Ann Inst Stat Math
https://doi.org/10.1007/s10463-018-0663-z

Unified estimation of densities on bounded and unbounded domains

Kairat Mynbaev[1] · **Carlos Martins-Filho**[2,3]

Received: 10 July 2017 / Revised: 31 January 2018

Abstract Kernel density estimation in domains with boundaries is known to suffer from undesirable boundary effects. We show that in the case of smooth densities, a general and elegant approach is to estimate an extension of the density. The resulting estimators in domains with boundaries have biases and variances expressed in terms of density extensions and extension parameters. The result is that they have the same rates at boundary and interior points of the domain. Contrary to the extant literature, our estimators require no kernel modification near the boundary and kernels commonly used for estimation on the real line can be applied. Densities defined on the half-axis and in a unit interval are considered. The results are applied to estimation of densities that are discontinuous or have discontinuous derivatives, where they yield the same rates of convergence as for smooth densities on $\mathbb{R}$.

Keywords Nonparametric density estimation · Hestenes' extension · Estimation in bounded domains · Estimation of discontinuous densities

We thank two anonymous referees for their comments and the Institute of Mathematics and Mathematical Modeling for hosting our research. Any remaining errors are the authors' responsibility. This project has been partially supported by N 0118PK00476 from the Ministry of Education and Science of the Republic of Kazakhstan.

✉ Carlos Martins-Filho
carlos.martins@colorado.edu; c.martins-filho@cgiar.org

Kairat Mynbaev
kairat_mynbayev@yahoo.com

1 International School of Economics, Kazakh-British Technical University, Tolebi 59, Almaty, Kazakhstan 050000

2 Department of Economics, University of Colorado, Boulder, CO 80309-0256, USA

3 IFPRI, 2033 K Street NW, Washington, DC 20006-1002, USA

Published online: 04 May 2018

1 Introduction

Kernel estimation of densities on the real line is a well-developed area. The core of the theory is a series of results covering smooth densities that do not exhibit extreme curvature. Let K denote a kernel, an integrable function on $\mathbb{R}$, which satisfies $\int_{\mathbb{R}} K(t)\mathrm{d}t = 1$, $h > 0$ be a bandwidth and f be a density on $\mathbb{R}$. Assuming that $\{X_i\}_{i=1}^n$ is an independent and identically distributed (IID) sample from f, the traditional Rosenblatt–Parzen kernel estimator of $f(x)$ is defined by $\hat{f}_R(x) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$. This estimator has three desirable characteristics: (1) there exists a great profusion of kernels that can be used to construct the estimator (usually Epanechnikov, Gaussian or triangular densities); they are usually symmetric and do not depend on the point (x) of estimation, or on the class of densities being estimated; (2) there is a simple link between the degree of smoothness of the density and the order of estimator's bias: if $f \in \mathcal{C}_b^s(\Omega)$ and the kernel is of order s, then $E\hat{f}_R(x) - f(x) = O(h^s)$.[1] The use of higher-order kernels in the case of smooth densities is also a standard feature; (3) the optimal bandwidth is of order $n^{-1/(2s+1)}$ for all estimation points, unless there are areas of extreme curvature or discontinuities.

In cases where the domain of f has a boundary, the main problem is bad estimator behavior in the vicinity of the boundary. This problem has called into being a range of estimation methods. Among the widely used ones are the reflection, the boundary kernel, the transformation and the local linear methods (see, *inter alia*, Schuster 1985; Jones 1993; Cheng 1994; Karunamuni and Alberts 2005; Malec and Schienle 2014; Wen and Wu 2015, and their references). Other methods have proposed the use of asymmetric kernels and kernel adjustments near the boundary (Chen 1999, 2000). Such techniques, alternatively, require variable bandwidths, two-step estimation procedures, separation of densities into subclasses that vanish or not at the boundary, densities that have derivatives of a certain sign at the boundary, etc. The difficulties in estimation near the boundary precluded researchers from identifying a core class of estimators for which analogs of the standard results mentioned above would be true. In particular, we have not seen in the literature results that would guarantee a better bias rate for densities of higher smoothness.

In this paper, we propose density estimators that permit a unified theoretical study of their properties under bounded or unbounded domains. We show that smoothness is all one needs to have a good bias rate, and for smooth densities the behavior at the boundary is irrelevant. (Derivatives at endpoints are one-sided derivatives.) For densities on the half-axis $[0, \infty)$ and on the unit interval $[0, 1]$ we introduce new estimators for which all standard facts given above hold. The usual symmetric kernels and constant bandwidths can be used across the domain, and for $f \in \mathcal{C}_b^s(\Omega)$ the biases of our estimators are of order $O(h^s)$. The bandwidth depends on the sample size in the same way as in case of estimation on the whole line. In the case of estimation of piece-wise continuous densities, with

[1] Let $s \in \mathbb{N}$ and $\Omega \subseteq \mathbb{R}$. The class of functions $f : \Omega \to \mathbb{R}$ which are s-times differentiable with $\left|f^{(s)}(x)\right| \le C$ for some $0 < C < \infty$ is denoted by $\mathcal{C}_b^s(\Omega)$. We say that the kernel K is of order $s \ge 2$ if $\int t^j K(t)\mathrm{d}t = 0$ for $j = 1, \ldots, s-1$ and $\int t^s K(t)\mathrm{d}t \ne 0$.

known discontinuity points, our estimators supply the required jumps at those points. As in some boundary correction estimators (Jones and Foster 1993; Cheng 1994), for densities in classes where $s \geq 1$ is our estimator is not necessarily nonnegative, because the estimation essentially involves higher-order kernels. Our theoretical results do not hold for densities or densities with derivatives with poles at endpoints.

Our estimators are based on Hestenes' extension (Hestenes 1941) of continuously differentiable functions from subsets $\Omega \subset \mathbb{R}$ to $\mathbb{R}$. Let D_f be the domain of the density f and denote by g its Hestenes' extension. (The definitions for the half-axis and intervals are given below in the respective sections.) The key observation is that g can be viewed as a linear combination of densities. The sample generated from f is used to estimate each of these densities, and the linear combination of the estimators estimates g. The restriction of the estimator of g to D_f estimates f. We show that the theory of estimation on a domain with boundaries for smooth densities in effect becomes a chapter in estimation on the whole line. The essential link between the proposed estimators $\hat{f}(x)$ of $f(x)$ and the properties of g is of type

$$E\hat{f}(x) - f(x) = \int_{\mathbb{R}} K(t)\,(g(x-ht) - g(x))\,\mathrm{d}t,\ x \in D_f.$$

This representation has eluded previous work and can be used for evaluating the asymptotic behavior of bias. Our estimation procedure does not require knowledge of g. There seems to be a slight loss in the speed of convergence as compared to convergence on the line because the same data are exploited more than once to estimate different parts of g. However, this loss does not affect the rate in $E\hat{f}(x) - f(x) = ch^s + o(h^s)$; it affects only the constant c, in comparison with the classical estimator for densities on the line. In Sect. 2, we start with estimation of a density on $[0, \infty)$. Section 3 treats densities on a bounded interval. In Sect. 4, the approach is extended to estimation of discontinuous densities. Section 5 provides two methods to satisfy zero boundary conditions, and Sect. 6 provides results from a Monte Carlo simulation. Section 7 concludes the paper and gives directions for future research. All proofs are collected in an "Appendix".

2 Estimation of densities defined on $[0, \infty)$

Let $w_1, \ldots, w_{s+1}$ be pairwise different positive numbers for $s = 0, 1, \ldots$. Of special interest are the decreasing sequence $w_i = 1/i$, $i = 1, \ldots, s+1$ (used by Hestenes 1941) and the increasing sequence $w_i = i$. Let the numbers $k_1, \ldots, k_{s+1}$ be defined from the following system

$$\sum_{i=1}^{s+1} (-w_i)^j k_i = 1,\ \ j = 0, \ldots, s. \tag{1}$$

Since this system has the Van der Monde determinant

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ -w_1 & -w_2 & \dots & -w_{s+1} \\ \dots & \dots & \dots & \\ (-w_1)^s & (-w_2)^s & \dots & (-w_{s+1})^s \end{vmatrix} \neq 0,$$

$k_1, \dots, k_{s+1}$ are uniquely defined. If $f \in \mathcal{C}_b^s$ on its domain $D_f = [0, \infty)$, its Hestenes' extension to $(-\infty, 0)$ is given by

$$\phi_s(x) = \sum_{j=1}^{s+1} k_j f(-w_j x), \; x < 0. \tag{2}$$

Note that if f is a density function, ϕ_s is not a density, but a linear combination of densities $w_j f(-w_j x)$ with coefficients k_j/w_j. Assuming that f has s right-hand derivatives $f(0+), \dots, f^{(s)}(0+)$ at zero ($s = 0$ means continuity), we see that the following sewing conditions at zero are satisfied due to (1):

$$\phi_s^{(m)}(0-) = \sum_{j=1}^{s+1} (-w_j)^m k_j f^{(m)}(0+) = f^{(m)}(0+), \quad m = 0, 1, \dots, s.$$

Now, define g_s on $\mathbb{R}$ by

$$g_s(x) = \begin{cases} f(x), & x \geq 0 \\ \phi_s(x), & x < 0 \end{cases}, \tag{3}$$

with g_s being s times differentiable. Moreover, if, for example, f belongs to the Sobolev space $W_p^s([0, \infty))$, then g_s belongs to $W_p^s(\mathbb{R})$, where $1 \leq p < \infty$ (see Burenkov 1998).

Suppose $f \in \mathcal{C}_b^s$, the kernel K is m times differentiable, where $m = 0, 1, \dots, s$, and let $\{X_i\}_{i=1}^n$ be an IID sample from f. We define the estimator of $f^{(m)}(x)$, for $x \geq 0$, by

$$\hat{f}_s^{(m)}(x) = \frac{1}{nh^{m+1}} \sum_{i=1}^n \left[K^{(m)}\left(\frac{x - X_i}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K^{(m)}\left(\frac{x + X_i/w_j}{h}\right) \right]. \tag{4}$$

When the kernel K is an even function and $m = s = 0$ in (4), $\hat{f}_0^{(0)}(x) \equiv \hat{f}_S(x)$ is the "reflection estimator" from Schuster (1985), i.e.,

$$\hat{f}_S(x) = \frac{1}{nh} \sum_{i=1}^n \left[K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + X_i}{h}\right) \right].$$

We note that Schuster's estimator does not depend on s, the index on $\mathcal{C}_b^s$. Thus, knowledge that $s > 0$ is not used in constructing his estimator, whereas it is central in the definition of $\hat{f}_s^{(m)}(x)$. The next assumption is used only for $m \geq 1$, when integration by parts is needed.

Assumption 1 (a) K is even, m times differentiable and $\max_{0\le j\le m-1}|K^{(j)}(t)||t| = o(1)$ as $|t|\to\infty$; (b) $\max_{0\le j\le m-1}|f^{(j)}(x)| = O(x)$ as $x\to\infty$.

The estimator in Eq. (4) can be constructed using kernels in the class $\{M_k(x)\}_{k\in\mathbb{N}}$ proposed by Mynbaev and Martins-Filho (2010), where

$$M_k(x) = -\frac{1}{C_{2k}^k}\sum_{|l|=1}^{k}\frac{(-1)^l C_{2k}^{l+k}}{|l|}K\left(\frac{x}{l}\right)$$

with $C_{2k}^l = \frac{2k!}{(2k-l)!l!}$ for $l = 0,\ldots,2k$. In this context, K is called the seed of M_k. These kernels are used together with an order $2k$ finite difference

$$\Delta_h^{2k} g_s(x) = \sum_{|l|=0}^{k}(-1)^{l+k}C_{2k}^{l+k}g_s(x-lh)$$

when Besov-type norms are employed to measure smoothness (see Mynbaev and Martins-Filho 2010; Mynbaev et al. 2016). We let $\hat{f}_{s,k}^{(m)}(x)$ denote the estimator defined in (4) with K replaced by M_k, i.e.,

$$\hat{f}_{s,k}^{(m)}(x) = \frac{1}{nh^{m+1}}\sum_{i=1}^{n}\left[M_k^{(m)}\left(\frac{x-X_i}{h}\right) + \sum_{j=1}^{s+1}\frac{k_j}{w_j}M_k^{(m)}\left(\frac{x+X_i/w_j}{h}\right)\right],\quad \text{for } x\ge 0. \tag{5}$$

Note that when K is even, $M_1(x) = K(x)$ and $\hat{f}_{s,1}^{(m)}(x) = \hat{f}_s^{(m)}(x)$.

Theorem 1 *Suppose $f\in C_b^s$ on $D_f = [0,\infty)$ and the kernel K is m times differentiable on $\mathbb{R}$ with $m = 0, 1, \ldots, s$. In case $m\ge 1$ suppose that Assumption 1 holds. Then,*

(1) *The bias of $\hat{f}_s^{(m)}(x)$ has the representation*

$$E\hat{f}_s^{(m)}(x) - f^{(m)}(x) = \int_{\mathbb{R}} K(t)\left[g_s^{(m)}(x-ht) - g_s^{(m)}(x)\right]\mathrm{d}t,\ x\in D_f. \tag{6}$$

(2) *If in Eq. (4) a kernel M_k with seed K is used, then*

$$E\hat{f}_{s,k}^{(m)}(x) - f^{(m)}(x) = \frac{(-1)^{k+1}}{C_{2k}^k}\int_{\mathbb{R}} K(t)\Delta_{ht}^{2k}g_s^{(m)}(x)\mathrm{d}t,\ x\in D_f. \tag{7}$$

The integral representations for biases obtained in Theorem 1 depend on the extension g_s, not the density f. Consequently, existing results for smooth functions (not densities) on $\mathbb{R}$ allow us to easily obtain bias estimates. If classical smoothness characteristics in terms of derivatives and Taylor expansions are used, then part (1) of Theorem 1 is relevant. This approach can be used for derivatives of orders $m\le s-1$ when the bias order is $O(h^{s-m})$ and guaranteed to tend to zero as $h\to 0$. If, on the other hand, smoothness is characterized in terms of finite differences and Besov spaces, then the second representation should be applied. It is appropriate for $m = s - 1$ or $m = s$ when the derivative of order s may have a residual fractional smoothness of order $0 < r < 1$.

For $1 \le p, q \le \infty$ and Ω an open subset of $\mathbb{R}$ put $\Delta_{h,\Omega}^{2k} f(x) = \Delta_h^{2k} f(x)$ if $[x - kh, x + kh] \subset \Omega$ and $\Delta_{h,\Omega}^{2k} f(x) = 0$ otherwise and let

$$\|f\|_{b^r_{p,q}(\Omega)} = \left\{ \int_{\mathbb{R}} \left[\frac{\left(\int_\Omega \left| \Delta_{h,\Omega}^{2k} f(x) \right|^p dx \right)^{1/p}}{|h|^r} \right]^q \frac{dh}{|h|} \right\}^{1/q}$$

where k is any integer satisfying $2k > r$, and in case $p = \infty$ and/or $q = \infty$ the integral(s) is (are) replaced by sup. Further, $\|f\|_{B^r_{p,q}(\Omega)} = \|f\|_{b^r_{p,q}(\Omega)} + \|f\|_{L_p(\Omega)}$. The Hestenes' extension is known to be bounded from $B^r_{p,q}(\Omega)$ to $B^r_{p,q}(\mathbb{R})$.

Assumption 2 For $0 \le m \le s$, $\left\| f^{(m)} \right\|_{B^r_{\infty,q}(0,\infty)} < \infty$ with some $r > 0$ and $1 \le q \le \infty$ and

$$\left(\int |K(t)|^{q'} |t|^{(r+1/q)q'} \mathrm{d}t \right)^{q'} < \infty$$

where $1/q + 1/q' = 1$.

Note that when $q = 1$, q' becomes infinity and the norm becomes sup norm.

Theorem 2 (1) *Let Assumption 1 hold when $m \ge 1$ and assume that $\int_{\mathbb{R}} K(t) t^j \mathrm{d}t = 0$, for $j = 1, \ldots, s - m - 1$, $\int_{\mathbb{R}} \left| K(t) t^{s-m} \right| \mathrm{d}t < \infty$ and $\left| f^{(s)}(x) \right| < C$ for all $x \ge 0$, then*

$$E \hat{f}_s^{(m)}(x) - f^{(m)}(x) = O(h^{s-m}) \text{ for all } x \in D_f. \tag{8}$$

(2) *Let f and K satisfy Assumption 2, then*

$$E \hat{f}_{s,k}^{(m)}(x) - f^{(m)}(x) = O(h^r) \text{ for all } x \in D_f. \tag{9}$$

Remark 1 In the density estimation literature, it is usually assumed that $f \in \mathcal{C}^2$. In this case, under the conditions in Theorem 2, when $m = 0$ we have $E \hat{f}_2(x) - f(x) = \frac{h^2}{2} f^{(2)}(x) \int_{\mathbb{R}} t^2 K(t) \mathrm{d}t + o(h^2)$ for all $x \in D_f$. This expression, similar to what is obtained for the classical Rosenblatt–Parzen estimator when $D_f = \mathbb{R}$, contrasts with what is obtained for Schuster's reflection estimator. In particular, for boundary points, viz., $x = ch$ for $0 \le c < \infty$, we have (see Marron and Ruppert 1994)

$$\begin{aligned} E \hat{f}_S(x) - f(x) &= \int K(t)(g_0(x - ht) - g_0(x)) \mathrm{d}t \\ &= \int_{-\infty}^{x/h} K(t) f(x - ht) \mathrm{d}t + \int_{x/h}^{\infty} K(t) f(-(x - ht)) \mathrm{d}t \\ &= \frac{h^2}{2} f^{(2)}(x) \int_{\mathbb{R}} t^2 K(t) \mathrm{d}t - 2h f^{(1)}(x) \left(c\mu_{0,-c} + \mu_{1,-c} \right) \\ &\quad + 2h^2 f^{(2)}(x)(c^2 \mu_{0,-c} + c\mu_{1,-c}) + o(h^2) \end{aligned}$$

where $\mu_{\ell,-c} = \int_{-\infty}^{-c} u^\ell K(u)\mathrm{d}u$ for $\ell = 0, 1, 2$. The additional bias terms result from the fact that Schuster's estimator does not use the additional smoothness (beyond continuity) of f in its construction.

Remark 2 Theorems 1 and 2, require the existence of densities and their derivatives up to order s at $x = 0$. As such, in the case of densities that either diverge to infinity, or have derivatives that diverge to infinity near the boundary, the bias order we have derived does not hold. In particular, when $f(x) \to \infty$ as $x \to 0$ (a pole at $x = 0$) the bias diverges to infinity at $x = 0$. If $f(0)$ is finite, but $f^{(1)}(x) \to \infty$ as $x \to 0$, the bias decays at a speed that is slower than h^s.

The following theorem provides the order of the variance of our proposed estimator. In this case, it will be necessary to consider the boundary $x = 0$ and interior points $(x > 0)$ separately. Let $k_0 = w_0 = -1$. The quantities

$$\Gamma = \int_{\mathbb{R}} F^2(t)\mathrm{d}t \text{ and } \Gamma_l = \int_0^\infty \left[\sum_{j=0}^{s+1} \frac{k_j}{w_j} F\left(\frac{t}{w_j}\right)\right]^2 \mathrm{d}t,$$

will appear in variance expressions inside the domain and at the left boundary, respectively.

Theorem 3 *Suppose the conditions in Theorem 2 hold,* $\sup_{x\geq 0} f(x) < \infty$, $\sup_x \left|K^{(m)}(x)\right| < \infty$ *and* $\int_{\mathbb{R}} \left|K^{(m)}(t)\right| \mathrm{d}t < \infty$. *In particular, when* $m \geq 1$ *let Assumption 1 hold. Denote* $F(t) = M_k^{(m)}(t)$. *Then, (I) for fixed* $x > 0$

$$V\left(\hat{f}_{s,k}^{(m)}(x)\right) = \frac{1}{nh^{2m+1}} \left\{f(x)\Gamma + o(1)\right\}, \tag{10}$$

(II) at the left boundary

$$V\left(\hat{f}_{s,k}^{(m)}(0)\right) = \frac{1}{nh^{2m+1}} \left\{f(0)\Gamma_l + o(1)\right\}. \tag{11}$$

The estimator $\hat{f}_s^{(m)}(x)$ *has a similar property with* $F(t) = K^{(m)}(t)$.

The proof is omitted because it is similar to, and simpler than, that of Theorem 5, which is given in full.

Remark 3 It is a direct consequence of Theorem 3 that, if $\int |K^{(m)}(u)|^{2+\delta}\mathrm{d}u < C$ for $m \leq s$ for some $\delta > 0$, by Lyapunov's central limit theorem we have for $x > 0$

$$\sqrt{nh^{2m+1}}\left(\hat{f}_s^{(m)}(x) - E\left(\hat{f}_s^{(m)}(x)\right)\right) \xrightarrow{d} N\left(0, f(x)\Gamma\right), \tag{12}$$

and for $x = 0$,

$$\sqrt{nh^{2m+1}}\left(\hat{f}_s^{(m)}(0) - E\left(\hat{f}_s^{(m)}(0)\right)\right) \xrightarrow{d} N\left(0, f(0)\Gamma_l\right). \tag{13}$$

Remark 4 For $m = 0$ and $f \in C^2$, Eq. (13) together with the expression for $E\hat{f}_2(x)$ in Remark 1 gives

$$\sqrt{nh}\left(\hat{f}_2(0) - \left(f(0) + \frac{h^2}{2} f^{(2)}(0+) \int_{\mathbb{R}} t^2 K(t)\mathrm{d}t + o(h^2)\right)\right) \xrightarrow{d} N\left(0,\ f(0)\Gamma_l\right),$$

which can be used to construct confidence intervals and conduct hypothesis testing as in McCrary (2008).

3 Estimation on a bounded interval

Let f be defined on $D_f = [0, 1]$, and let the vectors w, k be as before. We would like to extend f to the left of zero using (2). To obtain a common domain for the components of ϕ, we put $a = \min_i(1/w_i)$ and let

$$\phi_{1,s}(x) = \sum_{j=1}^{s+1} k_j f(-w_j x), \quad -a < x < 0.$$

The sewing conditions at 0 are satisfied as before. Put

$$\phi_{2,s}(x) = \sum_{j=1}^{s+1} k_j f(1 - w_j(x-1)), \quad 1 < x < 1 + a,$$

and define the extension by

$$g_s(x) = \begin{cases} \phi_{1,s}(x), & -a < x < 0 \\ f(x), & 0 \le x \le 1 \\ \phi_{2,s}(x), & 1 < x < 1 + a. \end{cases} \tag{14}$$

The sewing condition holds at $x = 1$:

$$\phi_{2,s}^{(j)}(1+) = \sum_{m=1}^{s+1} (-w_m)^j k_m f^{(j)}(1-) = f^{(j)}(1-), \quad j = 0, \ldots, s.$$

Suppose f is s times differentiable, m is an integer, $0 \le m \le s$, the kernel K is m times differentiable, and let $X_1, \ldots, X_n$ be an IID sample from f. The estimator of $f^{(m)}(x)$, $x \in [0, 1]$, is defined by

$$\hat{f}_s^{(m)}(x) = \frac{1}{nh^{m+1}} \left\{ \sum_{i=1}^{n} K^{(m)} \left(\frac{x - X_i}{h} \right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} \left[\sum_{X_i < aw_j} K^{(m)} \left(\frac{x + X_i/w_j}{h} \right) \right. \right.$$
$$\left. \left. + \sum_{X_i > 1 - aw_j} K^{(m)} \left(\frac{x - 1 + (X_i - 1)/w_j}{h} \right) \right] \right\}. \tag{15}$$

If $m = s = 0$, then $\hat{f}_0^{(0)}(x) = \hat{f}_S(x)$ the estimator suggested by Schuster (1985) in his equation (2.5).[2]

Theorem 4 *Let $f \in \mathcal{C}^s$ on $D_f = [0, 1]$ and let K be an m times differentiable kernel with finite support, $0 \le m \le s$. Let $h > 0$ be small (specifically, it should satisfy the condition* $\text{supp}K \subset (-a/h, a/h)$*). Then, the following statements hold:*

(1) *For a classical kernel K, Eq.* (6) *holds for the estimator given in* (15).
(2) *If in* (15) *K is replaced by M_k, then* (7) *holds for the estimator given in* (15).

Remark 5 Instead of requiring K to have compact support, one can define the extension so that it is sufficiently smooth and has compact support. Take a smooth function h such that $h(x) = 1$ on $(-a/2, 1 + a/2)$ and $h(x) = 0$ for x outside $(-a, 1 + a)$. Instead of (14) consider the extension $g_s^*(x) = h(x)g_s(x)$ and change (15) accordingly. Then the statement of Theorem 4 will be true for g_s^* without the assumption that K has compact support. When $m = 0$ and integration by parts is not necessary, the function h does not have to be smooth. g_s can be extended by zero outside $(-a, 1+a)$ or, equivalently, one can take $h(x) = 1$ on $(-a, 1 + a)$ and $h(x) = 0$ for x outside $(-a, 1 + a)$.

Theorem 5 *Under conditions of Theorem 4, the following is true.*

(1) *For a classical kernel K*
   (I) *for $x \in (0, 1)$ we have $V\left(\hat{f}_s^{(m)}(x)\right) = \frac{1}{nh^{2m+1}} \{f(x)\Gamma + o(1)\}$,*
   (II) *at the left boundary $V\left(\hat{f}_s^{(m)}(0)\right) = \frac{1}{nh^{2m+1}} \{f(0)\Gamma_l + o(1)\}$,*
   (III) *at the right boundary $V\left(\hat{f}_s^{(m)}(1)\right) = \frac{1}{nh^{2m+1}} \{f(1)\Gamma_r + o(1)\}$, where $\Gamma_r = \int_{-\infty}^{0} \left[\sum_{j=0}^{s+1} \frac{k_j}{w_j} F\left(\frac{t}{w_j}\right)\right]^2 \mathrm{d}t$.*
(2) *If M_k is used in place of K, then the same asymptotic expressions are true with $F(x) = M_k^{(m)}(x)$.*

4 Estimation of smooth pieces of densities

Ideas developed in the previous sections can be applied to estimation of densities with discontinuities or with discontinuous derivatives. Here we provide two results.

[2] Note that there is a typographical mistake in Schuster's expression. Using his notation, the last kernel in his equation (2.5) should be evaluated at $(x - 2d + X_i)/a$.

Cline and Hart (1991) used Schuster's symmetrization device to improve bias around a discontinuity point.

The first result in this section applies, for example, to the Laplace distribution which is continuous everywhere but has a discontinuous derivative at zero. The usual kernel density estimator on the whole line will inevitably have a large bias at zero. The suggestion is to estimate its smooth restrictions f_+ and f_- on the right half-axis $[0, \infty)$ and left half-axis $(-\infty, 0]$. Also, the first result in this section, together with the asymptotic distributional convergence in Remark 4, allows for the construction of a test for discontinuity as in McCrary (2008).

As a second example, consider a piece-wise constant density on the interval [0, 1]. The restriction of the density on each interval where it is constant is smooth and can be estimated using our approach. Obviously, the jumps of the estimators will estimate the jumps of the density.

f_+ and f_- do not need to have the same degree of smoothness. Suppose that the right part f_+ is s times differentiable and $0 \le m \le s$. The estimator of $f_+^{(m)}(x)$, $x \ge 0$, is defined by

$$\hat{f}_{+,s}^{(m)}(x) = \frac{1}{nh^{m+1}} \sum_{X_i \ge 0} \left[K^{(m)} \left(\frac{x - X_i}{h} \right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K^{(m)} \left(\frac{x + X_i/w_j}{h} \right) \right].$$

Theorem 6 *In Theorem 1 and in definition (3) let $f = f_+$ and $D_f = [0, \infty)$. If the conditions of Theorem 1 are satisfied for f and K, then (1) (6) and (7) are true and (2) for the variance of $\hat{f}_{+,s}^{(m)}(x)$ one has (10) for $x > 0$ and (11) for $x = 0$.*

Remark 6 As in Remark 3, a direct consequence of Theorem 6 is that, if $\int |K^{(m)}(u)|^{2+\delta} du < C$ for $m \le s$ for some $\delta > 0$, by Lyapunov's central limit theorem we have for $x > 0$

$$\sqrt{nh^{2m+1}} \left(\hat{f}_{+,s}^{(m)}(x) - E\left(\hat{f}_{+,s}^{(m)}(x) \right) \right) \xrightarrow{d} N\left(0, f_+(x)\Gamma\right), \tag{16}$$

and for $x = 0$,

$$\sqrt{nh^{2m+1}} \left(\hat{f}_{+,s}^{(m)}(0) - E\left(\hat{f}_{+,s}^{(m)}(0) \right) \right) \xrightarrow{d} N\left(0, f_+(0)\Gamma_l\right). \tag{17}$$

Equivalent results hold for $\hat{f}_{-,s}^{(m)}(x)$.

Remark 7 As a consequence of Remark 6, we have that for $m = 0$ and $f_+ \in \mathcal{C}_b^2$, Eq. (17) together with the expression for $E\hat{f}_{+,2}(x)$ obtained from using Remark 1, gives

$$\sqrt{nh} \left(\hat{f}_{+,2}(0) - \left(f_+(0) + \frac{h^2}{2} f_+^{(2)}(0+) \int_{\mathbb{R}} t^2 K(t) dt + o(h^2) \right) \right)$$
$$\xrightarrow{d} N\left(0, f_+(0)\Gamma_l\right),$$

and an equivalent expression holds for $\hat{f}_{-,2}(0)$. Consequently, if we define $\delta = \lim_{x \downarrow 0} f(x) - \lim_{x \uparrow 0} f(x) = f_+(0) - f_-(0)$ and $\hat{\delta}_{H,s} = \hat{f}_{+,2}(0) - \hat{f}_{-,2}(0)$, following the arguments in McCrary (2008) we have, for $x = 0$,

$$\sqrt{nh}(\hat{\delta}_{H,s} - \delta - B_\delta) \xrightarrow{d} N\left(0, (f_+(0) + f_-(0))\,\Gamma_l\right), \tag{18}$$

where $B_\delta = \frac{h^2}{2} f_+^{(2)}(0+) \int_{\mathbb{R}} t^2 K(t) \mathrm{d}t - \frac{h^2}{2} f_-^{(2)}(0-) \int_{\mathbb{R}} t^2 K(t) \mathrm{d}t + o(h^2)$.

Now suppose that the domain D_f of a density f contains a finite segment $[c, d]$ such that the restriction f_r of f onto $[c, d]$ is smooth. Denote

$$\phi_1(x) = \sum_{j=1}^{s+1} k_j f_r(c - w_j(x - c)), \quad c - a_1 < x < c,$$

the extension of f_r to the left of c and

$$\phi_2(x) = \sum_{j=1}^{s+1} k_j f_r(d - w_j(x - d)), \quad d < x < d + a_1,$$

the extension of f_r to the right of d. Here we choose $a_1 = a(d - c)$, to make sure that $c - w_j(x - c)$ and $d - w_j(x - d)$ belong to $[c, d]$. The extended restriction then is defined by

$$g_r(x) = \begin{cases} \phi_1(x), & c - a_1 < x < c, \\ f_r(x), & c \le x \le d, \\ \phi_2(x), & d < x < d + a_1. \end{cases} \tag{19}$$

Definition (15) guides us to define

$$\begin{aligned} \hat{f}_s^{(m)}(x) = \frac{1}{nh^{m+1}} \Bigg\{ &\sum_{c \le X_i \le d} K^{(m)}\left(\frac{x - X_i}{h}\right) \\ &+ \sum_{j=1}^{s+1} \frac{k_j}{w_j} \Bigg[\sum_{c < X_i < c + a_1 w_j} K^{(m)}\left(\frac{x - c + (X_i - c)/w_j}{h}\right) \\ &+ \sum_{d - a_1 w_j < X_i < d} K^{(m)}\left(\frac{x - d + (X_i - d)/w_j}{h}\right)\Bigg]\Bigg\}, \ x \in (c, d). \end{aligned}$$

Theorem 7 *Let f_r be s times differentiable, $0 \le m \le s$, and let K have compact support. For h sufficiently small (such that* $\operatorname{supp} K \subseteq (-a_1/h, a_1/h)$*), we have*

$$E \hat{f}_s^{(m)}(x) - f_r^{(m)}(x) = \int_{\mathbb{R}} K(u) \left[g_r^{(m)}(x - hu) - g_r^{(m)}(x)\right] \mathrm{d}u, \ c < x < d. \tag{20}$$

Further, for $x \in [c, d]$ we have $V\left(\hat{f}_s^{(m)}(x)\right) = \frac{1}{nh^{2m+1}} \{f(x)\Gamma + o(1)\}$, at the left boundary $V\left(\hat{f}_s^{(m)}(c)\right) = \frac{1}{nh^{2m+1}} \{f(c)\Gamma_l + o(1)\}$, and at the right boundary $V\left(\hat{f}_s^{(m)}(d)\right) = \frac{1}{nh^{2m+1}} \{f(d)\Gamma_r + o(1)\}$ where $F(t) = K^{(m)}(t)$ or $F(t) = M_k^{(m)}(t)$ depending on which kernel is used in the definition of $\hat{f}_s^{(m)}(x)$.

5 Estimators satisfying zero boundary conditions

For simplicity, we consider only densities on $D_f = [0, \infty)$. For estimator (4), we provide two modifications designed to satisfy zero boundary conditions for the estimator itself and/or its derivatives. In both cases, the bias rate is retained and the variance at zero becomes zero. The main difference between the estimators is in the number of derivatives that are guaranteed to vanish. Everywhere it is assumed that f is s times differentiable, $0 \le m \le s$ and the purpose is to estimate $f^{(m)}(x)$.

In the first result, we start with any estimator of the derivative with property (8). We assume that some consecutive derivatives of f, starting with $f^{(m)}(0+)$, are zero, and we want an estimator of $f^{(m)}$ which has at least as many derivatives vanishing at zero. Let l be an integer between m and s, and let ψ be a function on D_f with properties

$$\psi(0+) = \cdots = \psi^{(l-m)}(0+) = 0, \ \psi^{(l-m+1)}(0+) \ne 0, \tag{21}$$

$\psi(x) = 1$ for $x \ge 1$, $0 \le \psi(x) \le 1$ everywhere. If

$$f^{(m)}(0+) = \cdots = f^{(s)}(0+) = 0, \tag{22}$$

put $\alpha = 1$. Otherwise, let k be such that $f^{(m)}(0+) = \cdots = f^{(k-1)}(0+) = 0$, $f^{(k)}(0+) \ne 0$ and $m < k \le s$, and put $\alpha = \frac{s-m}{k-m}$. For any estimator $\hat{f}^{(m)}(x)$ of $f^{(m)}(x)$ define another estimator $\tilde{f}^{(m)}(x) = \psi(xh^{-\alpha})\hat{f}^{(m)}(x)$.

Theorem 8 *Let the estimator $\hat{f}^{(m)}(x)$ of $f^{(m)}(x)$ satisfy (8). Then $\tilde{f}^{(m)}(x)$ satisfies*

$$\tilde{f}^{(m)}(0+) = \frac{d}{dx}\tilde{f}^{(m)}(0+) = \cdots = \frac{d^{l-m}}{dx^{l-m}}\tilde{f}^{(m)}(0+) = 0, \tag{23}$$

$$E\tilde{f}^{(m)}(x) - f^{(m)}(x) = O(h^{s-m}) \text{ for all } x \in D_f, \tag{24}$$

and

$$V\left(\tilde{f}^{(m)}(x)\right) = \begin{cases} V\left(\hat{f}^{(m)}(x)\right), & x \ge h^{\alpha} \\ V\left(\hat{f}^{(m)}(x)\right)\psi^{(l-m+1)}(0+)(xh^{-\alpha})^{2(l-m+1)}, & x < h^{\alpha}. \end{cases} \tag{25}$$

In the second result, we modify estimator (4) so as to satisfy zero boundary conditions. Let ψ be a function with properties: ψ is m times differentiable on D_f, $\psi(x) = 1$ for $x \in (0, 2]$, $\psi(x) = 0$ for $x \ge 3$, $0 \le \psi(x) \le 1$ everywhere. Define for $x \ge 0$

$$\hat{f}^{(m)}(x) = \frac{1}{nh^{m+1}}\sum_{i=1}^{n}\psi(X_i/x)\left[K^{(m)}\left(\frac{x - X_i}{h}\right) + \sum_{j=1}^{s+1}\frac{k_j}{w_j}K^{(m)}\left(\frac{x + X_i/w_j}{h}\right)\right].$$

In this definition, we take $0/0 = 0$ and for $x > 0$, $x/0 = \infty$.

Theorem 9 *All derivatives of $\hat{f}^{(m)}(x)$ that exist vanish at zero and*

$$E\hat{f}^{(m)}(x) - f^{(m)}(x) = \int_{\mathbb{R}} K(t)\left[g_x^{(m)}(x-ht) - g_x^{(m)}(x)\right]\mathrm{d}t,\ x \in D_f, \qquad (26)$$

where g_x is the Hestenes' extension of $f_x(t) = f(t)\psi(t/x)$. Besides, for $x > 0$ $V\left(\hat{f}^{(m)}(x)\right)$ satisfies (10) (for $x = 0$ variance is zero).

6 Simulations

We conducted a series of simulations to provide some evidence of the finite sample performances of our estimators and to contrast them with that of some of the most commonly used estimators for densities with supports that are subsets of $\mathbb{R}$. We focus on two broad cases: first, we consider densities that are defined on $[0, \infty)$; second, we consider the case of a density with a discontinuity at $x = 0$. In the second case, we are particularly interested in the size of the jump at the point of discontinuity.

In the first case, we consider random variables with the following densities:

1. Normal density left-truncated at $x = 0$: $f_{TN}(x) = \frac{2}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$,
2. Gamma density: $f_G(x) = \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}\exp(-\frac{1}{\beta}x)$ with $\alpha = 2,\ \beta = 1$,
3. Chi-squared density: $f_\chi(x) = \frac{1}{2^{v/2}\Gamma(v/2)}x^{v/2-1}\exp(-\frac{1}{2}x)$ with $v = 5$,
4. Exponential density: $f_E(x) = \lambda\exp(-\lambda x)$ with $\lambda = 1$.

For each density, we generated samples of size $n = 250$, 500 and calculated the following estimators: $\hat{f}_R$, $\hat{f}_S$ and $\hat{f}_{s,k}$ for $k = 1, 2, 3$, $w_i = i, i^{-1}$ and $s = 1, 2$.[3] In each case, we used a Gaussian kernel, or a Gaussian seed kernel, as necessary. We also calculated the gamma kernel estimator of Chen (2000), which we denote by $\hat{f}_C$, and the generalized jackknife estimator proposed by Jones (1993), which we denote by $\hat{f}_J$.[4] For each estimator, we selected an optimal bandwidth by minimizing their integrated squared error, i.e., for an arbitrary estimator denoted by $\hat{f}(x; h)$ and an arbitrary density denoted by f, we choose

$$h_0 = \underset{h}{\operatorname{argmin}} \int_0^\infty (\hat{f}(u; h) - f(u))^2 \mathrm{d}u.$$

We then calculate the value of each estimator over a fixed grid on the interval $(0, 4)$ with step 10^{-1}. For each sample, and each estimator, an average root-squared error across the grid is calculated (RASE). The average of these RASE across all 1000 generated samples are reported in Table 1. Figure 1 gives a set of estimates for one of the generated samples of size $n = 250$ associated with the exponential density.

3 Results for $\hat{f}_{s,k}$ when $w_i = i^{-1}$ are not shown, as the performance of these estimators is generally dominated by the case where $w_i = i$. The full set of results, including experiments where $n = 1000$, is available from the authors upon request.

4 Specifically, we consider the estimator constructed using the kernel K_L defined on his equation (3.4).

Table 1 Average RASE × 10^2 $\hat{f}_R$ and $\hat{f}$ constructed using $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, $w_i = i$

	$\hat{f}_R$	$\hat{f}_C$	$\hat{f}_S$	$\hat{f}_J$	$\hat{f}_{s,1}$		$\hat{f}_{s,2}$		$\hat{f}_{s,3}$	
					$s=1$	$s=2$	$s=1$	$s=2$	$s=1$	$s=2$
$n = 250$										
f_{TN}	7.7974	3.6718	2.3328	2.7958	2.6234	2.9859	2.6436	3.2551	2.7526	3.3983
f_G	2.9336	2.7331	3.7670	2.7409	3.0848	2.9242	3.0915	3.0707	3.1172	3.1062
f_χ	1.3586	1.3469	1.5957	1.1949	1.2369	1.2979	1.1873	1.3300	1.1958	1.3256
f_E	9.7433	3.8200	3.8134	2.9546	2.9241	3.1386	2.9346	2.9570	2.9432	2.8879
$n = 500$										
f_{TN}	7.3139	2.9312	1.8576	2.1673	2.0838	2.2882	2.0629	2.5569	2.1377	2.6658
f_G	2.3973	2.1511	3.0778	2.1556	2.4199	2.2914	2.4140	2.3933	2.4345	2.4090
f_χ	1.0756	1.0382	1.2487	0.9261	0.9501	1.0206	0.9109	1.0306	0.9212	1.0226
f_E	9.1800	3.0807	3.1162	2.3264	2.2932	2.4399	2.2840	2.2589	2.2841	2.1926

As expected, the average RASE, for each estimator, and across all densities, decreases as n increases from 250 to 500. Except for data generated from the f_χ density, all estimators that are based on Hestenes' extension and constructed using the M_k kernels (including the case where $k = 1$ and $M_1 = K$) have smaller average RASE when $w_i = i$. Also, for estimators $\hat{f}_{s,2}$ and $\hat{f}_{s,3}$, choosing $s = 1$ reduces average RASE (compared to $s = 2$) for all densities, except f_G and f_E when $n = 500$.

Except for the case of the truncated normal density—f_{TN}—all Hestenes-based estimators outperform the estimator $\hat{f}_S$ proposed by Schuster (1985). The good performance of $\hat{f}_S$ in the case of f_{TN} is expected since, in this case, $f_{TN}^{(1)}(0) = 0$. It is also the case that $f_\chi^{(1)}(0) = 0$, but for this density all Hestenes-based estimators have smaller average RASE than $\hat{f}_S$. Also, except for the case of the gamma density—f_G—all Hestenes-based estimators outperforms the estimator $\hat{f}_C$ proposed by Chen (2000). A similar conclusion can also be reached regarding the relative performance of Hestenes-based estimators and the estimator $\hat{f}_J$ proposed by Jones (1993). Lastly, as expected, the traditional Rosenblatt–Parzen estimator has the poorest performance across all densities and all estimators, except for the case of f_χ, where $f_\chi(0) = 0$. The choice of kernel, or seed kernel, does not qualitatively impact the relative performance described above.

Although a complete theoretical treatment of the optimal choice of w_i, s and k for finite n is beyond the scope of this paper, the preliminary experimental evidence seems to support the use of $w_i = i$ and the choice of $s = 1$ relative to $s = 2$. Also, our results here confirm the simulation results in Mynbaev and Martins-Filho (2010) suggesting $k < 3$. Results for $s, k \geq 3$ (not reported here) suggest rapid deterioration of the performance of $\hat{f}_{s,k}$ as measured by average RASE. In summary, the simulation results suggest that Hestenes-based estimators can outperform the well-known estimators proposed by Schuster (1985), Jones (1993) and Chen (2000). In the few cases where this does not hold, additional information about the true density is needed to avail oneself of other estimators, while our estimators are universally applicable.

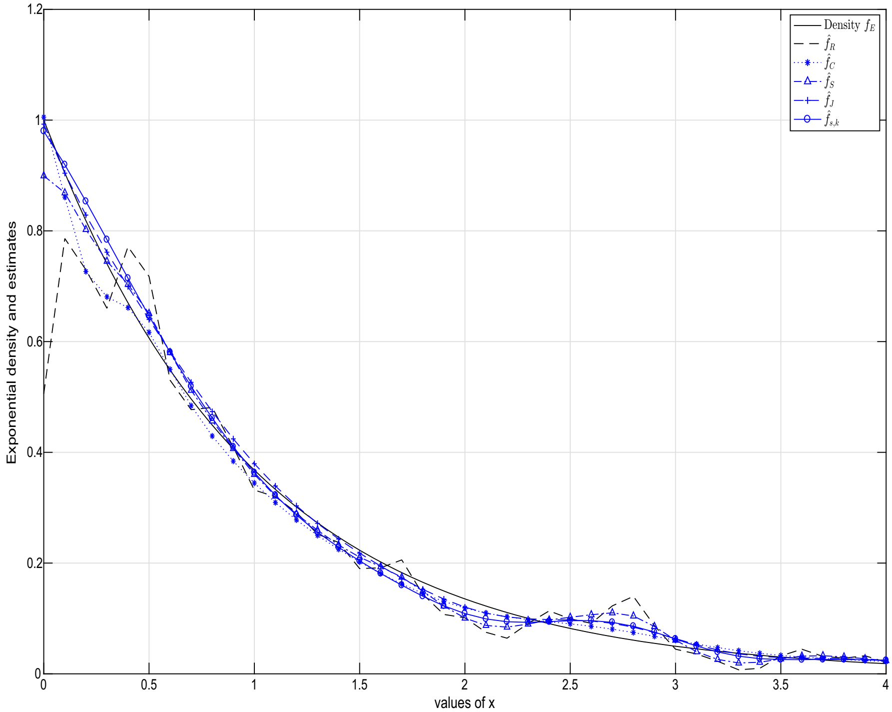

Fig. 1 Plots of the exponential density f_E with $\lambda = 1$ and the estimators $\hat{f}_R$, $\hat{f}_C$, $\hat{f}_S$, $\hat{f}_J$ and $\hat{f}_{s,k}$ for $s = 1, k = 2, w_i = i$. The bandwidths for each estimator are obtained by minimizing the integrated squared error. Excepting $\hat{f}_R$, the estimators differ mostly in the vicinity of zero

In the second broad case, we consider two densities that have a discontinuity at $x = 0$. The first is given by

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) & \text{if } x < 0 \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) & \text{if } x \geq 0, \end{cases} \tag{27}$$

where σ^2 controls the size of the jump. If $\sigma^2 = 1$ the density is continuous everywhere, and for $0 < \sigma^2 < 1$ the jump at $x = 0$ is given by $J_f(0) = f(0-) - f(0+) = \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\sigma} - 1\right) > 0$. The second is given by,

$$f(x) = \begin{cases} \frac{1}{2\sqrt{2\pi}\Phi(-\mu_1)} \exp\left(-\frac{1}{2}(x-\mu_1)^2\right) & \text{if } x < 0 \\ \frac{1}{2\sqrt{2\pi}(1-\Phi(-1))} \exp\left(-\frac{1}{2}(x-1)^2\right) & \text{if } x \geq 0, \end{cases} \tag{28}$$

where $\Phi(x)$ is the distribution function associated with a standard Gaussian density and μ_1 controls the size of the jump. If $\mu_1 = -1$ the density is continuous everywhere, and for $\mu_1 > -1$ the jump at $x = 0$ is given by

$$J_f(0) = f(0-) - f(0+)$$
$$= \frac{1}{2\sqrt{2\pi}} \left(\frac{1}{\Phi(-\mu_1)} \exp\left(-\frac{1}{2}\mu_1^2\right) - \frac{1}{1-\Phi(-1)} \exp(-1/2) \right) > 0. \quad (29)$$

The top two panels of Fig. 2 provide graphs of the first density for $\sigma^2 = 0.5$ and $\sigma^2 = 0.25$. The bottom two panels provide graphs of the second density for $\mu_1 = 0$ and $\mu_1 = 1.5$.

With knowledge of the point of discontinuity ($x = 0$), we assess the performance of two estimators for the size of the jump at $x = 0$. The first is based on a local linear density estimator proposed by Cheng (1994). As in McCrary (2008), for each sample $\{X_i\}_{i=1}^n$ we first compute a histogram with bin size $b = 2\hat{\sigma} n^{-1/2}$ where $\hat{\sigma} = \sqrt{n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2}$ and $\bar{X} = n^{-1}\sum_{i=1}^n X_i$. Using $2J$ bins with centers $G_j \in \{\cdots, -\frac{3}{2}b, -\frac{1}{2}b, \frac{1}{2}b, \frac{3}{2}b, \ldots\}$ and $J = \lfloor \frac{\max_i\{X_i\} - \min_i\{X_i\}}{b} \rfloor + 2$, we obtain the number of observations C_j that fall in each bin and define the standardized frequency counts $Y_j = \frac{1}{nb} C_j$. $\hat{f}_+^{LL}(0)$ is the local linear estimator obtained from regressing Y_j on $G_j > 0$ evaluated at $x = 0$ for the part of f defined on $[0, \infty)$. Similarly, $\hat{f}_-^{LL}(0)$ is the local linear estimator obtained from regressing Y_j on $G_j < 0$ for the part of f defined on $(-\infty, 0]$.

We consider two parameters that capture the jump discontinuity at $x = 0$: 1) $\delta = f_+(0) - f_-(0)$, where $f_-(0) = \lim_{x\uparrow 0} f(x)$ and $f_+(0) = \lim_{x\downarrow 0} f(x)$; 2) following McCrary (2008), $\theta = \log f_+(0) - \log f_-(0)$. Hence, we define the local linear estimators $\hat{\delta}_{LL} = \hat{f}_+^{LL}(0) - \hat{f}_-^{LL}(0)$ and $\hat{\theta}_{LL} = \log \hat{f}_+^{LL}(0) - \log \hat{f}_-^{LL}(0)$.

The second estimator we consider is our Hestenes-based estimator from Sect. 4. We define

$$\hat{f}_{+,s}(0) = \frac{1}{nh} \sum_{X_i \geq 0} \left[K\left(\frac{-X_i}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K\left(\frac{X_i/w_j}{h}\right) \right] \text{ and}$$
$$\hat{f}_{-,s}(0) = \frac{1}{nh} \sum_{X_i \leq 0} \left[K\left(\frac{-X_i}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K\left(\frac{X_i/w_j}{h}\right) \right]$$

for the part of f defined on $[0, \infty)$ and $(-\infty, 0]$, respectively. As such, we propose the estimators $\hat{\delta}_{H,s} = \hat{f}_{+,s}(0) - \hat{f}_{-,s}(0)$ and $\hat{\theta}_{H,s} = \log \hat{f}_{+,s}(0) - \log \hat{f}_{-,s}(0)$.

Tables 2 and 3 provide biases and variances for local linear and Hestenes-based estimators for the left and right limits of the densities given in Eqs. (27) and (28) at $x = 0$, as well as the root-mean-squared error (M) of δ_{LL}, $\delta_{H,s}$, θ_{LL} and $\theta_{H,s}$ at $x = 0$ for $n = 500, 1000$ based on 2000 samples. Given our Theorems 1 and 3, and Theorem 3 in Cheng (1994), we calculate each density estimator using optimal plug-in bandwidths that minimize the asymptotic mean integrated squared error at $x = 0$, i.e.,

$$h = n_*^{-1/5} \left(\frac{\int K^2(x)dx}{\mu_2^2 \int f^{(2)}(x)^2 dx} \right)^{1/5}, \quad \text{where } \mu_2 = \int x^2 K(x)dx.$$

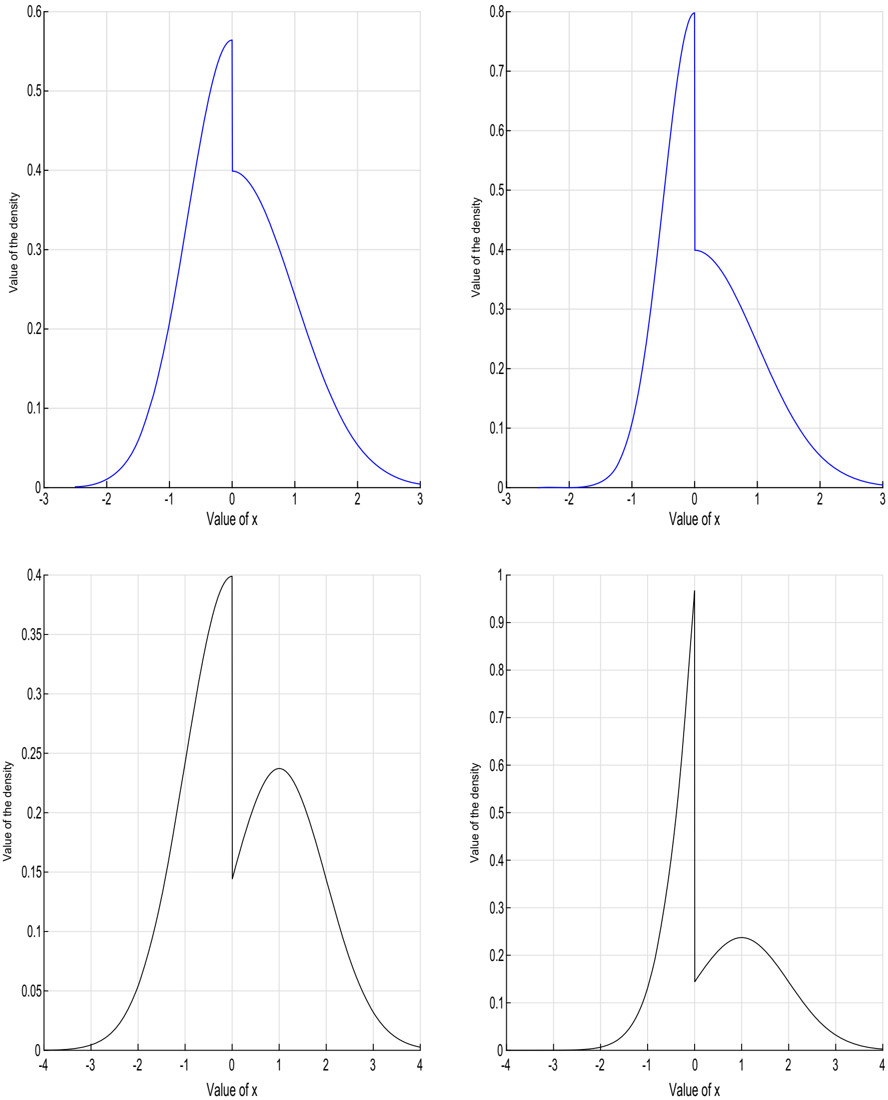

Fig. 2 Densities $f(x)$ with a discontinuity at $x = 0$. Top panels are for $f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) & \text{if } x < 0 \\ \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right) & \text{if } x \geq 0 \end{cases}$, with $\sigma^2 = 0.5$ on top left panel and $\sigma^2 = 0.25$ on top right panel. Bottom panels are for $f(x) = \begin{cases} \frac{1}{2\sqrt{2\pi}\,\Phi(-\mu_1)}\exp\left(-\frac{1}{2}(x-\mu_1)^2\right) & \text{if } x < 0 \\ \frac{1}{2\sqrt{2\pi}(1-\Phi(-1))}\exp\left(-\frac{1}{2}(x-1)^2\right) & \text{if } x \geq 0 \end{cases}$, with $\mu_1 = 0$ on bottom left panel and $\mu_1 = 1.5$ on bottom right panel

Note that in this case, n_* is either the number of observations on the positive or negative sides of $\mathbb{R}$, and $\int f^{(2)}(x)^2 dx$ is calculated using the corresponding expression

Table 2 Bias (B), Variance (V) and Root-Mean-Squared Error (M) for Local Linear and Hestenes Estimator

$s = 1$	$B(\hat{f}_{-}^{LL})$	$B(\hat{f}_{-,s})$	$B(\hat{f}_{+}^{LL})$	$B(\hat{f}_{+,s})$	$V(\hat{f}_{-}^{LL})$	$V(\hat{f}_{+}^{LL})$	$V(\hat{f}_{-,s})$	$V(\hat{f}_{+,s})$	$M(\hat{\theta}_{LL})$	$M(\hat{\theta}_{H,s})$	$M(\hat{\delta}_{LL})$	$M(\hat{\delta}_{H,s})$
$n = 500$												
GK												
$\sigma^2 = 0.25$	0.0652	0.0466	0.0181	0.0306	0.0059	0.0030	0.0031	0.0017	0.1734	0.1296	0.1101	0.0782
$\sigma^2 = 0.5$	0.0355	0.0455	0.0164	0.0296	0.0044	0.0033	0.0023	0.0018	0.1876	0.1401	0.0929	0.0723
TK												
$\sigma^2 = 0.25$	0.0651	0.0506	0.0154	0.0328	0.0070	0.0035	0.0030	0.0017	0.1923	0.1334	0.1202	0.0809
$\sigma^2 = 0.5$	0.0303	0.0487	0.0177	0.0344	0.0049	0.0035	0.0022	0.0017	0.1974	0.1375	0.0975	0.0716
$s = 2$												
GK												
$\sigma^2 = 0.25$	0.0627	0.1490	0.0209	0.0377	0.0061	0.0032	0.0055	0.0030	0.1785	0.1857	0.1117	0.1509
$\sigma^2 = 0.5$	0.0381	0.0845	0.0195	0.0373	0.0045	0.0030	0.0043	0.0029	0.1860	0.1809	0.0943	0.1027
TK												
$\sigma^2 = 0.25$	0.0637	0.1718	0.0159	0.0383	0.0068	0.0036	0.0050	0.0031	0.1932	0.1931	0.1192	0.1654
$\sigma^2 = 0.5$	0.0347	0.0954	0.0162	0.0379	0.0054	0.0035	0.0041	0.0030	0.2034	0.1847	0.1013	0.1067

Table 2 continued

$s=1$	$B(\hat{f}_{-}^{LL})$	$B(\hat{f}_{-,s})$	$B(\hat{f}_{+}^{LL})$	$B(\hat{f}_{+,s})$	$V(\hat{f}_{-}^{LL})$	$V(\hat{f}_{+}^{LL})$	$V(\hat{f}_{-,s})$	$V(\hat{f}_{+,s})$	$M(\hat{\theta}_{LL})$	$M(\hat{\theta}_{H,s})$	$M(\hat{\delta}_{LL})$	$M(\hat{\delta}_{H,s})$
$n=1000$												
GK												
$\sigma^2=0.25$	0.0574	0.0571	0.0146	0.0274	0.0037	0.0018	0.0020	0.0010	0.1391	0.1021	0.0900	0.0680
$\sigma^2=0.5$	0.0298	0.0448	0.0150	0.0275	0.0028	0.0018	0.0015	0.0010	0.1432	0.1083	0.0722	0.0575
TK												
$\sigma^2=0.25$	0.0554	0.0649	0.0130	0.0301	0.0043	0.0020	0.0019	0.0010	0.1457	0.1009	0.0935	0.0699
$\sigma^2=0.5$	0.0298	0.0448	0.0150	0.0275	0.0028	0.0018	0.0015	0.0010	0.1432	0.1083	0.0722	0.0575
$s=2$												
GK												
$\sigma^2=0.25$	0.0591	0.1379	0.0157	0.0256	0.0039	0.0018	0.0036	0.0017	0.1400	0.1613	0.0917	0.1369
$\sigma^2=0.5$	0.0292	0.0649	0.0150	0.0246	0.0026	0.0018	0.0025	0.0018	0.1412	0.1443	0.0706	0.0797
TK												
$\sigma^2=0.25$	0.0550	0.1568	0.0123	0.0242	0.0043	0.0021	0.0034	0.0019	0.1490	0.1779	0.0947	0.1536
$\sigma^2=0.5$	0.0268	0.0734	0.0131	0.0259	0.0029	0.0020	0.0024	0.0018	0.1531	0.1481	0.0758	0.0839

$w_i = i,\ f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) & \text{if } x < 0 \\ \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right) & \text{if } x \geq 0 \end{cases}$, GK and GT indicate the use of Gaussian and triangular kernels, respectively

Table 3 Bias, Variance and Root-Mean-Squared Error for Local Linear and Hestenes Estimator

$s = 1$	$B(\hat{f}_-^{LL})$	$B(\hat{f}_{-,s})$	$B(\hat{f}_+^{LL})$	$B(\hat{f}_{+,s})$	$V(\hat{f}_-^{LL})$	$V(\hat{f}_+^{LL})$	$V(\hat{f}_{-,s})$	$V(\hat{f}_{+,s})$	$M(\hat{\theta}_{LL})$	$M(\hat{\theta}_{H,s})$	$M(\hat{\delta}_{LL})$	$M(\hat{\delta}_{H,s})$
$n = 500$												
GK												
$\mu_1 = 0$	0.0207	0.0320	0.0100	0.0321	0.0031	0.0012	0.0017	0.0008	0.0803	0.0535	0.0688	0.0531
$\mu_1 = 1.5$	−0.0108	−0.0317	0.0089	0.0316	0.0195	0.0012	0.0106	0.0008	0.0861	0.0930	0.1468	0.1265
TK												
$\mu_1 = 0$	0.0100	0.0179	0.0008	0.0066	0.0065	0.0024	0.0033	0.0013	0.1929	0.0880	0.0955	0.0695
$\mu_1 = 1.5$	−0.0043	−0.0112	0.0013	0.0075	0.0461	0.0026	0.0193	0.0014	0.2331	0.0985	0.2226	0.1470
$s = 2$												
GK												
$\mu_1 = 0$	0.0185	0.0349	0.0106	−0.0056	0.0032	0.0013	0.0032	0.0014	0.0833	0.1352	0.0699	0.0804
$\mu_1 = 1.5$	−0.0140	0.0103	0.0085	−0.0072	0.0190	0.0013	0.0180	0.0014	0.0919	0.1328	0.1470	0.1434
TK												
$\mu_1 = 0$	0.0066	0.0047	0.0031	−0.0098	0.0065	0.0025	0.0060	0.0026	0.1957	0.2742	0.0943	0.0933
$\mu_1 = 1.5$	0.0013	0.0096	0.0018	−0.0105	0.0440	0.0024	0.0355	0.0025	0.2141	0.3093	0.2160	0.1958

Table 3 continued

$s=1$	$B(\hat{f}_{-}^{LL})$	$B(\hat{f}_{-,s})$	$B(\hat{f}_{+}^{LL})$	$B(\hat{f}_{+,s})$	$V(\hat{f}_{-}^{LL})$	$V(\hat{f}_{+}^{LL})$	$V(\hat{f}_{-,s})$	$V(\hat{f}_{+,s})$	$M(\hat{\theta}_{LL})$	$M(\hat{\theta}_{H,s})$	$M(\hat{\delta}_{LL})$	$M(\hat{\delta}_{H,s})$
$n=1000$												
GK												
$\mu_1=0$	0.0154	0.0282	0.0062	0.0230	0.0018	0.0007	0.0011	0.0005	0.0456	0.0301	0.0519	0.0409
$\mu_1=1.5$	−0.0102	−0.0259	0.0063	0.0231	0.0110	0.0007	0.0060	0.0005	0.0491	0.0549	0.1103	0.0959
TK												
$\mu_1=0$	0.0046	0.0130	0.0024	0.0056	0.0039	0.0013	0.0020	0.0008	0.0943	0.0510	0.0724	0.0542
$\mu_1=1.5$	−0.0057	−0.0110	−0.0007	0.0039	0.0218	0.0013	0.0113	0.0008	0.1014	0.0538	0.1530	0.1125
$s=2$												
GK												
$\mu_1=0$	0.0153	0.0250	0.0068	−0.0109	0.0018	0.0007	0.0018	0.0008	0.0462	0.0892	0.0518	0.0634
$\mu_1=1.5$	−0.0080	0.0123	0.0061	−0.0119	0.0105	0.0007	0.0104	0.0008	0.0488	0.0747	0.1088	0.1105
TK												
$\mu_1=0$	0.0026	−0.0001	−0.0001	−0.0087	0.0036	0.0014	0.0034	0.0014	0.1049	0.1301	0.0720	0.0710
$\mu_1=1.5$	−0.0056	0.0033	−0.0005	−0.0092	0.0219	0.0013	0.0202	0.0013	0.1047	0.1297	0.1548	0.1488

$w_i = i,\ f(x) = \begin{cases} \frac{1}{2\sqrt{2\pi}\Phi(-\mu_1)}\exp\left(-\frac{1}{2}(x-\mu_1)^2\right) & \text{if } x < 0 \\ \frac{1}{2\sqrt{2\pi}(1-\Phi(-1))}\exp\left(-\frac{1}{2}(x-1)^2\right) & \text{if } x \geq 0 \end{cases}$, GK and GT indicate the use of Gaussian and triangular kernels, respectively for $f^{(2)}(x)$ and limits of integration on the positive or negative sides of $\mathbb{R}$. In the case of the Hestenes-based estimators, we focus on the case where $w_i = i$ (results for the case where $w_i = 1/i$ are available upon request).

We start by observing some general regularities for Tables 2 and 3. First, for all density estimators, the bias and variance decrease when the sample size grows from $n = 500$ to $n = 1000$. Also, for both densities given by Eqs. (27) and (28) the local linear estimator has, in general, smaller bias than the Hestenes-based estimators. The latter, however, have smaller variances, with the exception of the case where the Hestenes-based estimators are calculated using $s = 2$ for density given by Eq. (28), where the variances are largely equal. The root-mean-squared errors (M) for $\hat{\theta}_{LL}$ and $\hat{\delta}_{LL}$ are greater than or equal to that for $\hat{\theta}_{H,s}$ for $s = 1$ for both densities and all sample sizes. When $s = 2$, this relationship is essentially reversed for all sample sizes when the density is given by Eq. (28). If the density is given by Eq. (27), then the root-mean-squared errors for $\hat{\delta}_{LL}$ and $\hat{\theta}_{LL}$ are generally smaller than those for $\hat{\delta}_{H,s}$ and $\hat{\theta}_{H,s}$ when $n = 1000$. The size of the jump discontinuity, as regulated by the values of σ^2 and μ_1, has little impact on the root-mean-squared error (M) of both $\hat{\theta}_{LL}$ and $\hat{\theta}_{H,s}$. In this case, samples that resulted in negative estimated densities at $x = 0$ were discarded, so that $\hat{\theta}_{LL}$ and $\hat{\theta}_{H,s}$ can be defined. For $\hat{\delta}_{LL}$ and $\hat{\delta}_{H,s}$, where the samples that produce negative estimated densities are not discarded, the size of the jump discontinuity, as regulated by μ_1, increases the root- mean-squared error of both $\hat{\delta}_{LL}$ and $\hat{\delta}_{H,s}$. This does not occur for the density given by Eq. (27), where no sample in our experiments produced negative density estimates. Overall, our simulations are largely inconclusive regarding the finite sample relative performance of the local linear (LL) and Hestenes-based estimators for the size of a jump discontinuity. The latter normally exhibiting smaller variances, while the former normally carrying smaller biases.

7 Summary and conclusions

We provided a set of easily implementable kernel estimators for densities defined on subsets of $\mathbb{R}$ that have boundaries. The use of Hestenes' extensions allows us to obtain theoretical representations for bias and variance of our proposed estimators that preserve the orders of traditional kernel estimators for densities defined on $\mathbb{R}$. In effect, the insights gained from using Hestenes' extensions make the study of suitably defined kernel estimators in sets that have boundaries a special case of the theory developed for densities defined on $\mathbb{R}$. Preliminary simulations reveal very good finite sample performance relative to a number of commonly used alternative estimators. Further work should investigate the possible existence of optimal choices for s and $w_1, \ldots, w_{s+1}$ under a suitably defined criterion. If possible, this would produce a *best* estimator in the class we have defined.

Appendix: Proofs

Proof of Theorem 1 (1) By the IID assumption

$$
\begin{aligned}
E\hat{f}_s^{(m)}(x) &= \frac{1}{h^{m+1}} E\left[K^{(m)}\left(\frac{x-X_1}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K^{(m)}\left(\frac{x+X_1/w_j}{h}\right)\right] \\
&= \frac{1}{h^{m+1}}\left[\int_0^\infty K^{(m)}\left(\frac{x-t}{h}\right) f(t)\mathrm{d}t\right. \\
&\quad \left. + \sum_{j=1}^{s+1} \frac{k_j}{w_j} \int_0^\infty K^{(m)}\left(\frac{x+t/w_j}{h}\right) f(t)\mathrm{d}t\right]. \qquad (30)
\end{aligned}
$$

In the first integral let $u = \frac{x-t}{h}$, in the others $u = \frac{x+t/w_j}{h}$. Then

$$
\begin{aligned}
E\hat{f}_s^{(m)}(x) &= \frac{1}{h^m}\left[-\int_{x/h}^{-\infty} K^{(m)}(u) f(x-hu)\mathrm{d}u\right. \\
&\quad \left. + \sum_{j=1}^{s+1} k_j \int_{x/h}^\infty K^{(m)}(u) f\left(-w_j(x-hu)\right)\mathrm{d}u\right] \\
&= \frac{1}{h^m}\left[\int_{-\infty}^{x/h} K^{(m)}(u) f(x-hu)\mathrm{d}u\right. \\
&\quad \left. + \int_{x/h}^\infty K^{(m)}(u) \sum_{j=1}^{s+1} k_j f\left(-w_j(x-hu)\right)\mathrm{d}u\right].
\end{aligned}
$$

In the first integral we have $x - hu > 0$ and $f(x - hu) = g_s(x - hu)$; in the second one $x - hu < 0$, so $\sum_{j=1}^{s+1} k_j f\left(-w_j(x-hu)\right) = g_s(x - hu)$. Hence,

$$
E\hat{f}_s^{(m)}(x) = \frac{1}{h^m} \int_{\mathbb{R}} K^{(m)}(u)\, g_s(x-hu)\mathrm{d}u. \qquad (31)
$$

By Assumption 1 $\left|K^{(j)}(u) g_s^{(m-1-j)}(x-hu)\right| = o(1)$, as $|u| \to \infty$ for $j = 0, \ldots, m-1$, $h > 0$. Therefore, integration by parts gives (31)

$$
\begin{aligned}
E\hat{f}_s^{(m)}(x) &= \sum_{j=0}^{m-1} \frac{1}{h^{m-j}} K^{(m-1-j)}(u) g_s^{(j)}(x-hu) \Big|_{-\infty}^{\infty} + \int_{\mathbb{R}} K(u) g_s^{(m)}(x-hu)\mathrm{d}u \\
&= \int_{\mathbb{R}} K(u)\, g_s^{(m)}(x-hu)\mathrm{d}u. \qquad (32)
\end{aligned}
$$

Since $\int_{\mathbb{R}} K(t)\mathrm{d}t = 1$, this implies (6).
(2) Plug the definition of M_k in (30) to get

$$E\hat{f}_{s,k}^{(m)}(x) = -\frac{1}{C_{2k}^k}\sum_{|l|=1}^{k}\frac{(-1)^l C_{2k}^{l+k}}{|l| l^m h^{m+1}}\left[\int_0^\infty K^{(m)}\left(\frac{x-t}{lh}\right)f(t)\mathrm{d}t\right.$$
$$\left. + \sum_{j=1}^{s+1}\frac{k_j}{w_j}\int_0^\infty K^{(m)}\left(\frac{x+t/w_j}{lh}\right)f(t)\mathrm{d}t\right]. \tag{33}$$

For $l < 0$, and after putting $\frac{x-t}{lh} = u$ and $\frac{x+t/w_j}{lh} = u$ on the first and second integrals

$$\int_0^\infty K^{(m)}\left(\frac{x-t}{lh}\right)f(t)\mathrm{d}t + \sum_{j=1}^{s+1}\frac{k_j}{w_j}\int_0^\infty K^{(m)}\left(\frac{x+t/w_j}{lh}\right)f(t)\mathrm{d}t$$
$$= -lh\left[\int_{x/(lh)}^\infty K^{(m)}(u)\, f(x-lhu)\mathrm{d}u\right.$$
$$\left. + \sum_{j=1}^{s+1}k_j\int_{-\infty}^{x/(lh)} K^{(m)}(u)\, f(-w_j(x-lhu))\mathrm{d}u\right]$$
$$= -lh\int_{\mathbb{R}} K^{(m)}(u)\, g_s(x-lhu)\mathrm{d}u.$$

Similarly, we have for $l > 0$

$$\int_0^\infty K^{(m)}\left(\frac{x-t}{lh}\right)f(t)\mathrm{d}t + \sum_{j=1}^{s+1}\frac{k_j}{w_j}\int_0^\infty K^{(m)}\left(\frac{x+t/w_j}{lh}\right)f(t)\mathrm{d}t$$
$$= lh\int_{\mathbb{R}} K^{(m)}(u)\, g_s(x-lhu)\mathrm{d}u.$$

Therefore, (33) gives

$$E\hat{f}_{s,k}^{(m)}(x) = -\frac{1}{C_{2k}^k}\sum_{|l|=1}^{k}\frac{(-1)^l C_{2k}^{l+k}}{(lh)^m}\int_{\mathbb{R}} K^{(m)}(u)\, g_s(x-lhu)\mathrm{d}u$$
$$\text{(integrating by parts as above)}$$
$$= -\frac{1}{C_{2k}^k}\sum_{|l|=1}^{k}(-1)^l C_{2k}^{l+k}\int_{\mathbb{R}} K(u)\, g_s^{(m)}(x-lhu)\mathrm{d}u.$$

Finally,

$$E\hat{f}_k^{(m)}(x) - f^{(m)}(x) = -\frac{1}{(-1)^k C_{2k}^k}\sum_{|l|=1}^{k}(-1)^{l+k}C_{2k}^{l+k}\int_{\mathbb{R}} K(u)\, g_s^{(m)}(x-lhu)\mathrm{d}u$$
$$-\frac{(-1)^k C_{2k}^k}{(-1)^k C_{2k}^k}\int_{\mathbb{R}} K(u)\, g_s^{(m)}(x)\mathrm{d}u = -\frac{1}{(-1)^k C_{2k}^k}\int_{\mathbb{R}} K(u)\Delta_{hu}^{2k} g_s^{(m)}(x)\mathrm{d}u$$

which is (7). □

Proof of Theorem 2 For part (1), we note that since $\int K(t)\mathrm{d}t = 1$ we have from Theorem 1

$$E\hat{f}_s^{(m)}(x) - f^{(m)}(x) = \int K(t)(g_s^{(m)}(x-ht) - g_s^{(m)}(x))\mathrm{d}t$$
$$= \int K(t)\left(g_s^{(m+1)}(x)(-ht) + \frac{1}{2!}g_s^{(m+2)}(x)(-ht)^2 + \cdots + \frac{1}{(s-m)!}g_s^{(s)}(x-th\tau)(-ht)^{s-m}\right)\mathrm{d}t$$

for some $\tau \in (0,1)$. If $\int_{\mathbb{R}} K(t)t^j\mathrm{d}t = 0$, for $j = 1,\ldots,s-m-1$, then

$$|E\hat{f}_s^{(m)}(x) - f^{(m)}(x)| \le \frac{h^{s-m}}{(s-m)!}\int |t|^{s-m}|K(t)||g^{(s)}(x-th\tau)|\mathrm{d}t \le Ch^{s-m},$$

where the last inequality follows from the assumptions that $\int_{\mathbb{R}}\left|K(t)t^{s-m}\right|\mathrm{d}t < \infty$, $\left|f^{(s)}(x)\right| < C$ for all $x \ge 0$ and the structure of $g^{(s)}$. For part 2), using (7) and Hölder's inequality we have

$$\left|E\hat{f}_{s,k}^{(m)}(x) - f^{(m)}(x)\right| = c\left|\int_{\mathbb{R}} K(t)\,|ht|^{r+1/q}\,\frac{\Delta_{ht}^{2k}g_s^{(m)}(x)}{|ht|^{r+1/q}}\mathrm{d}t\right|$$
$$\le c\left(\int_{\mathbb{R}}|K(t)|^{q'}\,|ht|^{(r+1/q)q'}\,\mathrm{d}t\right)^{1/q'} \times\left(\int_{\mathbb{R}}\left(\frac{\sup_x|\Delta_{ht}^{2k}g_s^{(m)}(x)|}{|ht|^r}\right)^q\frac{\mathrm{d}t}{|ht|}\right)^{1/q}$$

(changing variables on the second integral)

$$\le ch^r\left(\int|K(t)|^{q'}\,|t|^{(r+1/q)q'}\,\mathrm{d}t\right)^{1/q'}\left\|g_s^{(m)}\right\|_{b_{\infty,q}^r(\mathbb{R})}$$
$$= O(h^r).$$

In the last line, we used the bound $\left\|g_s^{(m)}\right\|_{B_{p,q}^r(\mathbb{R})} \le c\left\|f^{(m)}\right\|_{B_{p,q}^r(0,\infty)}$. □

Proof of Theorem 4 (1) Let I_A denote the indicator of a set A. Then, for an arbitrary function g, $\sum_{X_i<aw_j} g(X_i) = \sum_{i=1}^n I_{\{X_i<aw_j\}}g(X_i)$. Using indicators in (15) and the fact that $\{X_i\}_{i=1}^n$ is IID, we have

$$E\hat{f}_s^{(m)}(x) = \frac{1}{h^{m+1}}\left\{\int_0^1 K^{(m)}\left(\frac{x-t}{h}\right)f(t)\mathrm{d}t\right.$$

$$+\sum_{j=1}^{s+1}\frac{k_j}{w_j}\left[\int_0^{aw_j}K^{(m)}\left(\frac{x+t/w_j}{h}\right)f(t)\mathrm{d}t\right.$$
$$\left.\left.+\int_{1-aw_j}^{1}K^{(m)}\left(\frac{x-1+(t-1)/w_j}{h}\right)f(t)\mathrm{d}t\right]\right\}. \qquad (34)$$

Changing variables using $\frac{x-t}{h}=u$, $\frac{x+t/w_j}{h}=u$, $\frac{x-1+(t-1)/w_j}{h}=u$, we have

$$E\hat{f}_s^{(m)}(x)=\frac{1}{h^m}\left\{-\int_{x/h}^{(x-1)/h}K^{(m)}(u)\,f(x-hu)\mathrm{d}u\right.$$
$$+\sum_{j=1}^{s+1}k_j\left[\int_{x/h}^{(x+a)/h}K^{(m)}(u)\,f(-w_j(x-hu))\mathrm{d}u\right.$$
$$\left.\left.+\int_{(x-1-a)/h}^{(x-1)/h}K^{(m)}(u)\,f(1-w_j(x-hu-1))\mathrm{d}u\right]\right\}.$$

Applying (14), we have

$$E\hat{f}_s^{(m)}(x)=\frac{1}{h^m}\left\{\int_{(x-1)/h}^{x/h}K^{(m)}(u)\,f(x-hu)\mathrm{d}u\right.$$
$$+\int_{x/h}^{(x+a)/h}K^{(m)}(u)\sum_{j=1}^{s+1}k_jf(-w_j(x-hu))\mathrm{d}u$$
$$\left.+\int_{(x-1-a)/h}^{(x-1)/h}K^{(m)}(u)\sum_{j=1}^{s+1}k_jf(1-w_j(x-hu-1))\mathrm{d}u\right\}$$
$$=\frac{1}{h^m}\int_{(x-1-a)/h}^{(x+a)/h}K^{(m)}(u)\,g_s(x-hu)\mathrm{d}u. \qquad (35)$$

Regardless of $x\in[0,1]$, the interval $((x-1-a)/h,(x+a)/h)$ contains $(-a/h,a/h)$ which contains supp$K$ for all small h. Therefore,

$$E\hat{f}_s^{(m)}(x)=\frac{1}{h^m}\int_{\mathbb{R}}K^{(m)}(u)\,g_s(x-hu)\mathrm{d}u.$$

For this to hold formally, g_s should be extended outside $(-a,1+a)$ smoothly; the manner of extension does not affect the above integral. Finally, integration by parts and the condition $\int_{\mathbb{R}}K(t)\mathrm{d}t=1$ prove the statement.
(2) Since K is assumed to have finite support, we do not need Assumption 1. Calculations done in the proof of Theorem 2 after Eq. (33) include change of variables and integration by parts and can be easily repeated here. □

Proof of Theorem 5 Define

$$u_i = \frac{1}{h^{m+1}} \left\{ K^{(m)} \left(\frac{x - X_i}{h} \right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} \left[I_{\{X_i < aw_j\}} K^{(m)} \left(\frac{x + X_i/w_j}{h} \right) \right.\right.$$
$$\left.\left. + \; I_{\{X_i > 1 - aw_j\}} K^{(m)} \left(\frac{x - 1 + (X_i - 1)/w_j}{h} \right) \right] \right\}.$$

Then $V\left(\hat{f}_s^{(m)}(x)\right) = \frac{1}{n}[Eu_1^2 - (Eu_1)^2]$. It will be shown that Eu_1^2 is of order $h^{-(2m+1)}$ in all cases. Since $Eu_1 = O(1)$ by Theorem 4, it is enough to find the exact order of Eu_1^2. Letting $F = K^{(m)}$, denote

$$g = F\left(\frac{x - X_1}{h}\right), \quad g_i^l = I_{\{X_1 < aw_i\}} F\left(\frac{x + X_1/w_i}{h}\right),$$
$$g_i^r = I_{\{X_1 > 1 - aw_i\}} F\left(\frac{x - 1 + (X_1 - 1)/w_i}{h}\right).$$

g is used at internal points of the domain, g_i^l and g_i^r are used for correction at the left and right boundaries, respectively. Their contributions to variances reflect this. From

$$u_1 = \frac{1}{h^{m+1}} \left(g + \sum_{j=1}^{s+1} \frac{k_j}{w_j} \left(g_j^l + g_j^r \right) \right)$$

we see that Eu_1^2 contains (a) Eg^2, (b) Egg_i^l, (c) $Eg_i^l g_j^l$, (d) Egg_j^r, (e) $Eg_i^l g_j^r$, (f) $Eg_i^r g_j^r$.

(I) Let $x \in (0, 1)$.

(a) Replacing $\frac{x-t}{h} = u$, we have

$$\frac{1}{h} Eg^2 = \frac{1}{h} \int_0^1 F^2\left(\frac{x - t}{h}\right) f(t) \mathrm{d}t = \int_{(x-1)/h}^{x/h} F^2(u) f(x - hu) \mathrm{d}u. \quad (36)$$

Since $x/h \to \infty$, $(x-1)/h \to -\infty$ and f is bounded and continuous, in the equation

$$\frac{1}{h} Eg^2 = -\int_{x/h}^{\infty} F^2(u) f(x - hu) \mathrm{d}u - \int_{-\infty}^{(x-1)/h} F^2(u) f(x - hu) \mathrm{d}u$$
$$+ \int_{\mathbb{R}} F^2(u) f(x - hu) \mathrm{d}u$$

the first two integrals on the right tend to zero and the last integral tends to $f(x)\Gamma$ by the dominated convergence theorem. Thus,

$$\frac{1}{h}Eg^2 = f(x)\Gamma + o(1). \tag{37}$$

Similar arguments below will be omitted.

(b) Here we use boundedness of f, F and integrability of F:

$$\left|\frac{1}{h}Egg_i^l\right| = \left|\frac{1}{h}\int_0^{aw_i} F\left(\frac{x-t}{h}\right) F\left(\frac{x+t/w_i}{h}\right) f(t)\mathrm{d}t\right|$$

$$\left(\text{replacing } \frac{x+t/w_i}{h} = u \text{ and using dots in place of inconsequential arguments}\right)$$

$$= w_i \left|\int_{x/h}^{(x+a)/h} F(\cdots)\, F(u)\, f(\cdots)\mathrm{d}u\right| \le w_i \sup|fF| \int_{x/h}^{(x+a)/h} |F(u)|\mathrm{d}u \to 0. \tag{38}$$

(c) Denoting $\lambda = \min\{w_i, w_j\}$, we have

$$\frac{1}{h}Eg_i^l g_j^l = \frac{1}{h}\int_0^{a\lambda} F\left(\frac{x+t/w_i}{h}\right) F\left(\frac{x+t/w_j}{h}\right) f(t)\mathrm{d}t$$

$$\left(\text{replacing } \frac{x+t/w_i}{h} = u\right)$$

$$= w_i \int_{x/h}^{(x+a\lambda/w_i)/h} F(\cdots)\, F(u)\, f(\cdots)\mathrm{d}u \to 0. \tag{39}$$

(d) Replacing $\frac{x-1+(t-1)/w_i}{h} = u$, we get

$$\frac{1}{h}Egg_i^r = \frac{1}{h}\int_{1-aw_i}^{1} F\left(\frac{x-t}{h}\right) F\left(\frac{x-1+(t-1)/w_i}{h}\right) f(t)\mathrm{d}t$$

$$= w_i \int_{(x-1-a)/h}^{(x-1)/h} F(\cdots)\, F(u)\, f(\cdots)\mathrm{d}u \to 0. \tag{40}$$

(e) Replacing $\frac{x+t/w_i}{h} = u$

$$\frac{1}{h}Eg_i^l g_j^r = \frac{1}{h}\int_{1-aw_j}^{aw_i} F\left(\frac{x+t/w_i}{h}\right) F\left(\frac{x-1+(t-1)/w_j}{h}\right) f(t)\mathrm{d}t$$

$$= w_i \int_{[x+(1-aw_j)/w_i]/h}^{(x+a)/h} F(u)\, F(\cdots)\, f(\cdots)\mathrm{d}u \to 0. \tag{41}$$

Here we take into account that $aw_j \le 1$ for all j.

(f) Letting $\lambda = \min\{w_i, w_j\}$ we have

$$\frac{1}{h}Eg_i^r g_j^r = \frac{1}{h}\int_{1-a\lambda}^{1} F\left(\frac{x-1+(t-1)/w_i}{h}\right)$$

$$\times F\left(\frac{x-1+(t-1)/w_j}{h}\right) f(t)\mathrm{d}t$$

$$\left(\text{replacing } \frac{x-1+(t-1)/w_i}{h} = u\right)$$
$$= w_i \int_{(x-1-a\lambda/w_i)/h}^{(x-1)/h} F(u)\, F(\cdots)\, f(\cdots)\mathrm{d}u \to 0. \qquad (42)$$

The conclusion from (37) to (42) is that $Eu_1^2 = \frac{1}{h^{2m+1}}\{f(x)\Gamma + o(1)\}$ which proves statement I).

(II) Let $x = 0$.

(a) From (36)

$$\frac{1}{h}Eg^2 = \int_{-1/h}^{0} F^2(u) f(-hu)\mathrm{d}u \to f(0)\int_{-\infty}^{0} F^2(u)\mathrm{d}u$$
$$= f(0)\int_0^\infty F^2\left(\frac{u}{w_0}\right)\mathrm{d}u. \qquad (43)$$

(b) By (38)

$$\frac{1}{h}Egg_i^l = \frac{1}{h}\int_0^{aw_i} F\left(-\frac{t}{h}\right) F\left(\frac{t}{hw_i}\right) f(t)\mathrm{d}t \qquad \left(\text{replacing } \frac{t}{h} = u\right)$$
$$= \int_0^{aw_i/h} F(-u)\, F\left(\frac{u}{w_i}\right) f(hu)\mathrm{d}u \to f(0)$$
$$\times \int_0^\infty F(-u)\, F\left(\frac{u}{w_i}\right)\mathrm{d}u$$
$$= f(0)\int_0^\infty F\left(\frac{u}{w_0}\right) F\left(\frac{u}{w_i}\right)\mathrm{d}u. \qquad (44)$$

(c) By (39)

$$\frac{1}{h}Eg_i^l g_j^l = \frac{1}{h}\int_0^{a\lambda} F\left(\frac{t/w_i}{h}\right) F\left(\frac{t/w_j}{h}\right) f(t)\mathrm{d}t \qquad \left(\text{replacing } \frac{t}{h} = u\right)$$
$$= \int_0^{a\lambda/h} F\left(\frac{u}{w_i}\right) F\left(\frac{u}{w_j}\right) f(hu)\mathrm{d}u \to f(0)$$
$$\times \int_0^\infty F\left(\frac{u}{w_i}\right) F\left(\frac{u}{w_i}\right)\mathrm{d}u. \qquad (45)$$

(d) By (40)

$$\frac{1}{h}Egg_j^r = w_j \int_{(-1-a)/h}^{-1/h} F(\cdots)\, F(u)\, f(\cdots)\mathrm{d}u \to 0. \qquad (46)$$

(e) From (41)

$$\frac{1}{h}Eg_i^l g_j^r = \frac{1}{h}\int_{1-aw_j}^{aw_i} F\left(\frac{t/w_i}{h}\right) F\left(\frac{-1+(t-1)/w_j}{h}\right) f(t)\mathrm{d}t$$

$$\left(\text{here we replace } \frac{-1+(t-1)/w_j}{h} = u\right)$$

$$= w_j \int_{(-1-a)/h}^{[-1+(aw_i-1)/w_j]/h} F(\cdots) F(u) f(\cdots) du \to 0. \quad (47)$$

Here we remember that $aw_i \leq 1$.

(f) From (42)

$$\frac{1}{h} E g_i^r g_j^r = w_i \int_{(-1-a\lambda/w_i)/h}^{-1/h} F(u) F(\cdots) f(\cdots) du \to 0. \quad (48)$$

From (43) to (48), we conclude that

$$\begin{aligned} Eu_1^2 = \frac{1}{h^{2m+1}} \Bigg\{ & f(0) \int_0^\infty F^2\left(\frac{u}{w_0}\right) du \\ & + 2f(0) \sum_{i=1}^{s+1} \frac{k_i}{w_i} \int_0^\infty F\left(\frac{u}{w_0}\right) F\left(\frac{u}{w_i}\right) du \\ & + f(0) \sum_{i,j=1}^{s+1} \frac{k_i}{w_i} \frac{k_j}{w_j} \int_0^\infty F\left(\frac{u}{w_i}\right) F\left(\frac{u}{w_j}\right) du + o(1) \Bigg\} \\ = \frac{1}{h^{2m+1}} & \left\{ f(0) \int_0^\infty \left[\sum_{i=0}^{s+1} \frac{k_i}{w_i} F\left(\frac{u}{w_i}\right) \right]^2 du + o(1) \right\}. \end{aligned}$$

(III) Let $x = 1$.

(a) From (36)

$$\begin{aligned} \frac{1}{h} E g^2 &= \int_0^{1/h} F^2(u) f(1-hu) du \to f(1) \int_0^\infty F^2(u) du \\ &= f(1) \int_{-\infty}^0 F^2\left(\frac{u}{w_0}\right) du. \end{aligned} \quad (49)$$

(b) From the second line of (38)

$$\frac{1}{h} E g g_i^l = w_i \int_{1/h}^{(1+a)/h} F(\cdots) F(u) f(\cdots) du \to 0. \quad (50)$$

(c) From the last line of (39)

$$\frac{1}{h} E g_i^l g_j^l = w_i \int_{1/h}^{(1+a\lambda/w_i)/h} F(\cdots) F(u) f(\cdots) du \to 0. \quad (51)$$

(d) From (40)

$$\begin{aligned}\frac{1}{h}Egg_i^r &= \frac{1}{h}\int_{1-aw_i}^{1} F\left(\frac{1-t}{h}\right)\\ &\quad\times F\left(\frac{(t-1)/w_i}{h}\right) f(t)\mathrm{d}t \qquad \left(\text{replacing } \frac{t-1}{h}=u\right)\\ &= \int_{-aw_i/h}^{0} F(-u)\, F\left(\frac{u}{w_i}\right) f(1+hu)\mathrm{d}u\\ &\to f(1)\int_{-\infty}^{0} F(-u)\, F\left(\frac{u}{w_i}\right)\mathrm{d}u\\ &= f(1)\int_{-\infty}^{0} F\left(\frac{u}{w_0}\right) F\left(\frac{u}{w_i}\right)\mathrm{d}u. \end{aligned} \tag{52}$$

(e) From (41)

$$\begin{aligned}\frac{1}{h}Eg_i^l g_j^r &= \frac{1}{h}\int_{1-aw_j}^{aw_i} F\left(\frac{1+t/w_i}{h}\right) F\left(\frac{(t-1)/w_j}{h}\right) f(t)\mathrm{d}t\\ &\qquad\left(\text{replace } \frac{1+t/w_i}{h}=u\right)\\ &= w_i\int_{[1+(1-aw_j)/w_i]/h}^{(1+a)/h} F(u)\, F(\cdots)\, f(\cdots)\mathrm{d}u \to 0. \end{aligned} \tag{53}$$

(f) From (42)

$$\begin{aligned}\frac{1}{h}Eg_i^r g_j^r &= \frac{1}{h}\int_{1-a\lambda}^{1} F\left(\frac{t-1}{hw_i}\right)\\ &\quad\times F\left(\frac{t-1}{hw_j}\right) f(t)\mathrm{d}t \qquad \left(\text{replacing } \frac{t-1}{h}=u\right)\\ &= \int_{-a\lambda/h}^{0} F\left(\frac{u}{w_i}\right) F\left(\frac{u}{w_j}\right) f(1+hu)\mathrm{d}u\\ &\to f(1)\int_{-\infty}^{0} F\left(\frac{u}{w_i}\right) F\left(\frac{u}{w_j}\right)\mathrm{d}u. \end{aligned} \tag{54}$$

Collecting nonzero limits from (49), (52), (54)

$$\begin{aligned}Eu_1^2 = \frac{1}{h^{2m+1}}\Bigg\{ & f(1)\int_{-\infty}^{0} F^2\left(\frac{u}{w_0}\right)\mathrm{d}u\\ &+2f(1)\sum_{i=1}^{s+1}\frac{k_i}{w_i}\int_{-\infty}^{0} F\left(\frac{u}{w_0}\right) F\left(\frac{u}{w_i}\right)\mathrm{d}u\end{aligned}$$

$$+ f(1)\sum_{i,j=1}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\int_{-\infty}^{0}F\left(\frac{u}{w_i}\right)F\left(\frac{u}{w_i}\right)\mathrm{d}u+o(1)\Bigg\}$$

$$=\frac{1}{h^{2m+1}}\left\{f(1)\int_{-\infty}^{0}\left[\sum_{i=0}^{s+1}\frac{k_i}{w_i}F\left(\frac{u}{w_i}\right)\right]^2\mathrm{d}u+o(1)\right\}.$$

□

Proof of Theorem 6 (1) Instead of (30), we have

$$E\hat{f}_{+,s}^{(m)}(x)=\frac{1}{h^{m+1}}E\left\{I_{\{X_1\geq 0\}}\left[K^{(m)}\left(\frac{x-X_1}{h}\right)+\sum_{j=1}^{s+1}\frac{k_j}{w_j}K^{(m)}\left(\frac{x+X_1/w_j}{h}\right)\right]\right\}$$

$$=\frac{1}{h^{m+1}}\left[\int_0^{\infty}K^{(m)}\left(\frac{x-t}{h}\right)f_+(t)\mathrm{d}t+\sum_{j=1}^{s+1}\frac{k_j}{w_j}\int_0^{\infty}K^{(m)}\left(\frac{x+t/w_j}{h}\right)f_+(t)\mathrm{d}t\right].$$

Repeating calculations that led from (30) to (32), we get

$$E\hat{f}_{+,s}^{(m)}(x)=\int_{\mathbb{R}}K(s)\,g_{+,s}^{(m)}(x-hs)ds$$

(those calculations did not use the fact that f was a density).

(2) The proof is similar to that of Theorem 5.

□

Proof of Theorem 7 By the i.i.d. assumption

$$E\hat{f}_s^{(m)}(x)=\frac{1}{h^{m+1}}\left\{\int_c^d K^{(m)}\left(\frac{x-t}{h}\right)f_r(t)\mathrm{d}t+\sum_{j=1}^{s+1}\frac{k_j}{w_j}\left[\int_c^{c+a_1w_j}K^{(m)}\left(\frac{w_j(x-c)+(t-c)}{w_jh}\right)f_r(t)\mathrm{d}t+\int_{d-a_1w_j}^{d}K^{(m)}\left(\frac{w_j(x-d)+(t-d)}{w_jh}\right)f_r(t)\mathrm{d}t\right]\right\}.$$

The obvious changes of variables are:

$$\frac{x-t}{h}=u,\quad\frac{w_j(x-c)+(t-c)}{w_jh}=u,\quad\frac{w_j(x-d)+(t-d)}{w_jh}=u.$$

The mean value becomes

$$E\hat{f}_s^{(m)}(x) = \frac{1}{h^m}\left\{-\int_{(x-c)/h}^{(x-d)/h} K^{(m)}(u)\, f_r(x-hu)\mathrm{d}u \right.$$
$$+\sum_{j=1}^{s+1} k_j \left[\int_{(x-c)/h}^{(x-c+a_1)/h} K^{(m)}(u)\, f_r(c-w_j(x-c)+w_j hu)\mathrm{d}u\right.$$
$$\left.\left.+\int_{(x-d-a_1)/h}^{(x-d)/h} K^{(m)}(u)\, f_r(d-w_j(x-d)+w_j hu)\mathrm{d}u\right]\right\}.$$

Applying (19) we see that this is the same as

$$E\hat{f}_s^{(m)}(x) = \frac{1}{h^m}\left\{\int_{(x-d)/h}^{(x-c)/h} K^{(m)}(u)\, f_r(x-hu)\mathrm{d}u\right.$$
$$+\int_{(x-c)/h}^{(x-c+a_1)/h} K^{(m)}(u)\sum_{j=1}^{s+1} k_j f_r(c-w_j(x-hu-c))\mathrm{d}u$$
$$\left.+\int_{(x-d-a_1)/h}^{(x-d)/h} K^{(m)}(u)\sum_{j=1}^{s+1} k_j f_r(d-w_j(x-hu-d))\mathrm{d}u\right\}$$
$$=\frac{1}{h^m}\int_{(x-d-a_1)/h}^{(x-c+a_1)/h} K^{(m)}(u)\, g_r(x-hu)\mathrm{d}u.$$

Regardless of $x \in [c,d]$, the interval $((x-d-a)/h, (x-c+a)/h)$ contains $(-a/h, a/h)$ which contains supp$K$ for all small h. Therefore, also integrating by parts,

$$E\hat{f}_s^{(m)}(x) = \frac{1}{h^m}\int_{\mathbb{R}} K^{(m)}(u)\, g_r(x-hu)\mathrm{d}u = \int_{\mathbb{R}} K(u)\, g_r^{(m)}(x-hu)\mathrm{d}u.$$

The derivation of the expression for variance largely repeats that from Theorem 5. □

Proof of Theorem 8 (21) implies

$$\left(\frac{d}{dx}\right)^j \tilde{f}^{(m)}(x)\,|_{x=0+} = \sum_{i=0}^{j} C_j^i \left[\left(\frac{d}{dx}\right)^i \psi(xh^{-\alpha})\right]$$
$$\times\left[\left(\frac{d}{dx}\right)^{j-i} \hat{f}^{(m)}(x)\right]|_{x=0+} = 0,$$

for $j = 0, \ldots, l-m$, so (23) is satisfied.

To prove (24), consider two cases.

Case $xh^{-\alpha} \geq 1$. (24) follows trivially from (8) because $\tilde{f}^{(m)}(x) = \hat{f}^{(m)}(x)$.

Case $xh^{-\alpha} \le 1$. Obviously, in the equation

$$E\tilde{f}^{(m)}(x) - f^{(m)}(x) = \psi(xh^{-\alpha})\left[E\hat{f}^{(m)}(x) - f^{(m)}(x)\right] + \left[\psi(xh^{-\alpha}) - 1\right] f^{(m)}(x)$$

the first term on the right is $O(h^{s-m})$, and it remains to prove that $\left[\psi(xh^{-\alpha}) - 1\right] f^{(m)}(x) = O(h^{s-m})$. Suppose (22) is true, so that $\alpha = 1$. Then

$$\begin{aligned} f^{(m)}(x) &= f^{(m)}(0+) + \cdots + f^{(s)}(0+)\frac{x^{s-m}}{(s-m)!} + o(x^{s-m}) \\ &= o\left((h^{\alpha})^{s-m}\right) = o(h^{s-m}). \end{aligned} \tag{55}$$

Suppose (22) is wrong. Then $\alpha = \frac{s-m}{k-m}$ and

$$\begin{aligned} f^{(m)}(x) &= f^{(m)}(0+) + \cdots + f^{(k)}(0+)\frac{x^{k-m}}{(k-m)!} + o(x^{k-m}) \\ &= O\left(x^{k-m}\right) = O(h^{s-m}). \end{aligned} \tag{56}$$

(55) and (56) prove what we need.

To prove (25), consider two cases.
Case $xh^{-\alpha} \ge 1$. The first part of (25) is obvious because $\tilde{f}^{(m)}(x) = \hat{f}^{(m)}(x)$.
Case $xh^{-\alpha} \le 1$. From (35) it follows that

$$\begin{aligned} \psi(xh^{-\alpha}) &= \psi(0+) + \cdots + \psi^{(l-m+1)}(0+)\frac{\left(xh^{-\alpha}\right)^{(l-m+1)}}{(l-m+1)!} \\ &= \psi^{(l-m+1)}(0+)\left(xh^{-\alpha}\right)^{(l-m+1)} \end{aligned}$$

which proves the second part of (25). □

Proof of Theorem 9 For almost all samples $\min_i X_i > 0$ and for $0 < x \le \frac{1}{3}\min_i X_i$, one has $\psi(X_i/x) = 0$, $i = 1, \ldots, n$. Hence $\hat{f}^{(m)}(x)$ vanishes, together with all its derivatives, in the neighborhood of zero for almost all samples. Following (30), we see that the mean is

$$E\hat{f}^{(m)}(x) = \frac{1}{h^{m+1}}\int_0^\infty \left[K^{(m)}\left(\frac{x-t}{h}\right) + \sum_{j=1}^{s+1}\frac{k_j}{w_j}K^{(m)}\left(\frac{x+t/w_j}{h}\right)\right]\psi\left(\frac{t}{x}\right)f(t)\mathrm{d}t.$$

Here the function $f_x(t) = f(t)\psi(t/x)$ has support $\operatorname{supp} f_x \subseteq [0, 3x]$. Implementing changes applied after (31), including integration by parts, we obtain an analog of (32) with g_x instead of g. g_x is obtained by replacing f in (2)–(3) by f_x. The rest is familiar.

The statement about variance is obtained by repeating the corresponding part of the proof of Theorem 5. □

References

Burenkov, V. I. (1998). *Sobolev spaces on domains*. Leipzig: B. G. Teubner.

Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, *31*, 131–145.

Chen, S. X. (2000). Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, *52*, 471–480.

Cheng, M.-Y. (1994). *On boundary effects of smooth curve estimators*. Ph.D. thesis, University of North Carolina-Chapel Hill, Department of Statistics.

Cline, D. B. H., Hart, J. D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics: A Journal of Theoretical and Applied Statistics*, *22*, 69–84.

Hestenes, M. R. (1941). Extension of the range of a differentiable function. *Duke Mathematical Journal*, *8*, 183–192.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, *3*, 135–146.

Jones, M. C., Foster, P. J. (1993). Generalized jackknifing and higher order kernels. *Journal of Nonparametric Statistics*, *3*, 81–94.

Karunamuni, R. J., Alberts, T. (2005). A generalized reflection method of boundary correction in kernel density estimation. *The Canadian Journal of Statistics*, *33*, 497–509.

Malec, P., Schienle, M. (2014). Nonparametric kernel density estimation near the boundary. *Computational Statistics & Data Analysis*, *72*, 57–76.

Marron, J. S., Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society*: *Series B*, *56*, 653–671.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design. *Journal of Econometrics*, *142*, 698–714.

Mynbaev, K., Martins-Filho, C. (2010). Bias reduction in kernel density estimation via Lipschitz condition. *Journal of Nonparametric Statistics*, *22*, 219–235.

Mynbaev, K., Martins-Filho, C., Aipenova, A. (2016). A class of nonparametric density derivative estimators based on global Lipschitz conditions. *Advances in Econometrics*, *36*, 591–615.

Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics - Theory and Methods*, *14*, 1123–1136.

Wen, K., Wu, X. (2015). An improved transformation-based kernel density estimator of densities on the unit interval. *Journal of the American Statistical Association*, *110*, 773–783.