On Boundary Correction in Kernel Density Estimation *

R.J. Karunamuni[1] and T. Alberts[2]

[1]Department of Mathematical and Statistical Sciences
University of Alberta, Edmonton, Alberta, Canada, T6G 2G1

[2]Courant Institute of Mathematical Sciences
New York University, New York, NY, USA, 10009

Abstract

It is well known now that kernel density estimators are not consistent when estimating a density near the finite end points of the support of the density to be estimated. This is due to boundary effects that occur in nonparametric curve estimation problems. A number of proposals have been made in the kernel density estimation context with some success. As of yet there appears to be no single dominating solution that corrects the boundary problem for all shapes of densities. In this paper, we propose a new general method of boundary correction for univariate kernel density estimation. The proposed method generates a class of boundary corrected estimators. They all possess desirable properties such as local adaptivity and non-negativity. In simulation, it is observed that the proposed method perform quite well when compared with other existing methods available in the literature for most shapes of densities, showing a very important robustness property of the method. The theory behind the new approach and the bias and variance of the proposed estimators are given. Results of a data analysis are also given.

Keywords: Density estimation; Mean Squared Error; Kernel estimation; Reflection.

AMS Subject Classification: Primary: 62G05; Secondary: 42C10

*This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. This paper was as an invited presentation at the Fifth Biennial IISA International Conference on Statistics, Probability and Related Areas held at the University of Georgia, Athens, Georgia, from May 14-16, 2004.

1 Introduction

Let f denote a probability density function with support $[0, a], 0 < a \leq \infty$, and consider nonparametric estimation of f based on a random sample $X_1, \ldots, X_n$ from f. Then the traditional kernel estimator of f is given by

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1.1}$$

where K is some chosen unimodal density function, symmetric about zero, and h is the bandwidth ($h \to 0$ as $n \to \infty$). The basic properties of f_n at interior points are well-known, and under some smoothness assumptions these include, for $h \leq x \leq a - h$,

$$\mathrm{E}\, f_n(x) = f(x) + \frac{h^2}{2} f^{(2)}(x) \int t^2 K(t) dt + o(h^2),$$
$$\mathrm{Var}\, f_n(x) = (nh)^{-1} f(x) \int K^2(t) dt + o\left(\frac{1}{nh}\right),$$

see, e.g., Silverman (1986, Ch. 3) or Wand and Jones (1995, Ch. 2). The performance of f_n at boundary points, i.e. for $x \in [0, h) \cup (a - h, a]$, however, differs from the interior points due to so-called "boundary effects" that occur in nonparametric curve estimation problems. More specifically, the bias of $f_n(x)$ is of order $O(h)$ instead of $O(h^2)$ at boundary points. To remove those boundary effects in kernel density estimation, a variety of methods have been developed in the literature. Some well-known methods are summarized below:

(i) The reflection method (Cline and Hart, 1991; Schuster, 1985; Silverman, 1986).

(ii) The boundary kernel method (Gasser and Müller, 1979; Gasser, Müller and Mammitzsch, 1985; Jones, 1993; Müller, 1991; Zhang and Karunamuni, 2000).

(iii) The transformation method (Marron and Ruppert, 1994; Wand, Marron and Ruppert, 1991).

(iv) The pseudo-data method (Cowling and Hall, 1996).

(v) The local linear method (Cheng et al., 1997; Cheng, 1997; Zhang and Karunamuni, 1998).

(vi) Other methods (Zhang et al., 1999; Hall and Park, 2002).

The reflection method is specially designed for the case $f^{(1)}(0) = 0$ or $f^{(1)}(a) = 0$, where $f^{(1)}$ denotes the first derivative of f. The boundary kernel method is more general than the reflection method in the sense that it can adapt to any shape of density. However, a drawback of this method is that the estimates might be negative near the endpoints; especially when $f(0) \approx 0$ or $f(a) \approx 0$. To correct this deficiency of boundary kernel methods, some remedies have been proposed; see Jones (1993), Jones and Foster (1996), Gasser and Müller (1979) and Zhang and Karunamuni (1998). The local linear method is a special case of the boundary kernel method that is thought of by some as a simple, hard-to-beat default approach, partly because of "optimal" theoretical properties (Cheng et al., 1997) in the boundary kernel (without bandwidth variation) implicit in local linear fitting. The pseudo-data method of Cowling and Hall (1996) generates some extra data $X_{(i)}$'s using what they call the "three-point-rule", and then combine them with the original data X_i's to form a kernel type estimator. Marron and Ruppert's (1994) transformation method consists of a three-step process. First, a transformation g is selected from a parametric family so that the density of $Y_i = g(X_i)$ has a first derivative that is approximately equal to 0 at the boundaries of its support. Next, a kernel estimator with reflection is applied to the Y_i's. Finally, this estimator is converted by the change-of-variables formula to obtain an estimate of f. Among the other methods, two very promising recent ones are due to Zhang et al. (1999) and Hall and Park (2002). The former method is a combination of the methods of pseudo-data, transformation and reflection; whereas the latter method is based on what they call an "empirical translation correction".

In this paper we propose to investigate a class of estimators of the form

$$\tilde{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} \left\{ K\left(\frac{x + g_1(X_i)}{h}\right) + K\left(\frac{x - g_2(X_i)}{h}\right) \right\} \tag{1.2}$$

where h is the bandwidth, K is a kernel function and g_1 and g_2 are transformations that need to be determined. Clearly, when $g_i(t) = t, i = 1, 2$, then the estimator (1.2) reduces to the standard reflection estimator. Thus (1.2) can be thought of as a "generalized" version of the standard reflection method estimator. Alternatively, one can also argue that (1.2) is a "transformed-reflection" based estimator. Furthermore, it will be clear later that the estimator of Zhang et al. (1999) is a special case of (1.2). Estimator (1.2) is non-negative as long as the kernel K is non-negative. The transformations g_1 and g_2 are chosen so that the bias of (1.2) is of order $O(h^2)$ for all $x \geq 0$. In this paper, we study the particular case $g_1 = g_2$ in detail, but other choices of (g_1, g_2) are also discussed. They are all locally adaptive in that they depend on the point of estimation. Another desirable property of the proposed estimators is that they all reduce to the traditional kernel estimator (1.1) at the interior points. Most importantly, the proposed estimator improves the bias but holds on to the low variance, whereas ordinary reflection is well-known to have a bad bias but low variance (see Jones 1993, among others). Extensive simulations are carried out to compare the proposed estimator with the existing well-known methods. It is observed that the proposed estimator performs reasonably well compared to the existing estimators for most shapes of densities.

Section 2 contains the methodology, the development of the proposed estimator and the main results of the paper. Sections 3 and 4 present simulation studies and a data analysis, respectively. Final comments are given in Section 5 and the appendix contains proofs of the main results.

2 Methodology and Main Results

2.1 Preliminaries

For convenience, we shall assume that the unknown probability density function f has support $[0, \infty)$. Also, let the kernel function K in (1.2) be a non-negative, symmetric function with support $[-1, 1]$, and satisfying

$$\int K(t)dt = 1, \int tK(t)dt = 0, \text{ and } 0 < \int t^2 K(t)dt < \infty, \tag{2.1}$$

that is, K is a kernel of order 2. We assume that the transformations $g_i, i = 1, 2$, in (1.2) are non-negative, continuous and monotonically increasing functions defined on $[0, \infty)$. Further assume that g_i^{-1} exists, $g_i(0) = 0, g_i^{(1)}(0) = 1$, and that $g_i^{(2)}$ and $g_i^{(3)}$ exist and are continuous on $[0, \infty)$, where $g_i^{(j)}$ denotes the j^{th}-derivative of g_i, with $g_i^{(0)} = g_i$ and g_i^{-1} denoting the inverse function of $g_i, i = 1, 2$. Suppose that $f^{(j)}$, the j^{th}-derivative of f, exists and is continuous on $[0, \infty)$, $j = 0, 1, 2$, with $f^{(0)} = f$. Then the bias and variance of (1.2) are given by, for $x = ch, 0 \le c \le 1$,

$$
\begin{aligned}
\mathrm{E}\,\tilde{f}_n(x) - f(x) = h \Big\{ & 2f^{(1)}(0) \int_c^1 (t-c)K(t)dt - g_1^{(2)}(0)f(0) \int_c^1 (t-c)K(t)dt \\
& - g_2^{(2)}(0)f(0) \left(c + \int_c^1 (t-c)K(t)dt \right) \Big\} \\
+ \frac{h^2}{2} \Big\{ & f^{(2)}(0) \int_{-1}^1 t^2 K(t)dt \\
& - \left[g_1^{(3)}(0)f(0) + g_1^{(2)}(0)(f^{(1)}(0) - g_1^{(2)}(0)f(0)) \right] \int_c^1 (t-c)^2 K(t)dt \\
& - \left[g_2^{(3)}(0)f(0) + g_2^{(2)}(0)(f^{(1)}(0) - g_2^{(2)}(0)f(0)) \right] \int_{-1}^c (t-c)^2 K(t)dt \Big\} \\
+ o(h^2) &
\end{aligned} \tag{2.2}
$$

and

$$
\mathrm{Var}\,\tilde{f}_n(x) = \frac{f(0)}{nh} \left\{ 2 \int_c^1 K(t)K(2c-t)dt + \int_{-1}^1 K^2(t)dt \right\} + o\left(\frac{1}{nh}\right). \tag{2.3}
$$

The proofs of (2.2) and (2.3) are given in the Appendix. Note that the contribution of g_1 on the bias vanishes as $c \to 1$. We shall choose the transformations g_1 and g_2 so that the first order term in the bias expansions (2.2) is zero. It is clear that there are various possible choices available for the pair (g_1, g_2). Here we investigate the particular choice that $g_1 = g_2$ for convenience. For other possible choices, the reader is referred to the authors' technical report, Karunamuni and Alberts (2004).

Assume that $g_1 = g_2$. Let g denote the common transformation, i.e. $g_1 = g_2 = g$. Then g must satisfy

$$
g^{(2)}(0) = 2f^{(1)}(0) \int_c^1 (t-c)K(t)dt \Big/ f(0) \left(c + 2\int_c^1 (t-c)K(t)dt \right). \tag{2.4}
$$

Note that the right-hand side (RHS) of (2.4) depends on c; that is, the transformation g depends on the point of estimation inside the boundary region $[0, h)$. In this sense, the transformation g is locally adaptive. Combining (2.4) with the other assumptions given above, g should now satisfy the following three conditions:

$$\begin{aligned}
&\text{(i) } g : [0, \infty) \to [0, \infty), g \text{ is continuous, monotonically increasing} \\
&\quad \text{and } g^{(i)} \text{exists}, i = 1, 2, 3, \\
&\text{(ii) } g^{-1}(0) = 0, g^{(1)}(0) = 1, \\
&\text{(iii) } g^{(2)}(0) = \text{RHS of (2.4)}.
\end{aligned} \tag{2.5}$$

Functions satisfying the conditions of (2.5) can be easily constructed. We employ the following transformation in our investigation. For $0 \le c \le 1$, define

$$g(y) = y + \frac{1}{2} d k_c' y^2 + \lambda_0 (d k_c')^2 y^3, \tag{2.6}$$

where

$$d = f^{(1)}(0) \,/ f(0), \tag{2.7}$$

$$k_c' = 2 \int_c^1 (t - c) K(t) dt \Big/ \left(c + 2 \int_c^1 (t - c) K(t) dt \right), \tag{2.8}$$

and λ_0 is a positive constant such that $12\lambda_0 > 1$. This condition on λ_0 is necessary for $g(y)$ of (2.6) to be an increasing function of y. The function $g(y)$ at (2.6) is constructed as a cubic polynomial (because it is easy to decide whether its derivative, a quadratic polynomial, is always positive), with the only free parameter $g^{(2)}(0)$ determined as $g^{(2)}(0) = k_c' d$. Observe that $g(y)$ of (2.6) depends on c, and in order to display this local dependence we shall denote

g by $g_c, 0 \leq c \leq 1$, in what follows. As $c \to 1, k_c' \to 0$ and thus $g_c(y) \to y$ as $c \to 1$ for each y. This means that $\tilde{f}_n$ defined by (1.2) with $g_1 = g_2 = g_c$ of (2.6) reduces to the usual kernel estimator (1.1) at interior points, i.e. for $x \geq h$, $\tilde{f}_n(x)$ coincides with $f_n(x)$ of (1.1). For $g_1 = g_2 = g_c$ defined by (2.6), the bias and variance of (1.2) satisfy, as $c \to 1$, $\mathrm{E}\,\tilde{f}_n(x) - f(x) \to \frac{h^2}{2}\int_{-1}^{1} t^2 K(t)dt + o(h^2)$ and $\mathrm{Var}\,\tilde{f}_n(x) \to \frac{f(0)}{nh}\int_{-1}^{1} K^2(t)dt + o\left(\frac{1}{nh}\right)$, which are exactly the same expressions of the interior bias and variance of the traditional kernel estimator (1.1).

Remark: Suppose the transformation pair (g_1, g_2) in (1.2) is chosen so that $g_1^{(2)}(0) = 2d$ and $g_2^{(2)}(0) = 0$, where d is given by (2.7). Then g_1 and g_2 should satisfy the conditions (i) and (ii) of (2.5), and $g_1^{(2)}(0)$ and $g_2^{(2)}(0)$ must satisfy the above conditions. Two such transformations are

$$g_1(y) = y + dy^2 + \lambda_1 d^2 y^3,$$
$$g_2(y) = y + \lambda_2 k_c |d| y^3$$

where $k_c = 2k_c'$ with k_c' as in (2.8) and λ_1 and λ_2 are non-negative constants such that $3\lambda_1 \geq 1$ and $\lambda_2 \geq 0$. When $\lambda_2 = 0$, then the resulting estimator has the form

$$\frac{1}{nh}\sum_{i=1}^{n}\left\{K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + g_1(X_i)}{h}\right)\right\},$$

which is exactly the same estimator investigated in Zhang et al. (1999). Thus, their estimator is a very special case of (1.2). However, the estimator of Zhang et al. (1999) is not locally adaptive, whereas $\tilde{f}_n(x)$ is locally adaptive with the above pair (g_1, g_2).

2.2 Estimation of d

In order to implement the transformation g_c given by (2.6), one must replace $d = f^{(1)}(0)/f(0)$ with a pilot estimate, which requires the use of another density estimation method such as semiparametric, kernel or nearest neighbour method. In this paper, we employ the kernel method that was used in Zhang et al. (1999). Their method is easy to implement and is based on a simple idea that d can be written as the derivative of $\log f(x)$ evaluated at $x = 0$. We define an estimator of d, slightly modified from theirs, as

$$\hat{d} = (\log f_n^*(h_1) - \log f_n^*(0))/h_1 \tag{2.9}$$

with $h_1 = o(h)$, h as defined in (1.2). Further

$$f_n^*(h_1) = \frac{1}{nh_1}\sum_{i=1}^{n} K\left(\frac{h_1 - X_i}{h_1}\right) + \frac{1}{n^2} \tag{2.10}$$

and

$$f_n^*(0) = \max\left\{\frac{1}{nh_0}\sum_{i=1}^{n} K_{(0)}\left(\frac{-X_i}{h_0}\right), \frac{1}{n^2}\right\}, \tag{2.11}$$

with K being a normal, symmetric kernel function satisfying (2.1), $K_{(0)}$ is an endpoint order two kernel satisfying the conditions:

$$\int_{-1}^{0} K_{(0)}(t)dt = 1, \int_{-1}^{0} tK_{(0)}(t)dt = 0, \text{ and } 0 < \int_{-1}^{0} t^2 K_{(0)}(t)dt < \infty,$$

and $h_0 = b(0)h_1$, with $b(0)$ given by

$$b(0) = \left\{ \frac{\left(\int_{-1}^{1} t^2 K(t)dt\right)^2 \left(\int_{-1}^{0} K_{(0)}^2(t)dt\right)}{\left(\int_{-1}^{0} t^2 K_{(0)}(t)dt\right)^2 \left(\int_{-1}^{1} K^2(t)dt\right)} \right\}^{1/5} . \tag{2.12}$$

The factor $1/n^2$ in (2.10) and (2.11) is used to keep $f_n^*(h_1)$ and $f_n^*(0)$ bounded away from 0, and it does not affect the asymptotic statistical properties of $f_n^*(h_1)$ and $f_n^*(0)$. In fact, using an argument as in Lemma A.1 of Zhang et al. (1999) it is easy to show that

$$\mathrm{E}\left[|f_n^*(x) - f(x)|^3 \,|X_k = x_k, X_l = x_l\right] = O(h_1^6), \tag{2.13}$$

for any $x_k \geq 0, x_l \geq 0,\ 1 \leq k, l \leq n$ and $x = 0, h_1$, provided $h_1 = O(n^{-1/4})$ and $f^{(2)}$ is continuous near 0.

2.3 The Proposed Density Estimator

Our proposed estimator of $f(x)$ is defined as, for $x = ch, c \geq 0$,

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} \left\{ K\left(\frac{x + \hat{g}_c(X_i)}{h}\right) + K\left(\frac{x - \hat{g}_c(X_i)}{h}\right) \right\}, \tag{2.14}$$

where $\hat{g}_c(y)$ is given by (2.6) with d replaced by $\hat{d}$ of (2.9), and K and h are as in (1.2).

For $x \geq h$, $\hat{f}_n(x)$ reduce to the traditional kernel estimator $f_n(x)$ given by (1.1). Thus $\hat{f}_n$ is a natural boundary continuation of the usual kernel estimator (1.1). An important feature of the estimator (2.14) is that it is locally adaptive. Another desirable property is that it is non-negative, provided the kernel function K is non-negative - a property shared with other reflection estimators (see Jones and Foster, 1996) and transformation based estimators (see Wand et al., 1991; Marron and Ruppert, 1995; Zhang et al., 1999; and Hall and Park, 2002). Furthermore, it can be shown that $\hat{f}_n$ is itself a density function asymptotically; that is, it integrates to 1 as $n \to \infty$. The asymptotic bias and variance of $\hat{f}_n(x)$ are given in the following theorem, which is the main result of this paper.

Theorem 2.1. *Let $\hat{f}_n(x)$ be defined by* (2.14) *with a kernel function K satisfying* (2.1) *and a bandwidth $h = O\left(n^{-1/5}\right)$. Suppose h_1 in* (2.9) *is of the form $h_1 = o(h)$. Further, assume that $K^{(1)}$ exists and is bounded on $[-1, 1]$ and that $\int |K^{(1)}(t)|dt < \infty$. Assume that $f(0) > 0$ and that $f^{(2)}(0)$ exists and is continuous in a neighbourhood of* 0*. Then for $x = ch, 0 \le c \le 1$, we have*

$$\begin{aligned} \mathrm{E}\,\hat{f}_n(x) - f(x) = \frac{h^2}{2} \Big\{ & f^{(2)}(0) \int_{-1}^{1} t^2 K(t)dt \\ & - \left[g_c^{(3)}(0)f(0) + g_c^{(2)}(0)(f^{(1)}(0) - g_c^{(2)}(0)f(0))\right] \left(\int_{-1}^{1} t^2 K(t)dt + c^2 \right) \Big\} \\ & + o(h^2) \end{aligned} \tag{2.15}$$

and

$$\mathrm{Var}\,\hat{f}_n(x) = \frac{f(0)}{nh} \left\{ 2 \int_{c}^{1} K(t)K(2c - t)dt + \int_{-1}^{1} K^2(t)dt \right\} + o\left(\frac{1}{nh}\right), \tag{2.16}$$

where g_c is defined by (2.6).

3 Simulations and Discussion

To test the effectiveness of our estimator, we simulated its performance against other well-known methods. These included a boundary kernel and its close counterpart the local linear fitting method, the transformation and reflection based method given by Zhang et al. (1999), Jones & Foster's (1996) nonnegative adaptation estimator, and an estimator due to Hall & Park (2002) based on a transformation of the data "inside" the kernel.

In our simulation when a kernel of order $(0,2)$ was required we used the Epanechnikov kernel $K(t) = \frac{3}{4}(1-t^2)I_{[-1,1]}$, where I_A is the indicator function on the set A. It has been observed in Silverman (1986) that this kernel possesses the maximum efficiency, in the sense that it produces the minimal MISE given all else held equal.

In all simulations a sample size of $n = 200$ was used. The bandwidth chosen was the optimal global bandwidth of the regular kernel estimator (1.1), given by

$$h = \left\{ \frac{\int_{-1}^{1} K(t)^2 dt}{[\int_{-1}^{1} t^2 K(t) dt]^2 \int [f^{(2)}(x)]^2 dx} \right\}^{1/5} n^{-1/5} \tag{3.1}$$

as shown in Silverman (1986). The main reason for this choice is that it provides a fair basis for comparison among the different estimators without regards to bandwidth effects.

For estimation of d we chose $h_1 = hn^{-1/100}$, which is faster than h but not tremendously so. We also used the order two endpoint kernel

$$K_{(0)}(t) = 12(1+t)(t+1/2)I_{[-1,0]}. \tag{3.2}$$

Here $b(0) = 2$ and so $h_0 = 2h_1$.

Our first comparison estimator is the boundary kernel with bandwidth variation (BVF). It is defined (see Müller, 1991, and Zhang and Karunamuni, 1998) as

$$\hat{f}_B(x) = \frac{1}{nh_c} \sum_{i=1}^{n} K_{(c/b(c))} \left(\frac{x - X_i}{h_c} \right) \tag{3.3}$$

with c $= \min(x/h, 1)$. On the boundary, the bandwidth variation function $h_c = b(c)h$ is employed, here $b(c) = 2 - c$. We used the boundary kernel

$$K_{(c)}(t) = \frac{12}{(1+c)^4}(1+t)\left\{(1-2c)t + \frac{3c^2 - 2c + 1}{2}\right\} I_{[-1,c]}. \tag{3.4}$$

Note $K_{(1)}(t) = K(t)$, the Epanechnikov kernel as defined above. Moreover, Zhang and Karunamuni (1998) have shown that this kernel is optimal in the sense of minimizing the MSE in the class of all kernels of order (0,2) with exactly one change of sign in their support. The downside to the boundary kernel is that it is not necessarily non-negative, as will be seen on densities where $f(0) = 0$.

A simple modification of the boundary kernel gives us the local linear fitting method (LL). We use the kernel

$$K_{(c)}(t) = \frac{12(1-t^2)}{(1+c)^4(3c^2 - 18c + 19)}\{8 - 16c + 24c^2 - 12c^3 + t(15 - 30c + 15c^2)\}I_{[-1,c]} \tag{3.5}$$

in (3.3) and call the resulting estimator $\hat{f}_{LL}(x)$. In this case the bandwidth variation function is $b(c) = 1.86174 - .86174c$. Unfortunately, the LL method also suffers from the undesirable negativity property.

The method of Zhang et al. (1999) (hereafter Z,K&J) applies a transformation to the data and then reflects it across the left endpoint of the support of the density. The resultant kernel estimator is of the form

$$\hat{f}_{ZKJ}(x) = \frac{1}{nh}\sum_{i=1}^{n}\left\{K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + g_n(X_i)}{h}\right)\right\}, \tag{3.6}$$

where

$$g_n(y) = y + d_n y^2 + Ad^2 y^3.$$

This estimator is the motivation behind ours, but as we previously noted it is now a special case of our estimator. The main difference is that our g is dependent on the point of estimation, whereas this is not. The only requirement on A is that $3A > 1$, and in practice we used the recommended value of $A = .55$. However, d_n is again an estimate of $d = f'(0)/f(0)$, and the methodology used is the same as that given in Subsection 2.2.

The Jones and Foster (1996) (hereafter J&F) nonnegative adaptation estimator is defined as

$$\hat{f}_{JF}(x) = \bar{f}(x) \exp\left\{ \frac{\hat{f}(x)}{\bar{f}(x)} - 1 \right\}, \tag{3.7}$$

where

$$\bar{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \Big/ \int_{-1}^{c} K(t)dt. \tag{3.8}$$

Here $\bar{f}(x)$ is referred to as the cut-and-normalized kernel. This version of the J&F estimator takes $\hat{f}$ to be some boundary kernel estimate of f, and here we used $\hat{f}(x) = \hat{f}_B(x)$ as defined by (3.3) with (3.4).

The recent paper of Hall & Park (2002) (hereafter H&P) investigates the following estimator. It is defined as

$$\hat{f}_{HP}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i + \hat{\alpha}(x)}{h}\right) \Big/ \int_{-1}^{c} K(t)dt, \tag{3.9}$$

where $\hat{\alpha}(x)$ is a correction term given by

$$\hat{\alpha}(x) = h^2 \frac{\tilde{f}'(x)}{\bar{f}(x)} \rho\left(\frac{x}{h}\right)$$

with $\tilde{f}'(x)$ an estimate of $f^{(1)}(x)$, $\bar{f}(x)$ the cut-and-normalized kernel given by (3.8), and $\rho(u) = K(u)^{-1}\int_{v\le u} vK(v)dv$. To estimate $f^{(1)}(x)$ we used the endpoint kernel of order (1,2) (see Zhang and Karunamuni 1998) given by

$$K_{1,c}(t) = \frac{12(2c(1+t) - 3t^2 - 4t - 1)}{(c+1)^4} I_{[-1,c]}$$

and the corresponding kernel estimator

$$\tilde{f}'(x) = \frac{1}{nh_c^2}\sum_{i=1}^{n} K_{1,c/b(c)}\left(\frac{x - X_i}{h_c}\right),$$

where $h_c = b(c)h$ with the bandwidth variation function $b(c) = 2^{1/5}(1-c) + c$ for $0 \le c \le 1$. The idea behind Hall and Park's estimator is somewhat similar to ours in that it transforms the data inside the kernel based on the point of estimation, however it corrects the inconsistency inherent in regular kernel density estimation through normalization rather than with the addition of a second kernel term as in our estimator.

Each estimator was tested over various shapes of curves, but we have collapsed our results into four specific densities which are representative of the different possible combinations of the behaviour of f at 0. Density 1 handles the case $f(0) = 0$, whereas Densities 2, 3, & 4 illustrate what happens when $f(0) > 0$ but $f^{(1)}(0) = 0$, $f^{(1)}(0) > 0$ and $f^{(1)}(0) < 0$, respectively.

Figures 1 through 4 and Table 1 about here.

For each density we have calculated the bias of the estimators at zero (estimated value minus true value) over 1000 simulation and displayed the results in a boxplot. The variance of each estimator can be accurately gauged by the whiskers of the plot. In Table 1 we have also calculated the mean integrated squared error (MISE) over $[0, h)$. Together they give a thorough description of how each estimator performs over the boundary region.

In Density 1 both our estimator and H&P's estimate the value of $f(0)$ perfectly in that they always produce $f(0) = 0$. This is why the boxplots for the two reduce to horizontal bars. Both fair quite poorly in terms of MISE though, which is the result of both severely underestimating the function to the right of zero. The boundary kernel and LL method show low MISEs, but from Figure 1 we see they take on negative values at zero. The J&F method appears to be the best overall, holding onto a low bias and MISE while at the same time remaining always positive.

Our estimator performs best on Density 2, where $f^{(1)}(0) = 0$. The boxplot shows that at zero our estimator has a low bias and an extremely low variance. Further these properties are apparently preserved over the rest of the boundary region as the MISE over $[0, h)$ significantly outperforms the other estimators. The boundary kernel and LL methods perform similarly at zero and over the boundary, and are followed closely in terms of MISE by the J&H and H&P methods. The former is seen to have a relatively large bias and variance, while the latter has good bias but poor variance. The Z,K&J appears to have modestly low variance but unfortunately has the largest bias, which helps to rank it last in terms of MISE.

Density 3 is a tricky one to estimate with a small hump to the right of zero that is difficult to capture. At zero our estimator has a low bias, marginally higher than that of the H&P method, but our variance is lower. This appears to help in terms of MISE where our estimator puts in a smaller value than H&P. Both, however, lag behind the other four estimators in MISE.

Our estimator has its poorest showing on the exponential shape of Density 4. From the boxplot in Figure 4 we see that our estimator significantly underestimates the value of $f(0)$. The other estimators do the same but not as poorly. Clearly our estimator does preserve the low variance as we claimed, but the large bias dominates the mean squared error and is responsible for the substantial amount of MISE seen in Table 1. The best estimator among the six, in terms of both MSE at zero and MISE, is clearly Z,K&J. It is somewhat surprising that it does so well and ours does so poorly, the opposite of Density 2, especially given that they are each special cases of (1.2). This discrepancy highlights the importance that g_1 and g_2 can play in our method of estimation, and suggests that perhaps the Z,K&J estimator and our own could be used in tandem. For practical use, for example, the expected shape of the density could be used to select the appropriate estimator among the two.

4 Data Analysis

We have put our estimator to task on some well-known datasets in order to demonstrate its usefulness in practical applications. The first is the famous suicide data found in Table 2.1 of Silverman (1986). It consists of the lengths of 86 spells of psychiatric treatment undergone by patients used as controls in a study of suicide risks. This is a classical example of a dataset on which the traditional kernel estimator fails miserably over the boundary region. We have graphed the performance of our estimator and the kernel estimator given by (1.1) in Figure 5. The bandwidth is chosen subjectively to be $h = 60$, mainly for purposes of comparison with Figure 2.9(b) of Silverman's book. From the figure alone it is remarkably clear that our estimator removes a large part of the boundary effect inherent in kernel density estimation.

Figure 5 about here.

Our second dataset consists of 68 measurements of perpendicular distances of wooden stakes from a given path in a line transect survey (see Burnham et al., 1980, p. 62). The usual assumption in such surveys is that $f^{(1)}(0) = 0$, representing that the most likely objects to be spotted are those on the path. A common model for f is the half-normal density with the variance σ estimated by its MLE. We use this as a reference density for choosing the bandwidth (see Silverman 1986), which produces $h = 7.1822$. Figure 6 shows our estimator along with that of Zhang et al. (1999). The two capture roughly the same shape although our estimator predicts lower density near zero than the Z,K&J estimator does. To the eye it would also appear that our estimator more effectively captures the $f^{(1)}(0) = 0$ hypothesis.

Figure 6 about here.

5 Final Remarks

In this paper, we have introduced a very general method of boundary correction in kernel density estimation problems. The idea of the present paper stemmed from our preliminary work in Zhang, Karunamuni and Jones (1999) on the same problem. The preceding work is considered an important contribution in this area and is now seen a special case of the method proposed here. On the other hand, the present approach can be viewed as an f-dependent generalization of the reflection method that can improve the bias but hold on to a low variance. The proposed estimator possesses a number of desirable properties, including the non-negativity of the estimator. It is clear that no single existing estimator in the literature dominates all the others for all shapes of densities. Each estimator has certain advantages and works well at certain times. The performance of the proposed estimator, however, is generally very robust with respect to various shapes of densities. The proposed method seems to have inherited the best of both transformation and reflection methods.

We believe the present approach has the potential to produce even better estimators. In this paper, we have examined only one particular estimator, based on a particular choices of transformations with $g_1 = g_2$. One may further study other different transformations and investigate their corresponding estimators. In this sense, further research is needed in this direction.

Acknowledgements: We wish to thank the Editor, the Associate Editors and the referees for very constructive and useful comments that led to a much improved presentation.

A Appendix

Proof of (2.2)**:**

For $x = ch, 0 \le c \le 1$, we have, using a Taylor expansion of order 2,

$$
\begin{aligned}
\frac{1}{h}\,\mathrm{E}\,K\left(\frac{x+g_1(X_i)}{h}\right) &= \frac{1}{h}\int_0^\infty K\left(\frac{x+g_1(y)}{h}\right) f(y)dy \\
&= \int_c^1 K(t)\frac{f(g_1^{-1}((t-c)h))}{g_1^{(1)}(g_1^{-1}((t-c)h))}dt \\
= \int_c^1 K(t) \Bigg\{ & \frac{f(g_1^{-1}(0))}{g_1^{(1)}(g_1^{-1}(0))} + (t-c)h\left[\frac{g_1^{(1)}(g_1^{-1}(0))f^{(1)}(g_1^{-1}(0)) - g_1^{(2)}(g_1^{-1}(0))f(g_1^{-1}(0))}{g_1^{(1)}(0)(g_1^{(1)}(g_1^{-1}(0)))^2}\right] \\
&+ \frac{h^2(t-c)^2}{2(g_1^{(1)}(0))^3(g_1^{(1)}(g_1^{-1}(0)))^3}\Big[g_1^{(1)}(g_1^{-1}(0))f^{(2)}(g_1^{-1}(0)) - g_1^{(3)}(g_1^{-1}(0))f(g_1^{-1}(0)) \\
&\quad -3g_1^{(2)}(g_1^{-1}(0))\Big(g_1^{(1)}(g_1^{-1}(0))f^{(1)}(g_1^{-1}(0)) - g_1^{(2)}(g_1^{-1}(0))f(g_1^{-1}(0))\Big)\Big]\Bigg\} \\
&+ o(h^2) \\
= f(0)\int_c^1 K(t)dt &+ h(f^{(1)}(0) - g_1^{(2)}(0)f(0))\int_c^1 (t-c)K(t)dt \\
&+ \frac{h^2}{2}\left\{f^{(2)}(0) - g_1^{(3)}(0)f(0) - 3g_1^{(2)}(0)(f^{(1)}(0) - g_1^{(2)}(0)f(0))\right\}\int_c^1 (t-c)^2K(t)dt \\
&+ o(h^2). \qquad \text{(A.1)}
\end{aligned}
$$

Similarly, a Taylor expansion of order 2 on the function $f(g_2^{-1}(\cdot))/g_2^{(1)}(g_2^{-1}(\cdot))$ at $t = c$ gives

$$
\begin{aligned}
\frac{1}{h}\,\mathrm{E}\,K\left(\frac{x-g_2(X_i)}{h}\right) &= \int_{-1}^c K(t)\frac{f(g_2^{-1}((c-t)h))}{g_2^{(1)}(g_2^{-1}((c-t)h))} \\
&= f(0)\int_{-1}^c K(t)dt - h(f^{(1)}(0) - f(0)g_2^{(2)}(0))\int_{-1}^c (t-c)K(t)dt \\
&\quad + \frac{h^2}{2}\left\{f^{(2)}(0) - f(0)g_2^{(3)}(0) - 3g_2^{(2)}(0)(f^{(1)}(0) - f(0)g_2^{(2)}(0))\right\} \\
&\quad \times \int_{-1}^c (t-c)^2K(t)dt + o(h^2). \qquad \text{(A.2)}
\end{aligned}
$$

Using the properties of K, we have $\int_{-1}^c tK(t)dt = -\int_c^1 K(t)dt$ and $\int_{-1}^c K(t)dt = 1 -$

$\int_c^1 K(t)dt$. Also, by the existence and continuity of $f^{(2)}(\cdot)$ near 0, we have for $x = ch$,

$$\begin{aligned} f(0) &= f(x) - chf^{(1)}(x) + \frac{(ch)^2}{2} f^{(2)}(x) + o(h^2), \\ f^{(1)}(x) &= f^{(1)}(0) + chf^{(2)}(0) + o(h), \\ f^{(2)}(x) &= f^{(2)}(0) + o(1). \end{aligned} \tag{A.3}$$

Now combining (A.1) and (A.2) and using the properties of K along with (A.3), we have for $x = ch, 0 \le c \le 1$,

$$\begin{aligned} \mathrm{E}\, \tilde{f}_n(x) &= \frac{1}{h} \mathrm{E}\, K\left(\frac{x + g_1(X_i)}{h}\right) + \frac{1}{h} \mathrm{E}\, K\left(\frac{x - g_2(X_i)}{h}\right) \\ &= f(0) + h(f^{(1)}(0) - g_1^{(2)}(0)f(0)) \int_c^1 (t-c)K(t)dt \\ &\qquad - h(f^{(1)}(0) - g_2^{(2)}(0)f(0)) \int_{-1}^c (t-c)K(t)dt \\ &\quad + \frac{h^2}{2}\left\{f^{(2)}(0) - g_1^{(3)}(0)f(0) - 3g_1^{(2)}(0)(f^{(1)}(0) - g_1^{(2)}(0)f(0))\right\} \int_c^1 (t-c)^2 K(t)dt \\ &\quad + \frac{h^2}{2}\left\{f^{(2)}(0) - g_2^{(3)}(0)f(0) - 3g_2^{(2)}(0)(f^{(1)}(0) - g_2^{(2)}(0)f(0))\right\} \int_{-1}^c (t-c)^2 K(t)dt \\ &\quad + o(h^2) \\ &= f(x) + h\left\{2f^{(1)}(0) \int_c^1 (t-c)K(t)dt - g_1^{(2)}(0)f(0) \int_c^1 (t-c)K(t)dt \right. \\ &\qquad\qquad \left. - g_2^{(2)}(0)f(0)\left(c + \int_c^1 (t-c)K(t)dt\right)\right\} \\ &\qquad + \frac{h^2}{2}\left\{-c^2 f^{(2)}(0) + f^{(2)}(0) \int_{-1}^1 (t-c)^2 K(t)dt \right. \\ &\qquad\qquad - \left[g_1^{(3)}(0)f(0) + 3g_1^{(2)}(0)(f^{(1)}(0) - g_1^{(2)}(0)f(0))\right] \int_c^1 (t-c)^2 K(t)dt \\ &\qquad\qquad \left. - \left[g_2^{(3)}(0)f(0) + 3g^{(2)}(0)(f^{(1)}(0) - g_2^{(2)}(0)f(0))\right] \int_{-1}^c (t-c)^2 K(t)dt \right\} \\ &\quad + o(h^2). \end{aligned} \tag{A.4}$$

This completes the proof of (2.2).

Proof of (2.3)**:**

Observe that for $x = ch, 0 \le c \le 1$, we have

$$
\begin{aligned}
\operatorname{Var} \tilde{f}_n(x) &= \frac{1}{(nh)^2} \operatorname{Var} \left\{ \sum_{i=1}^{n} \left[K\left(\frac{x + g_1(X_i)}{h} \right) + K\left(\frac{x - g_2(X_i)}{h} \right) \right] \right\} \\
&= \frac{1}{nh^2} \operatorname{E} \left[K\left(\frac{x + g_1(X_i)}{h} \right) + K\left(\frac{x - g_2(X_i)}{h} \right) \right]^2 \\
&\quad - \frac{1}{nh^2} \left\{ \operatorname{E} \left[K\left(\frac{x + g_1(X_i)}{h} \right) + K\left(\frac{x - g_2(X_i)}{h} \right) \right] \right\}^2 \\
&= I_1 + I_2,
\end{aligned} \tag{A.5}
$$

where

$$
\begin{aligned}
I_1 &= \frac{1}{nh^2} \operatorname{E} \left[K\left(\frac{x + g_1(X_i)}{h} \right) + K\left(\frac{x - g_2(X_i)}{h} \right) \right]^2 \\
&= \frac{1}{nh^2} \int \left[K\left(\frac{x + g_1(y)}{h} \right) + K\left(\frac{x - g_2(y)}{h} \right) \right]^2 f(y)dy \\
&= \frac{1}{nh^2} \left[\int K^2 \left(\frac{x + g_1(y)}{h} \right) f(y)dy + \int K^2 \left(\frac{x - g_2(y)}{h} \right) f(y)dy \right] \\
&\qquad + \frac{2}{nh^2} \int K\left(\frac{x + g_1(y)}{h} \right) K\left(\frac{x - g_2(y)}{h} \right) f(y)dy \\
&= I_{1,1} + I_{1,2},
\end{aligned} \tag{A.6}
$$

where $I_{1,1}$ and $I_{1,2}$ denote the first and second terms, respectively, of the right-hand side of

(A.6). Using a Taylor expansion as in (A.1) and (A.2), it can be shown that

$$
\begin{aligned}
I_{1,1} &= \frac{1}{nh^2}\left[h\int_c^1 K^2(t)\frac{f(g_1^{-1}((t-c)h))}{g_1^{(1)}(g_1^{-1}((t-c)h))}dt \right.\\
&\quad \left. + h\int_{-1}^c K^2(t)\frac{f(g_2^{-1}((c-t)h))}{g_2^{(1)}(g_2^{-1}((c-t)h))}dt\right]\\
&= \frac{f(0)}{nh}\int_{-1}^1 K^2(t)dt + o\left(\frac{1}{nh}\right). \qquad \text{(A.7)}
\end{aligned}
$$

By the continuity property of $g_1^{(2)}$ and $g_2^{(2)}$ and by a Taylor expansion of order 2 on g_1 and g_2, we have

$$
\begin{aligned}
g_1((c-t)h) &= g_1(0) + (t-c)(-h)g_1^{(1)}(0) + O(h^2)\\
&= (c-t)h + O(h^2) \qquad \text{(A.8)}
\end{aligned}
$$

and

$$
\begin{aligned}
g_2((c-t)h) &= g_2(0) + (t-c)(-h)g_2^{(1)}(0) + O(h^2)\\
&= (c-t)h + O(h^2), \qquad \text{(A.9)}
\end{aligned}
$$

since $g_i(0) = 0$ and $g_i^{(1)}(0) = 1, i = 1, 2$. Using (A.8) and (A.9) and by a change of variables,

$x + g_1(y) = ht$, we obtain

$$
\begin{aligned}
I_{1,2} &= \frac{2}{nh^2} \int K\left(\frac{x+g_1(y)}{h}\right) K\left(\frac{x-g_2(y)}{h}\right) f(y)dy \\
&= \frac{2}{nh} \int_c^1 K(t) K\left(\frac{x - g_2(g_1^{-1}(ht-x))}{h}\right) f(g_1^{-1}(th-x))dt \\
&= \frac{2}{nh} \int_c^1 K(t) K\left(\frac{x-(t-c)h+O(h^2)}{h}\right) f(g_1^{-1}(th-x))dt \\
&= \frac{2}{nh} \int_c^1 K(t)K(2c-t+O(h))(f(0)+O(h))dt \\
&= \frac{2f(0)}{nh} \int_c^1 K(t)K(2c-t)dt + o\left(\frac{1}{nh}\right).
\end{aligned} \tag{A.10}
$$

From (A.1) and (A.2), it is easy to show that

$$
\begin{aligned}
I_2 &= \frac{1}{nh^2} \left\{ \mathrm{E}\left[K\left(\frac{x+g_1(X_i)}{h}\right) + K\left(\frac{x-g_2(X_i)}{h}\right) \right] \right\}^2 \\
&= o\left(\frac{1}{nh}\right).
\end{aligned} \tag{A.11}
$$

Now combine (A.5) to (A.11) to obtain, for $x = ch, 0 \leq c \leq 1$,

$$
\operatorname{Var} \tilde{f}_n(x) = \frac{f(0)}{nh} \left\{ 2\int_{-1}^{c} K(t)K(2c-t)dt + \int_{-1}^{1} K^2(t)dt \right\} + o\left(\frac{1}{nh}\right). \tag{A.12}
$$

This completes the proof of (2.3).

In order to prove Theorem 2.1, we first state a lemma.

Lemma A.1. *Let $\hat{d}$ be defined by (2.9), with h and h_1 as in Theorem 2.1. Suppose that $f(x) > 0$ for $x = 0, h$ and that $f^{(2)}$ is continuous in a neighbourhood of 0. Then*

$$
\mathrm{E}\left[|\hat{d} - d|^3 | X_i = x_i \right] = O(h_1^3)
$$

for any $x_i, x_j \geq 0$ *and integers* $1 \leq i, j \leq n$, *where* d *is given by* (2.7).

Proof. Similar to the proof of Lemma A.2 of Zhang et al. (1999).

Proof of Theorem 2.1

From (2.14) we have, for $x = ch, 0 \leq c \leq 1$,

$$\mathrm{E}\,\hat{f}_n(x) - f(x) = I_3 + I_4, \tag{A.13}$$

where

$$I_3 = \mathrm{E}\,\hat{f}_n(x) - \mathrm{E}\,\tilde{f}_n(x) \tag{A.14}$$

and

$$I_4 = \mathrm{E}\,\tilde{f}_n(x) - f(x), \tag{A.15}$$

where $\tilde{f}_n(x)$ is given by (1.2) with $g_1 = g_2 = g_c$, g_c as in (2.6), and $\hat{f}_n(x)$ given by (2.14). From (2.2) we obtain

$$\begin{aligned} I_4 =& \frac{h^2}{2}\Big\{ f^{(2)}(0)\int_{-1}^{1} t^2 K(t)dt \\ &- \left[g_c^{(3)}(0)f(0) + g_c^{(2)}(0)(f^{(1)}(0) - g_c^{(2)}(0)f(0))\right]\int_{-1}^{1}(t-c)^2 K(t)dt\Big\} \\ &+ o(h^2) \end{aligned} \tag{A.16}$$

where g_c is given by (2.6). From (A.14), we have

$$\begin{aligned}|I_3| &\le \frac{1}{nh}\sum_{i=1}^{n} \mathrm{E}\left|K\left(\frac{x+\hat{g}_c(X_i)}{h}\right) - K\left(\frac{x+g_c(X_i)}{h}\right)\right| \\ &+ \frac{1}{nh}\sum_{i=1}^{n} \mathrm{E}\left|K\left(\frac{x-\hat{g}_c(X_i)}{h}\right) - K\left(\frac{x-g_c(X_i)}{h}\right)\right| \\ &= I_5 + I_6, \end{aligned} \tag{A.17}$$

where I_5 and I_6 denote the first and second terms, respectively, of the RHS of (A.17). In order to prove (2.15), it is enough to show that $I_5 = o(h^2)$ and $I_6 = o(h^2)$. The proofs of the preceding two expressions are very similar. We give details of the proof of $I_6 = o(h^2)$ here. For notational convenience of the proofs, we shall denote g_c and $\hat{g}_c$ by g and $\hat{g}$, respectively, in what follows. By an application of Taylor's expansion of order 1 on K, we obtain

$$|I_6| = \frac{1}{nh}\sum_{i=1}^{n} \mathrm{E}\left|\left(\frac{g(X_i)-\hat{g}(X_i)}{h}\right) K^{(1)}\left(\frac{x-g(X_i)+\varepsilon(g(X_i)-\hat{g}(X_i))}{h}\right)\right|, \tag{A.18}$$

where $0 < \varepsilon < 1$ is a constant. Then for $y \ge 0$ and $12\lambda_0 \ge 1$,

$$\begin{aligned} g(y) &= y + \frac{1}{2}dk_c' y^2 + \lambda_0 (dk_c')^2 y^3 \\ &= y\left(1 + \frac{dk_c'}{2}y + \lambda_0(dk_c')^2 y^2\right) \\ &= y\left[\left(\sqrt{\lambda_0}dk_c' y + \frac{1}{4\sqrt{\lambda_0}}\right)^2 + 1 - \frac{1}{16\lambda_0}\right] \\ &\ge \left(1 - \frac{1}{16\lambda_0}\right) y. \end{aligned} \tag{A.19}$$

Thus, $g(y) \ge h$ for $y \ge \rho h$, where $\rho = \frac{16\lambda_0 - 1}{16\lambda_0} > 0$. Therefore, $\varepsilon\hat{g}(X_i) + (1-\varepsilon)g(X_i) \ge h$ for

$X_i \geq \rho h$. Since K vanishes outside $[-1,1]$, from (A.18) we obtain

$$
\begin{aligned}
|I_6| &\leq \frac{1}{nh}\sum_{i=1}^{n} \mathrm{E}\left|\left(\frac{g(X_i)-\hat{g}(X_i)}{h}\right) K^{(1)}\left(\frac{x-\varepsilon\hat{g}(X_i)-(1-\varepsilon)g(X_i)}{h}\right)\right| I\left[0 \leq X_i \leq \rho h\right] \\
&\leq \frac{C}{nh^2}\sum_{i=1}^{n} \mathrm{E}\left|\hat{g}(X_i)-g(X_i)\right| I\left[0 \leq X_i \leq \rho h\right], \qquad \text{(A.20)}
\end{aligned}
$$

where $C = \sup_{|t|\leq 1}|K^{(1)}(t)|$. Now observe that

$$
\begin{aligned}
&\mathrm{E}\left|\hat{g}(X_i)-g(X_i)\right| I\left[0 \leq X_i \leq \rho h\right] \\
&\quad = \mathrm{E}\left|\frac{k_c'}{2}(\hat{d}-d)X_i^2 + k_c'^2(\hat{d}^2-d^2)X_i^3\right| I\left[0 \leq X_i \leq \rho h\right]. \qquad \text{(A.21)} \\
&\quad \leq \frac{k_c'}{2}(\rho h)^2\, \mathrm{E}\,|\hat{d}-d| I\left[0 \leq X_i \leq \rho h\right] + k_c'^2(\rho h)^3\, \mathrm{E}\,|\hat{d}^2-d^2| I\left[0 \leq X_i \leq \rho h\right].
\end{aligned}
$$

By the Cauchy-Schwarz inequality and Lemma A.1 above, we have

$$
\mathrm{E}\left[|\hat{d}-d|^l | X_i = x_i, X_j = x_j\right] = O(h_1^l) \qquad \text{(A.22)}
$$

for $1 \leq l \leq 3$ and $1 \leq i, j \leq n$. From (A.22), we obtain

$$
\begin{aligned}
\mathrm{E}\,|\hat{d}-d| I\left[0 \leq X_i \leq \rho h\right] &= \mathrm{E}\left\{\mathrm{E}\left[|\hat{d}-d| I\left[0 \leq X_i \leq \rho h\right] \big| X_i = x_i\right]\right\} \\
&= \mathrm{E}\left\{I\left[0 \leq X_i \leq \rho h\right] \mathrm{E}\left[|\hat{d}-d| \big| X_i = x_i\right]\right\} \\
&\leq O(h_1)\, \mathrm{E}\left[0 \leq X_i \leq \rho h\right] \\
&= o(h^2) \qquad \text{(A.23)}
\end{aligned}
$$

where the last equality follows from the fact that $\lim_{n\to\infty} h^{-1}\,\mathrm{E}\left[0 \leq X_i \leq \rho h\right] = f(0)$. Similarly, we again obtain from (A.22) that

$$\begin{aligned}
\mathrm{E}\,|\hat{d}^2 - d^2| I\,[0 \le X_i \le \rho h] &= \mathrm{E}\,|\hat{d} - d||\hat{d} + d| I\,[0 \le X_i \le \rho h] \\
&= \mathrm{E}\,|\hat{d} - d||\hat{d} - d + 2d| I\,[0 \le X_i \le \rho h] \qquad \text{(A.24)} \\
&\le \mathrm{E}\,|\hat{d} - d| I\,[0 \le X_i \le \rho h] + 2|d|\,\mathrm{E}\,|\hat{d} - d| I\,[0 \le X_i \le \rho h]\,.
\end{aligned}$$

Now by combining (A.20) to (A.24), we obtain $I_6 = o(h^2)$. This completes the proof of (2.15).

We now prove (2.16). First we write

$$\begin{aligned}
\mathrm{Var}\,\hat{f}_n(x) &= \frac{1}{(nh)^2}\,\mathrm{Var}\left\{\sum_{i=1}^{n} K\left(\frac{x + \hat{g}(X_i)}{h}\right) + \sum_{i=1}^{n} K\left(\frac{x - \hat{g}(X_i)}{h}\right)\right\} \\
&= 4(I_7 + I_8 + I_9 + I_{10} + I_{11} + I_{12}), \qquad \text{(A.25)}
\end{aligned}$$

where

$$\begin{aligned}
Y_1 &= \sum_{i=1}^{n}\left[K\left(\frac{x + \hat{g}(X_i)}{h}\right) - K\left(\frac{x + g(X_i)}{h}\right)\right] \\
Y_2 &= \sum_{i=1}^{n}\left[K\left(\frac{x - \hat{g}(X_i)}{h}\right) - K\left(\frac{x - g(X_i)}{h}\right)\right] \\
Y_3 &= \sum_{i=1}^{n}\left[K\left(\frac{x + g(X_i)}{h}\right) + K\left(\frac{x - g(X_i)}{h}\right)\right]
\end{aligned}$$

and

$$I_7 = \frac{1}{(nh)^2}\,\mathrm{Var}\,Y_1, I_8 = \frac{1}{(nh)^2}\,\mathrm{Var}\,Y_2, I_9 = \frac{1}{(nh)^2}\,\mathrm{Var}\,Y_3,$$

$$I_{10} = 2\,\mathrm{Cov}(Y_1, Y_2), I_{11} = 2\,\mathrm{Cov}(Y_1, Y_3), I_{12} = 2\,\mathrm{Cov}(Y_2, Y_3)$$

From (2.3), we have

$$I_9 = \frac{f(0)}{nh}\left\{2\int_{-1}^{c} K(t)K(2c-t)dt + \int_{-1}^{1} K^2(t)dt\right\} + o\left(\frac{1}{nh}\right). \qquad \text{(A.26)}$$

Since we have $I_l = o((nh)^{-1}), l = 10, 11, 12$, by the covariance inequality, it follows that in order to prove (2.16) it is enough to show that $I_7 = o\left(\frac{1}{nh}\right)$ and $I_8 = o\left(\frac{1}{nh}\right)$. The proofs are very similar and therefore we only give details of the proof of $I_8 = o\left(\frac{1}{nh}\right)$. By an application of Taylor's expansion of order 1 on K, we obtain

$$\begin{aligned}
I_8 &\leq \frac{1}{(nh)^2}\,\mathrm{E}\left\{\sum_{i=1}^{n}\left[K\left(\frac{x-\hat{g}(X_i)}{h}\right) - K\left(\frac{x-g(X_i)}{h}\right)\right]\right\}^2 \\
&= \frac{1}{(nh)^2}\,\mathrm{E}\left\{\sum_{i=1}^{n}\left(\frac{g(X_i)-\hat{g}(X_i)}{h}\right)K^{(1)}\left(\frac{x-g(X_i)+\varepsilon(g(X_i)-\hat{g}(X_i))}{h}\right)\right\}^2 \\
&\leq \frac{2}{n^2h^4}\sum_{i=1}^{n}\mathrm{E}\left\{(\hat{g}(X_i)-g(X_i))^2\left[K^{(1)}\left(\frac{x-\varepsilon\hat{g}(X_i)-(1-\varepsilon)g(X_i)}{h}\right)\right]^2\right\} \\
&\leq \frac{C}{n^2h^4}\sum_{i=1}^{n}\mathrm{E}\left(\hat{g}(X_i)-g(X_i)\right)^2 I\left[0 \leq X_i \leq \rho h\right], \qquad \text{(A.27)}
\end{aligned}$$

using an argument similar to (A.20), where $0 < \varepsilon < 1, \rho = (12\lambda_0 - 1)/12\lambda_0$ and $C > 0$ are all constants independent of n. Now, again using (A.22), (A.23) and (A.24) above, we obtain (compare with (A.21))

$$
\begin{aligned}
&\mathrm{E}(\hat{g}(X_i) - g(X_i))^2 I\,[0 \le X_i \le \rho h] \\
&\qquad = \mathrm{E}\left\{\frac{k_c}{2}(\hat{d} - d)X_i^2 + k_c^2(\hat{d}^2 - d^2)X_i^3\right\}^2 I\,[0 \le X_i \le \rho h] \\
&\qquad \le \frac{k_c^2}{2}(\rho h)^2\,\mathrm{E}(\hat{d} - d)^2 I\,[0 \le X_i \le \rho h] \\
&\qquad\qquad + 2k_c^4(\rho h)^6\,\mathrm{E}(\hat{d}^2 - d^2)^2 I\,[0 \le X_i \le \rho h] \\
&\qquad \le O(h^4 h_1^2 h) + O(h^6 h_1^2 h) \\
&\qquad = o(h^7). \qquad\qquad (A.28)
\end{aligned}
$$

Now combine (A.27) and (A.28) to obtain $I_8 = o\left(\frac{1}{nh}\right)$. This completes the proof of (2.16).

Table 1: MISE Values over $[0, h)$

	Density 1	Density 2	Density 3	Density 4
New Estimator	0.000494	0.001826	0.009853	0.022152
Boundary Kernel	0.000234	0.002288	0.009414	0.015894
LL method	0.000271	0.002223	0.009365	0.014869
Z,K&J method	0.000426	0.002637	0.009219	0.014595
J&F method	0.000268	0.002439	0.009269	0.016964
H&P method	0.000374	0.002445	0.011264	0.015018

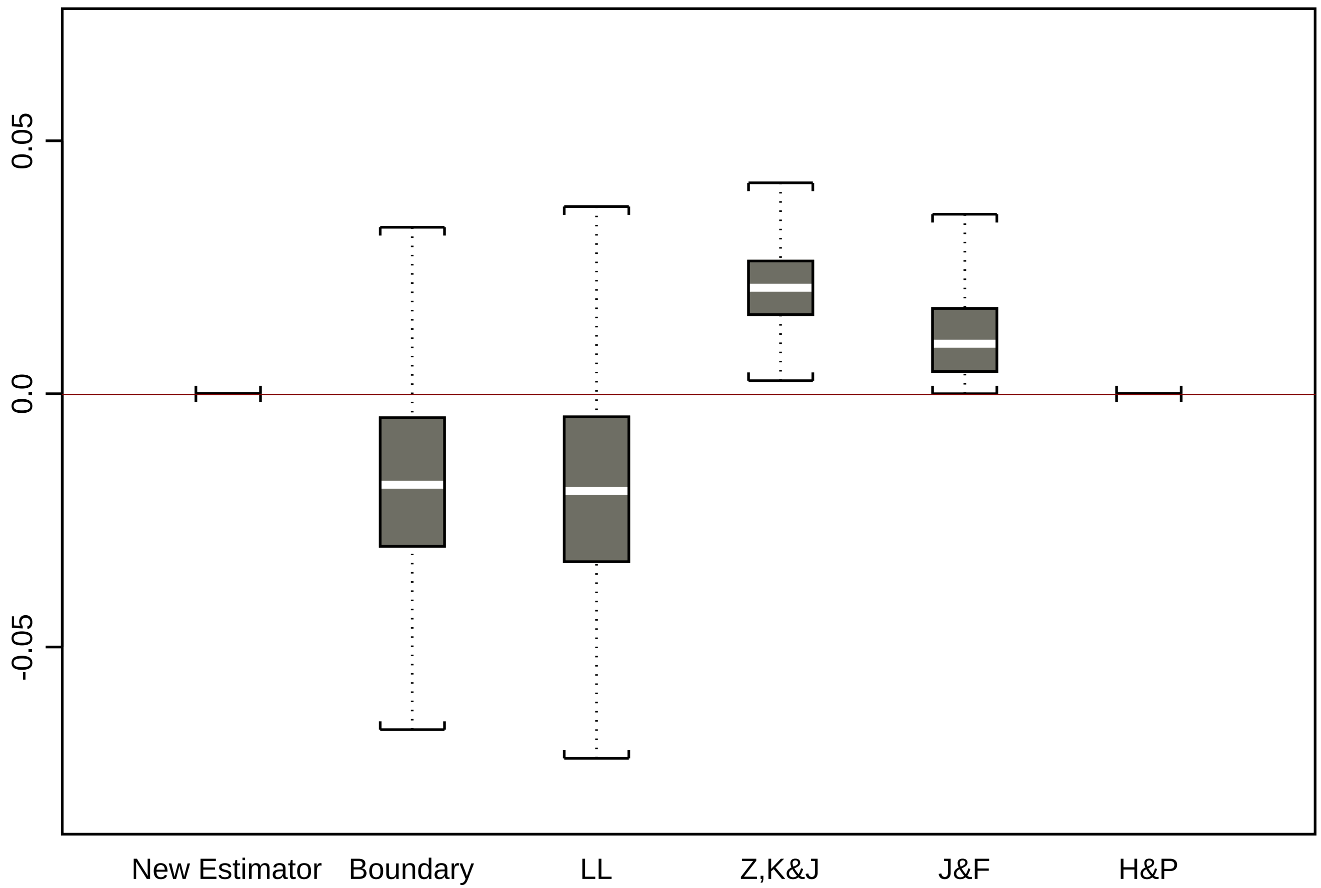

Figure 1: Bias of estimators for the Density 1: $f(x) = \frac{x^2}{2}e^{-x}$ over 1000 trials, with $h = .832109$ and $n = 200$.

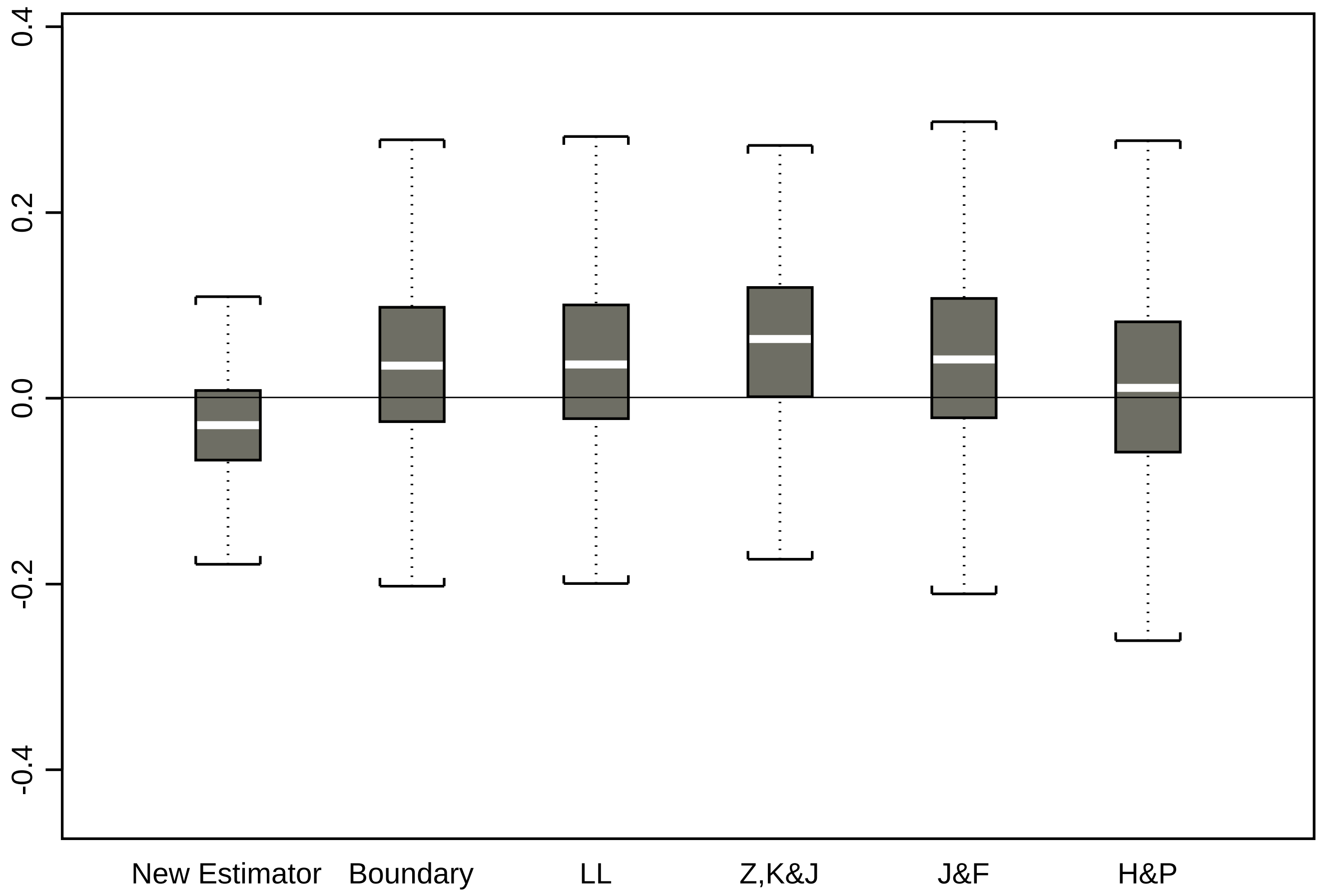

Figure 2: Bias of estimators for the Density 2: $f(x) = \frac{2}{\pi(1+x^2)}$ over 1000 trials, with $h = .690595$ and $n = 200$.

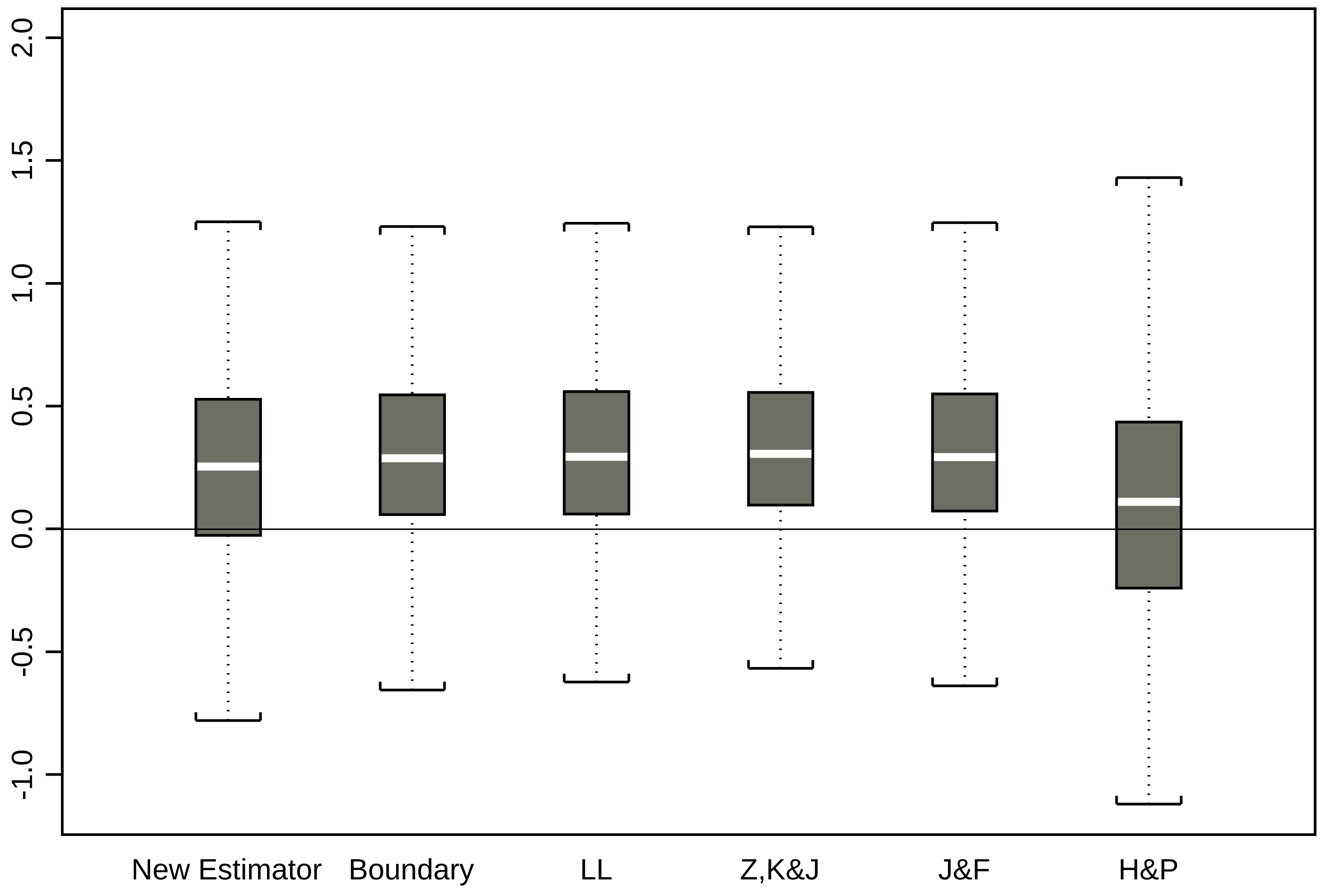

Figure 3: Bias of estimators for the Density 3: $f(x) = \frac{5}{4}(1 + 15x)e^{-5x}$ over 1000 trials, with $h = .139332$ and $n = 200$.

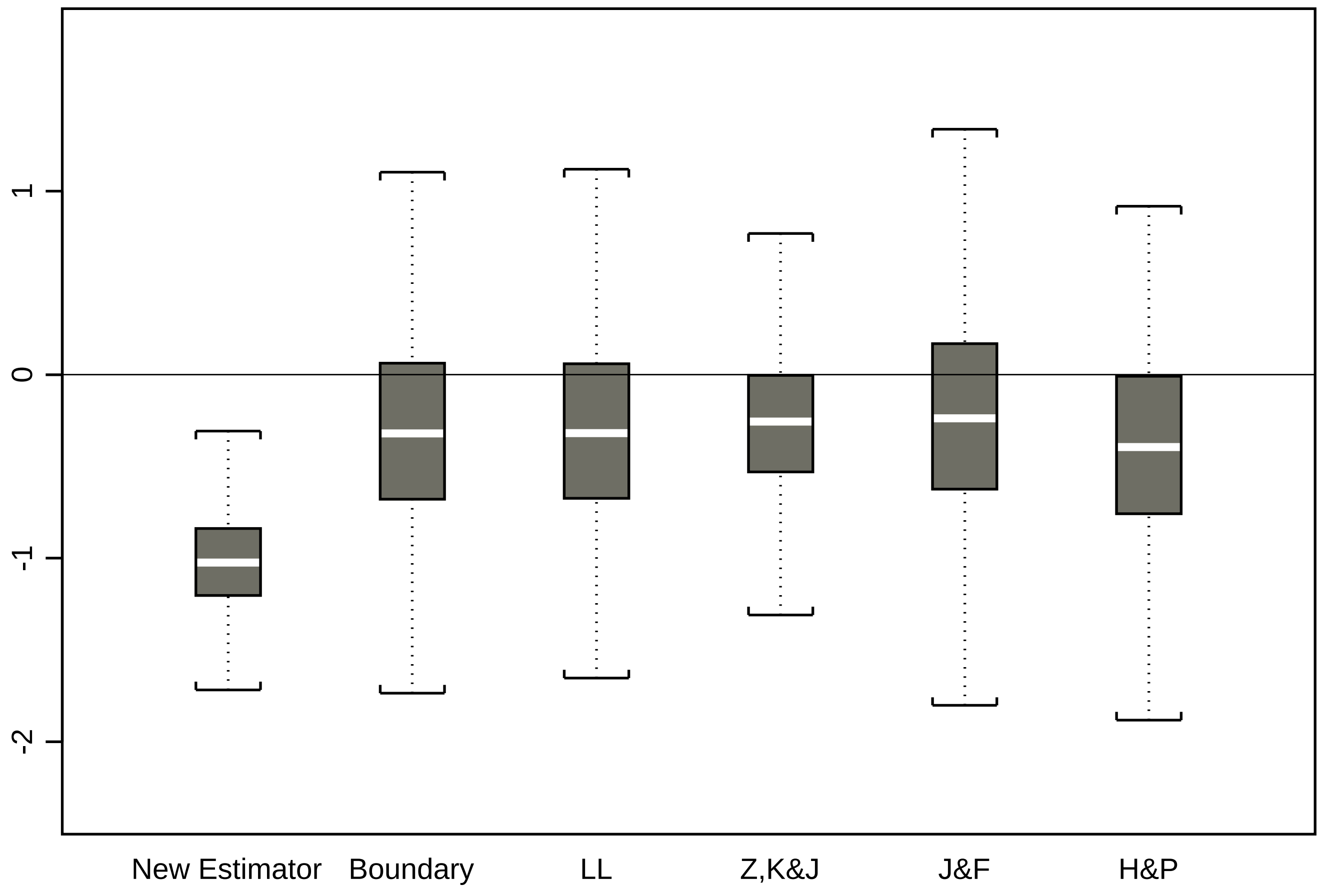

Figure 4: Bias of estimators for Density 4: $f(x) = 5e^{-5x}$ over 1000 trials, with $h = .136851$ and $n = 200$.

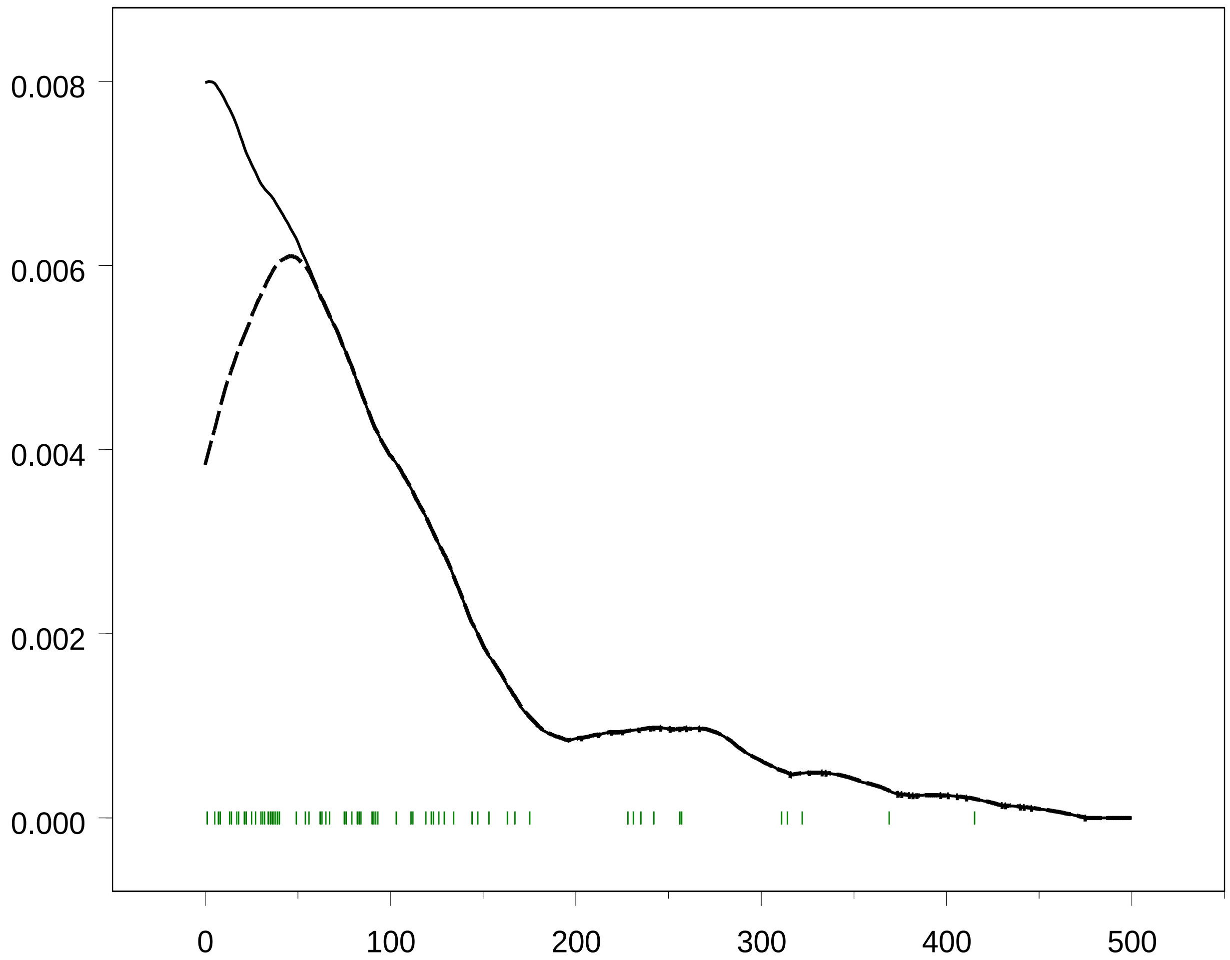

Figure 5: Density Estimates for the suicide data, with the rug showing the data. The solid line is our estimator, the dashed line is the regular kernel estimator. Both use bandwidth $h = 60$, as seen in Silverman (1986).

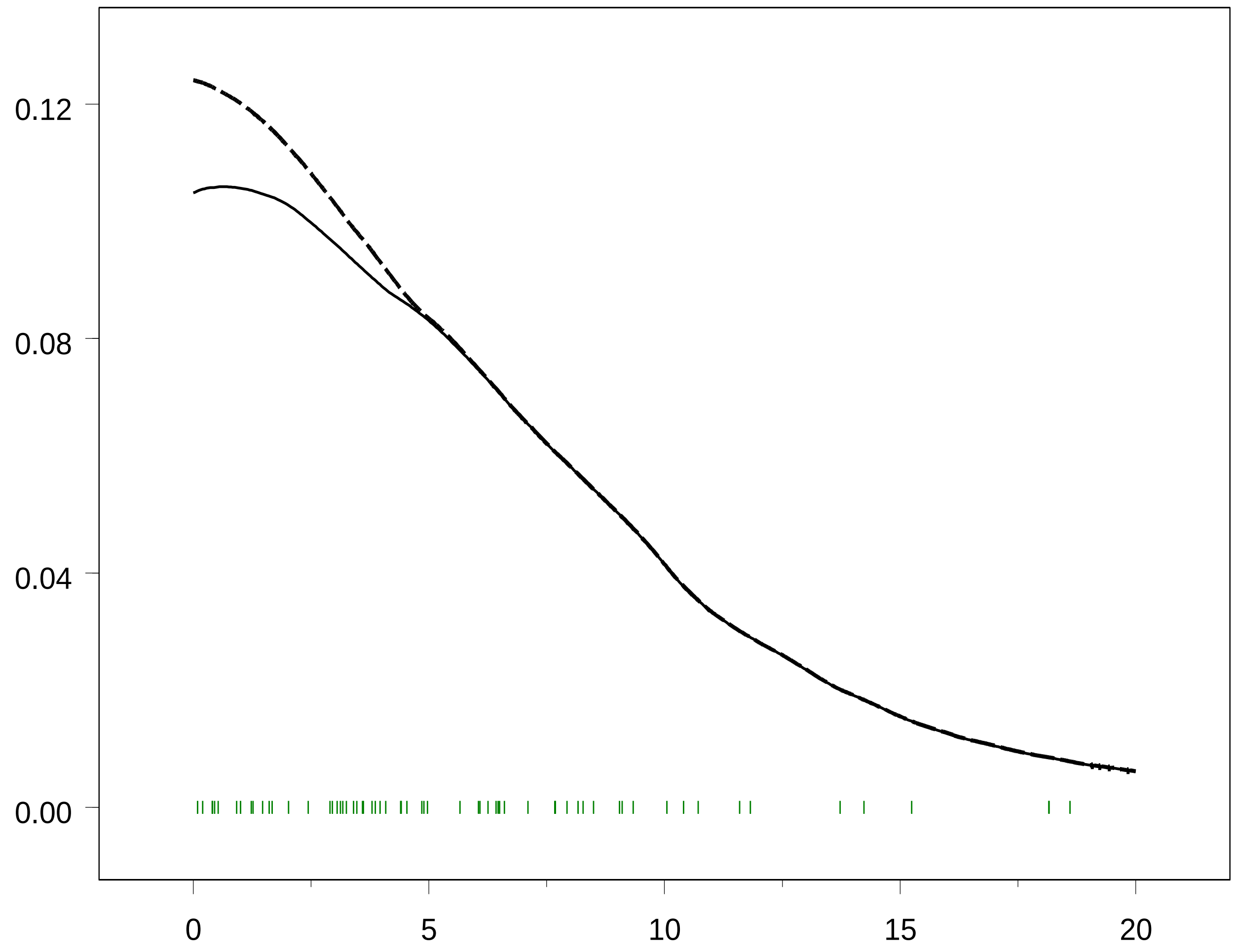

Figure 6: Density Estimates for the line transect data, with the rug showing the data. The solid line is our estimator, the dashed line is the estimator of Zhang et al. (1999). Both use bandwidth $h = 7.1822$.

References

[1] Burnham, K.P., Anderson, D.R. and Laake, J.L. (1980). Estimation of density from line transect sampling of biological populations. *Wildlife Monographs*, 72.

[2] Cheng, M.Y. (1997). Boundary-aware estimators of integrated density derivative. *Journal of the Royal Statistical Society Ser. B*, 59, 191-203.

[3] Cheng, M.Y., Fan, J. and Marron, J.S. (1997). On Automatic Boundary Corrections. *The Annals of Statistics*, 25, 1691-1708.

[4] Cline, D.B.H and Hart, J.D. (1991). Kernel Estimation of Densities of Discontinuous Derivatives. *Statistics*, 22, 69-84.

[5] Devroye, L. and Gyorfi, L. (1985). *Nonparametirc Density Estimation: The L1 View*, New York: John Wiley & Sons.

[6] Gasser, T. and Müller, H.G. (1979). Kernel Estimation of Regression Functions. In *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics 757, eds. T. Gasser and M. Rosenblatt, Heidelberg: Springer-Verlag, pp. 23-68.

[7] Gasser, T., Müller, H.G. and Mammitzsch, V. (1985). Kernels for Nonparametric Curve Estimation. *Journal of the Royal Statistical Society Ser. B.*, 47, 238-252.

[8] Hall, P. and Park, B.U. (2002). New methods for bias correction at endpoints and boundaries. *The Annals of Statistics*, 30, 1460-1479.

[9] Jones, M.C. (1993). Simple Boundary Correction for Kernel Density Estimation. *Statistics and Computing*, 3, 135-146.

[10] Jones, M.C. and Foster, P.J. (1996). A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation. *Statistica Sinica*, 6, 1005-1013.

[11] Karunamuni, R.J. and Alberts, T. (2004). On boundary correction in kernel density estimation. Unpublished Technical Report.

[12] Marron, J.S. and Ruppert, D. (1994). Transformations to Reduce Boundary Bias in Kernel Density Estimation. *Journal of the Royal Statistical Society Ser. B,* 56, 653-671.

[13] Müller, H.G. (1991). Smooth Optimum Kernel Estimators Near Endpoints. *Biometrika*, 78, 521-530.

[14] Prakasa Rao, B.L.S (1983). *Nonparametric Functional Estimation*, Orlando, FL: Academic Press.

[15] Schuster, E.F. (1985). Incorporating Support Constraints Into Nonparametric Estimators of Densities. *Communications in Statistics, Part A - Theory and Methods*, 14, 1123-1136.

[16] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

[17] Simonoff, J. (1996). *Smoothing Methods in Statistics*, New York: Springer-Verlag.

[18] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*, London: Chapman and Hall.

[19] Wand, M.P., Marron, J.S. and Ruppert, D. (1991). Transformations in Density Estimation (with discussion). *Journal of the American Statistical Association*, 86, 343-361.

[20] Zhang, S. and Karunamuni, R.J. (1998). On Kernel Density Estimation Near Endpoints. *Journal of Statistical Planning and Inference*, 70, 301-316.

[21] Zhang, S. and Karunamuni, R.J. (2000). On Nonparametric Density Estimation at the Boundary. *Nonparametric Statistics*, 12, 197-221.

[22] Zhang S., Karunamuni, R.J. & Jones, M.C. (1999). An Improved Estimator of the Density Function at the Boundary. *Journal of the American Statistical Association*, 448, 1231-1241.