A Locally Adaptive Transformation Method of Boundary Correction in Kernel Density Estimation

R.J. Karunamuni\textsuperscript{a} and T. Alberts\textsuperscript{b}

\textsuperscript{a}Department of Mathematical and Statistical Sciences
University of Alberta, Edmonton, Alberta, Canada, T6G 2G1;

\textsuperscript{b}Courant Institute of Mathematical Sciences,
New York University, New York, NY, U.S.A 10012

Abstract

Kernel smoothing methods are widely used in many research areas in statistics. However, kernel estimators suffer from boundary effects when the support of the function to be estimated has finite endpoints. Boundary effects seriously affect the overall performance of the estimator. In this article, we propose a new method of boundary correction for univariate kernel density estimation. Our technique is based on a data transformation that depends on the point of estimation. The proposed method possesses desirable properties such as local adaptivity and the non-negativity of the proposed estimator. Furthermore, unlike many other transformation methods available, the estimator is easy to implement. In a Monte Carlo study, the accuracy of the proposed estimator is numerically analyzed and compared with the existing methods of boundary correction. We find that the proposed estimator performs well for most shapes of densities. The theory behind the new methodology, along with the bias and variance of the proposed estimator, are presented. Results of a data analysis are also given.

Keywords: Density estimation; Mean Squared Error; Kernel estimation; Transformation.
1 Introduction

Nonparametric kernel density estimation is now popular and in wide use with great success in statistical applications. Kernel density estimates are commonly used to display the shape of a data set without relying on a parametric model, not to mention the exposition of skewness, multimodality, dispersion, and more in the data set. Early results on kernel density estimation are due to Rosenblatt (1956) and Parzen (1962). Since then, much research has been done in the area; see, e.g., the monographs of Silverman (1986), and Wand and Jones (1995).

Let $f$ denote a probability density function with support $[0, \infty)$ 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

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

where $K$ is some chosen unimodal density function, symmetric about zero with support traditionally on $[-1, 1]$, and $h$ is the bandwidth ($h \to 0$ as $n \to \infty$). The basic properties of $f_n$ are well-known, and under some smoothness assumptions these include, for $x \geq 0$,

$$\mathbb{E} f_n(x) = f(x) \int_{-c}^{c} K(t)dt - hf^{(1)}(x) \int_{-c}^{c} tK(t)dt + \frac{h^2}{2} f^{(2)}(x) \int_{-c}^{c} t^2 K(t)dt + o(h^2),$$

$$\text{Var} f_n(x) = \frac{f(x)}{nh} \int_{-1}^{c} K^2(t)dt + o \left( \frac{1}{nh} \right),$$

where $c = \min(x/h, 1)$. For $x \geq h$, so-called interior points, the bias of $f_n(x)$ is of order $O(h^2)$, whereas at boundary points, i.e. $x \in [0, h)$, the bias is only of order $O(h)$. In nonparametric curve estimation problems these are referred to as “boundary effects”. To remove those boundary effects a variety of methods have been developed in the literature. Some well-known methods are summarized below:
• The reflection method (Cline and Hart, 1991; Schuster, 1985; Silverman, 1986).

• The boundary kernel method (Gasser and Muller, 1979; Gasser, Muller and Mammitzoch, 1983; Jones, 1993; Muller, 1991; Zhang and Karunamuni, 2000).

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

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

• The local linear method (Cheng, Fan and Marron, 1997; Cheng, 1997; Zhang and Karunamuni, 1998).

• Other methods (Zhang, Karunamuni and Jones, 1999; Hall and Park, 2002).

The reflection method is specifically designed for the case \( f^{(1)}(0) = 0 \), 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 \( f(0) \approx 0 \). To correct this deficiency of boundary kernel methods, some remedies have been proposed; see Jones (1993), Jones and Foster (1996), Gasser and Muller (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, Fan and Marron, 1997) in the boundary kernel 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”, which are then combined 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 an estimate of $f$. Among other methods, two very promising recent ones are due to Zhang, Karunamuni and Jones (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”.

The boundary kernel and related methods usually have low bias but the price for that is an increase in variance. It has been observed that approaches involving only kernel modifications without regard to $f$ or data, such as the boundary kernel method, are always associated with larger variance. Furthermore, the corresponding estimates tend to take negative values near the boundary points. On the other hand, transformation-based boundary correction estimates are non-negative and generally have low variance, possibly due to non-negativity. It has gradually gotten through to researchers that this non-negativity property is important in practical applications and the approaches producing non-negative estimators are well worth exploring. In this article, we develop a new transformation approach for boundary correction. Our method consists of a straightforward transformation of data. It is easy to implement in practice compared to other existing transformation methods (compare with Marron and Ruppert, 1994). A distinct feature of our method is that it is locally adaptive; that is, the transformation depends on the point of estimation. Other desirable properties of the proposed estimator are: (i) it is non-negative everywhere, and (ii) it reduces to the traditional kernel estimator (1.1) at the interior points. Most importantly, our present approach improves the bias but holds on to the low variance. In simulations, we show that the proposed estimator performs quite well overall compared to the other existing methods of boundary correction, and therefore should be given due consideration for use in practical applications.

Section 2 contains the methodology and the main results. Section 3 and 4 present simulation studies and a data analysis, respectively. Final comments are given in Section 5.
2 Locally Adaptive Transformation Estimator

2.1 Methodology

For convenience, we shall assume that the unknown probability density function \( f \) has support \([0, \infty)\), and consider estimation of \( f \) based on a random sample \( X_1, \ldots, X_n \) from \( f \). Our transformation idea is based on transforming the original data \( X_1, \ldots, X_n \) to \( g(X_1), \ldots, g(X_n) \), where \( g \) is a non-negative, continuous and monotonically increasing function from \([0, \infty)\) to \([0, \infty)\). Based on the transformed data, we now define for \( x = ch, c \geq 0 \),

\[
\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - g(X_i)}{h}\right) / \int_{-1}^{c} K(t)dt,
\]

(2.1)

where \( h \) is the bandwidth (again \( h \to 0 \) as \( n \to \infty \)), \( K \) is a non-negative symmetric kernel function with support on \([-1, 1]\), satisfying \( \int K(t)dt = 1, \int tK(t)dt = 0 \), and \( 0 \neq \int t^2K(t)dt < \infty \). We can now state the following lemma, which exhibits the bias and variance of (2.1) under certain conditions on \( g \) and \( f \). The proof is given in the Appendix.

**Lemma 2.1.** Let \( \hat{f}_n \) be defined by (2.1). Assume that \( f^{(2)} \) and \( g^{(2)} \) exist and are continuous on \([0, \infty)\), where \( f^{(i)} \) and \( g^{(i)} \) denote the \( i^{th} \) derivative of \( f \) and \( g \), respectively, with \( i \geq 0 \), \( f^{(0)} = f, g^{(0)} = g \). Further assume that \( g^{-1}(0) = 0, g^{(1)}(0) = 1 \), where \( g^{-1} \) is the inverse function of \( g \). Then for \( x = ch, 0 \leq c \leq 1 \), we have

\[
E \hat{f}_n(x) - f(x) = \frac{-h}{\int_{-1}^{c} K(t)dt} \left\{ f(0)g^{(2)}(0) \int_{-1}^{c} (c - t)K(t)dt + f^{(1)}(0) \int_{-1}^{c} tK(t)dt \right\} \\
+ \frac{h^2}{2 \int_{-1}^{c} K(t)dt} \left\{ -f^{(2)}(0)c^2 \int_{-1}^{c} K(t)dt + \int_{-1}^{c} (t - c)^2K(t)dt \\
\times [f^{(2)}(0) - f(0)g^{(3)}(0) - 3g^{(2)}(0)(f^{(1)}(0) - f(0)g^{(2)}(0))] \right\} + o(h^2),
\]

(2.2)
and

\[
\text{Var } \hat{f}_n(x) = \frac{f(0)}{nh(f_{-1}^c K(t) dt)^2} \int_{-1}^{c} K^2(t) dt + o\left(\frac{1}{nh}\right)
= \frac{f(x)}{nh(f_{-1}^c K(t) dt)^2} \int_{-1}^{c} K^2(t) dt + o\left(\frac{1}{nh}\right),
\]

(2.3)

The last equality follows since \( f(0) = f(x) - ch f^{(1)}(x) + (ch)^2 f^{(2)}(x) / 2 + o(h^2) \) for \( x = ch \) (see (A.3)). Note that the leading term of the variance of \( \hat{f}_n \) is not affected by the transformation \( g \). As \( c \to 1 \), \( \text{Var } \hat{f}_n(x) \to f(x)(nh)^{-1} \int_{-1}^{1} K^2(t) dt + o((nh)^{-1}) \), which is exactly the expansion of the interior variance of the traditional estimator (1.1). We shall use our transformation \( g \) so that the first order term in the bias expansion (2.2) is zero. Assuming that \( f(0) > 0 \), it is enough to let

\[
g^{(2)}(0) = -f^{(1)}(0) \int_{-1}^{c} t K(t) dt / f(0) \int_{-1}^{c} (c - t) K(t) dt.
\]

(2.4)

Note that the right-hand side 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 is locally adaptive. In order to display this local dependence we shall denote \( g \) by \( g_c, 0 \leq c \leq 1 \), in what follows. Combining (2.4) with the additional assumptions in Lemma 2.1, \( g_c \) should now satisfy the following three conditions for each \( c, 0 \leq c \leq 1 \):

(i) \( g_c : [0, \infty) \to [0, \infty), g_c \) is continuous, monotonically increasing

and \( g_c^{(i)} \) exists, \( i = 1, 2, 3 \),

(ii) \( g_c^{-1}(0) = 0, g_c^{(1)}(0) = 1 \)

(iii) \( g_c^{(2)}(0) = -f^{(1)}(0) \int_{-1}^{c} t K(t) dt / f(0) \int_{-1}^{c} (c - t) K(t) dt. \)

(2.5)

6
Functions satisfying conditions (2.5) above can be easily constructed. We employ the following transformation in our analysis. For $0 \leq c \leq 1$, define

$$g_c(y) = y + \frac{d}{2} l_c y^2 + d^2 l_c^2 y^3,$$  \hspace{1cm} (2.6)$$

where

$$l_c = - \int_{-1}^{c} t K(t) dt / \int_{-1}^{c} (c - t) K(t) dt$$  \hspace{1cm} (2.7)$$

and

$$d = f^{(1)}(0) / f(0).$$  \hspace{1cm} (2.8)$$

Observe that $l_c \to 0$ as $c \to 1$, and therefore $g_c(y) \to y$ as $c \to 1$ for each $y$. This means that \( \hat{f}_n(x) \) defined by (2.1) with $g_c$ of (2.6) reduces to the usual kernel estimator $f_n(x)$ defined by (1.1) for interior points, i.e., for $x \geq h$, $f_n(x)$ coincides with $f_n(x)$. For $g_c$ defined by (2.6), the bias term of (2.2) takes the form

$$E \hat{f}_n(x) - f(x) = \frac{h^2}{2 \int_{-1}^{c} K(t) dt} \left\{ f^{(2)}(0) \int_{-1}^{c} (t^2 - 2tc) K(t) dt + 6 \left( \frac{f^{(1)}(0)}{f(0)} \right)^2 \int_{-1}^{c} (t - c)^2 K(t) dt (l_c^2 + l_c) \right\} + o(h^2),$$  \hspace{1cm} (2.9)$$

for $x = ch, 0 \leq c \leq 1$. Note that as $c \to 1$, $E \hat{f}_n(x) - f(x) \to \frac{h^2}{2} f^{(2)}(x) \int_{-1}^{1} t^2 K(t) dt + o(h^2)$, which is exactly the same expression of the interior bias of the traditional kernel estimator (1.1), since $f^{(2)}(x) = f^{(2)}(0) + o(1)$ for $x = ch$ (see (A.5)).
2.2 Estimation of \( g_c \)

In order to implement the transformation (2.6) in practice, one must replace \( d = f^{(1)}(0)/f(0) \) with a pilot estimate. This requires the use of another density estimation method, such as semiparametric, kernel or nearest neighbour method. Our experience is that proposed estimator of \( f \) given below is insensitive to the fine details of the pilot estimate of \( d \), and therefore any convenient estimate can be employed. A natural estimate would be a fixed kernel estimate with bandwidth chosen optimally at \( x = 0 \) as in Zhang, Karunamuni and Jones (1999). Other possible estimators of \( d \) are proposed in Choi and Hall (1999) and Park, Jeong, Jones and Kang (2003). Following Zhang, Karunamuni and Jones (1999), in this paper we use the kernel type pilot estimate of \( d \) given by

\[
\hat{d} = (\log f^*_n(h) - \log f^*_n(0))/h, \tag{2.10}
\]

where

\[
f^*_n(h) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{h - X_i}{h} \right) + \frac{1}{n^2} \tag{2.11}
\]

and

\[
f^*_n(0) = \max \left\{ \frac{1}{nh_0} \sum_{i=1}^{n} K_{(0)} \left( \frac{-X_i}{h_0} \right), \frac{1}{n^2} \right\}, \tag{2.12}
\]

where \( K \) is a usual symmetric kernel function as in (1.1), \( K_{(0)} \) is a so-called endpoint order two kernel satisfying

\[
\int_{-1}^{0} K_{(0)}(t)dt = 1, \int_{-1}^{0} tK_{(0)}(t)dt = 0, \text{ and } 0 < \int_{-1}^{0} t^2K_{(0)}(t)dt < \infty,
\]
and $h_0 = b(0)h$ with $b(0)$ given by

$$
 b(0) = \left\{ \frac{\left(\int_{-1}^{1} t^2 K(t) dt\right)^2 (\int_{-1}^{0} K(0)^2(t) dt)}{\left(\int_{-1}^{0} t^2 K(0)(t) dt\right)^2 (\int_{-1}^{1} K^2(t) dt)} \right\}. \tag{2.13}
$$

We now define

$$
 \hat{g}_c(y) = y + \frac{\hat{d}}{2} t_c y^2 + \frac{\hat{d}^2}{6} t_c^2 y^3,
$$

for $0 \leq c \leq 1$, as our estimator of $g_c(y)$ defined by (2.6), where $\hat{d}$ is given by (2.10). Note that $\hat{g}_c$ and $\hat{d}$ depend on $n$ but this dependence is suppressed for notational convenience.

### 2.3 The Proposed Estimator

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

$$
 \hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - \hat{g}_c(X_i)}{h}\right) / \int_{-1}^{c} K(t) dt, \tag{2.15}
$$

where $\hat{g}_c$ is given by (2.14), $K$ is as given in (2.1), and $h$ is the bandwidth.

For $x \geq h$, $\hat{f}_n(x)$ reduces to the traditional kernel estimator $f_n(x)$ given by (1.1). Thus $\hat{f}_n(x)$ is a natural boundary continuation of the usual kernel estimator (1.1). An important feature of the estimator (2.15) is that it is based on a locally adaptive transformation $\hat{g}_c$ given by (2.14), since $\hat{g}_c$ transforms the data depending on the point of estimation. Furthermore, it is important to remark here that the adaptive kernel estimator (2.15) is non-negative (provided $K$ is non-negative), a property shared with reflection based estimators (see Jones and Foster, 1996; Zhang, Karunamuni and Jones, 1999) and transformation based estimators (see Marron and Ruppert, 1994; Zhang, Karunamuni and Jones, 1999; Hall and Park, 2002), but not with most boundary kernel approaches.

The next theorem establishes the bias and variance of (2.15).
Theorem 2.1. Let \( \hat{f}_n(x) \) be defined by (2.15) with a kernel function \( K \) as given in (2.1) and a bandwidth \( h = O(n^{-1/5}) \). Further assume that \( K^{(1)} \) exists and is continuous on \([-1, 1]\) and that \( \int |K^{(1)}(t)|dt < \infty \). Assume that \( f(0) > 0 \) and \( f^{(2)} \) exists and is continuous in a neighbourhood of 0. Then for \( x = ch, 0 \leq c \leq 1 \), we have

\[
E\hat{f}_n(x) - f(x) = \frac{h^2}{2\int_{-1}^{c} K(t)dt} \left\{ f^{(2)}(0) \int_{-1}^{c} (t^2 - 2tc)K(t)dt + 6\frac{(f^{(1)}(0))^2}{f(0)} \int_{-1}^{c} (t-c)^2 K(t)dt(t_c^2 + l_c) \right\} + O(h^2),
\]

(2.16)

and

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

(2.17)

where \( l_c \) is given by (2.7).

2.4 Bandwidth Variation

From (2.16) we believe that the bias can be further reduced by choosing the bandwidth \( h \) to be locally adaptive along with \( g \). We do so by employing a bandwidth variation function (BVF) \( h_c = b(c)h \), where \( b : [0, 1] \rightarrow [0, \infty) \) satisfies \( b(1) = 1 \) and \( b(c) \geq c \). Then in the bias and variance expressions (2.2) and (2.3) of Lemma 2.1, the introduction of the BVF changes \( c \) to \( c/b(c) \). To compensate, the proposed form of the estimator becomes

\[
\hat{f}_n(x) \rightarrow \frac{1}{nh_c} \sum_{i=1}^{n} K\left(\frac{x - \hat{g}_c/b(c)(X_i)}{h_c}\right)/\int_{-1}^{c/b(c)} K(t)dt.
\]

(2.18)
Moreover, the bias equation (2.16) becomes

\[
E \hat{f}_n(x) - f(x) = \frac{h^2}{2} \frac{b(c)^2}{\int_{-1}^{c/b(c)} K(t)dt} \left\{ f^{(2)}(0) \int_{-1}^{c/b(c)} (t^2 - 2tc/b(c))K(t)dt + 6 \left( \frac{f^{(1)}(0)}{f(0)} \right)^2 \int_{-1}^{c/b(c)} (t - c/b(c))^2 K(t)dt (t^2_{c/b(c)} + t_{c/b(c)}) \right\} + O(h^2) \quad (2.19)
\]

and the variance is

\[
\text{Var} \hat{f}_n(x) = \frac{f(0)}{nb(c)h(\int_{-1}^{c/b(c)} K(t)dt)^2} \int_{-1}^{c/b(c)} K^2(t)dt + o \left( \frac{1}{nh} \right). \quad (2.20)
\]

The effect on the MSE of introducing \( b(c) \) is not entirely clear, especially given that the MSE depends so heavily on the unknown \( f \), but there are certain choices of \( b(c) \) that eliminate the coefficients on the bias or variance terms. Specifically we choose \( b(c) \) to satisfy one of the following:

\[
b(c)^2 = \int_{-1}^{c/b(c)} K(t)dt, \quad (2.21)
\]

or

\[
b(c) \left( \int_{-1}^{c/b(c)} K(t)dt \right)^2 = 1, \quad (2.22)
\]

where the first choice reduces the bias and the second reduces the variance. Given that \( K \) is non-negative with support on \([-1, 1]\) and integrates to one, a solution to each equation, satisfying \( b(1) = 1 \) and \( b(c) \geq c \), is easily seen to exist, although in practice it usually must be determined numerically. Moreover, it can be shown that the solution to the first (second) equation is monotone increasing (decreasing), so that a relatively smaller (larger) bandwidth value is used near the endpoint, but towards the interior points the bandwidth variation
approaches the regular, constant $h$. This has some intuitive appeal from a practical point of view since it controls the amount of smoothing applied to the data. Unfortunately, using the bias-reducing BVF (2.21) has the side effect of increasing the variance and vice-versa. In the former case we believe that the increase in variance is offset almost entirely by the smaller bias, but in the latter case the variance-reducing BVF (2.22) creates an unjustifiably large increase in bias. This suspicion is confirmed via simulations, and in practice we recommend only using a bandwidth variation function that satisfies (2.21).

3 Simulations and Discussion

To test the effectiveness of our estimator we simulated its performance, with and without bandwidth variation, against that of other well-known methods. Among these were some relatively new techniques, such as a recent estimator due to Hall and Park (2002) which is very similar to our own, and the transformation and reflection based method given by Zhang, Karunamuni and Jones (1999). We also competed against some more classical estimators; namely the boundary kernel and its close counterpart the local linear fitting method, Jones and Foster’s (1996) nonnegative adaptation estimator, and Cowling and Hall’s (1996) pseudo-data method.

The first of the competing estimators is due to Hall and Park (2002) (hereafter H&P). It is defined as, for $x = ch, 0 \leq c \leq 1$,

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

(3.1)

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

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

\[ \hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right) / \int_{-1}^{c} K(t) dt, \] (3.2)

which is referred to as the cut-and-normalized kernel, \( \hat{f}(x) \) is an estimate of \( f^{(1)}(x) \), and

\[ \rho(u) = K(u)^{-1} \int_{v \leq u} v K(v) dv. \]

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

\[ \hat{f}'(x) = \frac{1}{nh^2} \sum_{i=1}^{n} K_{1,c/b(c)} \left( \frac{x - X_i}{h_c} \right), \]

where \( h_c = b(c)h \) with the bandwidth variation function \( b(c) = 2^{1/5}(1-c) + c \) for \( 0 \leq c \leq 1 \).

Note that Hall and Park’s estimator is a special case of the estimator (2.1), but with \( \hat{g}_c(y) = y - \hat{\alpha}(ch) \), where \( x = ch \) for \( 0 \leq c \leq 1 \). This \( \hat{g} \), however, does not necessarily satisfy the conditions of (2.5).

The method of Zhang, Karunamuni and Jones (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

\[ \hat{f}_{ZKJ}(x) = \frac{1}{nh} \sum_{i=1}^{n} \left\{ K \left( \frac{x - X_i}{h} \right) + K \left( \frac{x + g_n(X_i)}{h} \right) \right\} \] (3.3)

where

\[ g_n(x) = x + \frac{x^2}{2} + \frac{Ax^3}{3} + \frac{x^4}{4} \]

13
Note this $g_n$ is very similar to our own, the main difference being it is not dependent on the point of estimation. 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 $d = f^{(1)}(0)/f(0)$, and the methodology used is the same as that given by (2.10) to (2.13).

The boundary kernel with bandwidth variation is defined (see Zhang and Karunamuni, 1998) as

$$
\hat{f}_B(x) = \frac{1}{nh_c} \sum_{i=1}^{n} K_{(c/b(c))} \left( \frac{x - X_i}{h_c} \right) \tag{3.4}
$$

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

$$
K_{(c)}(t) = \frac{12}{(1 + c)^4(1 + t)} \left\{ (1 - 2c)t + \frac{3c^2 - 2c + 1}{2} \right\} I_{[-1,c]} \tag{3.5}
$$

Zhang and Karunamuni 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.

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

$$
K_{(c)}(t) = \frac{12(1 - t^2)}{(1 + c)^4(3c^2 - 18c + 19)} \left\{ 8 - 16c + 24c^2 - 12c^3 + t(15 - 30c + 15c^2) \right\} I_{[-1,c]} \tag{3.6}
$$
in (3.4) and call the resulting estimator $\hat{f}_{LL}(x)$. In this case the bandwidth variation function is $b(c) = 1.86174 - .86174c$.

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

$$
\hat{f}_{JF}(x) = \bar{f}(x) \exp \left\{ \hat{f}(x) - 1 \right\} \tag{3.7}
$$
where \( \tilde{f}(x) \) is the cut-and-normalized kernel of (3.2). This version of the J&F estimator takes \( \hat{f} \) to be some boundary kernel estimate of \( f \), here we use \( \hat{f}(x) = \hat{f}_B(x) \) as defined by (3.5).

Cowling and Hall’s pseudo-data estimator generates data beyond the left endpoint of the support of the density. Their estimator (hereafter C&H) is defined as

\[
\hat{f}_{CH}(x) = \frac{1}{nh} \left[ \sum_{i=1}^{n} K\left( \frac{x - X_i}{h} \right) + \sum_{i=1}^{m} K\left( \frac{x - X_{(-i)}}{h} \right) \right]
\]

(3.8)

where \( m \) is an integer of larger order than \( nh \) but smaller order than \( n \); we chose \( m = n^{9/10} \). Further \( X_{(-i)} \) is given by their three-point rule

\[
X_{(-i)} = -5X_{(i/3)} - 4X_{(2i/3)} + \frac{10}{3}X_{(i)}
\]

where, for real \( t > 0 \), \( X_{(t)} \) linearly interpolates among the values \( 0, X_{(1)}, \ldots, X_{(n)} \), where \( X_{(i)} \) represents the \( i^{th} \) order statistic.

Throughout our simulations, whenever a kernel of order two was required we used the Epanechnikov kernel \( K(t) = \frac{3}{4}(1 - t^2)I_{[-1,1]} \). Note that outside of the boundary, the majority of the estimators presented reduce to the regular kernel method given by (1.1) with this \( K \). The only exception is Cowling and Hall’s, which is a close approximation to the kernel estimator away from the origin.

Each estimator was tested over various shapes of density 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 \( f(0) = 0 \) and Densities 2, 3, & 4 satisfy \( f(0) > 0 \) but \( f^{(1)}(0) = 0 \), \( f^{(1)}(0) > 0 \) and \( f^{(1)}(0) < 0 \), respectively.

In all simulations a sample size of \( n = 200 \) was used. The bandwidth was taken to be
the optimal global bandwidth of the regular kernel estimator (1.1), given by the formula

\[
h = \left\{ \frac{\int_{-1}^{1} K(t)^2 dt}{\left[ \int_{-1}^{1} t^2 K(t) dt \right]^2 \int [f^{(2)}(x)]^2 dx} \right\}^{1/5} n^{-1/5},
\]  
(3.9)

(see Silverman, 1986, Ch. 3). This particular choice was made so that it is possible to compare the different estimators without regard to bandwidth effects.

For each density we have calculated the bias, variance and MSE of the estimators at the points \( \{ \frac{k}{10} h : k = 0, \ldots, 10 \} \) over the boundary. The MISE is also given over the region \([0, h]\). The results are all averaged over 1000 iterations, and are presented in Tables 1 through 4. We have also graphed the performance of ten typical performances of each estimator over the boundary region and part of the interior. These are presented in Figures 1 through 4. In these we refer to our method as the new method, with and without the use of bandwidth variation.

Tables 1 - 4, and Figures 1 - 4 about here.

In Density 1, with \( f(0) = 0 \), our estimator behaves quite nicely. At zero both versions are the clear winners in terms of MSE, although Hall and Park’s estimator is not far behind. From Figure 1, we see that our method underestimates \( f \) to the right of zero, and this is further reflected in the MISE where Hall and Park’s estimator takes the lead. Still our method with and without bandwidth variation take second and third place, respectively, with an MISE of about half of the next-best J&F estimator. This is followed by the boundary kernel and LL method, the Z,K&J method and the C&H method. It should be noted that the the boundary kernel and LL method are, as is often the case, usually negative near zero. Most importantly though, the good performance of our estimators proves that the methodology developed in Section 2.1 is applicable even in the case \( f(0) = 0 \), which theoretically was a possible cause for concern.

For Density 2, with \( f^{(1)}(0) = 0 \) and \( f(0) > 0 \), our estimator is the clear winner in
terms of MSE at zero and the MISE over the boundary. Our estimator without bandwidth variation does especially well, and bests the other estimators by a significant amount. This is even obvious from Figure 2. Rather strangely though, the BVF proves to have a detrimental effect for this density. The variance increases as usual, but unfortunately so too does the bias. Still, our estimator with bandwidth variation puts in a better showing than the remaining competitors, in terms of both MSE at zero and the MISE. It is somewhat surprising that Hall and Park's estimator is much weaker for this density, especially given that it and our estimators are all special cases of the same general form (2.1).

Density 3 proves to be difficult for all of the estimators. This is most evident from Figure 3, in which we see a large amount of variability. In terms of MISE all the estimators perform similarly except for H&P, which does slightly worse than the rest, and C&H, which does poorly over the whole boundary. At zero, both versions of our estimator are the best overall but only by a slim margin this time. In terms of MISE it is the LL and J&F methods that are the winners, but again only by a small amount. Once again the bandwidth variation has a negative effect on our estimator, although in this case it is not very pronounced.

Our estimator has its poorest showing on the exponential shape of Density 4. Without bandwidth variation it is by far the worst estimator at zero, and even with the BVF it comes in at the second-to-last place. The improvement caused by the BVF is certainly worth noting though. In terms of MISE our estimator with BVF is comparable to the others (except C&H), and on the right half of the boundary region actually does quite well. The Z,K&J estimator is quite accurate at zero and over the entire boundary, and the performance of the boundary kernel, LL method, J&F and H&P estimators do not lag far behind. Even Cowling and Hall's estimator does well at zero, but pays the price for it in MISE over the rest of the boundary region.
4 Data Analysis

We tested our proposed estimator over two datasets. The first consists of 35 measurements of the average December precipitation in Des Moines, Iowa from 1961 to 1995. The bandwidth $h = 0.425$ was produced by cross-validation (see Bowman and Azzalini, 1997). Figure 5 shows our proposed estimator without BVF (solid line) along with Hall and Park’s (dashed line). Over most of the boundary region they are similar, but around zero they disagree on the shape of the density. Hall and Park’s estimator seems to believe there’s a local minimum to the right of zero, but the histogram and the rug show otherwise. A larger value for $h$ would probably diminish this behavior. We have observed that the boundary kernel, LL method and J&F estimators tend to linearly approximate the density over the boundary region, with little regard for curvature.

The second dataset is comprised of 365 wind speed measurements taken daily in 1993 at 2 AM in Eindhoven, the Netherlands (source: Royal Netherlands Meteorological Institute). Intuition would suggest that the underlying density $f$ should be similar to Density 3, with $f(0) > 0$ and $f^{(1)}(0) > 0$, so that there is substantial mass at zero but the mode is slightly to the right. We used bandwidth $h = 2.5$ which we chose subjectively. Values suggested by cross-validation typically seem to undersmooth this data, hence the subjective choice. We have observed that for this dataset the shape of the estimators is fairly insensitive to the choice of bandwidth, provided that it is large enough, and so we believe our choice is justified. Figure 6 shows our proposed estimator without BVF (solid line) and the LL method (dashed line) applied to the wind data. Both estimators capture the hypothesized shape of the density and are similar over the boundary region. However our estimator clearly produces a smoother curve.
5 Concluding Remarks

Marron and Ruppert’s (1994) excellent work on kernel density estimation was among the first to introduce a transformation technique in boundary correction (also see Wand, Marron and Ruppert, 1991). Their sophisticated transformation methodology, however, has been labelled “complicated” by some researchers; see, e.g., Jones and Foster (1996). In this paper, we presented an easy-to-implement, general transformation method given by (2.1). This class of estimators possesses certain desirable properties, such as being everywhere non-negative and having unit integral asymptotically, and can be tailored specifically to combat the boundary effects found in the traditional kernel estimator. The proposed estimator defined by (2.15) and the boundary corrected estimator of Hall and Park (2002) are special cases of the preceding general technique. Beyond this, both estimators are locally adaptive in the sense that the amount of transformation depends upon the point of estimation, and, even further, both transformations are estimated from the data itself. We believe the latter point is especially important, as it allows for all-important reductions in bias while, at the same time, maintaining low variance. The transformation that we have chosen at (2.14) is a convenient and computationally easy one, but by no means is it the result of an exhaustive search of functions satisfying the conditions (2.5). In fact, we conjecture that it is possible to construct different transformations satisfying these conditions but with much improved performance, in the sense that they will be more robust to the shape of the density.

Acknowledgements: This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. We also thank Drs. B.U. Park and S. Zhang for some personal correspondence.
A Appendix: Proofs

Proof of Lemma 2.1: For $x = ch, 0 \leq c \leq 1$, we have from (2.1)

$$
\begin{align*}
E \tilde{f}_n(x) \int_{-1}^{c} K(t) dt &= \frac{1}{h} E K \left( \frac{x - g(X_i)}{h} \right) \\
&= \frac{1}{h} \int_0^{\infty} K \left( \frac{x - g(y)}{h} \right) f(y) dy \\
&= \int_{-1}^{c} K(t) \frac{f(g^{-1}((c-t)h))}{g^{(1)}(g^{-1}((c-t)h))} dt.
\end{align*}
$$

(A.1)

Letting $q(t) = g^{-1}((c-t)h)$ we have

$$
q^{(1)}(t) = \frac{-h}{g^{(1)}(q(t))}, \quad \text{and} \quad q^{(2)}(t) = \frac{-h^2 g^{(2)}(q(t))}{g^{(1)}(q(t))^3},
$$

so then a Taylor expansion of order 2 on the function $f(q(\cdot))/g^{(1)}(q(\cdot))$ at $t = c$ gives

$$
\begin{align*}
\frac{f(q(t))}{g^{(1)}(q(t))} &= \frac{f(q(c))}{g^{(1)}(q(c))} - (t-c)h \left[ \frac{f^{(1)}(q(c))g^{(1)}(q(c)) - f(q(c))g^{(2)}(q(c))}{[g^{(1)}(q(c))]^3} \right] \\
&\quad + \frac{1}{2} h^2 (t-c)^2 \left\{ \left[ \frac{f^{(2)}(q(c))}{g^{(1)}(q(c))} - f^{(1)}(q(c)) \right] \frac{g^{(2)}(q(c))}{[g^{(1)}(q(c))]^2} \\
&\quad - f(q(c)) \frac{g^{(3)}(q(c))g^{(1)}(q(c)) - g^{(2)}(q(c))^2}{[g^{(1)}(q(c))]^3} \right\} \times [g^{(1)}(q(c))] \\
&\quad - 2 \left( f^{(1)}(q(c)) - f(q(c)) \right) \frac{g^{(2)}(q(c))}{g^{(1)}(q(c))} \frac{g^{(2)}(q(c))}{g^{(1)}(q(c))} \right\} + o(h^2) \\
&= f(0) - (t-c)h[f^{(1)}(0) - f(0)g^{(2)}(0)] \\
&\quad + \frac{1}{2} h^2 (t-c)^2 [f^{(2)}(0) - f(0)g^{(3)}(0) - 3g^{(2)}(0)(f^{(1)}(0) - f(0)g^{(2)}(0))] \\
&\quad + o(h^2),
\end{align*}
$$

(A.2)
with the last equality following from \( q(c) = g^{-1}(0) = 0 \) and \( g^{(1)}(q(c)) = g^{(1)}(0) = 1 \). Also, by the existence and continuity of \( f^{(2)}(\cdot) \) near 0, for \( x = ch \) we have

\[
f(0) = f(x) - chf^{(1)}(x) + \frac{(ch)^2}{2}f^{(2)}(x) + o(h^2), \tag{A.3}
\]

\[
f^{(1)}(x) = f^{(1)}(0) + chf^{(2)}(0) + o(h), \tag{A.4}
\]

and

\[
f^{(2)}(x) = f^{(2)}(0) + o(1). \tag{A.5}
\]

Now from (A.1) to (A.5), we obtain for \( x = ch, 0 \leq c \leq 1 \),

\[
\mathbb{E} \hat{f}_n(x) \int_{-1}^{c} K(t)dt = \int_{-1}^{c} K(t)dt (f(x) - chf^{(1)}(x) + \frac{(ch)^2}{2}f^{(2)}(x) + O(h^2))
\]

\[
- h \int_{-1}^{c} (t - c)K(t)dt (f^{(1)}(0) - f(0)g^{(2)}(0))
\]

\[
+ \frac{h^2}{2} \int_{-1}^{c} (t - c)^2K(t)dt \left\{ f^{(2)}(0) - f(0)g^{(3)}(0) \\
- 3g^{(2)}(0)(f^{(1)}(0) - f(0)g^{(2)}(0)) \right\}
\]

\[
- (ch)^2 f^{(2)}(0) \int_{-1}^{c} K(t)dt + o(h^2)
\]

\[
= f(x) \int_{-1}^{c} K(t)dt - h \left\{ f^{(1)}(0) - f(0)g^{(2)}(0) \int_{-1}^{c} tK(t)dt \\
+ f(0)g^{(2)}(0) \int_{-1}^{c} cK(t)dt \right\}
\]

\[
+ \frac{h^2}{2} \left\{ \left( \int_{-1}^{c} (t - c)^2K(t)dt \right) \left[ f^{(2)}(0) - f(0)g^{(3)}(0) \\
- 3g^{(2)}(0)(f^{(1)}(0) - f(0)g^{(2)}(0)) \right] \\
- 2c^2 f^{(2)}(0) \int_{-1}^{c} K(t)dt \right\} + O(h^2). \tag{A.6}
\]
The proof of (2.2) is now completed by dividing both sides of (A.6) by \( \int_{-1}^c K(t)dt \). Note that if we replace \( h \) by \( h_c = b(c)h \), then in (A.1) the change of variables transforms the \( c \)'s into \( c/b(c) \) from equations (A.1) through to (A.6).

In order to prove (2.3) we first note that from (2.1),

\[
\text{Var } \tilde{f}_n(x) = \left( \int_{-1}^c K(t)dt \right)^{-2} \text{Var} \left\{ \frac{1}{nh} \sum_{i=1}^n K \left( \frac{x - g(X_i)}{h} \right) \right\}
\]

\[
= \left( \int_{-1}^c K(t)dt \right)^{-2} \frac{1}{nh^2} \left\{ E K^2 \left( \frac{x - g(X_i)}{h} \right) - \left( E K \left( \frac{x - g(X_i)}{h} \right) \right)^2 \right\}
\]

\[
= \left( \int_{-1}^c K(t)dt \right)^{-2} (I_1 + I_2),
\]  

(A.7)

where

\[
I_1 = \frac{1}{nh^2} E K^2 \left( \frac{x - g(X_i)}{h} \right)
\]

\[
= \frac{1}{nh} \int_{-1}^c K^2(t) \frac{f(g^{-1}((c-t)h))}{g^{(1)}(g^{-1}((c-t)h))} dt
\]

\[
= \frac{1}{nh} f(0) \int_{-1}^c K^2(t) dt + o \left( \frac{1}{nh} \right),
\]  

(A.8)

from (A.2). Similarly,

\[
I_2 = -\frac{1}{nh^2} \left( E K \left( \frac{x - g(X_i)}{h} \right) \right)^2
\]

\[
= -\frac{1}{nh^2} \left( \frac{h}{nh} \int_{-1}^c K(t) \frac{f(g^{-1}((c-t)h))}{g^{(1)}(g^{-1}((c-t)h))} dt \right)^2
\]

\[
= o \left( \frac{1}{nh} \right),
\]  

(A.9)

again from (A.2). By combining (A.7) to (A.9) we now complete the proof of (2.3). Once again the effect of changing \( h \) to \( h_c \) is to change \( c \) to \( c/b(c) \).

**Lemma A.2.** Let \( \hat{d} \) be defined by (2.10). Suppose that \( f(x) > 0 \) for \( x = 0, h \) and that \( f^{(2)} \)
is continuous in a neighbourhood of 0. Then

\[ E \left[ (\hat{d} - d)^4 \right] |X_i = x_i] = O(h^4) \]

for any integer \( i, 1 \leq i \leq n \), where \( d \) is given by (2.8).

**Proof.** Follows from Lemma A.2 of Zhang, Karunamuni and Jones (1999).

**Proof of Theorem 2.1:** For \( x = ch, 0 \leq c \leq 1 \), we write

\[ E \hat{f}_n(x) - f(x) = I_3 + I_4, \]  

(A.10)

where

\[ I_3 = E \hat{f}_n(x) - E \hat{f}_n(x) \]  

(A.11)

and

\[ I_4 = E \hat{f}_n(x) - f(x), \]  

(A.12)

where \( \hat{f}_n(x) \) is given by (2.1). From Lemma 2.1, we obtain

\[ |I_4| = o(h^2) \]  

(A.13)

when \( g \) is chosen to satisfy (2.4). By an application of Taylor’s expansion of order 1 on \( K \),
we obtain from (A.11)

\[
|I_3| \leq \frac{(\int_{-1}^{c} K(t) dt)^{-1}}{nh} \sum_{i=1}^{n} \mathbb{E} \left| K \left( \frac{x - \hat{g}_c(X_i)}{h} \right) - K \left( \frac{x - g_c(X_i)}{h} \right) \right|
\]

\[
= \frac{(\int_{-1}^{c} K(t) dt)^{-1}}{nh} \sum_{i=1}^{n} \mathbb{E} \left| \left( g_c(X_i) - \hat{g}_c(X_i) \right) K^{(1)} \left( \frac{x - g_c(X_i) + \varepsilon (g_c(X_i) - \hat{g}_c(X_i))}{h} \right) \right|
\]

(A.14)

where \(0 < \varepsilon < 1\) is a constant. Note that for any constants \(d\) and \(l_c\), we have for any \(y \geq 0\),

\[
g_c(y) = y + \frac{d}{2} l_c y^2 + (dl_c)^2 y^3
\]

\[
= y \left( 1 + \frac{d}{2} l_c y + (dl_c)^2 y^2 \right)
\]

\[
= y \left[ \left( dl_c y + \frac{1}{4} \right)^2 + \frac{15}{16} \right]
\]

\[
\geq \frac{15}{16} y.
\]

(A.15)

Thus \(g_c(y) \geq h\) for \(y \geq \frac{16}{15} h\). Therefore, \(\varepsilon \hat{g}_c(X_i) + (1 - \varepsilon) g_c(X_i) \geq h\) for \(X_i \geq \frac{16}{15} h\). Since \(K\) vanishes outside \([-1, 1]\), from (A.14) we have

\[
|I_3| \leq \frac{(\int_{-1}^{c} K(t) dt)^{-1}}{nh} \sum_{i=1}^{n} \left\{ \mathbb{E} \left| \left( g_c(X_i) - \hat{g}_c(X_i) \right) \right| \right.
\]

\[
\times \left| K^{(1)} \left( \frac{x - \varepsilon \hat{g}_c(X_i) - (1 - \varepsilon) g_c(X_i)}{h} \right) \right| \mathbb{I}[0 \leq X_i \leq \rho h] \right\}
\]

\[
\leq \frac{C}{nh^2} \sum_{i=1}^{n} \mathbb{E} \left| \hat{g}_c(X_i) - g_c(X_i) \right| \mathbb{I}[0 \leq X_i \leq \rho h],
\]

(A.16)

where \(\rho = 16/15\) and \(C = \sup_{|t| \leq 1} |K^{(1)}(t)|\). Now by the definitions of \(g_c\) and \(\hat{g}_c\) (see (2.6)
and (2.14) we obtain

\[ E |\hat{g}_c(X_i) - g_c(X_i)| I[0 \leq X_i \leq \rho h] \]
\[ = E \left| \frac{l_c}{2}(\hat{d} - d)X_i^2 + l_c^2(\hat{d}^2 - d^2)X_i^3 \right| I[0 \leq X_i \leq \rho h] \]
\[ \leq \frac{|l_c|}{2}h^2 \rho^2 E |\hat{d} - d| I[0 \leq X_i \leq \rho h] \]
\[ + l_c^2(\rho h)^3 E |\hat{d}^2 - d^2| I[0 \leq X_i \leq \rho h]. \]  \tag{A.17}

The Cauchy-Schwartz inequality and Lemma A.2 yield that

\[ E \left| |\hat{d} - d|^k \right| X_i = x_i = O(h^k) \]  \tag{A.18}

for \( 1 \leq k \leq 4 \). From (A.18) we have

\[ E |\hat{d} - d| I[0 \leq X_i \leq \rho h] = E \left\{ E[|\hat{d} - d| I[0 \leq X_i \leq \rho h]|X_i = x_i] \right\} \]
\[ = E \left\{ I[0 \leq X_i \leq \rho h] E[|\hat{d} - d| |X_i = x_i] \right\} \]
\[ \leq O(h) E I[0 \leq X_i \leq \rho h] \]
\[ = O(h^2), \]  \tag{A.19}

where the last equality follows from the fact that \( \lim_{n \to \infty} h^{-1} E[0 \leq X_i \leq \rho h] = f(0) \) for each \( 0 \leq c \leq 1 \). Similarly, we again obtain from (A.18) that

\[ E |\hat{d}^2 - d^2| I[0 \leq X_i \leq \rho h] = E |\hat{d} - d| |\hat{d} + d| I[0 \leq X_i \leq \rho h] \]
\[ = E |\hat{d} - d| |\hat{d} - d + 2d| I[0 \leq X_i \leq \rho h] \]
\[ \leq E |\hat{d} - d|^2 I[0 \leq X_i \leq \rho h] + 2 |d| E |\hat{d} - d| I[0 \leq X_i \leq \rho h] \]
\[ = O(h^2). \]  \tag{A.20}
By combining (A.16) to (A.20), we now have $|I_3| = O(h^2)$. The proof is now completed by the preceding result, (A.13) and (A.10).

We now prove (2.17). First write

\[
\left( \int_{-1}^{c} K(t) dt \right)^2 \Var \hat{f}_n(x) = \frac{1}{(nh)^2} \Var \left\{ \sum_{i=1}^{n} K \left( \frac{x - \hat{g}_c(X_i)}{h} \right) \right\} \\
= \frac{1}{(nh)^2} \Var \left\{ \sum_{i=1}^{n} \left[ K \left( \frac{x - \hat{g}_c(X_i)}{h} \right) - K \left( \frac{x - g_c(X_i)}{h} \right) \right] \right\} \\
\leq 2(I_5 + I_6),
\]

(A.21)

where $g_c$ is given by (2.6),

\[
I_5 = \frac{1}{(nh)^2} \Var \left\{ \sum_{i=1}^{n} \left[ K \left( \frac{x - \hat{g}_c(X_i)}{h} \right) - K \left( \frac{x - g_c(X_i)}{h} \right) \right] \right\},
\]

(A.22)

and

\[
I_6 = \frac{1}{(nh)^2} \Var \sum_{i=1}^{n} K \left( \frac{x - g_c(X_i)}{h} \right).
\]

(A.23)

From Lemma 2.1, we have

\[
I_6 = \frac{f(0)}{nh} \int_{-1}^{c} K^2(t) dt + o \left( \frac{1}{nh} \right).
\]

(A.24)
Now consider $I_5$. By an application of Taylor’s expansion of order 1 on $K$, we obtain

$$I_5 \leq \frac{1}{(nh)^2} \mathbb{E} \left\{ \sum_{i=1}^{n} \left[ K \left( \frac{x - \hat{g}_c(X_i)}{h} \right) - K \left( \frac{x - g_c(X_i)}{h} \right) \right]^2 \right\}$$

$$= \frac{1}{(nh)^2} \mathbb{E} \left\{ \sum_{i=1}^{n} \frac{g_c(X_i) - \hat{g}_c(X_i)}{h} K^{(1)} \left( \frac{x - g_c(X_i) + \varepsilon(g_c(X_i) - \hat{g}_c(X_i))}{h} \right) \right\}^2$$

$$\leq \frac{2}{n^2 h^4} \sum_{i=1}^{n} \mathbb{E} \left\{ (g_c(X_i) - \hat{g}_c(X_i))^2 \left[ K^{(1)} \left( \frac{x - g_c(X_i) + \varepsilon(g_c(X_i) - \hat{g}_c(X_i))}{h} \right) \right]^2 \right\}$$

$$\leq \frac{C}{n^2 h^4} \sum_{i=1}^{n} \mathbb{E}(\hat{g}_c(X_i) - g_c(X_i))^2 I[0 \leq X_i \leq \rho h], \quad (A.25)$$

using an argument similar to (A.16) above, where $0 < \varepsilon < 1$, $\rho = 16/15$ and $C > 0$ are all constants independent of $n$. Similar to (A.17) we can write

$$\mathbb{E}(\hat{g}_c(X_i) - g_c(X_i))^2 I[0 \leq X_i \leq \rho h]$$

$$= \mathbb{E} \left[ \frac{l_c}{2} (\hat{d} - d) X_i^2 + l^2_c (\hat{d}^2 - d^2) X_i^3 \right]^2 I[0 \leq X_i \leq \rho h]$$

$$\leq \frac{l^2_c}{2} (\rho h)^4 \mathbb{E}(\hat{d} - d)^2 I[0 \leq X_i \leq \rho h] + 2l^4_c (\rho h)^6 \mathbb{E}(\hat{d}^2 - d^2)^2 I[0 \leq X_i \leq \rho h]$$

$$\leq O(h^4 h^2 h) + O(h^6 h^2 h)$$

$$= O(h^7). \quad (A.26)$$

Now combining (A.25) and (A.26), we obtain

$$I_5 = o \left( \frac{1}{nh} \right) \quad (A.27)$$

From (A.21), (A.24) and (A.27) we complete the proof.
References


Table 1: MSE & MISE values for density 1, \( f(x) = \frac{2\sqrt{x^2}}{\pi(x^2+1)}, \ x \geq 0 \), with \( n = 200, h = 0.476956 \) & 1000 iterations

<table>
<thead>
<tr>
<th>( c )</th>
<th>New Method</th>
<th>H&amp;P method</th>
<th>Z,K&amp;J method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>Var</td>
<td>MSE</td>
</tr>
<tr>
<td>0.0</td>
<td>0.000872</td>
<td>0.000014</td>
<td>0.000015</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.003242</td>
<td>0.000023</td>
<td>0.000023</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.004458</td>
<td>0.000036</td>
<td>0.000056</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.011321</td>
<td>0.000057</td>
<td>0.000185</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.020434</td>
<td>0.000103</td>
<td>0.000521</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.030794</td>
<td>0.000175</td>
<td>0.001123</td>
</tr>
<tr>
<td>0.6</td>
<td>-0.039783</td>
<td>0.000269</td>
<td>0.001851</td>
</tr>
<tr>
<td>0.7</td>
<td>-0.040563</td>
<td>0.000398</td>
<td>0.002043</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.041903</td>
<td>0.000556</td>
<td>0.00092</td>
</tr>
<tr>
<td>0.9</td>
<td>0.027441</td>
<td>0.000871</td>
<td>0.001624</td>
</tr>
<tr>
<td>1.0</td>
<td>0.051496</td>
<td>0.001136</td>
<td>0.003788</td>
</tr>
</tbody>
</table>

MISE 0.000488 0.000453 0.000350 0.001142

<table>
<thead>
<tr>
<th>( c )</th>
<th>Boundary Kernel</th>
<th>LL method</th>
<th>J&amp;F method</th>
<th>C&amp;H method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>Var</td>
<td>MSE</td>
<td>Bias</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.053579</td>
<td>0.000635</td>
<td>0.003506</td>
<td>-0.057007</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.029386</td>
<td>0.000485</td>
<td>0.001348</td>
<td>-0.033173</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.008409</td>
<td>0.000379</td>
<td>0.00045</td>
<td>-0.012381</td>
</tr>
<tr>
<td>0.3</td>
<td>0.000355</td>
<td>0.000318</td>
<td>0.000405</td>
<td>0.005372</td>
</tr>
<tr>
<td>0.4</td>
<td>0.023925</td>
<td>0.000301</td>
<td>0.000873</td>
<td>0.020099</td>
</tr>
<tr>
<td>0.5</td>
<td>0.035345</td>
<td>0.000329</td>
<td>0.001578</td>
<td>0.031847</td>
</tr>
<tr>
<td>0.6</td>
<td>0.043709</td>
<td>0.000401</td>
<td>0.002312</td>
<td>0.040706</td>
</tr>
<tr>
<td>0.7</td>
<td>0.049181</td>
<td>0.000518</td>
<td>0.002957</td>
<td>0.046839</td>
</tr>
<tr>
<td>0.8</td>
<td>0.052023</td>
<td>0.000667</td>
<td>0.003856</td>
<td>0.050492</td>
</tr>
<tr>
<td>0.9</td>
<td>0.052621</td>
<td>0.000886</td>
<td>0.003655</td>
<td>0.05198</td>
</tr>
<tr>
<td>1.0</td>
<td>0.051496</td>
<td>0.001136</td>
<td>0.003788</td>
<td>0.051496</td>
</tr>
</tbody>
</table>

MISE 0.000970 0.000979 0.000910 0.001240
Table 2: MSE & MISE values for density 2, $f(x) = \sqrt{2/\pi}e^{-x^2/2}$, $x \geq 0$, with $n = 200, h = .707481$ & 1000 iterations

<table>
<thead>
<tr>
<th>$c$</th>
<th>New Method with BVF (2.21)</th>
<th>H&amp;K</th>
<th>Z,K&amp;J</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias Var MSE</td>
<td>Bias Var MSE</td>
<td>Bias Var MSE</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.003343 0.003876 0.003775</td>
<td>0.019758 0.007893 0.008283</td>
<td>0.011316 0.015845 0.015973</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.009302 0.003473 0.003559</td>
<td>-0.008283 0.005747 0.005815</td>
<td>-0.001462 0.007893 0.008283</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.014859 0.003048 0.00333</td>
<td>-0.001986 0.004309 0.004313</td>
<td>-0.012174 0.006308 0.006186</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.019714 0.002598 0.002987</td>
<td>-0.010713 0.003251 0.003366</td>
<td>-0.020694 0.005739 0.004167</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.02353 0.002083 0.002636</td>
<td>-0.017415 0.002428 0.002731</td>
<td>-0.026948 0.002361 0.003087</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.02594 0.001616 0.002283</td>
<td>-0.021879 0.001783 0.001901</td>
<td>-0.030887 0.001554 0.002508</td>
</tr>
<tr>
<td>0.6</td>
<td>-0.026657 0.001236 0.001946</td>
<td>-0.021417 0.001319 0.001901</td>
<td>-0.032528 0.001128 0.002186</td>
</tr>
<tr>
<td>0.7</td>
<td>-0.025578 0.000971 0.001625</td>
<td>-0.024156 0.001015 0.001598</td>
<td>-0.031896 0.000917 0.002508</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.022944 0.000801 0.001337</td>
<td>-0.02232 0.000831 0.001329</td>
<td>-0.028943 0.000822 0.001659</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.019156 0.000737 0.001104</td>
<td>-0.019012 0.000744 0.001105</td>
<td>-0.023446 0.000777 0.001659</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.015226 0.000734 0.000966</td>
<td>-0.015226 0.000734 0.000966</td>
<td>-0.015226 0.000734 0.000966</td>
</tr>
<tr>
<td></td>
<td>MISE</td>
<td>0.001637</td>
<td>0.002037</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$c$</th>
<th>Boundary Kernel</th>
<th>LL method</th>
<th>J&amp;F method</th>
<th>C&amp;H method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias Var MSE</td>
<td>Bias Var MSE</td>
<td>Bias Var MSE</td>
<td>Bias Var MSE</td>
</tr>
<tr>
<td>0.0</td>
<td>0.052554 0.009785 0.012637</td>
<td>0.054271 0.009811 0.012757</td>
<td>0.059196 0.010828 0.014312</td>
<td>0.051713 0.015374 0.015621</td>
</tr>
<tr>
<td>0.1</td>
<td>0.030104 0.007893 0.008799</td>
<td>0.032721 0.007649 0.00872</td>
<td>0.034929 0.008529 0.009749</td>
<td>0.020491 0.015789 0.016209</td>
</tr>
<tr>
<td>0.2</td>
<td>0.016141 0.006148 0.006283</td>
<td>0.014895 0.005827 0.006049</td>
<td>0.015017 0.006557 0.006782</td>
<td>0.021245 0.014949 0.01541</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.003006 0.00464 0.004649</td>
<td>0.0007 0.004321 0.004322</td>
<td>-0.000696 0.004882 0.004882</td>
<td>0.019078 0.013009 0.013373</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.013902 0.00337 0.003564</td>
<td>-0.010011 0.003107 0.003207</td>
<td>-0.012409 0.003498 0.003652</td>
<td>0.015118 0.010239 0.010521</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.021273 0.002338 0.00279</td>
<td>-0.01753 0.002162 0.002467</td>
<td>-0.020364 0.002395 0.00281</td>
<td>0.010033 0.007316 0.007417</td>
</tr>
<tr>
<td>0.6</td>
<td>-0.025365 0.001542 0.002186</td>
<td>-0.021881 0.001463 0.001942</td>
<td>-0.02453 0.001562 0.00218</td>
<td>0.004629 0.004624 0.004646</td>
</tr>
<tr>
<td>0.7</td>
<td>-0.026648 0.000984 0.001685</td>
<td>-0.023597 0.000987 0.001544</td>
<td>-0.02621 0.000988 0.001675</td>
<td>-0.000158 0.002568 0.002568</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.024907 0.000664 0.001284</td>
<td>-0.022919 0.000713 0.001239</td>
<td>-0.024798 0.000664 0.001279</td>
<td>-0.004394 0.001343 0.001362</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.02105 0.00058 0.001023</td>
<td>-0.020126 0.000625 0.001013</td>
<td>-0.02106 0.00058 0.001022</td>
<td>-0.008199 0.000802 0.00087</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.015226 0.000734 0.000966</td>
<td>-0.015226 0.000734 0.000966</td>
<td>-0.015226 0.000734 0.000966</td>
<td>-0.011772 0.00072 0.000859</td>
</tr>
<tr>
<td></td>
<td>MISE</td>
<td>0.002736</td>
<td>0.002616</td>
<td>0.002915</td>
</tr>
</tbody>
</table>
Table 3: MSE & MISE values for density 3, $f(x) = (x^2 + 2x + \frac{1}{2})e^{-2x}, \ x \geq 0$, with $n = 200, h = .467044$ & 1000 iterations

<table>
<thead>
<tr>
<th>$c$</th>
<th>New Method with BVF (2.21)</th>
<th>H&amp;P method</th>
<th>Z,K&amp;J method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>Var</td>
<td>MSE</td>
</tr>
<tr>
<td>0.0</td>
<td>0.081894</td>
<td>0.01417</td>
<td>0.020876</td>
</tr>
<tr>
<td>0.1</td>
<td>0.047603</td>
<td>0.011734</td>
<td>0.014</td>
</tr>
<tr>
<td>0.2</td>
<td>0.02069</td>
<td>0.00946</td>
<td>0.009888</td>
</tr>
<tr>
<td>0.3</td>
<td>0.000184</td>
<td>0.007434</td>
<td>0.007434</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.014534</td>
<td>0.005728</td>
<td>0.005939</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.024578</td>
<td>0.004404</td>
<td>0.005008</td>
</tr>
<tr>
<td>0.6</td>
<td>-0.030558</td>
<td>0.003427</td>
<td>0.004361</td>
</tr>
<tr>
<td>0.7</td>
<td>-0.033211</td>
<td>0.002784</td>
<td>0.003887</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.03317</td>
<td>0.002403</td>
<td>0.003503</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.031113</td>
<td>0.002219</td>
<td>0.003187</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.02764</td>
<td>0.002166</td>
<td>0.00293</td>
</tr>
</tbody>
</table>

| MISE | 0.003194 | 0.003246 | 0.003719 | 0.003230 |

<table>
<thead>
<tr>
<th>$c$</th>
<th>Boundary kernel</th>
<th>LL method</th>
<th>J&amp;F method</th>
<th>C&amp;H method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>Var</td>
<td>MSE</td>
<td>Bias</td>
</tr>
<tr>
<td>0.0</td>
<td>0.090833</td>
<td>0.013616</td>
<td>0.021867</td>
<td>0.093012</td>
</tr>
<tr>
<td>0.1</td>
<td>0.051433</td>
<td>0.011092</td>
<td>0.013737</td>
<td>0.055425</td>
</tr>
<tr>
<td>0.2</td>
<td>0.019966</td>
<td>0.008874</td>
<td>0.009272</td>
<td>0.02533</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.004323</td>
<td>0.006962</td>
<td>0.006981</td>
<td>0.001964</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.022144</td>
<td>0.005357</td>
<td>0.005848</td>
<td>-0.0154</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.034167</td>
<td>0.004509</td>
<td>0.005226</td>
<td>-0.027447</td>
</tr>
<tr>
<td>0.6</td>
<td>-0.041014</td>
<td>0.003067</td>
<td>0.004749</td>
<td>-0.034818</td>
</tr>
<tr>
<td>0.7</td>
<td>-0.043268</td>
<td>0.002382</td>
<td>0.004254</td>
<td>-0.038133</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.041465</td>
<td>0.002003</td>
<td>0.003723</td>
<td>-0.037859</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.036144</td>
<td>0.001931</td>
<td>0.003235</td>
<td>-0.034136</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.02764</td>
<td>0.002166</td>
<td>0.00293</td>
<td>-0.02764</td>
</tr>
</tbody>
</table>

| MISE | 0.003202 | 0.003172 | 0.003178 | 0.005177 |
Table 4: MSE & MISE values for density 4, \( f(x) = 2e^{-2x}, x \geq 0 \), with \( n = 200, h = .342128 \) & 1000 iterations

<table>
<thead>
<tr>
<th>c</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>-0.407711 0.011661 0.177889</td>
<td>-0.251166 0.027113 0.090197</td>
<td>-0.163295 0.048959 0.075624</td>
<td>-0.111047 0.017739 0.03007</td>
<td>-0.01189 0.026948 0.02729</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>-0.298916 0.011244 0.105695</td>
<td>-0.190106 0.019992 0.056133</td>
<td>-0.135996 0.029036 0.047531</td>
<td>-0.032552 0.027776 0.028836</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>-0.207629 0.010534 0.053634</td>
<td>-0.13523 0.015958 0.03885</td>
<td>-0.111047 0.017739 0.03007</td>
<td>0.018198 0.026948 0.02729</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>-0.133684 0.009447 0.027319</td>
<td>-0.088411 0.012311 0.020127</td>
<td>-0.088525 0.011427 0.019264</td>
<td>0.046014 0.02325 0.025367</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>-0.075902 0.008115 0.013876</td>
<td>-0.049852 0.009622 0.012107</td>
<td>-0.068011 0.007949 0.012575</td>
<td>0.056288 0.005045 0.008214</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>-0.032339 0.006746 0.005374</td>
<td>-0.018801 0.00745 0.007804</td>
<td>-0.049081 0.006078 0.008487</td>
<td>0.060093 0.011836 0.015447</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>-0.004047 0.005491 0.005492</td>
<td>0.005698 0.005842 0.005874</td>
<td>-0.031253 0.005107 0.006084</td>
<td>0.058719 0.007565 0.011013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.022387 0.004546 0.005047</td>
<td>0.024524 0.004707 0.005309</td>
<td>-0.01382 0.004643 0.004834</td>
<td>0.056288 0.005045 0.008214</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.037731 0.003951 0.005374</td>
<td>0.038099 0.004029 0.005481</td>
<td>0.004021 0.004376 0.004392</td>
<td>0.053484 0.003896 0.007657</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.046151 0.003682 0.005812</td>
<td>0.04606 0.003707 0.005828</td>
<td>0.02359 0.004114 0.004671</td>
<td>0.050394 0.003585 0.006125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.047141 0.003718 0.00594</td>
<td>0.047141 0.003718 0.00594</td>
<td>0.047141 0.003718 0.00594</td>
<td>0.047141 0.003718 0.00594</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MISE 0.010568 0.006747 0.006007 0.005838

Boundary Kernel  
LL method  
J&F method  
C&H method

<table>
<thead>
<tr>
<th>c</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
<th>Bias Var MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>-0.122779 0.01461 0.059684</td>
<td>-0.124788 0.014327 0.058999</td>
<td>-0.087819 0.051707 0.059419</td>
<td>-0.117587 0.036915 0.050741</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>-0.072619 0.035804 0.041077</td>
<td>-0.077706 0.033972 0.040011</td>
<td>-0.044846 0.041083 0.043094</td>
<td>0.033051 0.047054 0.04817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>-0.031209 0.028046 0.02902</td>
<td>-0.038674 0.026011 0.027507</td>
<td>-0.009866 0.031745 0.031843</td>
<td>0.14495 0.05563 0.076509</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.002033 0.021336 0.02134</td>
<td>-0.007091 0.019431 0.019481</td>
<td>0.017742 0.023738 0.024053</td>
<td>0.218324 0.060359 0.108205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.027645 0.015674 0.016438</td>
<td>0.017606 0.01413 0.014439</td>
<td>0.038585 0.017089 0.018578</td>
<td>0.257093 0.060156 0.126262</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.046132 0.011061 0.013189</td>
<td>0.035955 0.010001 0.011294</td>
<td>0.053209 0.011786 0.014617</td>
<td>0.26512 0.053462 0.123767</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.057966 0.007496 0.010856</td>
<td>0.04846 0.006946 0.009295</td>
<td>0.062901 0.007791 0.011647</td>
<td>0.248463 0.041646 0.10338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.063586 0.004979 0.009022</td>
<td>0.055592 0.004865 0.007955</td>
<td>0.06564 0.005056 0.009365</td>
<td>0.214548 0.027689 0.07372</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.063403 0.00351 0.00753</td>
<td>0.057754 0.003622 0.006997</td>
<td>0.064185 0.003509 0.007629</td>
<td>0.171081 0.015163 0.044432</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.057802 0.00039 0.006431</td>
<td>0.055154 0.003264 0.006306</td>
<td>0.057961 0.003083 0.006442</td>
<td>0.126166 0.007198 0.023116</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.047141 0.003718 0.00594</td>
<td>0.047141 0.003718 0.00594</td>
<td>0.047141 0.003718 0.00594</td>
<td>0.086854 0.004 0.01154</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MISE 0.006358 0.005948 0.006785 0.025894
Figure 1: Ten typical estimates of density 1, \( f(x) = \frac{2\sqrt{x^2}}{\pi(x^2+1)}, x \geq 0 \), with the optimal global bandwidth \( h = .476956 \).
Figure 2: Ten typical estimates of density $2$, $f(x) = \sqrt{2/\pi} e^{-x^2/2}$, with the optimal global bandwidth $h = .707481$. 
Figure 3: Ten typical estimates of density $3, f(x) = \frac{1}{2}(2x^2 + 4x + 1)e^{-2x}$, with the optimal global bandwidth $h = 0.467044$. 
Figure 4: Ten typical estimates of density 4, \( f(x) = 2e^{-2x} \), with the optimal global bandwidth \( h = .342128 \).
Figure 5: Density estimates for 35 measurements of average December precipitation in Des Moines, Iowa from 1961-1965, shown on the rug, with \( h = 0.425 \). The solid line is our proposed estimator (without BVF), and the dashed line is Hall and Park's.
Figure 6: Density estimates for 365 measurements of wind speed at Eindhoven, the Netherlands, taken at 2 AM every day during 1993. The values are plotted on the bottom. The solid line is our proposed estimator (without BVF) and the dashed line is the LL estimator, both with bandwidth $h = 2.5$. 