(k) \wedge (s(k-1) + 2), \quad k = 2, \ldots, t\,.$$

Let us now give a few examples with $(P_2, Q_2) = (P_t, Q_t) = (P, Q)$, and (P, Q) taken from Table 1. Let us write $\alpha_k = \alpha,\ \beta_k = \beta,\ \gamma_{kj} = \gamma_j,\ m(k) = m$. This implies $s(k) = 2k \wedge m$.

Example 2.1. NN-type estimator: $\alpha(x,z) = \beta(x,z) = f(x)$, $\gamma_j(x) = (-1)^j \mu_j(K) f^{(j)}(x)/j!$, $m = 2$, $s(k) \equiv 2$.

Example 2.2. Abramson estimator: $\alpha(x,z) = \beta(x,z) = f(z)^{1/2}$, $\gamma_j(x) = (-1)^j \mu_j(K)[f(x)^{1-j/2}]^{(j)}/j!$, $m = 4$ and $s(k) = 2k \wedge 4$.

Example 2.3. TKDE estimator: $\alpha(x,z) = (F(z) - F(x))/(z - x)$, $\beta(x,z) = f(x)$, $f_{bk}^{\text{id}}(x; h_k) = f(x) \int K((F(z) - F(x))/h_k)) f(z)\, dz/h_k = f(x) \Rightarrow \gamma_j(x) = 0\ \forall j > 0$, $m = \infty$. This implies $s(k) = 2k$, as found by Ruppert and Cline (1994).

Example 2.4. TTKDE estimator: $\alpha(x,z) = \sum_0^q (z - x)^j f^{(j)}(x)/(j+1)!$, $\beta(x,z) = f(x)$, $\gamma_j(x) = 0$ for $j = 1, \ldots, q$ and $\gamma_{q+1}(x) = (-1)^j \mu_j(K) f^{(q+1)}(x)/(j! f(x)^{q+1})$. Hence, $m = 2[q/2] + 2$ and $s(k) = 2k \wedge (2[q/2] + 2)$. Here $[\cdot]$ denotes the integer part function and $\gamma_j(x)$ is calculated using the fact that $\gamma_{j,\text{TKDE}}(x) = 0$ for $j > 0$ and $\alpha_{\text{TTKDE}}(x,z)$ is defined as a Taylor series expansion (w.r.t. z) of $\alpha_{\text{TKDE}}(x,z)$.

Example 2.5. STKDE estimator: $\alpha(x,z) = f(x)/6 + 2f((x+z)/2)/3 + f(z)/6$, $\beta(x,z) = f(x)$, $\gamma_j(x) = 0$, $j = 1,2,3$ and $\gamma_4(x) = -\mu_4(K) f^{(4)}(x)/(24^2 f(x)^4)$, $m = 4$ and $s(k) = 2k \wedge 4$. Notice that $\gamma_1, \ldots, \gamma_4$ can easily be computed since $[\alpha(x,z)^{(0,j)}]_{z=x} = [\alpha(x,z)^{(0,j)}_{\text{TTKDE}}]_{z=x}$ for $j = 1,2,3$.

Example 2.6. JTKDE estimator: $\alpha(x,z) = (F(z) - F(x))/(f(x)(z - x))$, $\beta(x,z) = 1$, $\gamma_j(x) = 0\ \forall j > 0$, $m = \infty$ and $s(k) = 2k$, as derived by Hössjer and Ruppert (1993).

Example 2.7. JLN estimator: $\alpha(x,z) = 1$, $\beta(x,z) = f(x)/f(z)$, $\gamma_j(x) = (-1)^j \mu_j(K)[d^j f(x)/dz^j]_{z=x}/j! = 0\ \forall j > 0$, $m = \infty$ and $s(k) = 2k$.

Notice also that we may change functionals (P_k, Q_k). If for instance $(P_2, Q_2) =$ Abramson functional and $(P_3, Q_3) =$ TKDE functional we obtain $s(1) = 2$, $s(2) = 4$ and $s(3) = 6$.

3 Recursive formulas for bias

We will now derive recursive formulas for the bias b_k. We first specify the asymptotic expansion (1.4) in more detail. Write

$$\hat{f}_k(x;\mathbf{h}_k) = f_{bk}(x;\mathbf{h}_k) + W_k(x;\mathbf{h}_k) + R_k(x;\mathbf{h}_k)\,, \tag{3.1}$$

with f_{bk} the non-stochastic (biased) part of $\hat{f}_k$ and R_k a stochastic remainder term. The non-stochastic part is expanded as

$$f_{bk}(x;\mathbf{h}_k) = f(x) + b_k(x;\mathbf{h}_k) + r_k(x;\mathbf{h}_k)\,, \tag{3.2}$$

with r_k a non-stochastic remainder term. We will give recursive formulas for f_{bk} and b_k. When $k = 1$, standard asymptotic theory for kernel density estimates yields

$$\begin{aligned} f_{b1}(x;h_1) &= \int K(v)f(x+h_1v)\,dv \\ b_1(x;h_1) &= \tfrac{1}{2}\mu_2(K)f^{(2)}(x)h_1^2 \end{aligned} \tag{3.3}$$

Assume next that we know the form of b_{k-1} for some fixed $k \in \{2,\ldots,t\}$. In order to compute b_k, we first need to find the non-stochastic parts of $\hat{\alpha}_k$ and $\hat{\beta}_k$. These are defined as

(3.4)
$$\begin{aligned} \alpha_{bk}(x,z;\mathbf{h}_{k-1}) &= P_k(x,z;f_{b,k-1}) := \alpha_k(x,z) + b_{\alpha k}(x,z;\mathbf{h}_{k-1}) + r_{\alpha k}(x,z;\mathbf{h}_{k-1}) \\ \beta_{bk}(x,z;\mathbf{h}_{k-1}) &= Q_k(x,z;f_{b,k-1}) := \beta_k(x,z) + b_{\beta k}(x,z;\mathbf{h}_{k-1}) + r_{\beta k}(x,z;\mathbf{h}_{k-1})\,, \end{aligned}$$

with $b_{\alpha k}$ and $b_{\beta k}$ the main bias terms and $r_{\alpha k}$ and $r_{\beta k}$ non-stochastic remainders. Since $\alpha_{bk}(x,z;\mathbf{h}_{k-1}) = P_k(x,z;f_{b,k-1}) \approx P_k(x,z;f+b_{k-1})$, and b_{k-1} is small for large n, we will find $b_{\alpha k}$ through Taylor series expansion of the functional $g \to P_k(x,z;g)$ around $g = f$. Similarly, $b_{\beta k}$ is derived by Taylor expanding $g \to Q_k(x,z;g)$. We say that $g \to P_k(x,z;g)$ has Gateaux derivative $dP_k(x,z;g)$ at $g \in \mathscr{M}$ if for each $\eta \in \mathscr{M}$

$$\lim_{\varepsilon\to 0} \frac{P_k(x,z;g+\varepsilon\eta) - P_k(x,z;g)}{\varepsilon} = dP_k(x,z;g)(\eta)\,, \tag{3.5}$$

with $\eta \to dP_k(x,z;g)(\eta)$ a linear functional. (We refer to Fernholz (1983) for a discussion on Gateaux derivatives and related concepts.) Similarly, dQ_k is defined as the derivative of Q_k. Taking derivatives in (1.3), we obtain

$$\begin{aligned} dP_k(x,z;g)(\eta) &= \sum_{l=0}^{q_k} dP_{kl}(x,z;g)(\eta)(z-x)^l \\ dQ_k(x,z;g)(\eta) &= \sum_{l=0}^{q_k} dQ_{kl}(x,z;g)(\eta)(z-x)^l\,. \end{aligned} \tag{3.6}$$

Table 2 lists dP_k and dQ_k for the functionals from Table 1. Taylor expansion of P_k and Q_k now gives

$$(3.7)\qquad \begin{aligned} b_{\alpha k}(x,z;\mathbf{h}_{k-1}) &= dP_k(x,z;f)(b_{k-1}) \\ b_{\beta k}(x,z;\mathbf{h}_{k-1}) &= dQ_k(x,z;f)(b_{k-1}), \end{aligned}$$

with $b_k = b_k(\,\cdot\,;\mathbf{h}_k)$. Next, f_{bk} is computed recursively from α_{bk} and β_{bk} (cf. (2.4)),[2]

$$(3.8)\qquad f_{bk}(x;\mathbf{h}_k) = \frac{1}{h_k}\int \beta_{bk}(x,z;\mathbf{h}_{k-1})K\left(\frac{(x-z)\alpha_{bk}(x,z;\mathbf{h}_{k-1})}{h_k}\right)f(z)\,dz\,,$$

and b_k recursively from $b_{\alpha k}$ and $b_{\beta k}$ according to

$$(3.9)\qquad b_k(x;\mathbf{h}_k) = b_k^{\mathrm{id}}(x;h_k) + b_k^{\mathrm{ad}}(x;\mathbf{h}_k)\,,$$

Table 2. Examples of functional derivatives

Estimator	$dP(x,z;g)(\eta)$	$dQ(x,z;g)(\eta)$
KDE	0	0
NN-type	$\eta(x)$	$\eta(x)$
Abramson	$\frac{\eta(z)}{2g(z)^{1/2}}$	$\frac{\eta(z)}{2g(z)^{1/2}}$
TKDE	$\frac{1}{z-x}\int_x^z \eta(v)\,dv$	$\eta(x)$
TTKDE	$\sum_{j=0}^{q}\frac{(z-x)^j}{(j+1)!}\eta^{(j)}(x)$	$\eta(x)$
STKDE	$\frac{1}{6}\eta(x)+\frac{2}{3}\eta((x+z)/2)+\frac{1}{6}\eta(z)$	$\eta(x)$
JTKDE	$\frac{\int_x^z \eta(v)\,dv}{g(x)(z-x)} - \frac{\int_x^z g(v)\,dv}{(z-x)g(x)^2}\eta(x)$	0
JLN	0	$\frac{\eta(x)}{g(z)} - \frac{g(x)\eta(z)}{g(z)^2}$

where $b_k^{\mathrm{id}}(x;h_k) := \gamma_{k,s(k)}(x)h_k^{s(k)}$ comes from the ideal estimator (cf. (2.8)) and b_k^{ad} is the adaptive correction term. It has the form (see Lemma B.6 for a derivation)

(3.10)

$$b_k^{\mathrm{ad}}(x;\mathbf{h}_k) = 0, \quad s(k) = s(k-1)$$

$$b_k^{\mathrm{ad}}(x;\mathbf{h}_k) = \frac{\mu_2(K)}{2}\left[\frac{b_{\beta k}(x,z;\mathbf{h}_{k-1})f(z)}{\alpha_k(x,z)^3}\right]_{z=x}^{(0,2)} h_k^2 - \frac{3\mu_2(K)}{2} \times \left[\frac{b_{\alpha k}(x,z;\mathbf{h}_{k-1})\beta_k(x,z)f(z)}{\alpha_k(x,z)^4}\right]_{z=x}^{(0,2)} h_k^2, \quad s(k) = s(k-1)+2\,.$$

The ideal estimator corresponds to $b_{\alpha k} = b_{\beta k} \equiv 0$, and hence $f_{bk} = f_{bk}^{\mathrm{id}}$ and $b_k^{\mathrm{ad}} = 0$. Notice that the adaptive bias term vanishes when $s(k) = s(k-1)$ and the ideal bias term vanishes when $m(k) > s(k-1)+2$. Equations (3.4)

[2] The domain of integration in (3.8) is actually a subset of $\mathbb{R}$ to avoid tail effects (cf. Lemma B.6.)

and (3.8) together give f_{bk} in terms of $f_{b,k-1}$, and Eqs. (3.7), (3.9) and (3.10) b_k in terms of b_{k-1}. The recursive bias formulae will be exemplified in Sect. 6.

4 Recursive formula for the effective kernels

When computing the effective kernels $L_1,\dots,L_t$, we ignore the local variation of $f, \alpha_k(\cdot,\cdot)$ and $\beta_k(\cdot,\cdot)$ around x and (x,x) respectively. Asymptotically, this variation is only of secondary importance, so simpler kernels can be obtained by neglecting it. The cost of this simplification is larger remainder terms (intuitively, we have no theoretical result comparing the remainder terms) and some extra technicalities to define them. Let x' and u be numbers close to x. Given x and $\mathbf{h}_k$, define $(x',u) \to \bar{L}_k(x',u,x;\mathbf{h}_k)$ as the effective kernel we obtain at x' if $f(\cdot)$ is replaced by $f_x(\cdot) \equiv f(x)$, $\alpha_j(\cdot,\cdot)$ by $\alpha_j(x,x)$ and $\beta_j(\cdot,\cdot)$ by $\beta_j(x,x)$ for all $j \leqq k$. The extra x-argument of $\bar{L}_k$ indicates that this replacement depends on x. In analogy with (1.5), put also

$$\bar{W}_k(x',x;\mathbf{h}_k) = \frac{1}{n}\sum_{i=1}^{n}(\bar{L}_k(x',X_i,x;\mathbf{h}_k) - E\bar{L}_k(x',X,x;\mathbf{h}_k))\,. \tag{4.1}$$

After having computed $\bar{L}_k$, we put

$$L_k(x,u;\mathbf{h}_k) = \bar{L}_k(x,u,x;\mathbf{h}_k)\,. \tag{4.2}$$

For $k=1$, the local variation of f makes no difference, so we have

$$\bar{L}_1(x',u,x;h_1) = L_1(x',u;h_1) = \frac{1}{h_1}K\left(\frac{x'-u}{h_1}\right). \tag{4.3}$$

Suppose now that $\bar{L}_{k-1}$ has been computed for some $k \in \{2,\dots,t\}$. In order to find $\bar{L}_k$, we need asymptotic expansions of $\hat{\alpha}_k$ and $\hat{\beta}_k$:

$$\begin{aligned} \hat{\alpha}_k(x,z;\mathbf{h}_{k-1}) &= \alpha_{bk}(x,z;\mathbf{h}_{k-1}) + \bar{W}_{\alpha k}(x,z,x;\mathbf{h}_{k-1}) + R_{\alpha k}(x,z;\mathbf{h}_{k-1})\,, \\ \hat{\beta}_k(x,z;\mathbf{h}_{k-1}) &= \beta_{bk}(x,z;\mathbf{h}_{k-1}) + \bar{W}_{\beta k}(x,z,x;\mathbf{h}_{k-1}) + R_{\beta k}(x,z;\mathbf{h}_{k-1})\,, \end{aligned} \tag{4.4}$$

with $\bar{W}_{\alpha k}$ and $\bar{W}_{\beta k}$ main stochastic terms, $R_{\alpha k}$ and $R_{\beta k}$ stochastic remainders and α_{bk} and β_{bk} the non-stochastic parts, defined in (3.4). Here $\bar{W}_{\alpha k}(\cdot,\cdot,x;\mathbf{h}_{k-1})$ and $\bar{W}_{\beta k}(\cdot,\cdot,x;\mathbf{h}_{k-1})$ are computed by ignoring the local variation of f, α_j and β_j $(j \leqq k)$ around x. According to (3.1) we have $\hat{\alpha}_k(x,z;\mathbf{h}_{k-1}) \approx P_k(x,z;f_{b,k-1} + W_{k-1}(\cdot;\mathbf{h}_{k-1})) \approx P_k(x,z;f_{b,k-1} + \bar{W}_{k-1}(\cdot,x;\mathbf{h}_{k-1}))$. The last approximation follows, as we only consider W_{k-1} restricted to a small neighbourhood of x. Since $\bar{W}_{k-1}$ is small and $f_{b,k-1}$ close to f_x for x' around x and large n, we define

$$\begin{aligned} \bar{W}_{\alpha k}(x',z,x;\mathbf{h}_{k-1}) &= dP_k(x',z;f_x)(\bar{W}_{k-1}(\cdot,x,\mathbf{h}_{k-1}))\,, \\ \bar{W}_{\beta k}(x',z,x;\mathbf{h}_{k-1}) &= dQ_k(x',z;f_x)(\bar{W}_{k-1}(\cdot,x,\mathbf{h}_{k-1}))\,, \end{aligned} \tag{4.5}$$

Define then

$$\begin{aligned} \bar{L}_{\alpha k}(x',z,u,x;\mathbf{h}_{k-1}) &= dP_k(x',z;f_x)(\bar{L}_{k-1}(\cdot,u,x;\mathbf{h}_{k-1}))\,, \\ \bar{L}_{\beta k}(x',z,u,x;\mathbf{h}_{k-1}) &= dQ_k(x',z;f_x)(\bar{L}_{k-1}(\cdot,u,x;\mathbf{h}_{k-1}))\,. \end{aligned} \tag{4.6}$$

By the linearity of dP_k and dQ_k, it follows from (4.1),(4.5) and (4.6) that

$$(4.7)\quad \bar{W}_{\alpha k}(x',z,x;\mathbf{h}_{k-1}) = \frac{1}{n}\sum_{i=1}^{n}(\bar{L}_{\alpha k}(x',z,X_i,x;\mathbf{h}_{k-1}) - E\bar{L}_{\alpha k}(x',z,X,x;\mathbf{h}_{k-1}))\,,$$
$$\bar{W}_{\beta k}(x',z,x;\mathbf{h}_{k-1}) = \frac{1}{n}\sum_{i=1}^{n}(\bar{L}_{\beta k}(x',z,X_i,x;\mathbf{h}_{k-1}) - E\bar{L}_{\beta k}(x',z,X,x;\mathbf{h}_{k-1}))\,,$$

Here $\bar{L}_{\alpha k}$ and $\bar{L}_{\beta k}$ can be interpreted as the effective kernels corresponding to $\hat{\alpha}_k$ and $\hat{\beta}_k$. Observe that the expansions for $\bar{W}_{\alpha k}$ and $\bar{W}_{\beta k}$ in (4.7) are analogous to the expansion (4.1) for $\bar{W}_k$.

Table 3 displays functional derivatives when the local variation of g around x is ignored ($g_x \equiv g(x)$). We will show (Lemma B.5) that $\bar{L}_k$ can be computed from $\bar{L}_{\alpha k}$ and $\bar{L}_{\beta k}$ according to

$$(4.8)\quad \bar{L}_k(x',u,x;\mathbf{h}_k) = \frac{\alpha_k(x,x)}{h_k}K\left(\frac{(x'-u)\alpha_k(x,x)}{h_k}\right)$$
$$+\frac{f(x)}{h_k}\int \bar{L}_{\alpha k}(x',z,u,x;\mathbf{h}_{k-1})\check{K}\left(\frac{(x'-z)\alpha_k(x,x)}{h_k}\right)dz$$
$$+\frac{f(x)}{h_k}\int \bar{L}_{\beta k}(x',z,u,x;\mathbf{h}_{k-1})K\left(\frac{(x'-z)\alpha_k(x,x)}{h_k}\right)dz$$
$$:= \bar{L}_k^{\mathrm{id}}(x',u,x;h_k) + \sum_{v=1}^{2}\bar{L}_k^{\mathrm{ad},v}(x',u,x;\mathbf{h}_k)\,,$$

for $k = 2,\dots,t$ and $\check{K}(v) = vK'(v)$. Equations (4.6) and (4.8) together give $\bar{L}_k$ in terms of $\bar{L}_{k-1}$. This recursive scheme will be exemplified in Sect. 6. We see that $\bar{L}_k$ can be decomposed into an ideal and adaptive part, the first term in (4.8) representing the ideal part, and the last two the adaptive part. Notice that the ideal part here agrees with L_k^{id} in Sect. 2 (when $x' = x$) because of (2.9). The adaptive part is derived from U-statistics theory. The reason is that when (4.4) and (4.7) are inserted into (1.2), we obtain double sums after linearization.

Table 3. Examples of functional derivatives, local variation of g ignored

Estimator	$dP(x',z;g_x)(\eta)$	$dQ(x',z;g_x)(\eta)$
KDE	0	0
NN-type	$\eta(x')$	$\eta(x')$
Abramson	$\frac{\eta(z)}{2g(x)^{1/2}}$	$\frac{\eta(z)}{2g(x)^{1/2}}$
TKDE	$\frac{1}{z-x'}\int_{x'}^{z}\eta(v)\,dv$	$\eta(x')$
TTKDE	$\sum_{j=0}^{q}\frac{(z-x')^j}{(j+1)!}\eta^{(j)}(x')$	$\eta(x')$
STKDE	$\frac{1}{6}\eta(x')+\frac{2}{3}\eta((x'+z)/2)+\frac{1}{6}\eta(z)$	$\eta(x')$
JTKDE	$\frac{\int_{x'}^{z}\eta(v)\,dv}{g(x)(z-x')} - \frac{\eta(x')}{g(x)}$	0
JLN	0	$\frac{\eta(x')}{g(x)} - \frac{\eta(z)}{g(x)}$

5 Regularity conditions and main results

Before giving the main result (Theorem 5.1), we state a number of regularity conditions.

(i) The bandwidths $h_1 = h_1(n), \dots, h_t = h_t(n)$ are all of the same order as $n \to \infty$, i.e. for some $0 < C_0 \leqq 1$ and sequence $h = h(n)$, $C_0 \leqq h_k/h \leqq C_0^{-1}$ for all n and $k = 1, \dots, t$.

(ii) There exists a $\varepsilon_0 > 0$ such that $hn^{\varepsilon_0} \to 0$ and $hn^{1-\varepsilon_0} \to \infty$ as $n \to \infty$.

(iii) The bias exponents defined in (2.11) satisfy $s(1) \leqq s(2) \leqq \cdots \leqq s(t)$.

(iv) Let $\Omega = [\omega_1, \omega_2]$ be a closed interval and put $\Omega^\delta = [\omega_1 - \delta, \omega_2 + \delta]$. Then, for some $\delta_0 > 0$, $\|f^{(j)}\|_{\Omega^{\delta_0}} < \infty$ for $j = 0, 1, \dots, s(t) + 1 + \sum_{k=2}^{t} q_k$ and $\inf_{x \in \Omega^{\delta_0}} f(x) = \underline{f} > 0$.

(v) The kernel K is non-negative, symmetric and supported on $[-C_1, C_1]$ for some $0 < C_1 < \infty$. In addition, $\mu_0(K) = 1$ and K has $(3 \wedge 2 + \sum_{k=2}^{t} q_k)$ bounded derivatives.

(vi) $P_k^{(i,j)}(x,z;g)$ and $Q_k^{(i,j)}(x,z;g)$ depend only on $g, \dots, g^{(i+j)}$ restricted to $[x,z]$, $2 \leqq k \leqq t$. In particular, $P_{k0}(x,x;g) = U_k(g(x))$ for some $U_k : \mathbb{R} \to \mathbb{R}$. The function U_k is strictly positive and non-decreasing on the positive real line. This implies $\min_{2 \leqq k \leqq t} \inf_{x \in \Omega^{\delta_0}} \alpha_k(x,x) \geqq \min_{2 \leqq k \leqq t} U_k(\underline{f}) := \underline{\alpha} > 0$.

(vii) $\hat{\alpha}_k(x,z;\mathbf{h}_{k-1}) = P_k(x,z;\tilde{f}_{k-1})$ and $\hat{\beta}_k(x,z;\mathbf{h}_{k-1}) = Q_k(x,z;\tilde{f}_{k-1})$ where $\tilde{f}_{k-1} = \xi \circ \hat{f}_{k-1}$. The function $\xi : [0,\infty) \to [0,\infty)$ is non-decreasing, $\xi(0) := \xi_0 > 0$ and $\xi(v) = v$ for $v \geqq \xi_1$, with $\xi_0 < \xi_1 < \underline{f}$. Finally, ξ has $\sum_{k=2}^{t} q_k$ bounded derivatives.

(viii) $\hat{f}_k(x;\mathbf{h}_k) = \sum_{i=1}^{n} \tilde{\beta}_k(x,X_i;\mathbf{h}_{k-1}) K((x - X_i)\tilde{\alpha}_k(x,X_i;\mathbf{h}_{k-1})/h_k)/(nh_k)$, where $\tilde{\alpha}_k = \chi \circ \hat{\alpha}_k$ and $\tilde{\beta}_k = \chi \circ \hat{\beta}_k$. The function $\chi : [0,\infty) \to [0,\infty)$ is non-decreasing, $\chi(0) := \chi_0 > 0$ and $\chi(v) = v$ for $v \geqq \chi_1$, with $\chi_0 < \chi_1 < \underline{\alpha}$. Finally, χ has $\sum_{k=2}^{t} q_k$ bounded derivatives.

(ix) $P_k(x,x;g) = Q_k(x,x;g)$, and hence $dP_k(x,x;g)(\eta) = dQ_k(x,x;g)(\eta)$ as soon as the derivative exists.

(x) Suppose $g : \mathbb{R} \times \Theta \to \mathbb{R}$ for some $\Theta \subset \mathbb{R}^p$, and put $h_{kl}(x,z;\boldsymbol{\theta}) = P_{kl}(x,z;g(\cdot;\boldsymbol{\theta}))$. Let $\Xi \subset \mathbb{R}^2$ and $P(\Xi) \subset \mathbb{R}$ be defined by $P(\Xi) = \bigcup_{(x,z) \in \Xi} [x,z]$. Then, if g is bounded from below away from zero on $P(\Xi) \times \Theta$,

$$\|h_{kl}^{(i,j,\mathbf{d})}\|_{\Xi \times \Theta} \leqq C \sum_{\{\mathbf{v}_\mu\}} \prod_{\mu} \|g^{(\mathbf{v}_\mu)}\|_{P(\Xi) \times \Theta}, \tag{5.1}$$

for some finite constant C, where the sum ranges over all finite sequences $\{\mathbf{v}_\mu\}$, of vectors $v_\mu = (v_{\mu 1}, \mathbf{v}_{\mu 2}) \in \mathbb{N}^{p+1}$ with $\sum_\mu v_{\mu 1} = l + i + j$ and $\sum_\mu \mathbf{v}_{\mu 2} = \mathbf{d}$; and at most one $\mathbf{v}_\mu = 0$. The constant C may depend on the lower bound of g as well as the functional P_{kl}. Formula (5.1) is also true if P_{kl} is replaced by Q_{kl} in the definition of h_{kl}.

According to (i), all bandwidths have to be of the same order. This condition is somewhat restrictive, but it will be relaxed in Sect. 7.

We require K to have compact support in (v). This could be weakened to exponentially decaying tails (e.g. Gaussian or logistic K), by approximating such a kernel with a smoothly truncated kernel having compact support.

Condition (vi) can be verified for each of the functionals in Table 1. For instance, the Abramson functional has $U_k(y) = y^{1/2}$ and $P_k^{(i,j)}(x,z;g) = Q_k^{(i,j)}(x,z;g) = d^j g^{1/2}(z)/dz^j$, which only depends on $g(z),\ldots,g^{(j)}(z)$.

The definitions of $\hat{\alpha}_k$ and $\hat{\beta}_k$ in (vii), and of $\hat{f}_k$ in (viii), differ from the ones given in Sect. 1, since $\hat{f}_{k-1}$ is truncated from below by $\xi(\,\cdot\,)$ and $\hat{\alpha}_k$ and $\hat{\beta}_k$ from below by $\chi(\,\cdot\,)$. The first truncation is done to avoid derivatives of terms like $P_{kl}^{(i,j)}(x,z;\hat{f}_{k-1})$ becoming too large. The truncation of $\hat{\alpha}_k$ guarantees that only X_i close to x contribute to $\hat{f}_k(x;\mathbf{h}_k)$. (Actually, for the Abramson, TKDE and NN-functionals in Table 1, a truncation of $\hat{f}_{k-1}$ from below automatically gives a truncation of $\hat{\alpha}_k$ and $\hat{\beta}_k$ from below.)

Notice that (ix) implies (2.9), and also

$$\alpha_{bk}(x,x;\mathbf{h}_{k-1}) = \beta_{bk}(x,x;\mathbf{h}_{k-1})\,. \tag{5.2}$$

Whereas (2.9) is necessary for consistency of $\hat{f}_k$, (5.2) is a higher order analogue which guarantees that b_k^{ad} in (3.10) is of smaller order than $b_{k-1}, b_{\alpha k}$ and $b_{\beta k}$ (this will be seen in Lemma B.6).

Condition (x) can be viewed as a kind of product-rule of differentiation for P_{kl} and Q_{kl}. It gives smoothness conditions on $P_{kl}^{(i,j,\mathbf{d})}$ and $Q_{kl}^{(i,j,\mathbf{d})}$, in particular how these functions depend on $g^{(v_1,\mathbf{v}_2)}$, $v_1 \leqq l+i+j$, $|\mathbf{v}_2| \leqq |\mathbf{d}|$, restricted to the set $P(\Xi)\times\Theta$. If Θ is a single point, $\Xi = \{(x,z)\}$ and $P(\Xi) = [x,z]$, (x) states that $P_{kl}^{(i,j)}(x,z;g)$ and $Q_{kl}^{(i,j)}(x,z;g)$ only depend on g and its partial derivatives up to order $l+i+j$ restricted to the interval $[x,z]$. Some consequences of (x), important in the proofs, are given in Appendix D.

Theorem 5.1 *Assume* (i)–(x). *Then, for* $k = 1,\ldots,t$,

$$\hat{f}_k(x;\mathbf{h}_k) = f(x) + b_k(x;\mathbf{h}_k) + W_k(x;\mathbf{h}_k) + r_k(x;\mathbf{h}_k) + R_k(x;\mathbf{h}_k)\,, \tag{5.3}$$

with $b_1,\ldots,b_t$ *defined recursively in* (3.3), (3.7), (3.9) *and* (3.10), W_k *defined in* (1.5) *and* $L_1,\ldots,L_t$ *defined recursively in* (4.2), (4.6) *and* (4.8). *Finally,* r_k *and* R_k *are remainder terms, defined in* (3.2) *and* (3.1) *respectively, with*

$$\sup_{x\in\Omega} |r_k(x;\mathbf{h}_k)| = o(h^{s(k)})\,, \tag{5.4}$$

and

$$\left\| \sup_{x\in\Omega} |R_k(x;\mathbf{h}_k)| \right\|_{L^p} = O((nh)^{-1/2} n^{-\varepsilon}) \quad \forall p > 0 \tag{5.5}$$

for some $\varepsilon > 0$.

Remark. 5.1. We may also allow stochastic bandwidths $\hat{h}_1,\ldots,\hat{h}_t$. Typically, $\hat{h}_k$ is then an estimator of h_k, with h_k chosen optimally according to some risk criterion. In such cases, h_k depends on f and has to be estimated. It is possible to extend Theorem 5.1 to this case, using techniques of Hall and Marron (1988). Since this would include extra technicalities we have not included these calculations in the paper.

Remark. 5.2. The requirement of $s(t)+1+\sum_{k=2}^{t} q_k$ derivatives in (iv) can be lowered to $s(t)+\varepsilon+\sum_{k=2}^{t} q_k$ derivatives for any $\varepsilon > 0$, provided we sharpen (x) to hold also for derivatives of non-integral order. In the proof of Theorem 5.1 in the appendix, we just change $\bar{J}_k + 1$ in (B.11) and (B.15) to $\bar{J}_k + \varepsilon$ and $J_0 - q_k + 1$ in (B.8) to $J_0 - q_k + \varepsilon$.

6 Examples of bias and effective kernels

Assume $P_2 = \cdots = P_t = P$ and $Q_2 = \cdots = Q_t = Q$ throughout this section, with P and Q taken from Table 1. We also write $\alpha_k = \alpha$, $\beta_k = \beta$, $\gamma_{kj} = \gamma_j$, $m(k) = m$, $q_k = q$ for $k \in \{2,\ldots,t\}$, and we put $K_h(v) = K(v/h)/h$. To simplify notation, we will omit $\mathbf{h}_k$ as argument and also x for the effective kernels, so $b_k(x;\mathbf{h}_k) = b_k(x)$, $\bar{L}_{\alpha k}(x',z,u,x;\mathbf{h}_{k-1}) = \bar{L}_{\alpha k}(x',z,u)$ and so on. All the effective kernels we study have the form

$$
\begin{aligned}
(6.1)\qquad \bar{L}_k(x',u,x;\mathbf{h}_k) &= \bar{K}_k(x'-u)\,,\\
L_k(x,u;\mathbf{h}_k) &= \bar{K}_k(x-u)\,,
\end{aligned}
$$

where $\bar{K}_k$ may depend on x, but this will not be made explicit in the notation. Notice that $\bar{K}_1 = K_{h_1}$. To simplify the exposition, we have provided $\bar{K}_2$ for all estimators in Table 4.

Table 4. Effective kernel $\bar{K}_2$ after first iteration, with $\tilde{K}(v) = vK'(v) + K(v)$ and $\check{K}(v) = vK'(v)$

Estimator	$\bar{K}_2$
KDE	K_{h_2}
NN-type	$K_{h_2/f(x)}$
Abramson	$K_{h_2/f(x)^{1/2}} + K_{h_1} * \tilde{K}_{h_2/f(x)^{1/2}}/2$
TKDE	$K_{h_1} + K_{h_2/f(x)} - K_{h_1} * K_{h_2/f(x)}$
TTKDE	$K_{h_2/f(x)} - \sum_{l=1}^{q} \mu_l(K)(K^{(l)})_{h_1} h_2^l/(l!\,f(x)^l h_1^l)$
STKDE	$K_{h_2/f(x)} + 5K_{h_1}/6 + 2K_{h_1} * \check{K}_{h_2/(2f(x))}/3 + K_{h_1} * \check{K}_{h_2/f(x)}/6$
JTKDE	$K_{h_1} + K_{h_2} - K_{h_1} * K_{h_2}$
JLN	$K_{h_1} + K_{h_2} - K_{h_1} * K_{h_2}$

Example 6.1. NN-type: Example 2.1 and Eqs. (3.9) and (3.10) give $b_k(x) = \mu_2(K) f^{(2)}(x) h_k^2/(2f(x)^2)$ for $k \geqq 2$. By (4.6), $\bar{L}_{\alpha k}(x',z,u) = \bar{L}_{\beta k}(x',z,u) = \bar{K}_{k-1}(x'-u)$, so (4.8) implies $\bar{K}_k = K_{h_k/f(x)}$ for $k \geqq 2$.

Example 6.2. Abramson: $b_{\alpha k}(x,z) = b_{\beta k}(x,z) = b_{k-1}(z)/(2f(z)^{1/2})$, so the adaptive bias term becomes $b_k^{\mathrm{ad}}(x) = -\mu_2(K)[b_{k-1}(x)/f(x)]^{(2)} h_k^2/2$. Combining this with Example 2.2 we obtain $b_2(x) = \mu_4(K)[1/f(x)]^{(4)}h_2^4/24 - \mu_2(K)^2[f^{(2)}(x)/f(x)]^{(2)}h_1^2h_2^2/4$ and $b_k(x) = \mu_4(K)[1/f(x)]^{(4)}h_k^2/24$ for $k \geqq 3$. For the effective kernels we have $\bar{L}_{\alpha k}(x',z,u) = \bar{L}_{\beta k}(x',z,u) = \bar{K}_{k-1}(z-u)/(2f(x)^{1/2})$, so (4.8) gives $\bar{K}_k = K_{h_k/f(x)^{1/2}} + \tilde{K}_{h_k/f(x)^{1/2}} * \bar{K}_{k-1}/2$, with $\tilde{K}(v) = vK'(v) + K(v)$, $\tilde{K}_h(v) = \tilde{K}(v/h)/h$, and $*$ denotes convolution. In particular, $\bar{K}_2 = K_{h_2/f(x)^{1/2}} + K_{h_1} * \tilde{K}_{h_2/f(x)^{1/2}}/2$ and $\bar{K}_3 = K_{h_3/f(x)^{1/2}} + K_{h_2/f(x)^{1/2}} * \tilde{K}_{h_3/f(x)^{1/2}}/2 + K_{h_1} * \tilde{K}_{h_2/f(x)^{1/2}} * \tilde{K}_{h_3/f(x)^{1/2}}/4$.

Example 6.3. TKDE: $b_{\alpha k}(x,z) = \int_x^z b_{k-1}(v)\,dv/(z-x)$, $b_{\beta k}(x,z) = b_{k-1}(x)$. We have $m = \infty$, so the leading bias term reduces to $b_k(x) = b_k^{\text{ad}}(x) = -3\mu_2(K)f(x)[b_{\alpha k}(x,z)f(z)/\alpha(x,z)^4]_{z=x}^{(0,2)}h_k^2/2 = -\mu_2(K)[b_{k-1}^{(2)}(x)/f(x)^2 - 3f^{(1)}(x)b_{k-1}^{(1)}(x)/f(x)^3 + (3f^{(1)}(x)^2/f(x)^4 - f^{(2)}(x)/f(x)^3)b_{k-1}(x)]\,h_k^2/2 := B_{\text{TKDE}}(b_{k-1})(x)h_k^2$, with $b \to B_{\text{TKDE}}(b)$ a differential operator. Equation (4.6) gives $\bar{L}_{\alpha k}(x',z,u) = \int_{x'}^z \bar{K}_{k-1}(v-u)\,dv/(z-x')$ and $\bar{L}_{\beta k}(x',z,u) = \bar{K}_{k-1}(x'-u)$, which implies, using (4.8) and integration by parts, $\bar{K}_k = K_{h_k/f(x)} + \bar{K}_{k-1} - \bar{K}_{k-1} * K_{h_k/f(x)}$. This yields for instance $\bar{K}_2 = K_{h_1} + K_{h_2/f(x)} - K_{h_1} * K_{h_2/f(x)}$. The formulas for b_k and $\bar{L}_k$ were derived by Hössjer and Ruppert (1995).

Example 6.4. TTKDE: $b_{\alpha k}(x,z) = \sum_{l=0}^q b_{k-1}^{(l)}(x)(z-x)^l/(l+1)!$ and $b_{\beta k}(x,z) = b_{k-1}(x)$. Since β and $b_{\beta k}$ are the same as for the TKDE-functional, and α and $b_{\alpha k}$ are Taylor expansions of the corresponding TKDE-quantities, if follows that $b_k^{\text{ad}}(x)$ is the same as for TKDE when $q \geqq 2$. Combining this with γ_j in Example 2.4, we obtain, when $q = 3$, $b_2(x) = \mu_4(K)f^{(4)}(x)h_2^4/(24f(x)^4) + b_{2,\text{TKDE}}(x)$ and $b_k(x) = \mu_4(K)f^{(4)}(x)h_k^4/(24f(x)^4)$ for $k \geqq 3$. When $q = 5$ we have $b_2(x) = b_{2,\text{TKDE}}(x)$, $b_3(x) = \mu_6(K)f^{(6)}(x)h_3^6/(720f(x)^6) + b_{3,\text{TKDE}}(x)$ and $b_k(x) = \mu_6(K)f^{(6)}(x)h_k^6/(720f(x)^6)$ for $k \geqq 4$. The effective kernels take the form $\bar{L}_{\alpha k}(x',z,u) = \sum_{l=0}^q \bar{K}_{k-1}^{(l)}(x'-u)(z-x')^l/(l+1)!$ and $\bar{L}_{\beta k}(x',z,u) = \bar{K}_{k-1}(x'-u)$, which implies $\bar{K}_k = K_{h_k/f(x)} - \sum_{l=1}^q \mu_l(K)\bar{K}_{k-1}^{(l)}h_k^l/(l!f(x)^l)$. For instance, $\bar{K}_2 = K_{h_2/f(x)} - \sum_{l=1}^q \mu_l(K)(K^{(l)})_{h_1}h_2^l/(l!f(x)^lh_1^l)$. Values of b_2 and $\bar{K}_2$ were given by Hössjer and Ruppert (1994).

Example 6.5. STKDE: Notice that the first three partial derivatives of α, β, $b_{\alpha k}$ and $b_{\alpha k}$ w.r.t. z are the same as for the TKDE-functional. In combination with Example 2.5 this gives $b_k = -\mu_4(K)f^{(4)}(x)h_k^4/(24^2f(x)^4) + B_{\text{TKDE}}(b_{k-1})(x)h_k^2$. For the effective kernels we obtain $\bar{L}_{\alpha k}(x',z,u) = \bar{K}_{k-1}(x'-u)/6 + 2\bar{K}_{k-1}((x'+z)/2-u)/3 + \bar{K}_{k-1}(z-u)/6$ and $\bar{L}_{\beta k}(x',z,u) = \bar{K}_{k-1}(x'-u)$. Insertion into (4.8) implies $\bar{K}_k = K_{h_k/f(x)} + 5\bar{K}_{k-1}/6 + 2\bar{K}_{k-1} * \check{K}_{h_k/(2f(x))}/3 + \bar{K}_{k-1} * \check{K}_{h_k/f(x)}/6$.

Example 6.6. JTKDE: $b_{\alpha k}(x,z) = \int_x^z b_{k-1}(v)\,dv/((z-x)f(x)) - \alpha(x,z)b_{k-1}(x)/f(x)^2$, $b_{\beta k}(x,z) = 0$. Since $m = \infty$, $b_k(x) = b_k^{\text{ad}}(x) = -3\mu_2(K)[b_{\alpha k}(x,z)f(z)/\alpha(x,z)^4]_{z=x}^{(0,2)}h_k^2/2 = f(x)^2B_{\text{TKDE}}(b_{k-1})(x) := B_{\text{JTKDE}}(b_{k-1})(x)$. For the stochastic part, $\bar{L}_{\alpha k}(x',z) = \int_{x'}^z \bar{K}_{k-1}(v-u)\,dv/((z-x')f(x)) - \bar{K}_{k-1}(x'-u)/f(x)$ and $\bar{L}_{\beta k}(x',z) = 0$. This implies $\bar{K}_k = K_{h_k} + \bar{K}_{k-1} - K_{h_k} * \bar{K}_{k-1}$. In particular, $\bar{K}_2 = K_{h_1} + K_{h_2} - K_{h_1} * K_{h_2}$ and $\bar{K}_3 = K_{h_1} + K_{h_2} + K_{h_3} - K_{h_1} * K_{h_2} - K_{h_1} * K_{h_3}$

$-K_{h_2} * K_{h_3} + K_{h_1} * K_{h_2} * K_{h_3}$. These formulas for b_k and $\bar{K}_k$ were obtained by Hössjer and Ruppert (1993).

Example 6.7. JLN: $b_{\alpha k}(x,z) = 0$, $b_{\beta k}(x,z) = b_{k-1}(x)/f(z) - f(x)b_{k-1}(z)/f(z)^2$. Since $m = \infty, b_k(x) = b_k^{\text{ad}}(x) = -\mu_2(K)f(x)[b_{k-1}(x)/f(x)]^{(2)}h_k^2/2 := B_{\text{JLN}}(b_{k-1})(x)h_k^2$, for instance $b_2(x) = -\mu_2(K)^2 f(x)[f^{(2)}(x)/f(x)]^{(2)}h_1^2 h_2^2/4$. The effective kernels satisfy $\bar{L}_{\alpha k}(x', z, u) = 0$ and $\bar{L}_{\beta k}(x', z, u) = (\bar{K}_{k-1}(x' - u) - \bar{K}_{k-1}(z - u))/f(x)$, which implies $\bar{K}_k = K_{h_k} + \bar{K}_{k-1} - K_{h_k} * \bar{K}_{k-1}$, the same formula as for the JTKDE, so $\bar{K}_2$ and $\bar{K}_3$ have the same form as in Example 6.6. The formulas for b_2 and $\bar{K}_2$ when $h_1 = h_2$ were derived by Jones et al. (1995).

Remark. 6.1. If $m < \infty$, then $s(k) = 2k \wedge m$, as noted in Sect. 2. Hence, the bias order agrees with the one for the ideal estimator (i.e. $s(k) = m$) when $k \geqq m/2$. In addition, b_k is exactly the same as for the ideal estimator when $k \geqq m/2 + 1$.

Remark. 6.2. It follows by induction w.r.t. k that $\text{supp}(L_k(x, \cdot\,; \mathbf{h}_k))$ is $C_1(h_1 + \sum_{j=2}^{k} h_j/\alpha_j(x,x))$, with C_1 defined in (v). This indicates the varying bandwidth structure of $\hat{f}_k$.

Remark. 6.3. If $m = \infty$, it follows that $b_k^{\text{id}} = 0$ in (3.9). Combining (3.7) and (3.10) then gives $b_k(\cdot) = B_k(b_{k-1})(\cdot)h_k^2$, where the differential operator $b \to B_k(b)$ is defined by

$$B_k(b)(x) = \frac{\mu_2(K)}{2}\left[\frac{dQ_k(x,z;f)(b)f(z)}{P_k(x,z;f)^3}\right]_{z=x}^{(0,2)} - \frac{3\mu_2(K)}{2}\left[\frac{dP_k(x,z;f)(b)Q_k(x,z;f)f(z)}{P_k(x,z;f)^4}\right]_{z=x}^{(0,2)} .$$

Notice that B_k only depends on f and (P_k, Q_k). In Examples 6.4, 6.6 and 6.7 it reduces to $B_{\text{TKDE}}, B_{\text{JTKDE}}$ and B_{JLN}, respectively.

7 Different bandwidth orders

So far, we have assumed that all bandwidths are of the same order in (i). We now change this condition to

(ia) The bandwidths $h_1, \ldots, h_{t-1}$ are all of the same order as $n \to \infty$, i.e. for some $0 < C_0 \leqq 1$ and sequence $h = h(n)$, $C_0 \leqq h_k/h \leqq C_0^{-1}$ for all n and $k = 1, \ldots, t-1$.
(ib) For some $\varepsilon_2 > 0$, $h_t/h = O(n^{-\varepsilon_2})$.
(ic) The bias exponent in Step $t-1$ satisfies $s(t-1) = m(t)$.
(id) $m(t) \geqq 4$ and $h_t \gg h^{m(t)/(m(t)-2)}$.

Condition (ib) states that h_t is of a smaller order than $h_1, \ldots, h_{t-1}$, but not too much smaller, according to (id). Notice that (ic) implies

$$t \geqq \frac{m(t)}{2} + 1 ,$$

since $s(1) = 2$ and $s(k) \leqq s(k-1) + 2$. Another consequence of (ic) is $s(t) = s(t-1)$, which implies

$$b_t(x; \mathbf{h}_t) = b_t^{\mathrm{id}}(x; h_t)\,,$$

because of (3.10). Define also

$$W_t^{\mathrm{ad}}(x; \mathbf{h}_t) = \frac{1}{n} \sum_{i=1}^{n} (L_t^{\mathrm{ad}}(x, X_i; \mathbf{h}_t) - EL_t^{\mathrm{ad}}(x, X; \mathbf{h}_t))\,, \tag{7.1}$$

where $L_t^{\mathrm{ad}}(x, u; \mathbf{h}_t) = \sum_{v=1}^{2} \bar{L}_t^{\mathrm{ad},v}(x, u, x; \mathbf{h}_t)$ (cf. (4.8)). We then have the following variant of Theorem 5.1.

Theorem 7.1 *Assume* (ia)–(id) *and* (ii)–(x)*. Then*

$$\hat{f}_t(x; \mathbf{h}_t) = f(x) + b_t^{\mathrm{id}}(x; h_t) + W_t^{\mathrm{id}}(x; h_t) + r_t(x; \mathbf{h}_t) + \bar{R}_t(x; \mathbf{h}_t)\,,$$

with $b_t^{\mathrm{id}}, W_t^{\mathrm{id}}$ *and* r_t *defined in* (3.9), (2.5) *and* (3.2), *respectively, and* $\bar{R}_t = R_t + W_t^{\mathrm{ad}}$. *Moreover,*

$$\sup_{x \in \Omega} |r_t(x; \mathbf{h}_t)| = o(h^{m(t)})\,, \tag{7.2}$$

and for some $\varepsilon > 0$

$$\left\| \sup_{x \in \Omega} |\bar{R}_t(x; \mathbf{h}_t)| \right\|_{L^p} = O((nh_t)^{-1/2} n^{-\varepsilon}) \quad \forall p > 0\,. \tag{7.3}$$

Example 7.1. Abramson estimator. Put $t = 3$, $(P_3, Q_3) =$ Abramson functional and (P_2, Q_2) any functional with $m(2) \geqq 4$. Then (ib) and (id) reduce to $h^2 \ll h_3 \ll h$ (where the last relation is sharpened by the factor $n^{-\varepsilon_2}$). We have $b_3^{\mathrm{id}}(x; h_3) = \mu_4(K)[1/f(x)]^{(4)} h_3^4/24$ and $L_3^{\mathrm{id}}(x, u; h_3) = K_{h_3/f(x)^{1/2}}(x - u)$.

Example 7.2. TTKDE with $q \geqq 2$. The same assumptions as in Example 7.1, but with $(P_3, Q_3) =$ TTKDE functional. Then $b_3^{\mathrm{id}}(x; h_3) = \mu_4(K) f^{(4)}(x) h_3^4/(24 f(x)^4)$ and $L_3^{\mathrm{id}}(x, u; h_3) = K_{h_3/f(x)}(x - u)$.

Example 7.3. STKDE. The same assumptions as in Example 7.1, but with $(P_3, Q_3) =$ STKDE functional. Then $b_3^{\mathrm{id}}(x; h_3) = -\mu_4(K) f^{(4)}(x) h_3^4/(24^2 f(x)^4)$ and $L_3^{\mathrm{id}}(x, u; h_3) = K_{h_3/f(x)}(x - u)$.

Examples 7.2 and 7.3 represent varying bandwidth estimators with simple asymptotic mean squared error (AMSE). This makes it possible to develop automatic bandwidth selectors for these estimators based on so called plug-in rules.

Remark. 7.1. The result of Theorem 7.1 is surprising: It is always possible to construct an adaptive estimator that is asymptotically equivalent to the ideal one if $m(t) < \infty$. The basic trick is to let the bandwidth $h_1, \ldots, h_{t-1}$ be of larger order than h_t, and compensate this by choosing a larger t. This means that we should avoid irregularities (large variances) for the preliminary estimators $\hat{f}_1, \ldots, \hat{f}_{t-1}$, since these irregularities will otherwise be transferred to later iterations. Even though this causes a larger bias in each step, we can iterate more times instead. We conjecture that (ia) can be weakened, for instance so that $h_1 \gg h_2 \gg \cdots \gg h_{t-1}$, and still have asymptotic equivalence with the ideal estimators. In this way we allow more irregularities/smaller bandwidth for each iteration. We imposed (ia) in order to utilize the proof of Theorem 5.1 as much as possible in Theorem 7.1 (since everything is the same until the last iteration).

Remark. 7.2. We may also assume $h_1, \ldots, h_{t-1} \ll h_t$. Formulas (3.9) and (3.10) then imply $b_t \sim b_t^{\text{id}}$, and this can be achieved already for $t = m(t)/2$. We conjecture $W_t(x; \mathbf{h}_t) = O_p((n \min(h_1, \ldots, h_t))^{-1/2})$ in general, which is of a larger order of magnitude than $W_t^{\text{id}}(x; h_t) = O_p((nh_t)^{-1/2})$. This is the case for the TKDE, TTKDE ($q \geqq 1$), JTKDE and the JLN estimator. However, for the Abramson estimator we may actually sharpen this to $W_t = O_p((nh_t)^{-1/2})$, even though $h_1, \ldots, h_{t-1} \ll h_t$. The reason is that $K_{h_1}, K_{h_1/f(x)^{1/2}}, \ldots, K_{h_{t-1}/f(x)^{1/2}}, \tilde{K}_{h_2/f(x)^{1/2}}, \ldots, \tilde{K}_{h_{t-1}/f(x)^{1/2}}$ only appear in convolutions with either K_{h_t} or $K_{h_t/f(x)^{1/2}}$. This explains why Hall and Marron (1988) could choose h_1 of smaller order than h_2 for the Abramson estimator, and obtain an estimator $\hat{f}_2$ with the same leading bias, and a variance of the same order as for the ideal estimator $\hat{f}_2^{\text{id}}$. Hall and Marron prove that $\text{Var}(\hat{f}_2) \sim C\,\text{Var}(\hat{f}_2^{\text{id}})$ for some constant $C > 1$, so this adaptive estimator has efficiency strictly less than one compared to the ideal estimator.

8 Outlook

By putting $p = 2$ in (5.5), we may easily compute the leading terms of both $E(\hat{f}_k(x) - f(x))^2$ (AMSE) and $\int_\Omega E(\hat{f}_k(x) - f(x))^2\,dx$ (AIMSE).

The representation in Theorems 5.1 and 7.1 can also be derived for weakly dependent data (under the appropriate regularity conditions). Technically, we just have to replace Rosentahl's inequality for martingale differences in the proof with the corresponding inequality for mixingales.

The varying location estimator of Samiuddin and El-Sayyad (1990) and the varying location and scale estimator of Jones et al. (1994) are not included in the class (1.2). An interesting research topic would be to derive recursive formulas for $s(k)$, b_k and L_k for a larger class of estimators including these two examples. Indeed, McKay (1993) has obtained a bias formula (generalizing the one in Hall (1990)) for such a class of estimators.

Appendix A

Properties of TKDE and NN estimators

The TKDE is usually calculated in several steps. Let us illustrate this for $t = 2$. Define $\hat{F}_1(x; h_1) = \int_{-\infty}^{x} \hat{f}_1(u; h_1)\,du$ as the c.d.f. computed from the pilot estimate. Transform the data into $Y_i = \hat{F}_1(X_i; h_1)$ and compute

$$\hat{f}_Y(y; \mathbf{h}_2) = \frac{1}{nh_2} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h_2}\right),$$

as an estimate of the transformed density. Since $\hat{F}_1$ is a monotone function, we may transform back $\hat{f}_Y$ to obtain

$$\hat{f}_2(x; \mathbf{h}_2) = \hat{F}_1'(x; h_1)\hat{f}_Y(\hat{F}_1(x; h_1); \mathbf{h}_2)$$

as an estimate of $f(x)$. If the last two displays are combined we have

$$\hat{f}_2(x; \mathbf{h}_2) = \frac{\hat{f}_1(x; h_1)}{nh_2} \sum_{i=1}^{n} K\left(\frac{\hat{F}_1(x; h_1) - \hat{F}_1(X_i; h_1)}{h_2}\right),$$

which coincides with (1.1), if we take P_2 and Q_2 as the TKDE-functionals in Table 1.

The NN-estimator is defined as

$$\hat{f}_N(x) = \frac{l}{2nd_l(x)},$$

where $d_l(x)$ is the distance from x to the lth nearest of $X_1, \ldots, X_n$. Here $l = l(n)$ is a sequence of numbers. If now K is the uniform kernel supported on $[-1, 1]$ and $h_2 = l/(2n)$ we have

$$\text{(A.1)} \quad \hat{f}_N(x) = \frac{1}{nd_l(x)} \sum_{i=1}^{n} K\left(\frac{x - X_i}{d_l(x)}\right) = \frac{\hat{f}_N(x)}{nh_2} \sum_{i=1}^{n} K\left(\frac{\hat{f}_N(x)(x - X_i)}{h_2}\right).$$

Thus $\hat{f}_N(x)$ can be formulated as a varying bandwidth estimator with itself as pilot estimate. If we choose $\hat{f}_1$ as a pilot instead of $\hat{f}_N$ in the RHS of (A.1), we obtain a special case of (1.1), with $P_2(x, z; \hat{f}_1) = Q_2(x, z; \hat{f}_1) = \hat{f}_1(x)$.

Appendix B

Proof of Theorem 5.1

Theorem 5.1 will be proved by induction w.r.t. k, $k = 1, \ldots, t$. Before proving the theorem in a series of lemmas, let us introduce some notation. Put $C_2 = C_1/(C_0\chi_0)$, with C_1 and χ_0 as defined in (v) and (viii). Then $|x - X_i| \geqq C_2 h$ implies

$$\frac{|x - X_i|\chi(\hat{\alpha}_k(x, X_i; \mathbf{h}_{k-1}))}{h_k} \geqq \frac{C_2 h \chi_0}{C_0^{-1} h} = C_1,$$

and hence

$$\text{(B.1)} \qquad K\left(\frac{|x - X_i|\chi(\hat{\alpha}_k(x, X_i; \mathbf{h}_{k-1}))}{h_k}\right) = 0 \quad \text{when } |x - X_i| \geqq C_2 h.$$

Choose numbers $0 < \delta_t < \cdots < \delta_1 < \delta_0$. The behaviour of $\hat{f}_k$ will be studied on Ω^{δ_k}. We will assume that n is so large (h so small) that

(B.2) $$\delta_k + C_2 h < \delta_{k-1}, \quad k = 1,\dots,t\,.$$

By (B.1), this means that for $x \in \Omega^{\delta_k}$ and $X_i \notin \Omega^{\delta_{k-1}}$, the corresponding term in (viii) does not contribute to $\hat{f}_k(x;\mathbf{h}_k)$.

Put

(B.3) $$\bar{C}_k = (k-1)C_2 + C_1 C_0^{-1}, \quad k = 1,\dots,t\,.$$

Then

(B.4)
$$\bar{L}_k(x', \cdot, x; \mathbf{h}_k) \text{ is supported on } [x' - \bar{C}_k h, x' + \bar{C}_k h] \text{ for any } x',\ x \in \Omega^{\delta_k}\,.$$

and

(B.5) $$\bar{L}_{\alpha k}(x', z, \cdot, x; \mathbf{h}_{k-1}) \text{ and } \bar{L}_{\beta k}(x', z, \cdot, x; \mathbf{h}_{k-1}) \text{ are supported on } [x' - \bar{C}_k h, x' + \bar{C}_k h] \text{ for any } x,\ x' \in \Omega^{\delta_k},\ |z - x'| \leqq C_2 h\,.$$

Notice that $\bar{L}_1(x', u, x; \mathbf{h}_1) = K((x'-u)/h_1)/h_1$, which is zero for $|x' - u| \leqq C_1 h_1 \leqq \bar{C}_1 h$, because of (i). The rest follows by induction w.r.t. k, making use of (4.6) and (4.8), noticing that $\alpha_k(x;x) > \chi_0$ in (4.8) and finally, observing that $\bar{L}_{\alpha k}(x', z, u, x; \mathbf{h}_{k-1})$ (and $L_{\beta k}(x', z, u, x; \mathbf{h}_{k-1})$) only depend on $\bar{L}_{k-1}(\cdot, u, x; \mathbf{h}_{k-1})$ restricted to $[x', z]$ (this follows from (vi)). In addition to (B.2), assume that n is so large that $\delta_1 + \bar{C}_1 h < \delta_0$. Then

(B.6) $$\delta_k + \bar{C}_k h < \delta_0, \quad k = 1,\dots,t\,.$$

This implies that only those data with $X_i \in \Omega^{\delta_0}$ contribute to $W_k(x;\mathbf{h}_k)$ in (1.5) for $x \in \Omega^{\delta_k}$.

Define the regions

$$\begin{aligned}
\Lambda_k &= \{(x,z);\ x \in \Omega^{\delta_k},\ |z - x| \leqq C_2 h\}\,,\\
\bar{\Lambda}_k &= \{(x', u, x);\ x', x \in \Omega^{\delta_k},\ |u - x'| \leqq \bar{C}_k h\}\,,\\
\tilde{\Lambda}_k &= \{(x', z, u, x);\ x', x \in \Omega^{\delta_k},\ |z - x'| \leqq C_2 h,\ |u - x'| \leqq \bar{C}_k h\}\,,\\
\check{\Lambda}_k &= \{(x,u);\ x \in \Omega^{\delta_k},\ |u - x| \leqq \bar{C}_k h\}\,,
\end{aligned}$$

consisting of the relevant values of x, x', z and u at each iteration. Introduce also

(B.7) $$\begin{aligned}
J_k &= s(t) - s(k) + \sum_{l=k+1}^{t} q_l\,,\\
\tilde{J}_k &= s(t) - s(k-1) + \sum_{l=k+1}^{t} q_l\,,\\
\bar{J}_k &= \sum_{l=k+1}^{t} q_l\,.
\end{aligned}$$

Theorem 5.1 will be proved by establishing the following 9 conditions recursively w.r.t. k:

(B.8) $$\|\alpha_k^{(i,j)}\|_{\Lambda_k} \text{ and } \|\beta_k^{(i,j)}\|_{\Lambda_k} = O(1), \quad 0 \leqq i+j \leqq J_0 - q_k + 1\,,$$

(B.9) $$\|b_{\alpha k}^{(i,j)}\|_{\Lambda_k} \text{ and } \|b_{\beta k}^{(i,j)}\|_{\Lambda_k} = O(h^{s(k-1)}), \quad 0 \leqq i+j \leqq \tilde{J}_k\,,$$

(B.10) $$\|r_{\alpha k}^{(i,j)}\|_{\Lambda_k} \text{ and } \|r_{\beta k}^{(i,j)}\|_{\Lambda_k} = o(h^{s(k-1)}), \quad 0 \leqq i+j \leqq \tilde{J}_k\,,$$

(B.11) $$\|\bar{L}_{\alpha k}^{(i,0,0,d)}\|_{\bar{\Lambda}_k} \text{ and } \|\bar{L}_{\beta k}^{(i,0,0,d)}\|_{\bar{\Lambda}_k} = O(h^{-(1+i)}), \quad 0 \leqq i+d \leqq \bar{J}_k + 1\,,$$

(B.12) $$\left\|\|R_{\alpha k}^{(i)}\|_{\Lambda_k}\right\|_{L^p} \text{ and } \left\|\|R_{\beta k}^{(i)}\|_{\Lambda_k}\right\|_{L^p} = O((nh)^{-1/2}h^{-i}n^{-\varepsilon}),$$
$$0 \leqq i \leqq \bar{J}_k \text{ and any } p > 0\,,$$

(B.13) $$\|b_k^{(i)}\|_{\Omega^{\delta_k}} = O(h^{s(k)}), \quad 0 \leqq i \leqq J_k\,,$$

(B.14) $$\|r_k^{(i)}\|_{\Omega^{\delta_k}} = o(h^{s(k)}), \quad 0 \leqq i \leqq J_k\,,$$

(B.15) $$\|\bar{L}_k^{(i,0,d)}\|_{\bar{\Lambda}_k} = O(h^{-(1+i)}), \quad 0 \leqq i+d \leqq \bar{J}_k + 1\,,$$

(B.16) $$\left\|\|R_k^{(i)}\|_{\Omega^{\delta_k}}\right\|_{L^p} = O((nh)^{-1/2}h^{-i}n^{-\varepsilon}),$$
$$0 \leqq i \leqq \bar{J}_k \text{ and any } p > 0\,.$$

We also require $b_{\alpha k}^{(i,j)}, b_{\beta k}^{(i,j)}, r_{\alpha k}^{(i,j)}$ and $r_{\beta k}^{(i,j)}$ to be continuous over Λ_k if $i+j = \tilde{J}_k$ in (B.9) and (B.10). Likewise, $b_k^{(J_k)}$ and $r_k^{(J_k)}$ in (B.13) and (B.14) are required to be continuous over Ω^{δ_k}.

Schematically, the proof looks like that shown in Fig. 1.

Lemma B.1 *Equations* (B.13)–(B.16) *hold for* $k = 1$.

Proof. Recall formulas (3.3) and (4.3) for b_1 and $\bar{L}_1$. Moreover, (3.2) and (3.8) imply $r_1(x; h_1) = \int K(t)(f(x+th_1) - f(x) - f^{(1)}(x)th_1 - f^{(2)}(x)t^2h_1^2/2)\,dt$. Finally, $R_1(x; h_1) = 0$, $J_1 = s(t) - 2 + \sum_{k=2}^{t} q_k$ and $\bar{J}_1 = \sum_{k=2}^{t} q_k$. The lemma follows from (iv) and (v). □

Lemma B.2 *Suppose* $\hat{f}_k$ *has the asymptotic representation* (3.1), *with* b_k, r_k, $\bar{L}_k$ *and* R_k *satisfying* (B.13)–(B.16) *and* $1 \leqq k \leqq t-1$. *Then* $\xi(\hat{f}_k)$ *has the same expansion*

(B.17) $$\xi(\hat{f}_k(x; \mathbf{h}_k)) := \tilde{f}_k(x; \mathbf{h}_k) = f_{bk}(x; \mathbf{h}_k) + W_k(x; \mathbf{h}_k) + \tilde{R}_k(x; \mathbf{h}_k)\,,$$

with $\tilde{R}_k$ *satisfying* (B.16).

Proof. Put $\bar{R}_k(x; \mathbf{h}_k) = \xi(\hat{f}_k(x; \mathbf{h}_k)) - \hat{f}_k(x; \mathbf{h}_k)$. It suffices to prove that $\bar{R}_k$ satisfies (B.16), since $\tilde{R}_k = R_k + \bar{R}_k$. Write $\hat{f}_k(x; \mathbf{h}_k) = f_{bk}(x; \mathbf{h}_k) + V_k(x; \mathbf{h}_k)$,

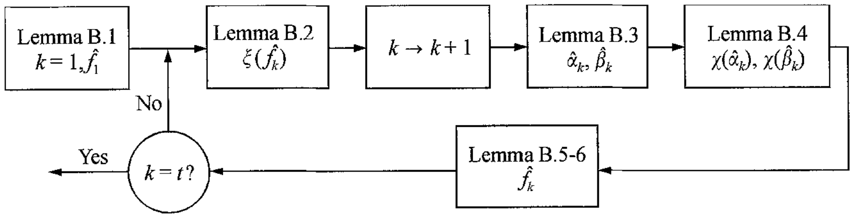

Fig. 1.

with $V_k = W_k + R_k$ the stochastic part. We will establish below that

$$\left\| \|W_k^{(i)}\|_{\Omega^{\delta_k}} \right\|_{L^p} \leqq C(nh)^{-1/2} h^{-i} n^{\varepsilon} , \tag{B.18}$$

for $0 \leqq i \leqq \bar J_k$, any $\varepsilon > 0$ and any $p > 0$ (and ε is independent of p). By expanding derivatives of $\hat f_k(\,\cdot\,; \mathbf{h}_k)$, it follows from (ii), (B.13), (B.14), (B.16), (B.18) and the smoothness of ξ (cf. (vii)) that

$$\left\| \|\bar R_k^{(i)}\|_{\Omega^{\delta_k}} \right\|_{L^p} \leqq Ch^{-i} ,$$

for $0 \leqq i \leqq \bar J_k$, provided ε in (B.18) is chosen small enough compared to ε_0 in (ii). By (vii), $\xi(v) = v$ for $v \geqq \xi_1$ and $\xi_1 < \underline{f}$. Put $\zeta = (\underline{f} - \xi_1)/2$. It follows from (ii), (iv), (B.13) and (B.14) that $\inf_{x \in \Omega^{\delta_k}} f_{bk}(x; \mathbf{h}_k) > \underline{f} - \zeta$ for n large enough. Hence,

$$|V_k(x; \mathbf{h}_k)| \leqq \zeta \Rightarrow \xi(\hat f_k(x; \mathbf{h}_k)) = \hat f_k(x; \mathbf{h}_k) .$$

By choosing p large enough and using Markov's inequality, it follows from (ii), (B.16) and (B.18) with $i = 0$ that

$$P(\|V_k\|_{\Omega^{\delta_k}} > \zeta) \leqq C\zeta^{-p}(nh)^{-p/2} n^{\varepsilon p} \leqq C(\zeta, \gamma) n^{-\gamma}$$

for any $\gamma > 0$, provided p is first chosen large enough and ε then small enough. Let A be the set $\{\|V_k\|_{\Omega^{\delta_k}} > \zeta\}$. Then, by Cauchy–Schwartz inequality,

$$\begin{aligned} \left\| \|\bar R_k^{(i)}\|_{\Omega^{\delta_k}} \right\|_{L^p} &= \left\| \|\bar R_k^{(i)}\|_{\Omega^{\delta_k}} 1_A \right\|_{L^p} \\ &\leqq \left\| \|\bar R_k^{(i)}\|_{\Omega^{\delta_k}} \right\|_{L^{2p}} \|1_A\|_{L^{2p}} \\ &\leqq Ch^{-i} n^{-\gamma/(2p)} = O((nh)^{-1/2} h^{-i} n^{-\varepsilon}) \end{aligned} \tag{B.19}$$

for $0 \leqq i \leqq \bar J_k$ and any $\varepsilon > 0$. The last line holds provided we choose γ large enough given p and ε. It remains to establish (B.18), and it suffices to consider the case $p \geqq 2$. The technique, based on Rosentahl's moment inequality for sums of martingale differences, is taken from Hall and Marron (1988). Rosentahl's inequality states: If $Z_1, \ldots, Z_n$ are zero mean martingale differences (which means $E(Z_j \mid Z_1,\ldots,Z_{j-1}) = 0$) and $\tilde{p} \geqq 2$, then

$$E\left|\sum_{j=1}^{n} Z_j\right|^{\tilde{p}} \leqq C(\tilde{p})\left[\left\{\sum_{j=1}^{n} E(Z_j^2)\right\}^{\tilde{p}/2} + \sum_{j=1}^{n} E|Z_j|^{\tilde{p}}\right],$$

where $C(\tilde{p})$ does not depend on n. By (4.2), (B.4) and (B.15),

(B.20) $\quad L_k(x,\cdot\,;\mathbf{h}_k)$ is supported on $[x-\bar{C}_k h, x+\bar{C}_k h]$ for any $x \in \Omega^{\delta_k}$,

and

(B.21) $$\|L_k^{(i)}\|_{\bar{\Lambda}_k} = O(h^{-(1+i)}) \text{ for } 0 \leqq i \leqq \bar{J}_k + 1 .$$

Recalling (1.5), we will use $Z_j = (L^{(i)}(x,X_j;\mathbf{h}_k) - EL^{(i)}(x,X;\mathbf{h}_k))/n$ in Rosentahl's inequality, with $i \leqq \bar{J}_k$. By (B.20) and (B.21),

$$E|Z_j|^{\tilde{p}} \leqq Ch^{-\tilde{p}(1+i)+1}n^{-\tilde{p}}$$

holds uniformly for all $x \in \Omega^{\delta_k}$. Let now Γ be a equispaced finite grid in Ω^{δ_k} with $|\Gamma| = O(n^s)$ elements and put $\xi = (nh)^{-1/2}h^{-i}n^{\varepsilon}\rho$, where ε is the same number as in (B.18) and $\rho > 0$. Then

$$\begin{aligned} P(\|W_k^{(i)}\|_\Gamma > \xi) &\leqq \sum_{x\in\Gamma} \|W_k^{(i)}(x;\mathbf{h}_k)\|_{L^{\tilde{p}}}^{\tilde{p}}\xi^{-\tilde{p}} \leqq |\Gamma| \sup_{x\in\Gamma} \|W_k^{(i)}(x;\mathbf{h}_k)\|_{L^{\tilde{p}}}^{\tilde{p}}\xi^{-\tilde{p}} \\ &\leqq C|\Gamma|\{(h^{-2(1+i)+1}n^{-1})^{\tilde{p}/2} + h^{-\tilde{p}(1+i)+1}n^{-\tilde{p}+1}\}\xi^{-\tilde{p}} \\ &\leqq C|\Gamma|n^{-\varepsilon\tilde{p}}\rho^{-\tilde{p}} . \end{aligned}$$

Given ε, choose first s and then $\tilde{p} > s/\varepsilon$. Since $i \leqq \bar{J}_k$, (B.21) implies that $L_k^{(i)}$ (and hence also $W_k^{(i)}$) is Lipschitz continuous w.r.t. x. If s is chosen large enough we therefore have

$$P(\|W_k^{(i)}\|_{\Omega^{\delta_k}} > (nh)^{-1/2}h^{-i}n^{\varepsilon}\rho) \leqq C\rho^{-\tilde{p}} ,$$

which implies (B.20) for any $p < \tilde{p}$. □

Lemma B.3 *Suppose* $2 \leqq k \leqq t$, *that* $\tilde{f}_{k-1}$ *has the asymptotic expansion* (B.17), *and that* (B.13)–(B.16) *hold (for* $k-1$, *with* $\tilde{R}_{k-1}$ *in place of* R_{k-1} *in* (B.16)). *Then* $\hat{\alpha}_k$ *and* $\hat{\beta}_k$, *defined in* (viii), *have the asymptotic expansions* (3.4), (4.4), *with* (B.8)–(B.12) *being satisfied.*

Proof. We concentrate on $\hat{\alpha}_k$, since the treatment of $\hat{\beta}_k$ is completely analogous.

We start by proving (B.8). Write $\alpha_k(x,z) = \sum_{l=0}^{q_k}\alpha_{kl}(x,z)(z-x)^l$, with $\alpha_{kl}(x,z) = P_{kl}(x,z;f)$. Clearly, it suffices to prove $\|\alpha_{kl}^{(i,j)}\|_{\Lambda_k}^{l} = O(1)$, for $0 \leqq i+j \leqq J_0 - q_k + 1$. But this follows from (D.1), with $g = f$ and $\Xi = \Lambda_k$. Notice that $P(\Xi) \subset \Omega^{\delta_{k-1}}$ because of (B.2), and $l+i+j \leqq J_0 + 1$. Hence, because of (iv), $\|g^{(v)}\|_{P(\Xi)} = O(1)$ for $0 \leqq v \leqq l+i+j$.

In order to prove (B.9), write $b_{\alpha k}(x,z;\mathbf{h}_{k-1}) = \sum_{l=0}^{q_k} b_{\alpha kl}(x,z;\mathbf{h}_{k-1})(z-x)^l$, with $b_{\alpha kl}(x,z;\mathbf{h}_{k-1}) = dP_{kl}(x,z;f)(b_{k-1})$. Apply (D.2) with $g_0 = f$, $\eta = b_{k-1}(\cdot,\mathbf{h}_{k-1})$, $\Xi = \Lambda_k$ and $P(\Xi) \subset \Omega^{\delta_{k-1}}$. Notice that $\|g_0^{(v)}\|_{P(\Xi)} = O(1)$ and

$\|\eta^{(v)}\|_{P(\Xi)} = O(h^{s(k-1)})$ for $0 \leqq v \leqq l+i+j$, because of (iv) and (B.13), since $l+i+j \leqq \tilde{J} + q_k = J_{k-1}$.

To establish (B.11), write

$$\bar{L}_{\alpha k}(x',z,u,x;\mathbf{h}_{k-1}) = \sum_{l=0}^{q_k} \bar{L}_{\alpha kl}(x',z,u,x;\mathbf{h}_{k-1})(z-x')^l\,, \tag{B.22}$$

with $\bar{L}_{\alpha kl}(x',z,u,x;\mathbf{h}_{k-1}) = dP_{kl}(x',z;f_x)(\bar{L}_{k-1}(\,\cdot\,,u,x;\mathbf{h}_{k-1}))$. It suffices to prove

$$\|\bar{L}_{\alpha kl}^{(i,0,0,d)}\|_{\tilde{\Lambda}_k} = O(h^{-(1+l+i)}), \quad 0 \leqq i+d \leqq \bar{J}_k + 1\,. \tag{B.23}$$

Then (B.11) will follow by applying the product-rule of differentiation on each term of (B.22), using the fact that $|z-x'| \leqq C_2 h$ for any element of $\tilde{\Lambda}_k$. To prove (B.23), apply (D.5), with $\theta_1 = x$, $g(\,\cdot\,,x) = f_x(\,\cdot\,)$, $\theta_2 = (u,x)$, $\eta(\,\cdot\,,(u,x)) = \bar{L}_{k-1}(\,\cdot\,,u,x;\mathbf{h}_{k-1})$, $\Xi = \Lambda_k$, $\Theta_{1x'z} = \Omega^{\delta_k}$ and $\Theta_{2x'z} = [x' - \bar{C}_k h, x' + \bar{C}_k h] \times \Omega^{\delta_k}$. Define Υ, Υ_1 and Υ_2 as in (D.5). Then (iv) implies $\|g^{(v_1,v_2)}\|_{\Upsilon_1} = O(1)$ for $0 \leqq v_1 + v_2 \leqq \bar{J}_k + l + 1$. Observe next that $\Upsilon_2 = \tilde{\Lambda}_k$. Therefore, (B.15) (with $k-1$ instead of k) implies $\|\eta^{(v_1,(0,v_2))}\|_{\Upsilon_2} = O(h^{-(1+v_1)})$, whenever $0 \leqq v_1 + v_2 \leqq \bar{J}_{k-1} + 1$. Here $(0,v_2)$ indicates differentiation w.r.t. θ_2. The last two estimates can now be plugged into (D.5). Then (B.23) follows, since $\bar{J}_k + l \leqq \bar{J}_{k-1}$ and $\bar{L}_{\alpha kl}(x',z,u,x;\mathbf{h}_{k-1}) = \tilde{h}(x',z,x,(u,x))$, with $\tilde{h}$ as defined in (D.5).

For (B.10), we decompose $r_{\alpha k}$ into two terms: By (2.2), (3.2) and (3.7), we obtain

$$\begin{aligned} r_{\alpha k}(x,z;\mathbf{h}_k) &= \alpha_{bk}(x,z;\mathbf{h}_{k-1}) - \alpha_k(x,z) - b_{\alpha k}(x,z;\mathbf{h}_{k-1}) = dP_k(x,z;f)(r_k) \\ &\quad + (P_k(x,z;f_{b,k-1}) - P_k(x,z;f) - dP_k(x,z;f)(f_{b,k-1} - f)) \\ &:= \sum_{v=1}^{2} r_{\alpha kv}(x,z;\mathbf{h}_k)\,. \end{aligned} \tag{B.24}$$

We have to establish (B.10) for each term $r_{\alpha kv}$. For $(r_{\alpha k1})$ this follows similarly as (B.9) was proved for $b_{\alpha k}$, using (B.14) instead of (B.13). For $r_{\alpha k2}$, use (D.3), with $g_0 = f$, $g_1 = f_{b,k-1}$ and $g_1 - g_0 = b_{k-1} + r_{k-1}$. Use then (iv), (B.13) and (B.14).

Finally, consider $R_{\alpha k}$. By (vii), (B.17) and (4.4), we may write

$$\begin{aligned} R_{\alpha k}(x,z;\mathbf{h}_k) &= P_k(x,z;\tilde{f}_{k-1}) - P_k(x,z;f_{b,k-1}) - dP_k(x,z;f_x)(\bar{W}_{k-1}) \\ &= dP_k(x,z;f_x)(W_{k-1} - \bar{W}_{k-1}) \\ &\quad + (dP_k(x,z;f_{b,k-1})(W_{k-1}) - dP_k(x,z;f_x)(W_{k-1})) \\ &\quad + dP_k(x,z;f_{b,k-1})(\tilde{R}_{k-1}) \\ &\quad + (P_k(x,z;\tilde{f}_{k-1}) - P_k(x,z;f_{b,k-1}) \\ &\qquad - dP_k(x,z;f_{b,k-1})(\tilde{f}_{k-1} - f_{b,k-1})) \\ &:= \sum_{v=1}^{4} R_{\alpha kv}(x,z;\mathbf{h}_{k-1})\,, \end{aligned} \tag{B.25}$$

with $\bar{W}_{k-1} = \bar{W}_{k-1}(\cdot, x; \mathbf{h}_{k-1})$ and $W_{k-1} = W_{k-1}(\cdot; \mathbf{h}_{k-1})$. We have to establish (B.12) for each term $R_{\alpha k\nu}$. We do this in detail only for $\nu = 1$. Write $R_{\alpha k1}(x,z;\mathbf{h}_{k-1}) = \sum_{l=0}^{q_k} R_{\alpha kl1}(x,z;\mathbf{h}_{k-1})(z-x)$, with $R_{\alpha kl1}(x,z;\mathbf{h}_{k-1}) = dP_{kl}(x,z;f_x)(W_{k-1} - \bar{W}_{k-1})$. We only have to prove

(B.26)

$$\left\| \| (R_{\alpha kl1}^{(i)}) \|_{\Lambda_k} \right\|_{L^p} = O((nh)^{-1/2} h^{-(i+l)} n^{-\varepsilon}) \quad 0 \leqq i \leqq \bar{J}_k \text{ and any } p > 0 .$$

Apply (D.5), with $\Xi = \Lambda_k$, $\theta_1 = \theta_2 = x$, $g(\cdot, x) = f_x(\cdot)$, $\eta(\cdot, x) = W_{k-1}(\cdot; \mathbf{h}_{k-1}) - \bar{W}_{k-1}(\cdot, x; \mathbf{h}_{k-1})$ and $\Theta_{1xz} = \Theta_{2xz} = \{x\}$. With $\tilde{h}$ as defined in (D.5) we then have $R_{\alpha kl1}(x,z;\mathbf{h}_{k-1}) = \tilde{h}(x,z,x,x)$, so that (B.26) will follow if we prove

$$\text{(B.27)} \qquad \left\| \| h^{(i,0,d_1,d_2)} \|_{\Upsilon} \right\|_{L^p} = O((nh)^{-1/2} h^{-(i+d_1+d_2+l)} n^{-\varepsilon}) ,$$

for any $p > 0$, when $0 \leqq i + d_1 + d_2 \leqq \bar{J}_k$, and Υ is defined in (D.5). Let $\Upsilon_1 = \bigcup_{(x,z)\in\Xi} [x,z] \times \{x\}$ and notice that $\|g^{(\nu_1,\nu_2)}\|_{\Upsilon_1} = O(1)$ for $0 \leqq \nu_1 + \nu_2 \leqq \bar{J}_k + l$ because of (iv). This will prove (B.27), in conjunction with (D.5) and the statement

(B.28)

$$\left\| \| \eta^{(\nu_1,\nu_2)} \|_{\Upsilon_2} \right\|_{L_p} = O((nh)^{-1/2} h^{-(\nu_1+\nu_2-l)} n^{\varepsilon}) = O((nh)^{-1/2} h^{-(\nu_1+\nu_2)} n^{-\varepsilon})$$

for all sufficiently small $\varepsilon > 0$, $0 \leqq \nu_1 + \nu_2 \leqq \bar{J}_k + l$ and $\Upsilon_2 = \Upsilon_1$. The last relation in (B.28) follows from (ii) if we choose ε small enough compared to ε_0. In order to prove the first relation in (B.28), notice that

$$\eta(x', x) = \frac{1}{n} \sum_{j=1}^{n} (\tilde{L}_{k-1}(x', X_j, x; \mathbf{h}_{k-1}) - E\tilde{L}_{k-1}(x', X, x; \mathbf{h}_{k-1})) ,$$

with $\tilde{L}_{k-1}(x', u, x; \mathbf{h}_{k-1}) = \bar{L}_{k-1}(x', u, x'; \mathbf{h}_{k-1}) - \bar{L}_{k-1}(x', u, x; \mathbf{h}_{k-1})$. It follows from (B.15) (with $k-1$ instead of k) that

$$\text{(B.29)} \qquad \|\tilde{L}_{k-1}^{(\nu_1,\nu_2)}\|_{\tilde{\Upsilon}} = \begin{cases} O(h^{-(\nu_1+\nu_2)}), & 0 \leqq \nu_1 + \nu_2 \leqq \bar{J}_{k-1} , \\ O(h^{-(\bar{J}_{k-1}+2)}), & \nu_1 + \nu_2 = \bar{J}_{k-1} + 1 , \end{cases}$$

with $\tilde{\Upsilon} = \{(x', u, x),\ x \in \Omega^{\delta_k},\ |x' - x| \leqq C_2 h \text{ and } |u - x'| \leqq \bar{C}_k h\}$. But (B.28) now follows from (B.29) and (B.4), in the same way as (B.18) was proved, using Rosentahl's inequality.

Returning to the last three terms of (B.25), we use (D.3) for $R_{\alpha k2}$, with $g_0 = f_x$, $g_1 = f_{b,k-1}$ and $\eta = W_{k-1}$. (Actually, $g_0(\cdot) = g_0(\cdot, x)$, so we use a generalization of (D.3), as (D.5) was stated as a generalization of (D.2).) For $R_{\alpha k3}$, use (D.2), with $g_0 = f_{b,k-1}$ and $\eta = \tilde{R}_{k-1}$. Finally, apply (D.4) for $R_{\alpha k4}$, with $g_0 = f_{b,k-1}$ and $g_1 = \tilde{f}_{k-1}$. □

Lemma B.4 *Suppose* $2 \leqq k \leqq t$ *and that* $\hat{\alpha}_k$ *has the asymptotic representation* (4.4) *and* (3.4), *with* α_k, $b_{\alpha k}$, $r_{\alpha k}$, $\bar{L}_{\alpha k}$ *and* $R_{\alpha k}$ *satisfying* (B.8)–(B.12).

Assume also the same for $\hat{\beta}_k$. *Then* $\chi(\hat{\alpha}_k)$ *and* $\chi(\hat{\beta}_k)$ *have the expansions*

(B.30)
$$\chi(\hat{\alpha}_k(x,z;\mathbf{h}_k)) := \tilde{\alpha}_k(x,z;\mathbf{h}_k) = \alpha_{bk}(x,z) + \bar{W}_{\alpha k}(x,z,x;\mathbf{h}_k) + \tilde{R}_{\alpha k}(x,z;\mathbf{h}_k)\,,$$
$$\chi(\hat{\beta}_k(x,z;\mathbf{h}_k)) := \tilde{\beta}_k(x,z;\mathbf{h}_k) = \beta_{bk}(x,z) + \bar{W}_{\beta k}(x,z,x;\mathbf{h}_k) + \tilde{R}_{\beta k}(x,z;\mathbf{h}_k)\,,$$

with $\tilde{R}_{\alpha k}$ *and* $\tilde{R}_{\beta k}$ *satisfying* (B.12).

Proof. The proof is analogous to the proof of Lemma B.2. □

Lemma B.5 *Suppose* $2 \leqq k \leqq t$ *and that* $\chi(\hat{\alpha}_k)$ *and* $\chi(\hat{\beta}_k)$ *have the expansions in* (B.30), *with* (B.8)–(B.12) *satisfied,* ((B.12) *with* $\tilde{R}_{\alpha k}$ *and* $\tilde{R}_{\beta k}$ *in place of* $R_{\alpha k}$ *and* $R_{\beta k}$). *Then* $\hat{f}_k$, *defined in* (viii), *has the expansion* (3.1), *and* (B.15)–(B.16) *hold.*

Proof. Formula (B.15) follows from (B.11) and differentiation w.r.t. x' and x in (4.8). The rest of the proof consists of establishing (B.16). We omit $\mathbf{h}_{k-1}$ and $\mathbf{h}_k$ in the notation for simplicity.[3] By (viii),

$$\text{(B.31)} \qquad \hat{f}_k(x) = \frac{1}{nh_k}\sum_{i=1}^{n} \tilde{\beta}_k(x,X_i)K\left(\frac{(x-X_i)\tilde{\alpha}_k(x,X_i)}{h_k}\right).$$

Perform a Taylor expansion of each term in (B.31), using the expansions in (B.30):

(B.32)
$$\begin{aligned}\tilde{\beta}_k(x,z)K\left(\frac{(x-z)\tilde{\alpha}_k(x,z)}{h_k}\right) &= \beta_{bk}(x,z)K\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right)\\ &\quad+\frac{\beta_{bk}(x,z)}{\alpha_{bk}(x,z)}\bar{W}_{\alpha k}(x,z,x)\check{K}\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right)\\ &\quad+\bar{W}_{\beta k}(x,z,x)K\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right)\\ &\quad+\sum_{v=1}^{4}\varepsilon_v(x,z)\,,\end{aligned}$$

with $\check{K}(v) = vK'(v)$, and $\varepsilon_1,\ldots,\varepsilon_4$ are remainder terms, defined by

$$\text{(B.33)} \quad \varepsilon_1(x,z) = \frac{\beta_{bk}(x,z)}{\alpha_{bk}(x,z)}\tilde{R}_{\alpha k}(x,z)\check{K}\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right),$$
$$\varepsilon_2(x,z) = \tilde{R}_{\beta k}(x,z)K\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right),$$
$$\varepsilon_3(x,z) = \tilde{\beta}_k(x,z)\left(K\left(\frac{(x-z)\tilde{\alpha}_k(x,z)}{h_k}\right) - K\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right)\right.$$

[3] Given x, we assume $|z-x| \leqq C_2h$ throughout the proof. This is justified because of (B.1), even though quantities like $K((x-z)\alpha_k(x,z)/h_k)$ may be nonzero for other values of z. For the same reason, we also assume $|X_i-x|, |X_j-x| \leqq C_2h$

$$- \frac{\tilde{\alpha}_k(x,z) - \alpha_{bk}(x,z)}{\alpha_{bk}(x,z)} \check{K}\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right)\Bigg) ,$$

$$\varepsilon_4(x,z) = (\tilde{\beta}_k(x,z) - \beta_{bk}(x,z)) \frac{\tilde{\alpha}_k(x,z) - \alpha_{bk}(x,z)}{\alpha_{bk}(x,z)} \check{K}\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right) .$$

Put also

$$R_{kv}(x) = \frac{1}{nh_k} \sum_{i=1}^{n} \varepsilon_v(x, X_i), \quad v = 1,2,3,4 .$$

Insert the expansion (B.32) into (B.31), and use (4.7) and (3.8) to obtain

(B.34)
$$\begin{aligned} \hat{f}_k(x) = f_{bk}(x) &+ \frac{1}{nh_k} \sum_{i=1}^{n} (l_1(x,X_i) - E(l_1(x,X))) \\ &+ \frac{1}{n^2 h_k} \sum_{i,j=1}^{n} (l_2(x,X_i,X_j) - E(l_2(x,X_i,X) \,|\, X_i)) \\ &+ \frac{1}{n^2 h_k} \sum_{i,j=1}^{n} (l_3(x,X_i,X_j) - E(l_3(x,X_i,X) \,|\, X_i)) \\ &+ \sum_{v=1}^{4} R_{kv}(x) \end{aligned}$$

with

$$\begin{aligned} l_1(x,z) &= \beta_{bk}(x,z) K\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right) , \\ l_2(x,z,u) &= \frac{\beta_{bk}(x,z)}{\alpha_{bk}(x,z)} \bar{L}_{\alpha k}(x,z,u,x) \check{K}\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right) , \\ l_3(x,z,u) &= \bar{L}_{\beta k}(x,z,u,x) K\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right) . \end{aligned}$$

Define next

(B.35)
$$\begin{aligned} R_{k5}(x) &= \frac{1}{nh_k} \sum_{i=1}^{n} (l_1(x,X_i) - E(l_1(x,X)) \\ &\quad - \frac{1}{nh_k} \sum_{i=1}^{n} (\bar{L}_k^{\mathrm{id}}(x,X_i,x) - E(\bar{L}_k^{\mathrm{id}}(x,X,x)) , \\ R_{k6}(x) &= \frac{1}{n^2 h_k} \sum_{i,j=1}^{n} (l_2(x,X_i,X_j) - E(l_2(x,X_i,X) \,|\, X_i)) \\ &\quad - \frac{1}{nh_k} \sum_{j=1}^{n} (\bar{L}_k^{\mathrm{ad},1}(x,X_j,x) - E(\bar{L}_k^{\mathrm{ad},1}(x,X,x))) , \\ R_{k7}(x) &= \frac{1}{n^2 h_k} \sum_{i,j=1}^{n} (l_3(x,X_i,X_j) - E(l_3(x,X_i,X) \,|\, X_i)) \\ &\quad - \frac{1}{nh_k} \sum_{j=1}^{n} (\bar{L}_k^{\mathrm{ad},2}(x,X_j,x) - E(\bar{L}_k^{\mathrm{ad},2}(x,X,x))) , \end{aligned}$$

with $\bar{L}_k^{\mathrm{id}}$, $\bar{L}_k^{\mathrm{ad},1}$ and $\bar{L}_k^{\mathrm{ad},2}$ defined in (4.8). It now follows from (4.1), (4.8), (B.34) and (B.35) that

$$\hat{f}_k(x) = f_{bk}(x) + \bar{W}_k(x,x) + \sum_{v=1}^{7} R_{kv}(x) .$$

Hence, it suffices to prove for each $v = 1, \dots, 7$,

$$\text{(B.36)} \quad \left\| \|R_{kv}^{(i)}\|_{\Omega^{\delta_k}} \right\|_{L^p} = O((nh)^{-1/2} h^{-i} n^{-\varepsilon}), \quad 0 \leqq i \leqq \bar{J}_k \text{ and any } p > 0 .$$

We start with the case $v = 1, 2, 3, 4$. Suppose we can show
(B.37)

$$\left\| \|\varepsilon_v^{(i)}\|_{\Lambda_k} \right\|_{L^p} = O((nh)^{-1/2} h^{-i} n^{-\varepsilon}), \quad 0 \leqq i \leqq \bar{J}_k \text{ and any } p > 0,\ v = 1,2,3,4 .$$

Let $\psi \in C^\infty(\mathbb{R})$ be a positive function with compact support on $[-2, 2]$ and $\psi(x) = 1$ for $x \in [-1, 1]$. Then

$$\text{(B.38)} \qquad \|R_{kv}^{(i)}\|_{\Omega^{\delta_k}} \leqq \|\varepsilon_v^{(i)}\|_{\Lambda_k} \left\| \frac{1}{nh_k} \sum_{j=1}^{n} \psi\left(\frac{\cdot - X_j}{C_2 h} \right) \right\|_{\Omega^{\delta_k}} .$$

As in Hall and Marron (1988), one shows

$$\text{(B.39)} \qquad \left\| \left\| \frac{1}{nh_k} \sum_{j=1}^{n} \psi\left(\frac{\cdot - X_j}{C_2 h} \right) \right\|_{\Omega^{\delta_k}} \right\|_{L^p} = O(1) \quad \text{for any } p > 0 ,$$

using Rosentahl's inequality. Cauchy–Schwarz inequality and (B.37)–(B.39) prove (B.36). It remains to prove (B.37). For $v = 1, 2$, this follows easily from (B.12). For $v = 3$, put

$$H(\kappa) = H(\kappa, x, z) = K\left(\frac{(x - z)\kappa}{h_k} \right) .$$

Then

(B.40)

$$\begin{aligned} \varepsilon_3(x, z, h_k) &= \tilde{\beta}_k (H(\tilde{\alpha}_k) - H(\alpha_{bk}) - (\tilde{\alpha}_k - \alpha_{bk}) H'(\alpha_{bk})) \\ &= \tilde{\beta}_k \int_{\alpha_{bk}}^{\tilde{\alpha}_k} (\tilde{\alpha}_k - \kappa) H^{(2)}(\kappa)\, d\kappa , \end{aligned}$$

where we have omitted x and z in the notation for simplicity. Define $\tilde{V}_{\alpha k}(x, z) = \tilde{\alpha}_k(x, z) - \alpha_{bk}(x, z) = \bar{W}_{\alpha k}(x, z, x) + \tilde{R}_{\alpha k}(x, z)$ and $\tilde{V}_{\beta k}$ similarly. By differentiating w.r.t. x repeatedly in (B.40), one can show that

$$\text{(B.41)} \qquad \|\varepsilon_3^{(i)}\|_{\Lambda_k} \leqq C \sum_{v_1, v_2, v_3, v_4} h^{-v_1} \|\tilde{V}_{\alpha k}^{(v_2)}\|_{\Lambda_k} \|\tilde{V}_{\alpha k}^{(v_3)}\|_{\Lambda_k} \|\tilde{\beta}_k^{(v_4)}\|_{\Lambda_k} ,$$

where the sum ranges over all non-negative v_1, v_2, v_3, v_4 with $v_1 + v_2 + v_3 + v_4 = i$. Similarly as for (B.18), it is possible to show that

$$\text{(B.42)} \qquad \left\| \|\tilde{V}_{\alpha k}^{(i)}\|_{\Lambda_k} \right\|_{L^p} , \left\| \|\tilde{V}_{\beta k}^{(i)}\|_{\Lambda_k} \right\|_{L^p} \leqq C (nh)^{-1/2} h^{-i} n^{\varepsilon}$$

for any $p > 0$ and $0 \leqq i \leqq \bar{J}_k$. Formula (B.37) for $v = 3$ now follows from (B.41) and (B.42), (B.8)–(B.10), (ii) and Hölder's inequality. When $v = 4$ we have the estimate

$$\|\varepsilon_4^{(i)}\|_{\Lambda_k} \leqq C \sum_{v_1, v_2, v_3} h^{-v_1} \|\tilde{V}_{\alpha k}^{(v_2)}\|_{\Lambda_k} \|\tilde{V}_{\beta k}^{(v_3)}\|_{\Lambda_k} ,$$

summing over non-negative indices with $v_1+v_2+v_3=i$, and then the rest follows as for $v=3$.

For $v=5$, (B.36) is proved similarly as (B.28) was proved in Lemma B.3, using an estimate like (B.29) for the effective kernel involved. The main ingredient for the cases $v=6,7$ is Rosentahl's inequality for degenerate U-statistics of order 2, see Hall and Marron (1988) for such a proof. (Put $l(X_i,X_j) = l_2(x,X_i,X_j) - \bar{L}_k^{\mathrm{ad},1}(x,X_j,x)$, $i \neq j$. Then $l(X_i,X_j) - E(l(X_i,X_j)\,|\,X_i)$ can be approximated by a degenerate kernel $\tilde{l}(X_i,X_j)$, satisfies $E(\tilde{l}(X_i,X_j)\,|\,X_j) = E(\tilde{l}(X_i,X_j)\,|\,X_i) = 0$ a.s.) □

Lemma B.6 *Suppose* $2 \leqq k \leqq t$ *and that* α_{bk} *and* β_{bk} *have the expansions given in* (3.4), *with* (B.8)–(B.10) *being satisfied. Then* f_{bk}, *defined in* (3.8), *has the expansion* (3.2). *The main bias term* b_k *is defined in* (3.9)–(3.10), *and* (B.13)–(B.14) *hold.*

Proof[4]. As in the proof of Lemma B.5, we omit h_{k-1} and h_k in the notation. We assume $s(k)=s(k-1)+2$. The case $s(k+1)=s(k)$ is similar, but simpler. Formula (B.13) follows by the definition of b_k in (3.9) and (3.10) in conjunction with (iv), (B.8) and (B.9). Notice that J_k derivatives of b_k are required, and hence J_k+2 derivatives of $b_{\alpha k}$ and $b_{\beta k}$. However, since $\tilde{J}_k = J_k+2$ when $s(k)=s(k-1)+2$, $b_{\alpha k}$ and $b_{\beta k}$ have the required number of derivatives.
We now turn to r_k. In order to establish (B.14), we first need some expansions. Put $\bar{b}_{\alpha k} = b_{\alpha k} + r_{\alpha k}$ and $\bar{b}_{\beta k} = b_{\beta k} + r_{\beta k}$, so that $\alpha_{bk} = \alpha_k + \bar{b}_{\alpha k}$ and

(B.43)
$$\beta_{bk}(x,z) = \beta_k(x,z) + \bar{b}_{\beta k}(x,z)\,.$$

Perform the Taylor expansion

(B.44)
$$K\left(\frac{(x-z)\alpha_{bk}(x,z)}{h_k}\right) = K\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right) + \frac{\bar{b}_{\alpha k}(x,z)}{\alpha_k(x,z)}\check{K}\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right) + \left(\frac{\bar{b}_{\alpha k}(x,z)}{\alpha_k(x,z)}\right)^2 \hat{K}\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right) + \varepsilon(x,z)\,,$$

with ε a remainder term, $\check{K}(v) = vK'(v)$ and $\hat{K}(v) = v^2K^{(2)}(v)/2$. Inserting (B.43) and (B.44) into (3.8) gives

(B.45)
$$f_{bk}(x) = \frac{1}{h_k}\int \beta_k(x,z)K\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right)f(z)\,dz + \frac{1}{h_k}\int \bar{b}_{\beta k}(x,z)K\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right)f(z)\,dz + \frac{1}{h_k}\int \frac{\beta_k(x,z)\bar{b}_{\alpha k}(x,z)}{\alpha_k(x,z)}\check{K}\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right)f(z)\,dz$$

[4] As in the proof of Lemma B.5, we tacitly assume $|z-x| \leqq C_2h$ to avoid tail effects. This is no restriction because of (B.1), and implies that all integrals w.r.t. z have bounded domain of integration, in particular in the definition of f_{bk}

$$+\frac{1}{h_k}\int \frac{\bar{b}_{\alpha k}(x,z)\bar{b}_{\beta k}(x,z)}{\alpha_k(x,z)}\check{K}\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right)f(z)\,dz$$

$$+\frac{1}{h_k}\int \frac{\beta_k(x,z)\bar{b}_{\alpha k}(x,z)^2}{\alpha_k(x,z)^2}\hat{K}\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right)f(z)\,dz$$

$$+\frac{1}{h_k}\int \frac{\bar{b}_{\alpha k}(x,z)^2\bar{b}_{\beta k}(x,z)}{\alpha_k(x,z)^2}\hat{K}\left(\frac{(x-z)\alpha_k(x,z)}{h_k}\right)f(z)\,dz$$

$$+\frac{1}{h_k}\int \beta_{bk}(x,z)\varepsilon(x,z)f(z)\,dz$$

$$:=\sum_{v=1}^{5}T_v(x)+\sum_{v=6}^{7}r_{kv}(x)\,,$$

with r_{k6} and r_{k7} remainder terms. Making use of Theorem 1 in Hall (1990) we expand the first five terms of (B.45) as follows:

$$\text{(B.46)}\quad T_1(x)=f(x)+\gamma_{kj}(x)h_k^{s(k)}+r_{k1}(x)\,,$$

$$T_2(x)=\frac{\bar{b}_{\beta k}(x,x)f(x)}{\alpha_k(x,x)}+\frac{\mu_2(K)}{2}\left[\frac{b_{\beta k}(x,z)f(z)}{\alpha_k(x,z)^3}\right]_{z=x}^{(0,2)}h_k^2$$

$$+\frac{\mu_2(K)}{2}\left[\frac{r_{\beta k}(x,z)f(z)}{\alpha_k(x,z)^3}\right]_{z=x}^{(0,2)}h_k^2+r_{k2}(x)\,,$$

$$T_3(x)=-\frac{\bar{b}_{\alpha k}(x,x)f(x)}{\alpha_k(x,x)}-\frac{3\mu_2(K)}{2}\left[\frac{b_{\alpha k}(x,z)\beta_k(x,z)f(z)}{\alpha_k(x,z)^4}\right]_{z=x}^{(0,2)}h_k^2$$

$$-\frac{3\mu_2(K)}{2}\left[\frac{r_{\alpha k}(x,z)\beta_k(x,z)f(z)}{\alpha_k(x,z)^4}\right]_{z=x}^{(0,2)}h_k^2+r_{k3}(x)\,,$$

$$T_4(x)=-\frac{\bar{b}_{\alpha k}(x,x)\bar{b}_{\beta k}(x,x)f(x)}{\alpha_k(x,x)^2}+r_{k4}(x)\,,$$

$$T_5(x)=\frac{\bar{b}_{\alpha k}(x,x)^2f(x)}{\alpha_k(x,x)^2}+r_{k5}(x)\,,$$

where we have used (2.9) (which follows from (ix)), $\mu_0(\check{K})=-1$, $\mu_2(\check{K})=-3\mu_2(K)$ and $\mu_0(\hat{K})=1$. Inserting the expansions (B.46) into (B.45) we obtain,

(B.47)

$$r_k(x)=f_{bk}(x)-f(x)-b_k(x)$$

$$=\frac{\mu_2(K)}{2}\left[\frac{r_{\beta k}(x,z)f(z)}{\alpha_k(x,z)^3}\right]_{z=x}^{(0,2)}h_k^2-\frac{3\mu_2(K)}{2}\left[\frac{r_{\alpha k}(x,z)\beta_k(x,z)f(z)}{\alpha_k(x,z)^4}\right]_{z=x}^{(0,2)}h_k^2$$

$$+\sum_{v=1}^{7}r_{kv}(x):=\sum_{v=0}^{7}r_{kv}(x)\,,$$

where we have used the definitions of r_k and b_k in (3.2), (3.9) and (3.10). Notice that several terms cancel since

$$\bar{b}_{\alpha k}(x,x) = \bar{b}_{\beta k}(x,x) \,,$$

which follows from (ix), (2.9) and (5.2).

It remains to establish (B.14) for each of the terms in (B.47). For r_{k0}, this is proved in the same way as (B.13) was for b_k, making use of (iv), (B.8) and (B.10). $r_{k1}, \ldots, r_{k5}$ are all remainder estimates in various Taylor series expansions based on Theorem 1 in Hall (1990). We omit the details, but since J_k derivatives w.r.t. x are required for each r_{kv}, and we make Taylor expansions of T_1–T_5 up to order 2, we must require $J_k + 2$ continuous derivatives for the functions f, α_k, β_k, $\bar{b}_{\alpha k}$ and $\bar{b}_{\beta k}$ appearing in T_1–T_5. But this follows from (iv) and (B.8)–(B.10) (see the remark after (B.8)–(B.16)).

To handle r_{k6}, perform the change of variables $v = \alpha_k(x,z)(x-z)$ in the integral defining r_{k6} and then differentiate under the integral sign J_k times. Finally, for r_{k7}, notice first that $\varepsilon(x,z)$ is a remainder term in a Taylor expansion, and therefore

$$\varepsilon(x,z) = \frac{1}{2}(\alpha_{bk} - \alpha_k)^3 \int_0^1 (1-\rho)^2 \left(\frac{x-z}{h_k}\right)^3 \times K^{(3)}\left(\frac{(x-z)(\alpha_k + \rho(\alpha_{bk} - \alpha_k))}{h_k}\right) d\rho \,,$$

with $\alpha_k = \alpha_k(x,z)$ and $\alpha_{bk} = \alpha_{bk}(x,z)$. Insert this identity into the integral r_{k7}, change order of integration between $d\rho$ and dz and change variables $v = (\alpha_k + \rho(\alpha_{bk} - \alpha_k))(x-z)$ for each fixed ρ. Finally, differentiate w.r.t. x up to J_k times, and move the differentiation operator under the inner integral. The rest is similar to r_{k6}. □

Appendix C

Proof of Theorem 7.1

Since $h_1, \ldots, h_{t-1}$ are all of the same order, Theorem 5.1 implies (B.13)–(B.16) for $k = t-1$ and (B.8)–(B.12) for $k = t$. Using this, we will prove, for some $\varepsilon > 0$ and all $p > 0$,

(C.1) $$\big\| \| W_t^{\mathrm{ad}} \|_{\Omega^{\delta_t}} \big\|_{L^p} = O((nh_t)^{-1/2} n^{-\varepsilon}) \,,$$

and

(C.2) $$\big\| \| R_t \|_{\Omega^{\delta_t}} \big\|_{L^p} = O((nh_t)^{-1/2} n^{-\varepsilon}) \,.$$

We will also prove

(C.3) $$\| r_t \|_{\Omega^{\delta_t}} = o(h^{m(t)}) \,.$$

The theorem then follows from (C.1)–(C.3). By making the change-of-variables $v = (x'-z)\alpha_k(x,x)/h_k$ in (4.8) and differentiating under the integral sign, it follows from (B.11) (for $k = t$) that

$$\|(\bar{L}_t^{\mathrm{ad}})^{(i,0,d)}\|_{\bar{\Lambda}_k} = O(h^{-(1+i)}) \quad 0 \leqq i+d \leqq \bar{J}_t + 1 = 1 \,,$$

which implies

$$\text{(C.4)} \qquad \|(L_t^{\mathrm{ad}})^{(i)}\|_{\check{\Lambda}_k} = O(h^{-(1+i)}) \quad i = 0, 1 \, .$$

It follows from (4.8) and (B.5) that

(C.5) $\quad L_k^{\mathrm{ad}}(x, \cdot\,; \mathbf{h}_t)$ is supported on $[x - \bar{C}_t h, x + \bar{C}_t h]$ for any $x \in \Omega^{\delta_t}$.

As in the proof of Lemma B.2, (C.4) and (C.5) imply

$$\text{(C.6)} \qquad \big\| \|W_t^{\mathrm{ad}}\|_{\Omega^{\delta_t}} \big\|_{L^p} = O((nh)^{-1/2} n^{\varepsilon})$$

for any $p > 0$ and $\varepsilon > 0$. This implies (C.1), if we choose ε in (C.6) small enough compared to ε_2 in (ib).

Formula (C.2) is proved as in Lemma B.5, provided we make some small adjustments for the fact $h_t \ll h$. For instance, (B.37) and (B.38) become (notice that $\bar{J}_t = 0$)

$$\big\| \|\varepsilon_v\|_{\Lambda_t} \big\|_{L^p} = O((nh_t)^{-1/2} n^{-\varepsilon}) \quad \text{for any } p > 0$$

and

$$\|R_{tv}\|_{\Omega^{\delta_t}} \leqq \|\varepsilon_v\|_{\Lambda_t} \left\| \frac{1}{nh_t} \sum_{j=1}^{n} \psi\left(\frac{\cdot - X_j}{C_1 h_t / \chi_0} \right) \right\|_{\Omega^{\delta_t}} ,$$

for $v = 1, 2, 3, 4$. The rest of the proof is analogous to Lemma B.5.

Finally, (C.3) is derived as in Lemma B.6. For the remainder terms $r_{t0} - r_{t7}$ in that proof we obtain

$$\|r_{tv}\|_{\Omega^{\delta_t}} = \begin{cases} o(h_t^{m(t)}) & v = 1, \\ o(h^{s(t-1)} h_t^2) & v = 0, 2, 3, \\ o(h^{2s(t-1)}) & v = 4, 5, \\ O(h^{3s(t-1)}) & v = 6, 7 \, . \end{cases}$$

All the quantities above are $o(h_t^{m(t)})$. For $v = 0, 2, 3$ this follows from (ic) and (id). When $v = 4, 5$, (ic) and (id) imply $h^{2s(t-1)} = h^{2m(t)} \ll h_t^{2(m(t)-2)} = O(h^{m(t)})$.

Appendix D

Some consequences of the regularity conditions

We will state some consequences of (5.1) in Sect. 5, that are used in the proof of Lemma B.3. First some notation. For a function $g : \mathbb{R} \to \mathbb{R}$ that is v times differentiable on a set $\Upsilon \subset \mathbb{R}$, define

$$\|g\|_{v,\Upsilon} = \sum_{\{v_\mu\}} \prod_{\mu} \|g^{(v_\mu)}\|_{\Upsilon} \, ,$$

where the sum ranges over the finite collection of sequences $0 \leqq \nu_1 \leqq \cdots \leqq \nu_r$ with $\nu_2 > 0$ if $r > 1$ and $\sum_\mu \nu_\mu = \nu$. Similarly, given two functions g_0 and g_1, we write

$$\|\{g_0, g_1\}\|_{\nu,\Upsilon} = \sum_{\{\nu_\mu\}} \prod_\mu \max(\|g_0^{(\nu_\mu)}\|_\Upsilon, \|g_1^{(\nu_\mu)}\|_\Upsilon)\,.$$

Recall the definitions of Ξ and $P(\Xi)$ from (x), and the function $h(x,z;\boldsymbol{\theta}) = P_{kl}(x,z;g(\,\cdot\,;\boldsymbol{\theta}))$.

(I) Let $p = 0$, so that $g(x;\Theta) = g(x)$. Then (5.1) reduces to

$$\text{(D.1)} \qquad \|P_{kl}^{(i,j)}(\,\cdot\,,\,\cdot\,;g)\|_\Xi \leqq C \sum_{\{\nu_\mu\}} \prod_\mu \|g^{(\nu_\mu)}\|_{P(\Xi)}\,,$$

where the sum is taken over all finite sequences $\{\nu_\mu\}$ with $\sum_\mu \nu_\mu = l + i + j$ and at most one $\nu_\mu > 0$.

(II) Let $p = 1$, with $\Theta = \{0\}$ and $g(x;\theta) = g_0(x) + \theta\eta(x)$. Then $dP_{kl}^{(i,j)}(x,z;g_0)(\eta) = h_{kl}^{(i,j,1)}(x,z;0)$. Notice that $g^{(\nu,d)}$ equals $g_0^{(\nu)}$ when $d = 0$ and $\eta^{(\nu)}$ if $d = 1$. Application of (5.1) yields

$$\begin{aligned}
\text{(D.2)} \qquad \|dP_{kl}^{(i,j)}(\,\cdot\,,\,\cdot\,;g_0)(\eta)\|_\Xi &= \|h_{kl}^{(i,j,1)}\|_{\Xi\times\Theta} \\
&\leqq C \sum_{\{\nu_\mu\}} \|\eta^{(\nu_1)}\|_{P(\Xi)} \prod_{\mu\geqq 2} \|g_0^{(\nu_\mu)}\|_{P(\Xi)} \\
&\leqq C \sum_{\nu=0}^{l+i+j} \|\eta^{(\nu)}\|_{P(\Xi)} \|g_0\|_{l+i+j-\nu,P(\Xi)}\,,
\end{aligned}$$

with $\sum_\mu \nu_\mu = l + i + j$, and at most one $\nu_\mu > 0$ for $\mu \geqq 2$.

(III) Let $p = 2$, $\boldsymbol{\theta} = (\theta_1, \theta_2)$, $\Theta = (\{0\}, [0,1])$, $g(x;\boldsymbol{\theta}) = g_0(x) + \theta_1\eta(x) + \theta_2(g_1(x) - g_0(x))$. Then, $dP_{kl}^{(i,j)}(x,z;g_1)(\eta) - dP_{kl}^{(i,j)}(x,z;g_0)(\eta) = \int_0^1 h_{kl}^{(i,j,1,1)}(x,z;0,\theta_2)\,d\theta_2$. Let $g_{\theta_2}(x) = g_0(x) + \theta_2(g_1(x) - g_0(x))$. Observe that $g^{(\nu,d_1,d_2)}$ equals $g_{\theta_2}^{(\nu)}$ if $d_1 = d_2 = 0$, $\eta^{(\nu)}$ if $d_1 = 1$, $d_2 = 0$, $(g_1 - g_0)^{(\nu)}$ if $d_1 = 0, d_2 = 1$ and 0 if $d_1 + d_2 \geqq 2$. Applying (5.1) gives

$$\begin{aligned}
\text{(D.3)} \quad \|dP_{kl}^{(i,j)}(\,\cdot\,,\,\cdot\,;g_1)(\eta) - dP_{kl}^{(i,j)}(\,\cdot\,,\,\cdot\,;g_0)(\eta)\|_\Xi &\leqq \|h_{kl}^{(i,j,1,1,)}\|_{\Xi\times\Theta} \\
&\leqq C \sup_{0\leqq\theta_2\leqq 1} \sum_{\{\nu_\mu\}} \|\eta^{(\nu_1)}\|_{P(\Xi)} \|(g_1 - g_0)^{(\nu_2)}\|_{P(\Xi)} \prod_{\mu\geqq 3} \|g_{\theta_2}^{(\nu_\mu)}\|_{P(\Xi)} \\
&\leqq C \sum_{\nu_1,\nu_2} \|\eta^{(\nu_1)}\|_{P(\Xi)} \|(g_1 - g_0)^{(\nu_2)}\|_{P(\Xi)} \|\{g_0, g_1\}\|_{l+i+j-\nu_1-\nu_2,P(\Xi)}
\end{aligned}$$

with $\sum_\mu \nu_\mu = l + i + j$.

(IV) Choose $p = 1$ and $\Theta = [0,1]$. Given functions g_0 and g_1, put $g(x;\theta) = g_0(x) + \theta(g_1(x) - g_0(x)) := g_\theta(x)$. Then $P_{kl}^{(i,j)}(x,z;g_1) - P_{kl}^{(i,j)}(x,z;g_0) - dP_{kl}^{(i,j)}(x,z;g_0)(g_1 - g_0) = \int_0^1 (1-\theta) h^{(i,j,2)}(x,z;\theta)\,d\theta$. Notice also that $g^{(\nu,d)}$

equals $g^{(v)}$ if $d = 0$, $(g_1 - g_0)^{(v)}$ if $d = 1$ and 0 if $d \geqq 2$. Hence,

(D.4)

$$\begin{aligned}
&\|P_{kl}^{(i,j)}(\cdot,\cdot;g_1) - P_{kl}^{(i,j)}(\cdot,\cdot;g_0) - dP_{kl}^{(i,j)}(\cdot,\cdot;g_0)(g_1 - g_0)\|_{\Xi} \\
&\leqq \frac{1}{2}\|h^{(i,j,2)}\|_{\Xi\times\Theta} \leqq \sup_{0\leqq\theta\leqq 1} C \sum_{\{v_\mu\}} \Bigg(\|(g_1 - g_0)^{(v_1)}\|_{P(\Xi)} \\
&\times \|(g_1 - g_0)^{(v_2)}\|_{P(\Xi)} \prod_{\mu\geqq 3} \|g_\theta^{(v_\mu)}\|_{P(\Xi)} \Bigg) \\
&\leqq C \sum_{v_1,v_2} \|(g_1 - g_0)^{(v_1)}\|_{P(\Xi)} \|(g_1 - g_0)^{(v_2)}\|_{P(\Xi)} \|\{g_0, g_1\}\|_{l+i+j-v_1-v_2,P(\Xi)} ,
\end{aligned}$$

the sum ranging over all sequences $\{v_\mu\}$ with $\sum_\mu v_\mu = l + i + j$ and at most one $v_\mu > 0$.

Equations (D.1)–(D.4) can also be extended to the case when g_0, g_1 or η depend on finite dimensional parameters. We illustrate this for (D.2): Assume $g_0 = g_0(\cdot;\boldsymbol{\theta}_1)$, $\eta = \eta(\cdot;\boldsymbol{\theta}_2)$, with $\boldsymbol{\theta}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\theta}_2 \in \mathbb{R}^{p_2}$. Put $\tilde{h}(x,z,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = dP_{kl}(x,z;g(\cdot,\boldsymbol{\theta}_1))(\eta(\cdot,\boldsymbol{\theta}_2))$. We consider domains of $\tilde{h}$ such that for each fixed $(x,z) \in \Xi$, $\boldsymbol{\theta}_1 \in \Theta_{1xz} \subset \mathbb{R}^{p_1}$ and $\boldsymbol{\theta}_2 \in \Theta_{2xz} \subset \mathbb{R}^{p_2}$. Then it follows from (5.1), similarly as for (D.2), that

$$\|\tilde{h}^{(i,j,\mathbf{d}_1,\mathbf{d}_2)}\|_{\Upsilon} \leqq C \sum_{\{\mathbf{v}_\mu\}} \|\eta^{(\mathbf{v}_1)}\|_{\Upsilon_2} \prod_{\mu\geqq 2} \|g^{(\mathbf{v}_\mu)}\|_{\Upsilon_1} , \tag{D.5}$$

where $\Upsilon = \bigcup_{[x,z]\in\Xi} (x,z) \times \Theta_{1xz} \times \Theta_{2xz}$, $\Upsilon_1 = \bigcup_{(x,z)\in\Xi} [x,z] \times \Theta_{1xz}$ and $\Upsilon_2 = \bigcup_{(x,z)\in\Xi} [x,z] \times \Theta_{2xz}$. The sum in (D.5) ranges over the finite collection of sequences $\{\mathbf{v}_\mu = (v_{\mu 1}, \mathbf{v}_{\mu 2})\}_{\mu=1}^r$ with $\sum_\mu v_{\mu 1} = l + i + j$, $\mathbf{v}_{12} = \mathbf{d}_2$, $\sum_{\mu\geqq 2} \mathbf{v}_{\mu 2} = \mathbf{d}_1$, and at most one $\mathbf{v}_\mu$ equal to zero for $\mu \geqq 2$.

Acknowledgement. The author would like to thank three anonymous referees for very useful suggestions on an earlier version of this paper that made the presentation clearer.

References

Abramson, L.S.: On bandwidth variation in kernel estimates – a square root law. Ann. Stat. **10**, 1217–1223 (1982)

Breiman, L., Meisel, W., Purcell, E.: Variable kernel estimates of probability densities. Technometrics **19**, 135–144 (1977)

Fernholz, L.T.: Von Mises Calculus for Statistical Functionals (Lect. Notes Statist. 19) New York: Springer 1983

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

Hall, P. Hu, T.-C., Marron, J.S.: Improved variable window estimators of probability densities. Ann. Statist. **23**, 1–10 (1995)

Hall, P., Marron, J.S.: Variable window width kernel estimates of probability densities. Probab. Theory Relat. Fields **80**, 37–49 (1988)

Hössjer, O., Ruppert, D.: Transformation kernel density estimators using stochastic bandwidths. Technical Report 1993:12, University of Lund, Dept. of Mathematical Statistics (1993)

Hössjer, O., Ruppert, D.: Taylor series approximations of kernel density estimators. J. Nonparametric Statistics **4**, 165–177 (1994)

Hössjer, O., Ruppert, D.: Asymptotics for the transformation kernel density estimator. Technical Report 1993:10, University of Lund, Dept. of Mathematical Statistics. Ann. Stat. **23**, 1198–1222 (1995)

Jones, M.C.: Variable kernel density estimates and variable kernel density estimates. Austral. J. Statist. **32**, 361–71. Correction, **33**, 119 (1990)

Jones, M.C., Linton, O., Nielsen, J.P.: A simple and effective bias reduction method for density and regression estimation. Biometrika **82**, 327–338 (1995)

Jones, M.C., McKay, I.J., Hu, T.-C.: Variable location and scale kernel density estimation. Ann. Inst. Statist. Math. **46**, 521–535 (1994)

Jones, M.C., Signorini, D.F.: A comparison of higher order bias kernel density estimators to appear in J. Amer. Statist. Assoc. (1996)

Loftsgaarden, D.O., Quesenberry, C.P.: A nonparametric estimate of a multivariate density function. Ann. Math. Statist. **36**, 1049–1051 (1965)

McKay, I.J.: Variable kernel density methods in density estimation. Ph.D. thesis, Queens University, (1993)

McKay, I.J.: Bias reduction in variable kernel density estimates. (submitted, 1995)

Marron, S., Wand, M.: Exact mean integrated squared error. Ann. Statist. **20**, 712–736 (1992)

Ruppert, D., Cline, D.: Bias reduction in kernel density estimation by smoothed empirical transformations. Ann. Statist. **22**, 185–210 (1994)

Samiuddin, M., El-Sayyad, G.M.: On nonparametric density estimates. Biometrika **77**, 865–874 (1990)

Terrell, G.R., Scott, D.W.: Variable kernel density estimation. Ann. Statist. **20**, 1236–1263 (1992)