Unified estimation of densities on bounded and unbounded domains

Kairat Mynbaev

International School of Economics
Kazakh-British Technical University
Tolebi 59
Almaty 050000, Kazakhstan
email: kairat_mynbayev@yahoo.com

and

Carlos Martins-Filho

Department of Economics
University of Colorado
Boulder, CO 80309-0256, USA
email: carlos.martins@colorado.edu

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: c.martins-filho@cgiar.org

March, 2017

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 extend the density and estimate its extension. The resulting estimators in domains with boundaries have biases and variances expressed in terms of density extensions. 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 and phrases. Nonparametric estimation, Hestenes extension, estimation in bounded domains, estimation of discontinuous densities.

AMS-MS Classification. 62F12, 62G07, 62G20.

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 $\int_{\mathbb{R}} K(t)dt = 1$, $h > 0$ 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, they are from the Gaussian or Epanechnikov families); 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 $E\hat{f}(x) - f(x) = O(h^s)$.[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 called into being a range of estimation methods. Among the widely used ones are the reflection method, the boundary kernel method, the transformation method and the local linear method (see, *inter alia,* Schuster (1985), 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. Such techniques necessarily require variable bandwidths, 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 of estimation near the boundary precluded researchers from identifying a core class 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 estimation procedures that permit a unified theoretical study of their properties

[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 $\left|f^{(s)}(x)\right| \leq C$ for some $0 < C < \infty$ is denoted by $\mathcal{C}_b^s(\Omega)$. We say that the kernel K is of order $s \geq 2$ if $\int t^j K(t)dt = 0$ for $j = 1, \cdots, s-1$ and $\int t^s K(t)dt \neq 0$.

under bounded and 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 hold. 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 known discontinuity points, our estimators supply the required jumps at those points. As in $\mathbb{R}$, the estimator for densities in classes where $s > 2$ is not necessarily nonnegative, because the estimation involves higher-order kernels. Our method does not cover densities with poles at endpoints.

Our estimation method is based on Hestenes' extension (Hestenes (1941)). Let D_f be the domain of the density f and denote by g its Hestenes extension (the definitions for the half-axis and interval 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 estimator $\hat{f}(x)$ of $f(x)$ and the properties of g is of type

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

This representation has eluded previous work, and can be used for evaluating 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 is 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 section 2, we start with estimation of a density on $[0,\infty)$. Section 3 treats densities on a bounded interval. In section 4, the approach is extended to estimation of discontinuous densities. Section 5 provides two methods to satisfy zero boundary conditions.

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

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

$$\sum_{i=1}^{s+1} (-w_i)^j k_i = 1,\ \ j = 0, ..., s. \tag{2.1}$$

Since this system has the Van-der-Monde determinant

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

$k_1, ..., k_{s+1}$ are uniquely defined. If f has $[0, \infty)$ as its domain, its Hestenes extension for $x < 0$ is defined by

$$\phi_s(x) = \sum_{j=1}^{s+1} k_j f(-w_j x),\ \ x < 0. \tag{2.2}$$

ϕ_s is not a density, but it is 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 $f(0+), ..., f^{(s)}(0+)$ ($s = 0$ means continuity), we see that the following sewing conditions at zero are satisfied due to (2.1):

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

Now define g on $\mathbb{R}$ by

$$g(x) = \left\{ \begin{array}{l} f(x),\ \ x \geq 0 \\ \phi_s(x),\ \ x < 0 \end{array} \right. , \tag{2.3}$$

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

Suppose f is s times differentiable, $m = 0, 1, \cdots, s$, the kernel K is m times differentiable, and let $\{X_i\}_{i=1}^n$ be an IID sample from f. The estimator of $f^{(m)}(x)$, for $x > 0$, is defined by

$$\hat{f}^{(m)}(x) = \frac{1}{nh^{m+1}} \sum_{i=1}^{n} \left[K^{(m)}\left(\frac{x - X_i}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K^{(m)}\left(\frac{x + X_i / w_j}{h}\right) \right]. \tag{2.4}$$

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

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

The next assumption is used only for $m \geq 1$, when integration by parts is needed.

Assumption 2.1. *a) K is even, m times differentiable and $\max_{1\leq j\leq m-1}|K^{(j)}(t)||t| = o(1)$ as $|t| \to \infty$; b) $\max_{1\leq j\leq m-1}|f^{(j)}(x)| = O(x)$ as $x \to \infty$.*

The estimator in (2.4) 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 $l = 0, \cdots, 2k$. In this context, K is called the seed of M_k. These kernel are used together with an order $2k$ finite difference

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

when Besov type norms are employed to measure smoothness (see Mynbaev and Martins-Filho (2010)). We let $\hat{f}_k^{(m)}(x)$ denote the estimator defined in (2.4) with K replaced by M_k. Note that if K is even, then $M_1(x) = K(x)$ and $\hat{f}_1^{(m)}(x) = \hat{f}^{(m)}(x)$.

Theorem 2.1. *Suppose f is s-times differentiable on $D_f = [0, \infty)$ and the kernel K is m times differentiable on $\mathbb{R}$ with $m = 0, 1, \cdots, s$. In case $m \geq 1$ suppose that Assumption 2.1 holds. Then,*

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

$$E\hat{f}^{(m)}(x) - f^{(m)}(x) = \int_{\mathbb{R}} K(t)\left[g^{(m)}(x - ht) - g^{(m)}(x)\right]dt,\ x \in D_f. \tag{2.5}$$

2) If in (2.4) a kernel M_k with seed K is used, then

$$E\hat{f}_k^{(m)}(x) - f^{(m)}(x) = \frac{(-1)^{k+1}}{C_{2k}^k}\int_{\mathbb{R}} K(t)\Delta_{ht}^{2k}g^{(m)}(x)dt,\ x \in D_f. \tag{2.6}$$

Proof. 1) By the IID assumption

$$
\begin{aligned}
E\hat{f}^{(m)}(x) &= \frac{1}{h^{m+1}} E\left[K^{(m)}\left(\frac{x-X_1}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K^{(m)}\left(\frac{x+X_1/w_j}{h}\right)\right] \\
&= \frac{1}{h^{m+1}}\left[\int_0^\infty K^{(m)}\left(\frac{x-t}{h}\right) f(t)dt + \sum_{j=1}^{s+1} \frac{k_j}{w_j}\int_0^\infty K^{(m)}\left(\frac{x+t/w_j}{h}\right) f(t)dt\right]. \qquad (2.7)
\end{aligned}
$$

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

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

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

$$
E\hat{f}^{(m)}(x) = \frac{1}{h^m}\int_{\mathbb{R}} K^{(m)}(u)\, g(x-hu)du. \qquad (2.8)
$$

By Assumption 2.1 $\left|K^{(j)}(u) g^{(m-1-j)}(x-hu)\right| = o(1)$, as $|u| \to \infty$ for $j = 0, ..., m-1$, $h > 0$. Therefore, integration by parts gives the following expression for (2.8)

$$
\begin{aligned}
E\hat{f}^{(m)}(x) &= \sum_{j=0}^{m-1} \frac{1}{h^{m-j}} K^{(m-1-j)}(u) g^{(j)}(x-hu)\Big|_{-\infty}^{\infty} + \int_{\mathbb{R}} K(u) g^{(m)}(x-hu)du \\
&= \int_{\mathbb{R}} K(u)\, g^{(m)}(x-hu)du. \qquad (2.9)
\end{aligned}
$$

Since $\int_{\mathbb{R}} K(t)dt = 1$, this implies (2.5).

2) Plug the definition of M_k in (2.7) to get

$$
E\hat{f}_k^{(m)}(x) = -\frac{1}{C_{2k}^k}\sum_{|l|=1}^{k} \frac{(-1)^l C_{2k}^{l+k}}{|l| l^m h^{m+1}}\left[\int_0^\infty K^{(m)}\left(\frac{x-t}{lh}\right) f(t)dt + \sum_{j=1}^{s+1}\frac{k_j}{w_j}\int_0^\infty K^{(m)}\left(\frac{x+t/w_j}{lh}\right) f(t)dt\right]. \qquad (2.10)
$$

For $l < 0$,

$$\begin{aligned}
& \int_0^\infty K^{(m)}\left(\frac{x-t}{lh}\right) f(t)dt + \sum_{j=1}^{s+1} \frac{k_j}{w_j} \int_0^\infty K^{(m)}\left(\frac{x+t/w_j}{lh}\right) f(t)dt \\
& \text{(replacing } \frac{x-t}{lh} = u \text{ and } \frac{x+t/w_j}{lh} = u) \\
= & -lh\left[\int_{x/(lh)}^\infty K^{(m)}(u) f(x-lhu)du + \sum_{j=1}^{s+1} k_j \int_{-\infty}^{x/(lh)} K^{(m)}(u) f(-w_j(x-lhu))du\right] \\
= & -lh \int_{\mathbb{R}} K^{(m)}(u) g(x-lhu)du.
\end{aligned}$$

Similarly, we have for $l > 0$

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

Therefore, (2.10) gives

$$\begin{aligned}
E\hat{f}_k^{(m)}(x) &= -\frac{1}{C_{2k}^k} \sum_{|l|=1}^k \frac{(-1)^l C_{2k}^{l+k}}{(lh)^m} \int_{\mathbb{R}} K^{(m)}(u) g(x-lhu)du \\
& \quad \text{(integrating by parts as above)} \\
&= -\frac{1}{C_{2k}^k} \sum_{|l|=1}^k (-1)^l C_{2k}^{l+k} \int_{\mathbb{R}} K(u) g^{(m)}(x-lhu)du.
\end{aligned}$$

Finally,

$$\begin{aligned}
E\hat{f}_k^{(m)}(x) - f^{(m)}(x) &= -\frac{1}{(-1)^k C_{2k}^k} \sum_{|l|=1}^k (-1)^{l+k} C_{2k}^{l+k} \int_{\mathbb{R}} K(u) g^{(m)}(x-lhu)du \\
-\frac{(-1)^k C_{2k}^k}{(-1)^k C_{2k}^k} \int_{\mathbb{R}} K(u) g^{(m)}(x)du &= -\frac{1}{(-1)^k C_{2k}^k} \int_{\mathbb{R}} K(u) \Delta_{hu}^{2k} g^{(m)}(x)du
\end{aligned}$$

which is (2.6). □

The integral representations for biases obtained in Theorem 2.1 depend on the extension g, 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 2.1 is relevant. This approach can be used for derivatives of orders $m \leq s-1$ 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 $1 \leq p, q \leq \infty$ and Ω an open subset of $\mathbb{R}$ put $\Delta^{2k}_{h,\Omega} f(x) = \Delta^{2k}_h f(x)$ if $[x-kh, x+kh] \subset \Omega$ and $\Delta^{2k}_{h,\Omega} f(x) = 0$ otherwise and let

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

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 extensionis known to be bounded from $B^r_{p,q}(\Omega)$ to $B^r_{p,q}(\mathbb{R})$.

Assumption 2.2. *For $0 \leq m \leq s$, $\left\|f^{(m)}\right\|_{b^r_{\infty,q}(0,\infty)} < \infty$ with some $r > 0$ and $1 \leq q \leq \infty$ and*

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

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

Theorem 2.2. *1) Let Assumption 2.1 hold and assume that $\int_{\mathbb{R}} K(t) t^j dt = 0$, for $j = 1, ..., s-m-1$, $\int_{\mathbb{R}} |K(t) t^{s-m}| dt < \infty$ and $\left|f^{(s)}(x)\right| < C$ for all $x > 0$, then*

$$E\hat{f}^{(m)}(x) - f^{(m)}(x) = O(h^{s-m}) \text{ for all } x \in D_f. \tag{2.11}$$

2) Let f and K satisfy Assumption 2.2, then

$$E\hat{f}^{(m)}_k(x) - f^{(m)}(x) = O(h^r) \text{ for all } x \in D_f. \tag{2.12}$$

Proof. For part 1), we note that since $\int K(t) dt = 1$ we have from Theorem 2.1

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

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

$$|E\hat{f}^{(m)}(x) - f^{(m)}(x)| \leq \frac{h^{s-m}}{(s-m)!} \int |t|^{s-m} |K(t)| |g^{(s)}(x - th\tau)| dt \leq C h^{s-m},$$

where the last inequality follows from the assumptions that $\int_{\mathbb{R}} |K(t)t^{s-m}|\,dt < \infty$, $\left|f^{(s)}(x)\right| < C$ for all $x > 0$ and the structure of $g^{(s)}$. For part 2), using (2.6) and Hölder's inequality we have

$$\begin{aligned}\left|E\hat{f}_k^{(m)}(x) - f^{(m)}(x)\right| &= c\left|\int_{\mathbb{R}} K(t)\,|ht|^{r+1/q}\,\frac{\Delta_{ht}^{2k}g^{(m)}(x)}{|ht|^{r+1/q}}dt\right| \\ &\le c\left(\int_{\mathbb{R}} |K(t)|^{q'}\,|ht|^{(r+1/q)q'}\,dt\right)^{1/q'}\left(\int_{\mathbb{R}}\left(\frac{\sup_x|\Delta_{ht}^{2k}g^{(m)}(x)|}{|ht|^r}\right)^q\frac{dt}{|ht|}\right)^{1/q} \\ &\quad \text{(changing variables on the second integral)} \\ &\le ch^r\left(\int |K(t)|^{q'}\,|t|^{(r+1/q)q'}\,dt\right)^{1/q'}\left\|g^{(m)}\right\|_{b^r_{\infty,q}(\mathbb{R})} = O(h^r).\end{aligned}$$

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

From now on we give just expressions for bias and variance leaving consequences of type (2.11) and (2.12) to the reader.

Theorem 2.3. *Suppose that f is continuous, $\sup_x|f(x)| < \infty$, $\sup_x\left|K^{(m)}(x)\right| < \infty$, $\int_{\mathbb{R}}\left|K^{(m)}(t)\right|^2 dt$ and $\int_{\mathbb{R}}\left|K^{(m)}(t)\right| dt < \infty$. In case $m > 0$ also let Assumption 2.1 hold. Denote $F(t) = M_k^{(m)}(t)$. Then*

$$V\left(\hat{f}_k^{(m)}(x)\right) = \frac{1}{nh^{2m+1}}\left\{f(x)\int_{\mathbb{R}} F^2(t)dt + o(1)\right\},\ x > 0. \tag{2.13}$$

The estimator $\hat{f}^{(m)}(x)$ has a similar property with $F(t) = K^{(m)}(t)$ in place of $M_k^{(m)}(t)$.

Proof. Denoting $u_i = \frac{1}{h^{m+1}}\left[F\left(\frac{x-X_i}{h}\right) + \sum_{j=1}^{s+1}\frac{k_j}{w_j}F\left(\frac{x+X_i/w_j}{h}\right)\right]$ we have $\hat{f}_k^{(m)}(x) = \frac{1}{n}\sum_{i=1}^n u_i$ and given the IID assumption

$$V\left(\hat{f}_k^{(m)}(x)\right) = \frac{1}{n}\left[Eu_1^2 - (Eu_1)^2\right]. \tag{2.14}$$

Let $w_0 = -\frac{1}{2},\ k_0 = 0,\ w_{-1} = -1,\ k_{-1} = -1$ (the value of w_0 does not matter). Then

$$\begin{aligned}Eu_1^2 &= E\left[\frac{1}{h^{m+1}}\sum_{j=-1}^{s+1}\frac{k_j}{w_j}F\left(\frac{x+X_1/w_j}{h}\right)\right]^2 \\ &= \frac{1}{h^{2m+2}}\sum_{i,j=-1}^{s+1}\frac{k_i}{w_i}\frac{k_j}{w_j}\int_0^\infty F\left(\frac{x+t/w_i}{h}\right)F\left(\frac{x+t/w_j}{h}\right)f(t)dt. \end{aligned} \tag{2.15}$$

Denote by $B(x,r) = \{t : |x-t| \le r\}$ the ball centered at x with radius r. For any positive ε there is $a > 0$ such that $\int_{|u|>a}|F(u)|\,du < \varepsilon$. Note that

$$B_i \equiv \left\{t : \left|\frac{x+t/w_i}{h}\right| \le a\right\} = \{t : |w_i x + t| \le |w_i|\,ha\} = B(-w_i x, |w_i|\,ha).$$

Since all w_i are different and $x > 0$, the balls B_i do not overlap for small h. Let h be that small and denote $F_i(t) = F((x + t/w_i)/h)$. Then using $\mathbb{R} = B_i \cup B_i^c$ (here c stands for the complement) we have

$$\begin{aligned} \left| \int_0^\infty F_i(t) F_j(t) f(t) dt \right| &\leq \sup |f| \int_{B_i \cup B_i^c} |F_i(t) F_j(t)| dt \\ &= c_1 \left[\int_{B_i \cap (B_j \cup B_j^c)} |F_i(t) F_j(t)| dt + \int_{B_i^c} |F_i(t) F_j(t)| dt \right] \\ &\quad \text{(using } B_i \cap B_j = \varnothing,\ B_i \cap B_j^c \subset B_j^c) \\ &\leq c_2 \left(\int_{B_j^c} |F_j(t)| dt + \int_{B_i^c} |F_i(t)| dt \right). \end{aligned}$$

Further,

$$\int_{B_i^c} |F_i(t)| dt = \int_{|w_i x + t| > |w_i| h a} \left| F\left(\frac{x + t/w_i}{h} \right) \right| dt = |w_i| h \int_{|u| > a} |F(u)| du = O(\varepsilon h).$$

It follows that $\int_0^\infty F_i(t) F_j(t) f(t) dt = o(h)$ and from (2.15)

$$\begin{aligned} Eu_1^2 &= \frac{1}{h^{2m+2}} \left[\sum_{i=-1}^{s+1} \left(\frac{k_i}{w_i} \right)^2 \int_0^\infty F_i^2(t) f(t) dt + o(h) \right] \\ &= \frac{1}{h^{2m+2}} \left[\int_0^\infty F^2\left(\frac{x-t}{h} \right) f(t) dt + \sum_{i=1}^{s+1} \left(\frac{k_i}{w_i} \right)^2 \int_0^\infty F^2\left(\frac{x + t/w_i}{h} \right) f(t) dt + o(h) \right] \\ &= \frac{1}{h^{2m+1}} \left[\int_{-\infty}^{x/h} F^2(u) f(x - hu) du + \sum_{i=1}^{s+1} \left(\frac{k_i}{w_i} \right)^2 \int_{x/h}^\infty F^2(u) f(-w_i(x - hu)) dt + o(1) \right]. \end{aligned}$$

Defining

$$g^*(x) = \begin{cases} f(x), & x > 0 \\ \sum_{i=1}^{s+1} \left(\frac{k_i}{w_i} \right)^2 f(-w_j x), & x < 0 \end{cases} \tag{2.16}$$

we can write

$$Eu_1^2 = \frac{1}{h^{2m+1}} \left[\int_{\mathbb{R}} F^2(u) g^*(x - hu) du + o(1) \right]. \tag{2.17}$$

g^* is not a smooth extension of f but it is integrable on $\mathbb{R}$ and continuous at $x > 0$. Therefore $\int_{\mathbb{R}} F^2(u) g^*(x-hu) du \to g^*(x) \int_{\mathbb{R}} F^2(u) du = f(x) \int_{\mathbb{R}} F^2(u) du$. Theorem 2.2, (2.14) and (2.17) finish the proof. $\square$

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 (2.2). To obtain a common domain for the components of ϕ, we put $a = \min_i(1/w_i)$ and let

$$\phi_1(x) = \sum_{j=1}^{s+1} k_j f(-w_j x), \ -a < x < 0.$$

The sewing conditions at 0 are satisfied as before. Put

$$\phi_2(x) = \sum_{j=1}^{s+1} k_j f(1 - w_j(x-1)), \ 1 < x < 1 + a,$$

and define the extension by

$$g(x) = \begin{cases} \phi_1(x), & -a < x < 0 \\ f(x), & 0 < x < 1 \\ \phi_2(x), & 1 < x < 1 + a \end{cases} \tag{3.1}$$

The sewing condition holds at $x = 1$:

$$\phi_2^{(j)}(1+) = \sum_{m=1}^{s+1} (-w_m)^j k_m f^{(j)}(1-) = f^{(j)}(1-), \ j = 0, ..., s.$$

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

$$\begin{aligned} \hat{f}^{(m)}(x) \quad &= \quad \frac{1}{nh^{m+1}} \left\{ \sum_{i=1}^{n} K^{(m)}\left(\frac{x - X_i}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} \left[\sum_{X_i < aw_j} K^{(m)}\left(\frac{x + X_i/w_j}{h}\right) \right.\right. \\ &+ \quad \left.\left. \sum_{X_i > 1 - aw_j} K^{(m)}\left(\frac{x - 1 + (X_i - 1)/w_j}{h}\right) \right] \right\}. \end{aligned} \tag{3.2}$$

Theorem 3.1. *Let f be s times differentiable 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 (2.5) *is true.*

2) If in (3.2) *K is replaced by M_k, then* (2.6) *is true.*

Proof. 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 (3.2) and the fact that $\{X_i\}_{i=1}^n$ is IID, we have

$$\begin{aligned} E\hat{f}^{(m)}(x) &= \frac{1}{h^{m+1}}\left\{\int_0^1 K^{(m)}\left(\frac{x-t}{h}\right) f(t)dt + \sum_{j=1}^{s+1}\frac{k_j}{w_j}\left[\int_0^{aw_j} K^{(m)}\left(\frac{x+t/w_j}{h}\right) f(t)dt\right.\right. \\ &+ \left.\left.\int_{1-aw_j}^1 K^{(m)}\left(\frac{x-1+(t-1)/w_j}{h}\right) f(t)dt\right]\right\}. \end{aligned} \tag{3.3}$$

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

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

Applying (3.1), we have

$$\begin{aligned} E\hat{f}^{(m)}(x) &= \frac{1}{h^m}\left\{\int_{(x-1)/h}^{x/h} K^{(m)}(u) f(x-hu)du + \int_{x/h}^{(x+a)/h} K^{(m)}(u)\sum_{j=1}^{s+1} k_j f(-w_j(x-hu))du\right. \\ &+ \left.\int_{(x-1-a)/h}^{(x-1)/h} K^{(m)}(u)\sum_{j=1}^{s+1} k_j f(1-w_j(x-hu-1))du\right\} \\ &= \frac{1}{h^m}\int_{(x-1-a)/h}^{(x+a)/h} K^{(m)}(u)\, g(x-hu)du. \end{aligned} \tag{3.4}$$

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,

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

For this to hold formally, g 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 $\int_{\mathbb{R}} K(t)dt=1$ prove the statement.

2) Since K is assumed to have finite support, we do not need Assumption 2.1. Calculations done in the proof of Theorem 2.2 afterequation (2.10) include change of variables and integration by parts and can be easily repeated here. □

Remark. 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 (3.1) consider the extension $g^*(x)=h(x)g(x)$ and change

(3.2) accordingly. Then the statement of Theorem 3.1 will be true for g^* 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 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 3.2. *Under conditions of Theorem 3.1 the following is true.*

1) For a classical kernel K

$$V\left(\hat{f}^{(m)}(x)\right) = \frac{1}{nh^{2m+1}}\left\{f(x)\int_{\mathbb{R}} F^2(t)dt + O(h)\right\},\ x \in D_f.$$

where $F(x) = K^{(m)}(x)$.

2) If M_k is used in place of K, then the same asymptotic expression is true with $F(x) = M_k^{(m)}(x)$.

Proof. 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 we can use equations similar to (2.14) and (2.15).

We want to show that in the analog of (2.15) all cross-products vanish if h is small. Suppose supp$K \subset (-b, b)$ for some $b > 0$. In the cross-products we have integrands that include products of functions

$$f_0(t) = K^{(m)}\left(\frac{x - t}{h}\right),\ f_j(t) = K^{(m)}\left(\frac{x + t/w_j}{h}\right),\ g_j(t) = K^{(m)}\left(\frac{x - 1 + (t - 1)/w_j}{h}\right).$$

It is easy to see that

$$\text{supp}f_0 \subseteq B(x, hb),\ \text{supp}f_j \subseteq B(-w_j x, w_j hb),\ \text{supp}\, g_j \subseteq B(1 - w_j(x - 1), w_j hb).$$

Since the radii of the balls here tend to zero, for small h they do not overlap if their centers are different. It is easy to show that all the centers are indeed different using the conditions that $0 < x < 1$, $w_j > 0$ for all j and all w_j are different.

Thus, for small h

$$\begin{aligned} Eu_1^2 &= \frac{1}{h^{2m+2}}\left\{\int_0^1 F^2\left(\frac{x-t}{h}\right)f(t)dt+\sum_{j=1}^{s+1}\left(\frac{k_j}{w_j}\right)^2\left[\int_0^{aw_j}F^2\left(\frac{x+t/w_j}{h}\right)f(t)dt\right.\right. \\ &\quad \left.\left.+\int_{1-aw_j}^1 F^2\left(\frac{x-1+(t-1)/w_j}{h}\right)f(t)dt\right]\right\}. \end{aligned}$$

Define

$$g^*(x)=\begin{cases}\sum_{i=1}^{s+1}\left(\frac{k_i}{w_i}\right)^2 f(-w_i x), & -a<x<0\\ f(x), & x\in(0,1)\\ \sum_{i=1}^{s+1}\left(\frac{k_i}{w_i}\right)^2 f(1-w_i(x-1)), & 1<x<1+a\end{cases} \tag{3.5}$$

Repeating the changes of variables that led from (4.2) to (5.1) and recalling that supp$K\subset(-a/h,a/h)$ we obtain

$$Eu_1^2=\frac{1}{h^{2m+1}}\int_{(x-1-a)/h}^{(x+a)/h}F^2(u)\,g(x-hu)du=\frac{1}{h^{2m+1}}\int_{\mathbb{R}}F^2(u)\,g(x-hu)du.$$

Overall, using Theorem 3.1.1)

$$\begin{aligned} V\left(\hat f^{(m)}(x)\right) &= \frac{1}{n}\left[\frac{1}{h^{2m+1}}\int_{\mathbb{R}}F^2(u)\,g(x-hu)du-\left(\frac{1}{h^m}\int_{\mathbb{R}}K^{(m)}(u)\,g(x-hu)du\right)^2\right] \\ &= \frac{1}{nh^{2m+1}}\left(f(x)\int_{\mathbb{R}}F^2(u)\,du+O(h)\right). \end{aligned}$$

The proof for M_k is the same. □

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. 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 $(-\infty,0)$. 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 $f_+^{(m)}(x)$, $x > 0$, is defined by

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

Theorem 4.1. *In Theorem 2.1 and in definition* (2.3) *let* $f = f_+$ *and* $D_f = (0, \infty)$. *If the conditions of Theorem 2.1 are satisfied for* f *and* K, *then 1)* (2.5) *and* (2.6) *are true and 2) for the variance of* $\hat{f}_+^{(m)}(x)$ *one has* (2.13).

Proof. 1) Instead of (2.7) we have

$$\begin{aligned} E\hat{f}_+^{(m)}(x) &= \frac{1}{h^{m+1}} E\left\{ I_{\{X_1>0\}} \left[K^{(m)}\left(\frac{x - X_1}{h}\right) + \sum_{j=1}^{s+1} \frac{k_j}{w_j} K^{(m)}\left(\frac{x + X_1/w_j}{h}\right) \right] \right\} \\ &= \frac{1}{h^{m+1}} \left[\int_0^\infty K^{(m)}\left(\frac{x - t}{h}\right) f_+(t)dt + \sum_{j=1}^{s+1} \frac{k_j}{w_j} \int_0^\infty K^{(m)}\left(\frac{x + t/w_j}{h}\right) f_+(t)dt \right]. \end{aligned}$$

Repeating calculations that led from (2.7) to (2.9) we get

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

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

2) Similarly, replacing in (4.3) f by f_+ and repeating the argument of Theorem 2.3 we see that

$$V\left(\hat{f}_+^{(m)}(x)\right) = \frac{1}{nh^{2m+1}} \left\{ f_+(x) \int_{\mathbb{R}} F^2(t)dt + o(1) \right\},\ x > 0,$$

where $F(t) = K^{(m)}(t)$. □

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

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

the extension of f_r to the left of c and

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

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

$$g_r(x) = \begin{cases} \phi_1(x), & c - a_1 < x < c, \\ f_r(x), & c < x < d, \\ \phi_2(x), & d < x < d + a_1. \end{cases} \tag{4.1}$$

Definition (3.2) guides us to define

$$\begin{aligned} \hat{f}^{(m)}(x) &= \frac{1}{nh^{m+1}} \Bigg\{ \sum_{c<X_i<d} K^{(m)}\left(\frac{x - X_i}{h}\right) \\ &+ \sum_{j=1}^{s+1} \frac{k_j}{w_j} \Bigg[\sum_{c<X_i<c+a_1w_j} K^{(m)}\left(\frac{x - c + (X_i - c)/w_j}{h}\right) \\ &+ \sum_{d-a_1w_j<X_i<d} K^{(m)}\left(\frac{x - d + (X_i - d)/w_j}{h}\right)\Bigg]\Bigg\},\ x \in (c,d). \end{aligned}$$

Theorem 4.2. *Let f_r be s times differentiable, $0 \le m \le s$, and let K have compact support. For h sufficiently small (such that* $\mathrm{supp}K \subseteq (-a_1/h, a_1/h)$*) we have*

$$E\hat{f}^{(m)}(x) - f_r^{(m)}(x) = \int_{\mathbb{R}} K(u)\left[g_r^{(m)}(x - hu) - g_r^{(m)}(x)\right] du,\ c < x < d. \tag{4.2}$$

Further,

$$V\left(\hat{f}^{(m)}(x)\right) = \frac{1}{nh^{2m+1}}\left\{ f(x)\int_{\mathbb{R}} F^2(t)dt + O(h)\right\},\ c < x < d,$$

where $F(t) = K^{(m)}(t)$ or $F(t) = M_k^{(m)}(t)$ depending on which kernel is used in the definition of $\hat{f}^{(m)}(x)$.

Proof. By the i.i.d. assumption

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

The obvious changes of variables are:

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

The mean value becomes

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

Applying (4.1) we see that this is the same as

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

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

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

The derivation of the expression for variance largely repeats that from Theorem 3.2. Define

$$\begin{aligned}
u_i &= \frac{1}{h^{m+1}}\left\{I_{\{c<X_i<d\}}K^{(m)}\left(\frac{x-X_i}{h}\right)\right.\\
&+\sum_{j=1}^{s+1}\frac{k_j}{w_j}\left[I_{\{c<X_i<c+a_1w_j\}}K^{(m)}\left(\frac{x-c+(X_i-c)/w_j}{h}\right)\right.\\
&\left.\left.+I_{\{d-a_1w_j<X_i<d\}}K^{(m)}\left(\frac{x-d+(X_i-d)/w_j}{h}\right)\right]\right\}.
\end{aligned}$$

Then we can use equations similar to (2.14) and (2.15).

We want to show that in the analog of (2.15) all cross-products vanish if h is small. Suppose supp$K \subset (-b, b)$ for some $b > 0$. In the cross-products we have integrands that include products of functions

$$\begin{aligned} f_0(t) &= K^{(m)}\left(\frac{x-t}{h}\right), \; f_j(t) = K^{(m)}\left(\frac{x-c+(X_i-c)/w_j}{h}\right), \\ g_j(t) &= K^{(m)}\left(\frac{x-d+(X_i-d)/w_j}{h}\right). \end{aligned}$$

It is easy to see that

$$\text{supp} f_0 \subseteq B(x, hb), \; \text{supp} f_j \subseteq B(c - w_j(x-c), w_j hb), \; \text{supp}\, g_j \subseteq B(d - w_j(x-d), w_j hb).$$

Since the radii of the balls here tend to zero, for small h they do not overlap if their centers are different. It is easy to show that all the centers are indeed different using the conditions that $c < x < d$, $w_j > 0$ for all j and all w_j are different.

Thus, for small h

$$\begin{aligned} Eu_1^2 &= \frac{1}{h^{2m+2}}\left\{\int_c^d F^2\left(\frac{x-t}{h}\right) f_r(t)dt \right. \\ &\quad + \sum_{j=1}^{s+1}\left(\frac{k_j}{w_j}\right)^2 \left[\int_c^{c+a_1 w_j} F^2\left(\frac{x-c+(t-c)/w_j}{h}\right) f_r(t)dt \right. \\ &\quad \left.\left. + \int_{d-a_1 w_j}^{d} F^2\left(\frac{x-d+(t-d)/w_j}{h}\right) f_r(t)dt\right]\right\}. \end{aligned}$$

Define

$$g_r^*(x) = \begin{cases} \sum_{i=1}^{s+1}\left(\frac{k_i}{w_i}\right)^2 f(c - w_i(x-c)), & c - a_1 < x < c \\ f(x), & x \in (c, d) \\ \sum_{i=1}^{s+1}\left(\frac{k_i}{w_i}\right)^2 f(d - w_i(x-d)), & d < x < d + a_1 \end{cases} \tag{4.3}$$

Repeating the changes of variables that led from (4.2) to (5.1) and recalling that supp$K \subset (-a_1/h, a_1/h)$ we obtain

$$Eu_1^2 = \frac{1}{h^{2m+1}}\int_{(x-d-a_1)/h}^{(x-c+a_1)/h} F^2(u)\, g_r^*(x-hu)du = \frac{1}{h^{2m+1}}\int_{\mathbb{R}} F^2(u)\, g_r^*(x-hu)du.$$

Overall, combining this with the bias estimate 4.2 we obtain

$$\begin{aligned} V\left(\hat{f}^{(m)}(x)\right) &= \frac{1}{n}\left[\frac{1}{h^{2m+1}}\int_{\mathbb{R}} F^2(u)\, g_r^*(x-hu)du - \left(\frac{1}{h^m}\int_{\mathbb{R}} K^{(m)}(u)\, g_r^*(x-hu)du\right)^2\right] \\ &= \frac{1}{nh^{2m+1}}\left(f_r(x)\int_{\mathbb{R}} F^2(u)\, du + O(h)\right). \end{aligned}$$

□

5 Estimators satisfying zero boundary conditions

For simplicity we consider only densities on $D_f = (0, \infty)$. For estimator (2.4) 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. The first estimator also reduces the variance near 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 $f^{(m)}(x)$.

Let l be an integer between m and s and let ψ be a function on D_f with properties

$$\psi(0+) = ... = \psi^{(l-m)}(0+) = 0,\ \psi^{(l-m+1)}(0+) \neq 0, \tag{5.1}$$

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

$$f^{(m)}(0+) = ... = f^{(s)}(0+) = 0, \tag{5.2}$$

put $\alpha = 1$. Otherwise, let k be the least integer such that $m \le k \le s$, $f^{(k)}(0+) \neq 0$ 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 5.1. *Let the estimator $\hat{f}^{(m)}(x)$ of $f^{(m)}(x)$ satisfy* (2.11)*. Then $\tilde{f}^{(m)}(x)$ satisfies*

$$\tilde{f}^{(m)}(0+) = ... = \tilde{f}^{(l)}(0+) = 0, \tag{5.3}$$

$$E\tilde{f}^{(m)}(x) - f^{(m)}(x) = O(h^{s-m}) \text{ for all } x \in D_f. \tag{5.4}$$

and

$$V\left(\tilde{f}^{(m)}(x)\right) = \begin{cases} V\left(\hat{f}^{(m)}(x)\right), & x \ge h^{\alpha} \\ V\left(\hat{f}^{(m)}(x)\right)(xh^{-\alpha})^{2(l-m+1)}, & x < h^{\alpha}. \end{cases} \tag{5.5}$$

Proof. (5.1) implies

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

for $j = 0, ..., l-m$, so (5.3) is satisfied.

To prove (5.4), consider two cases.

Case $xh^{-\alpha} \ge 1$. (5.4) follows trivially from (2.11) 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 (5.2) is true, so that $\alpha = 1$. Then

$$\begin{aligned} f^{(m)}(x) &= f^{(m)}(0+) + ... + f^{(s)}(0+)\frac{x^{s-m}}{(s-m)!} + o(x^{s-m}) \\ &= o\left((h^{\alpha})^{s-m}\right) = o(h^{s-m}). \end{aligned} \tag{5.6}$$

Suppose (5.2) is wrong. Then $\alpha = \frac{s-m}{k-m}$ and

$$\begin{aligned} f^{(m)}(x) &= f^{(m)}(0+) + ... + f^{(s)}(0+)\frac{x^{k-m}}{(k-m)!} + o(x^{k-m}) \\ &= O\left(x^{k-m}\right) = O(h^{s-m}). \end{aligned} \tag{5.7}$$

(5.6) and (5.7) prove what we need.

To prove (5.5), consider two cases.

Case $xh^{-\alpha} \ge 1$. The first part of (5.5) is obvious because $\tilde{f}^{(m)}(x) = \hat{f}^{(m)}(x)$.

Case $xh^{-\alpha} \le 1$. From (5.1) it follows that

$$\psi(xh^{-\alpha}) = \psi(0+) + ... + \psi^{(l-m+1)}(0+)\frac{(xh^{-\alpha})^{(l-m+1)}}{(l-m+1)!} \sim \left(xh^{-\alpha}\right)^{(l-m+1)}$$

which proves the second part of (5.5). □

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

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

Theorem 5.2. *All derivatives of* $\hat{f}^{(m)}(x)$ *that exist vanish at zero and*

$$E\hat{f}^{(m)}(x) - f^{(m)}(x) = \int_{\mathbb{R}} K(t)\left[g_x^{(m)}(x - ht) - g_x^{(m)}(x)\right] dt, \ x \in D_f, \tag{5.8}$$

where g_x *is the Hestenes extension of* $f_x(t) = f(t)\psi(t/x)$. *Besides,* $V\left(\hat{f}^{(m)}(x)\right)$ *satisfies* (2.13).

Proof. For almost all samples $\min_i X_i > 0$ and for $0 < x < \frac{1}{3}\min_i X_i$ one has $\psi(X_i/x) = 0$, $i = 1, ..., n$. Hence $\hat{f}^{(m)}(x)$ vanishes, together with all its derivatives, in the neighborhood of zero for almost all samples. Following (2.7) we see that the mean is

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

Here the function $f_x(t) = f(t)\psi(t/x)$ has support $\mathrm{supp} f_x \subseteq [0,3]$. Implementing changes applied after (2.8), including integration by parts, we obtain an analog of (2.9) with g_x instead of g. g_x is obtained by replacing f in (2.2)-(2.3) by f_x. The rest is familiar.

Repeating the calculations from the proof of Theorem 2.3 with f_x in place of f and using $f_x(x) = f(x)$ we obtain (2.13). □

6 Simulations

We conducted a series of simulations to provide some evidence on 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 first case we considered 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 $\alpha = \beta = 1$,
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 $\hat{f}_k$ for $k = 1, 2, 3$, $w_i = i, i^{-1}$ and $s = 1, 2$. In each case we used a Gaussian kernel and a Gaussian seed kernel, as necessary. We also calculated the Gamma kernel estimator of Chen (2000), which we denote by $\hat{f}_C$. For each estimator we selected an optimal bandwidth by minimizing the integrated squared error over a fixed grid on the interval $(0, 4)$ with step 10^{-3}. Figure 1 gives a set of estimates for one of the generated samples of size $n = 250$ associated with the Gamma density. For each sample, and each estimator, a root average squared error (RASE) over the grid is calculated. The average of RASE across all 1000 generated samples are reported on Table 1.

As expected, the 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 RASE when $w_i = i$. Also, for estimators $\hat{f}_2$ and $\hat{f}_3$, choosing $s = 1$ reduces RASE (compared with $s = 2$) for all densities, except f_G and f_E when $n = 500$. Except for the case of the truncated normal density - f_{TN} - there is always a Hestenes' based estimator that outperforms the estimator $\hat{f}_S$ proposed by Schuster (1985). In fact, if we choose $w_i = i$, all Hestenes' based estimators have smaller RASE than that of $\hat{f}_S$. For all densities, there is always a Hestenes' based estimator that outperforms the estimator $\hat{f}_C$ proposed by Chen (2000). In addition, as in the comparison with $\hat{f}_S$, if we choose $w_i = i$, all Hestenes' based estimators have smaller RASE than that of $\hat{f}_C$ for all densities. Lastly, as expected, the traditional Rosenblatt-Parzen estimator has the poorest performance across all densities and all estimators. The choice of kernel, or seed-kernel, does not qualitatively impact the relative performance described above. This preliminary experimental evidence seems to support the use of $w_i = i$ and the choice of $s = 1$ relative to $s = 2$. Results for $s = 3$ and $k > 3$ (not reported) suggest rapid deterioration of performance. The evidence also suggests that Hestenes-based estimators outperform the well known estimators proposed by Schuster (1985) and Chen (2000).

In the second broad case, we consider a density that has a point of discontinuity at $x = 0$. Specifically,

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) & \text{if } x < 0 \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) & \text{if } x \geq 0 \end{cases} \tag{6.1}$$

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 $J_f(0) = f(0-) - f(0+) = \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\sigma} - 1\right)$. The left and right panels of Figure 2 provide graphs of this density for $\sigma^2 = 0.5$ and $\sigma^2 = 0.25$.

With knowledge of the point of discontinuity ($x = 0$) we evaluate two estimators for f. The first, $\hat{f}_H$, based on Hestenes' extension, is composed of the estimator

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

for the part f_+ of f defined on $[0, \infty)$ and $\hat{f}_-$ is the analog estimator for the f_- part of f defined on $(-\infty, 0]$. The jump $J_f(0)$ is estimated by $J_{\hat{f}}(0) = \hat{f}_-(0-) - \hat{f}_+(0+)$. The second estimator is composed of a Rosenblatt-Parzen estimator

$$\hat{f}_{R+}(x) = \frac{1}{nh} \sum_{X_i > 0} K\left(\frac{x - X_i}{h}\right)$$

for f_+ and $\hat{f}_{R-}$ is the analog estimator for f_-. $J_f(0)$ is estimated by $J_{\hat{f}_R}(0) = \hat{f}_{R-}(0-) - \hat{f}_{R+}(0+)$.

For each density ($\sigma^2 = 0.25, 0.5$) we generated samples of size $n = 250, 500$ and calculated the Rosenblatt-Parzen and Hestenes estimator for $w_i = i, i^{-1}, i/(s+1)$ and $s = 1, 2$ using a Gaussian kernel. For each estimator we selected an optimal bandwidth by minimizing the integrated squared error over a fixed grid on the interval $(-2.5, 3)$ with step 5×10^{-3} for the case where $\sigma^2 = 0.5$ and $(-1.5, 3)$ with step 5×10^{-3} for the case where $\sigma^2 = 0.25$. For each sample, and each estimator, a root average squared error (RASE) over the grid is calculated. The average of RASE across all 1000 generated samples are reported on Table 2 and 3. Average estimated jumps (across all generated samples) for both estimators, as well as their deviation from the true jump, are also provide for both estimators.

We observe the following general regularities. First, all estimators have smaller average RASE when $n = 500$ compared to $n = 250$ for both $\sigma^2 = 0.5$ and $\sigma^2 = 0.25$. All versions of $\hat{f}_H$ have smaller average RASE than $\hat{f}_R$ and $\hat{f}_H$ performs better when $s = 1$ and $w_i = i$ for both $\sigma^2 = 0.5$ and $\sigma^2 = 0.25$. For each choice of s, the estimator $\hat{f}_H$ performs similarly when $w_i = i^{-1}$ or $w_i = i/(s+1)$.

The estimated jump is much closer to the true jump under $\hat{f}_H$ than under $\hat{f}_R$. $J_{\hat{f}_R}$ severely underestimates the true jump for both $\sigma^2 = 0.5$ and $\sigma^2 = 0.25$. $J_{\hat{f}_H}$, calculated under all versions of the estimators, is much closer to the true value of the jump, but for both $\sigma^2 = 0.5$ and $\sigma^2 = 0.25$ there seems to be some evidence of a slight overestimation of the jump. Except for the case where $s = 1$ and $w_i = i$, jump estimators become more accurate when $n = 500$ compared to $n = 250$. However, results are mixed regarding the impact of s and w_i on the accuracy of jump estimates. For an arbitrary sample of size $n = 500$ and $\sigma^2 = 0.25$, Figure 3 provides a graph for $\hat{f}_R$ and $\hat{f}_H$ for $s = 2$ and $w_i = i^{-1}$.

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 $w_1, \cdots, w_{s+1}$ under a suitably defined criterion. If possible, this would produce a *best* estimator in the class we have defined.

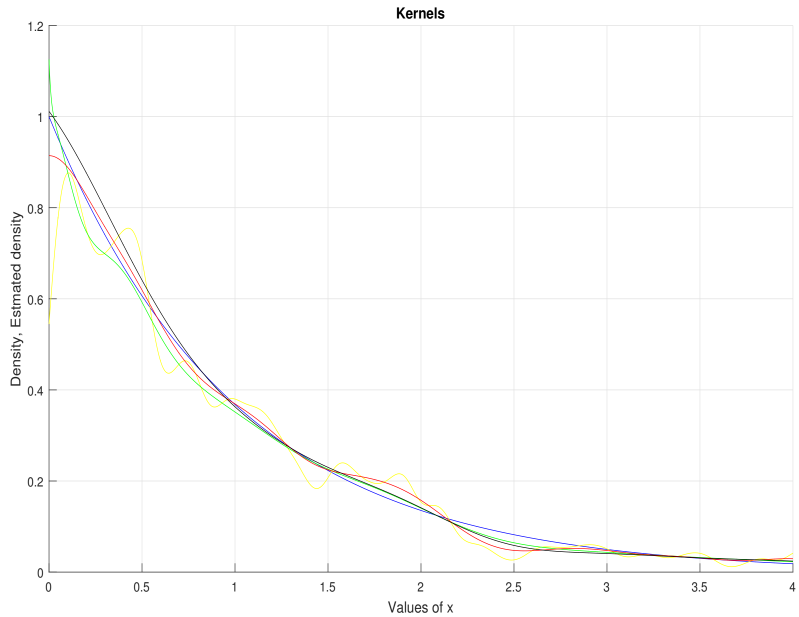

Figure 1: Density $f_G(x)$ (blue), $\hat{f}_R$ (yellow), $\hat{f}_C$ (green), $\hat{f}_S$ (red) and $\hat{f}_k$ (black) for $k = 2$, $s = 1$, $w_i = i$.

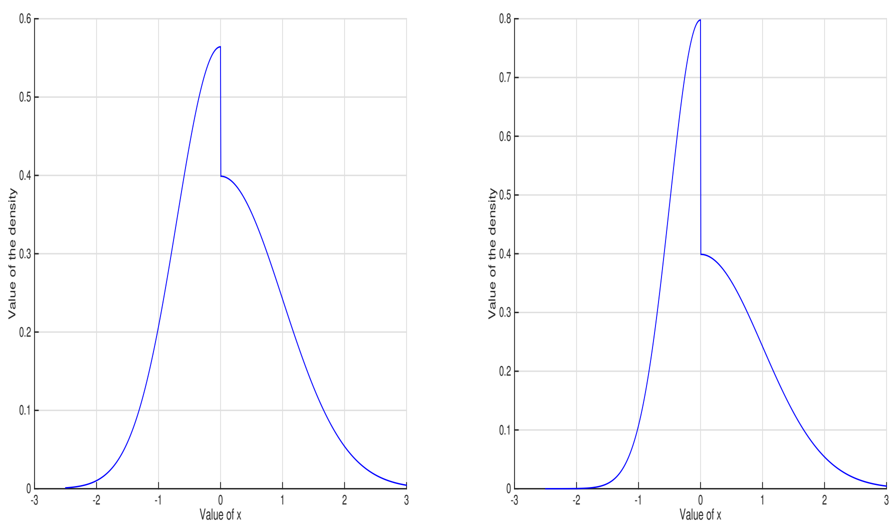

Figure 2: Density $f(x)$ with a discontinuity at $x = 0$. Left panel is for $\sigma^2 = 0.5$, right panel is for $\sigma^2 = 0.25$

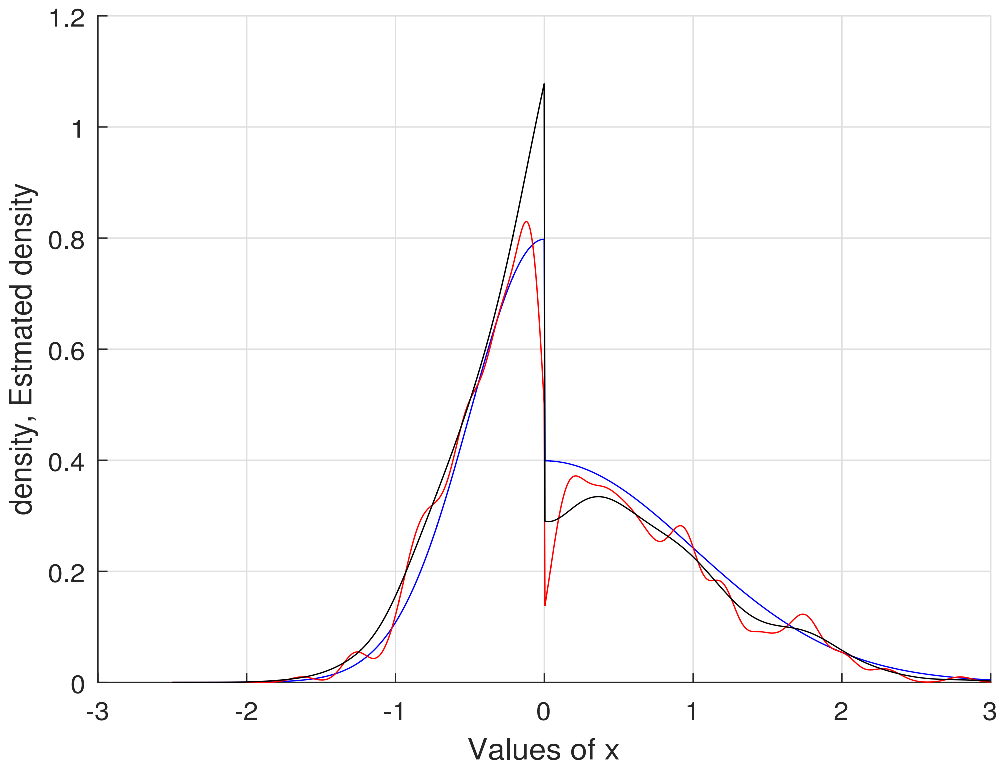

Figure 3: True density $f(x)$ (blue), $\hat{f}_R$ (red) and $\hat{f}_H$ for $s = 2$ and $w_i = i^{-1}$ (black) for $\sigma^2 = 0.25$

Table 1. Average RASE$\times 10^2$

$\hat{f}_R$ and $\hat{f}$ constructed using $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

		$\hat{f}_R$	$\hat{f}_C$	$\hat{f}_S$	$\hat{f}$				$\hat{f}_2$				$\hat{f}_3$			
s					1		2		1		2		1		2	
w_i					i^{-1}	i	i^{-1}	i	i^{-1}	i	i^{-1}	i	i^{-1}	i	i^{-1}	i
$n = 250$	f_{TN}	6.140	3.363	**2.291**	2.948	2.903	4.773	2.814	2.959	2.470	5.637	3.034	3.017	2.554	6.124	3.183
	f_G	7.161	3.517	3.390	3.311	3.229	5.034	3.123	3.097	**2.888**	5.486	2.960	3.052	2.889	5.940	2.892
	f_χ	1.264	1.321	1.466	1.105	1.189	1.304	1.207	0.980	1.118	1.247	1.229	**0.963**	1.128	1.411	1.223
	f_E	7.316	3.429	3.322	3.194	3.080	5.016	3.024	3.001	**2.776**	5.515	2.870	2.987	2.786	6.061	2.811
$n = 500$	f_{TN}	5.281	2.670	**1.820**	2.247	2.210	3.628	2.145	2.229	1.919	4.128	2.320	2.275	1.961	4.458	2.420
	f_G	6.091	2.706	2.661	2.468	2.415	3.712	2.374	2.290	2.205	3.851	2.197	2.262	2.203	4.204	**2.126**
	f_χ	1.002	1.023	1.149	0.869	0.911	0.999	0.945	0.745	0.856	0.906	0.940	**0.725**	0.863	1.012	0.931
	f_E	6.369	2.745	2.693	2.468	2.421	3.637	2.359	2.270	2.182	3.895	2.176	2.248	2.179	4.299	**2.110**

Table 2. Average RASE and Jump Estimate

$\hat{f}_R$ and $\hat{f}_H$ for $\sigma^2 = 0.5$

Experiment (s,w_i,n)	Average Squared Error		Jump Estimate		True Jump: $J = 0.1652$	
	$\hat{f}_R$	$\hat{f}_H$	$J_{\hat{f}_R}$	$J_{\hat{f}_H}$	$J_{\hat{f}_R} - J$	$J_{\hat{f}_H} - J$
(s=1,$w_i = i^{-1}$,250)	0.0500	0.0316	0.0782	0.1820	-0.0870	0.0168
(s=2,$w_i = i^{-1}$,250)	0.0508	0.0566	0.0773	0.1714	-0.0879	0.0061
(s=1,$w_i = i$,250)	0.0509	0.0273	0.0718	0.1679	-0.0935	0.0026
(s=2,$w_i = i$,250)	0.0508	0.0294	0.0777	0.2094	-0.0875	0.0441
(s=1,$w_i = i/(s+1)$,250)	0.0505	0.0314	0.0807	0.1869	-0.0845	0.0217
(s=2,$w_i = i/(s+1)$,250)	0.0508	0.0510	0.0808	0.1735	-0.0845	0.0082
(s=1,$w_i = i^{-1}$,500)	0.0431	0.0236	0.0775	0.1757	-0.0877	0.0104
(s=2,$w_i = i^{-1}$,500)	0.0432	0.0398	0.0793	0.1637	-0.0860	-0.0016
(s=1,$w_i = i$,500)	0.0435	0.0211	0.0781	0.1817	-0.0871	0.0164
(s=2,$w_i = i$,500)	0.0430	0.0230	0.0765	0.1977	-0.0888	0.0324
(s=1,$w_i = i/(s+1)$,500)	0.0428	0.0236	0.0780	0.1785	-0.0873	0.0132
(s=2,$w_i = i/(s+1)$,500)	0.0432	0.0371	0.0777	0.1570	-0.0875	-0.0083

Table 3. Average RASE and Jump Estimate

$\hat{f}_R$ and $\hat{f}_H$ for $\sigma^2 = 0.25$

Experiment (s,w_i,n)	Average Squared Error		Jump Estimate		True Jump: $J = 0.3989$	
	$\hat{f}_R$	$\hat{f}_H$	$J_{\hat{f}_R}$	$J_{\hat{f}_H}$	$J_{\hat{f}_R} - J$	$J_{\hat{f}_H} - J$
(s=1,$w_i = i^{-1}$,250)	0.0578	0.0356	0.1906	0.4322	-0.2083	0.0332
(s=2,$w_i = i^{-1}$,250)	0.0579	0.0625	0.1923	0.4097	-0.2067	0.0107
(s=1,$w_i = i$,250)	0.0581	0.0316	0.1882	0.4261	-0.2107	0.0271
(s=2,$w_i = i$,250)	0.0581	0.0341	0.1901	0.4863	-0.2089	0.0873
(s=1,$w_i = i/(s+1)$,250)	0.0579	0.0355	0.1919	0.4318	-0.2071	0.0329
(s=2,$w_i = i/(s+1)$,250)	0.0582	0.0561	0.1939	0.4282	-0.2050	0.0293
(s=1,$w_i = i^{-1}$,500)	0.0497	0.0276	0.1937	0.4214	-0.2053	0.0224
(s=2,$w_i = i^{-1}$,500)	0.0493	0.0459	0.1957	0.4002	-0.2033	0.0013
(s=1,$w_i = i$,500)	0.0488	0.0243	0.1968	0.4403	-0.2022	0.0413
(s=2,$w_i = i$,500)	0.0493	0.0266	0.1943	0.4742	-0.2047	0.0753
(s=1,$w_i = i/(s+1)$,500)	0.0496	0.0278	0.1957	0.4246	-0.2033	0.0257
(s=2,$w_i = i/(s+1)$,500)	0.0493	0.0420	0.1900	0.4025	-0.2089	0.0035

References

Burenkov, V. I., 1998. Sobolev spaces on domains. B. G. Teubner, Leipzig.

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

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.

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 and Data Analysis 72, 57–76.

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

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 od densities on the unit interval. Journal of the American Statistical Association 110, 773–783.