UNIVERSITY OF ALBERTA

SOME CONTRIBUTIONS TO EMPIRICAL BAYES THEORY AND FUNCTIONAL ESTIMATION

BY

SHUNPU ZHANG ©

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN

STATISTICS

DEPARTMENT OF MATHEMATICAL SCIENCES

EDMONTON, ALBERTA

FALL 1997

0-612-23100-3

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Shunpu Zhang

TITLE OF THESIS:

SOME CONTRIBUTIONS TO EMPIRICAL BAYES THEORY AND FUNCTIONAL ESTIMATION

DEGREE FOR WHICH THESIS WAS PRESENTED: Doctor of Philosophy

YEAR THE DEGREE GRANTED: 1997

Shunpu Zhang

Permanent Address:
323-10045-83 Avenue
Edmonton, Alberta
Canada T6E 2C3

Date: Aug 7, 1997

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Some Contributions to Empirical Bayes Theory and Functional Estimation** submitted by **Shunpu Zhang** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Statistics**.

R. J. Karunamuni (Supervisor)

A. N. Al-Hussaini

Y. Wu

P. P. Talwar

Y. S. Wong

R. S. Singh (External Examiner)

Date: Aug 1/97

TO MY WIFE YING LI AND DAUGHTER
DAISY ZHANG

ABSTRACT

The empirical Bayes approach is applicable to statistical decision problems when one is experienced with an independent sequence of Bayes decision problems each having similar probability structure. It has been argued that much can be gained by using the data available from the first n decision problems for the $(n+1)^{st}$ decision problem. There has been a great deal of work done on empirical Bayes problems in the case of uncontaminated data. In this thesis (Chapters 1, 2 and 3) we extend empirical Bayes rules for the case where the observed data are contaminated (errors in variables). Specifically, we study squared error loss estimation and linear loss two-action problems. We construct both Bayes and empirical Bayes rules. Asymptotic optimality and rates of convergence of the proposed empirical Bayes rules are investigated uniformly over a class of prior distributions for two types of error distributions.

In Chapters 4, 5 and 6 of this thesis we consider the problem of nonparametric density estimation at the boundary region. Compared to interior points estimation, this is rather formidable due to " the boundary effect" that occurs at the boundary. We extend the local polynomial fitting method to the case of den-

sity estimation, and in particular for estimating a density at the boundary region (Chapters 4 and 5). Optimal end-point kernels are obtained. The implementation of bandwidth variation functions is extensively discussed. Furthermore, a new way of removing the boundary effect is proposed. Chapter 6 proposes another new method of boundary correction for kernel density estimation. The technique is a kind of generalised reflection method involving reflecting a transformation of the data. In simulations, this new method is seen to clearly outperform an earlier generalised reflection idea. It also, overall, has advantages over boundary kernel methods and a non-negative adaptation thereof, although the latter are competitive in some situations.

ACKNOWLEDGMENTS

I would like to thank all professors who have helped me during my study at the University of Alberta. In particular, I would like to express my sincere thanks to my supervisor Professor Rohana J. Karunamuni. Without his guidance, patience and constant encouragement, this work would not have been possible.

I am also indebt to my family for their love and support throughout my graduate studies.

I gratefully acknowledge the Department of Mathematical Sciences for continuing support during my studies, the Ph.D Scholarship of the University of Alberta, J Gordin Kaplan Graduate Student Award and the Faculty of Graduate Studies and Research of the University of Alberta.

TABLE OF CONTENTS

Introduction

The work in this thesis is presented in six papers which have been prepared for publication.

In Chapter 1, " Bayes and Empirical Bayes Estimation with Errors in Variables," we consider the following estimation problem. Suppose the parameter θ is distributed according to some (prior) distribution G, and one is to estimate θ based on a random variable X with X, given θ, being distributed according to the continuous one-parameter exponential family. That is,

$$f_{X|\theta}(x) = u(x)c(\theta)e^{\theta x}, \quad -\infty < x < \infty,$$

where $u(x) > 0$ if and only if $x > -\infty$ and $c(\theta) = (\int e^{\theta x}u(x)dx)^{-1}$. Let the loss function be squared error loss. But assume that X is not directly observable, and because of measurement error or the nature of environment, one can only observe

$$Y = X + \epsilon,$$

where the random disturbance or the random error ϵ is independent of (X, θ) and has a known distribution F_ϵ. Under squared error loss, the Bayes estimator based on the contaminated data $Y = y$ is the posterior mean $E(\theta|Y = y)$; i.e.,

$$\begin{aligned} \delta_G(y) &= E(\theta|Y = y) = \int_\Omega \theta f_{Y|\theta}(y)dG(\theta)/f_Y(y) \\ &= \frac{\int_\Omega \theta \int_{-\infty}^{\infty} f_{X|\theta}(y-x)dF_\epsilon(x)dG(\theta)}{f_Y(y)}, \end{aligned}$$

where G is the prior distribution on Ω which is the support of G, and

$$f_Y(y) = \int_\Omega f_{Y|\theta}(y)dG(\theta)$$

with

$$f_{Y|\theta}(y) = \int_{-\infty}^{\infty} f_{X|\theta}(y-x)dF_\epsilon(x).$$

If we assume that $\int_\Omega |\theta| \int_{-\infty}^{\infty} f_{X|\theta}(y-x)dF_\epsilon(x)dG(\theta) < \infty$ uniformly in y, then by Fubini's theorem we obtain

$$\delta_G(y) = \frac{\int_{-\infty}^{\infty} f_X^{(1)}(y-x)dF_\epsilon(x) - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} f_X(y-x)dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_X(y-x)dF_\epsilon(x)},$$

where

$$f_X(x) = \int_\Omega C(\theta)u(x)e^{\theta x}dG(\theta).$$

When G is unknown, δ_G is not available. Assume that we have obtained a random sample of contaminated data $Y_1, Y_2, ..., Y_n$, where Y_i is distributed according to the same marginal distribution F_Y with density f_Y. Using these data, we are able to construct an empirical Bayes (EB) estimator of δ_G by using the deconvolution method. We prove that the proposed EB estimator is asymptotically optimal uniformly over a class of priors in the sense that the Bayes risk of the EB estimator converge to the Bayes risk of δ_G which achieves the minimum Bayes risk w.r.t. G over all estimators for which the risk is finite.

In Chapter 2, "Empirical Bayes Estimation for the Continuous One-Parameter Exponential Family with Errors in Variables," we improve the results obtained in Chapter 1. By constructing an improved EB estimator of δ_G, the Bayes estimator of θ, we obtain uniform convergence rates of the proposed EB estimator over a class of priors for different types of error distributions. Since the convergence rate of the EB estimator is extremely slow when the error distribution is 'supersmooth', we propose two models in which the convergence rate can be improved significantly.

In Chapter 3, "Empirical Bayes Two-Action Problem for the Continuous One-Parameter Exponential Family with Errors in Variables," we study the empirical Bayes linear-loss two-action problem for the continuous one-parameter exponential family when the observed data are contaminated (errors in variables). A new empirical Bayes testing rule is constructed, and its asymptotic optimality uniformly over a class of prior distributions is established. Uniform rates of convergence of the corresponding regret (excess risk), which depend on the type of the error distribution, are also obtained for two types of error distributions. Our results are compared with the 'pure' observed data results of the literature.

In Chapter 4, "On Kernel Density Estimation near End-Points with Application to Line Transect Sampling," we consider the problem of estimating a density at the boundary region and its application to line transect sampling. Line transect estimation of population density of objects, such as animals or plants, is intimately related to the estimation problem of $f(0)$, the value of the detection density at the left end-point 0. Nonparametric estimation of $f(0)$ is rather formidable due to boundary effects that occur in nonparametric curve estimation. It is well known that the usual kernel density estimates require modifications when estimating the density near endpoints of the support. Here we investigate the local polynomial smoothing technique as a possible alternative method for the problem. This method has shown a number of advantages over other popular nonparametric estimation methods in the case of regression function estimation. By mimicking the techniques for regression function, we obtain a local polynomial density estimator. It is observed that our density estimator also possesses desirable properties such as automatic adaptability for boundary effects near end-points. We also obtain an " optimal kernel" of order (0,2) in order to estimate the density at the end-points as a solution of a variational problem. Various bandwidth variation schemes are discussed and investigated in a Monte Carlo study.

In Chapter 5, "On Nonparametric Density Estimation at the Boundary," we generalize the results of Chapter 4 to higher order kernel case. We propose a new and intuitive method of removing boundary effects in density estimation. Our idea, which replaces the unwanted terms in the bias expansion by their estimators, offers a new way of constructing boundary kernels. Further, we show that the class of boundary kernels obtained from the local polynomial fitting method is a special case of ours. Furthermore, one easy way of constructing the optimal end-point kernels is proposed. We also discuss the choice of bandwidth variation functions at the boundary region. The performance of our results are numerically analyzed in a Monte Carlo study.

In Chapter 6, we propose a new method of boundary correction for kernel density estimation. The technique is a kind of generalised reflection method involving reflecting a transformation of the data. The transformation depends on a pilot estimate of the logarithmic derivative of the density at the boundary. In simulations, the new method is seen to clearly outperform an earlier generalised reflection idea. It also, overall, has advantages over boundary kernel methods and a non-negative adaptation thereof, although the latter are competitive in some situations. We also present the theory underlying the new methodology.

Chapter 1

Bayes and Empirical Bayes Estimation with Errors in Variables

1. Introduction

Consider the following estimation problem. Suppose θ is distributed according to some (prior) distribution G, and one is to estimate θ based on a random variable X with X, given θ, being distributed according to some distribution $F_{X|\theta}$ with Lebesgue density $f_{X|\theta}$. Let the loss function be squared error loss. But assume that X is not directly observable and because of measurement error or the nature of environment, one can only observe

$$Y = X + \epsilon \tag{1.1}$$

where the random disturbance or the random error ϵ is independent of X. Assume that ϵ has a known distribution F_ϵ. We investigate the problem of estimation of θ based on Y with squared error loss. In this paper, it is of our interest to develop both Bayes (in the case when G is known) and empirical Bayes (in the case when G is unknown) estimators for the preceding problem. In Section 2 below, we obtain the Bayes estimator of θ w.r.t. G. In Section 3, an empirical Bayes (EB)

A version of this chapter has been published. Shunpu Zhang and Rohana J. Karunamuni. *Statistics and Probability Letters.* Vol. 33, 23-34, 1997.

estimator is constructed when X, given θ, is distributed according to the continuous one-parameter exponential family; that is, for some a $\geq -\infty$,

$$f_{X|\theta}(x) = u(x)c(\theta)e^{\theta x} \tag{1.2}$$

where $u(x) > 0$ if and only if $x > a$ and $c(\theta) = (\int e^{\theta x}u(x)dx)^{-1}$.

Empirical Bayes estimators are constructed based on the data gathered from n independent repetitions of the same component problem (-Robbins(1956,1964)). Under the present model, then the problem occurs independently with the same unknown G throughout, there is a sequence of independent random vectors $(\theta_i, X_i, Y_i), i = 1, 2, ...,$ where the random variables θ_i's and X_i's are unobservable, and the θ_i's are i.i.d. with the same distribution G. Conditional on θ_i, X_i is distributed according to $f_{X|\theta_i}$. Only Y_i's are observable, where $Y_i = X_i + \epsilon_i$ with ϵ_i and X_i are independent. For the $(n+1)^{st}$-problem, (empirical Bayes) estimator $\delta_n(y)$ depends on Y_1, ...,Y_n and $Y_{n+1} = Y$, $n \geq 1$.

There are a number of practical situations in which one may face with the type of problem described above. For example, Y could be the measurement made on an item manufactured using certain equipment. (Usually, more than one measurements are made on successive items.) If one wishes to estimate some parameter θ of the equipment, subject to squared error loss, one may have available measurements $Y_1, Y_2, ..., Y_n$ on items manufactured using the same type of equipment in the past.

The history of the standard empirical Bayes estimation problem is such that the only problem that seems to have been considered thus far is the situation where the random variables X_i are observed without an error. The literature is too extensive to warrant a complete listing here. For empirical Bayes estimation in the family (1.2) see, for instance, Yu(1970), Hannan and Macky(1971), Lin(1975), Efron and Morris(1973), Singh(1976, 1979), Van Houwelingen and

Stijnen(1983) and Singh and Wei(1992). For additional references, the reader is referred to the monograph of Maritz and Lwin(1989).

The proposed EB estimator and its asymptotic optimality are given in Section 3. Proofs of the main results are deferred to Section 4. Results of a simulation study are given in Section 5. Section 6 contains concluding remarks.

2. The Bayes Estimator

Under squared error loss, the Bayes estimator based on the contaminated data $Y = y$ is the posterior mean $E(\theta|Y = y)$; i.e.,

$$\delta_G(y) = E(\theta|Y = y) = \int_\Omega \theta f_{Y|\theta}(y) dG(\theta)/f_Y(y),$$

where G is the prior distribution on Ω and

$$f_Y(y) = \int_\Omega f_{Y|\theta}(y) dG(\theta) \tag{2.1}$$

with

$$f_{Y|\theta}(y) = \int_{-\infty}^{y-a} f_{X|\theta}(y-x) dF_\epsilon(x).$$

Then

$$\delta_G(y) \quad = \quad \frac{\int_\Omega \theta \int_{-\infty}^{y-a} f_{X|\theta}(y-x) dF_\epsilon(x) dG(\theta)}{f_Y(y)}.$$

If we assume that $\int_\Omega |\theta| \int_{-\infty}^{y-a} f_{X|\theta}(y-x) dF_\epsilon(x) dG(\theta) < \infty$ uniformly in y and if $f_{X|\theta}$ is given by (1.2), then by Fubini's theorem we obtain

$$\begin{aligned} \delta_G(y) &= \frac{\int_{-\infty}^{y-a} \int_\Omega \theta f_{X|\theta}(y-x) dG(\theta) dF_\epsilon(x)}{f_Y(y)} \\ &= \frac{\int_{-\infty}^{y-a} f_X^{(1)}(y-x) dF_\epsilon(x) - \int_{-\infty}^{y-a} \frac{u^{(1)}(y-x)}{u(y-x)} f_X(y-x) dF_\epsilon(x)}{f_Y(y)}, \end{aligned} \tag{2.2}$$

where

$$f_X(x) = \int_\Omega C(\theta)u(x)e^{\theta x}dG(\theta) \tag{2.3}$$

with $u(x)$ is as given in (1.1) and $u^{(1)}(x)$ denotes the first derivative of $u(x)$. When $a = -\infty$, then (2.2) becomes

$$\delta_G(y) = \frac{\int_{-\infty}^{\infty} f_X^{(1)}(y-x)dF_\epsilon(x) - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} f_X(y-x)dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_X(y-x)dF_\epsilon(x)}. \tag{2.4}$$

Example 2.1. Consider the exponential family in (1.2) with $u(x) = e^{-x^2/2}$ and $C(\theta) = (2\pi)^{-1/2}e^{-\theta^2/2}$; that is, for each $-\infty < \theta < \infty$, $f_{X|\theta}(x) = (2\pi)^{-1/2}e^{-(x-\theta)^2/2}$, where $-\infty < x < \infty$. Then $\Omega = (-\infty, \infty)$ and $a = -\infty$. Then the Bayes estimator $\delta_G(y)$ given by (2.4) is equal to

$$\begin{aligned}\delta_G(y) &= \frac{-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}(y-x-\theta)e^{-\frac{(y-x-\theta)^2}{2}}dG(\theta)dF_\epsilon(x)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{(y-x-\theta)^2}{2}}dG(\theta)dF_\epsilon(x)} \\ &\quad + \frac{\int_{-\infty}^{\infty}(y-x)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{(y-x-\theta)^2}{2}}dG(\theta)dF_\epsilon(x)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{(y-x-\theta)^2}{2}}dG(\theta)dF_\epsilon(x)}.\end{aligned} \tag{2.5}$$

Further, suppose that the prior on Ω is $G_0 = Normal(0,1)$. Then (2.5) reduces to

$$\begin{aligned}\delta_{G_0}(y) &= \frac{-\int_{-\infty}^{\infty} \frac{1}{4\sqrt{\pi}}(y-x)e^{-\frac{(y-x)^2}{4}}dF_\epsilon(x) + \int_{-\infty}^{\infty}(y-x)\frac{1}{2\sqrt{\pi}}e^{-\frac{(y-x)^2}{4}}dF_\epsilon(x)}{\int_{-\infty}^{\infty} \frac{1}{2\sqrt{\pi}}e^{-\frac{(y-x)^2}{4}}dF_\epsilon(x)} \\ &= \frac{\int_{-\infty}^{\infty}(y-x)e^{-\frac{(y-x)^2}{4}}dF_\epsilon(x)}{2\int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{4}}dF_\epsilon(x)}.\end{aligned}$$

Example 2.2. Consider the scale exponential family with Lebesgue densities given by $f_{X|\theta}(x) = \theta e^{-\theta x}$ for $x > 0$, where $\theta > 0$. Then $\Omega = (0, \infty)$ and $a = 0$. Then the Bayes estimator $\delta_G(y)$ given by (2.2) is equal to

$$\delta_G(y) = -\frac{\int_{-\infty}^{y}\int_0^{\infty} \theta^2 e^{-(y-x)\theta}dG(\theta)dF_\epsilon(x)}{\int_{-\infty}^{y}\int_0^{\infty} \theta e^{-(y-x)\theta}dG(\theta)dF_\epsilon(x)}. \tag{2.6}$$

Further, suppose that the prior on Ω is $G_0 = Gamma(\alpha, \beta), \alpha > 0, \beta > 0$. Then the Bayes estimator (2.6) reduces to

$$\begin{aligned}\delta_{G_0}(y) &= \frac{\int_{-\infty}^{y} \beta(\beta+1)\frac{\alpha^\beta}{(y-x+\alpha)^{\beta+2}} dF_\epsilon(x)}{\int_{-\infty}^{y} \beta\frac{\alpha^\beta}{(y-x+\alpha)^{\beta+1}} dF_\epsilon(x)} \\ &= \frac{(\beta+1)\int_{-\infty}^{y}(y-x+\alpha)^{-(\beta+2)} dF_\epsilon(x)}{\int_{-\infty}^{y}(y-x+\alpha)^{-(\beta+1)} dF_\epsilon(x)}.\end{aligned}$$

3. An Empirical Bayes Estimator

In this section we shall consider the case where the prior G is not completely known. Assume that a sequence of contaminated observations $Y_1, Y_2, ..., Y_n$ is available, where Y_i's are i.i.d. according to the marginal distribution F_Y with density f_Y given by (2.1) when $f_{X|\theta}$ is given by (1.2). At the $(n+1)^{st}$ problem, the estimator δ_n is allowed to depend on all of the past observations as well as the $(n+1)^{st}$ observation. Hence, δ_n is a measurable function of $Y_1, Y_2, ..., Y_n$ and $Y_{n+1} = Y$. In order to construct δ_n, we shall first construct estimators of f_X and $f_X^{(1)}$ based on $Y_1, Y_2, ..., Y_n$, where f_X is given by (2.3) and $f_X^{(1)}$ is the first derivative of f_X. Let ϕ_ϵ denote the characteristic function of the error variable ϵ. Let $\hat{\phi}_n$ denote the empirical characterisric function defined by $\hat{\phi}_n(t) = \frac{1}{n}\sum_{j=1}^{n}\exp(itY_j)$. For a nice kernel K, let ϕ_K be its Fourier transform with $\phi_K(0)$=1. For $x > 0$, then we define

$$f_n^{(l)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-itx)(-it)^l\phi_K(th_n)\frac{\hat{\phi}_n(t)}{\phi_\epsilon(t)}dt \tag{3.1}$$

as our (kernel) estimator of $f^{(l)}(x)$, $l = 0, 1$, where h_n is the bandwidth ($h_n \to 0$ as $n \to \infty$). Here we assume that $|t^l\phi_K(th_n)/\phi_\epsilon(t)|$ is integrable on $(-\infty, \infty)$. The construction of (3.1) is due to Stefanski and Carroll(1990). A similar construction is studied by Fan(1991a,b; 1992) and Zhang(1990). In the special case when l=0, denote $f_n^{(l)}$ by f_n. Under the model (1.1), the estimator (3.1) can be rewritten in the kernel form

$$f_n^{(l)}(x) = \frac{1}{nh_n^{l+1}} \sum_{j=1}^{n} K_{nl}(\frac{x - Y_j}{h_n}), \tag{3.2}$$

where

$$K_{nl}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) \frac{(-it)^l \phi_K(t)}{\phi_\epsilon(t/h_n)} dt.$$

In view of (2.2), we shall estimate δ_G through f_X and $f_X^{(1)}$. Let

$$\delta_n(y) = \frac{\int_{-\infty}^{y-a} f_n^{(1)}(y-x)dF_\epsilon(x) - \int_{-\infty}^{y-a} \frac{u^{(1)}(y-x)}{u(y-x)} f_n(y-x)dF_\epsilon(x)}{\hat{f}_n(y)} \tag{3.3}$$

with

$$\hat{f}_n(y) = \begin{cases} \int_{-\infty}^{y-a} f_n(y-x)dF_\epsilon(x) & \text{if } |\int_{-\infty}^{y-a} f_n(y-x)dF_\epsilon(x)| > \Delta_n \\ \Delta_n & \text{otherwise,} \end{cases} \tag{3.4}$$

and $f_n^{(l)}$ being the type of kernel estimator of $f_n^{(l)}$ given by (3.1) for $l = 0, 1$, where Δ_n is a sequence of positive numbers such that $\Delta_n \to 0$ as $n \to \infty$. Let $R(\delta_n, G)$ denote the Bayes risk of δ_n given by (3.3) w.r.t. G. Then $R(\delta_n, G) = E(\delta_n - \theta)^2$, where the expectation E is over the random variables $Y_1, Y_2, ..., Y_n, Y_{n+1}$ and θ. For the Bayes estimator δ_G given by (2.2), $R(\delta_G, G)$ achieves the minimum Bayes risk w.r.t.G. That is, $R(\delta_G, G) = \inf_d R(d, G)$, where the infinimum is over all estimators d for which $R(d, G) < \infty$. For convenience, denote $R(\delta_G, G)$ by $R(G)$. Then $R(G)$ is the Bayes envelope value of the problem. This motivates the use of the excess risk (regret)

$$R(\delta_n, G) - R(G) = E(\delta_n - \theta)^2 - E(\delta_G - \theta)^2$$

as a measure of goodness of the estimator δ_n. Restricting G to those with finite Bayes risk, the excess risk satisfies (Lemma 2.1 of Singh(1979))

$$0 \leq R(\delta_n, G) - R(G) = E(\delta_n - \delta_G)^2. \tag{3.5}$$

Empirical Bayes estimator δ_n is said to be asymptotically optimal if $\lim_{n\to\infty} R(\delta_n, G) = R(G)$ (-Robbins(1956,1964)). To state main results of this paper, which establish the asymptotic optimality of the EB estimator δ_n given by (3.3) under various conditions, we need following assumptions on the kernel and the error ϵ:

(A1) $K(\cdot)$ is bounded, continuous and $\int_{-\infty}^{\infty} |x|^2 K(x)dx < \infty$.

(A2) The Fourier transform ϕ_K of K is a symmetric function and satisfies $\phi_K(t) = 1 + O(|t|^2)$, as $t \to 0$.

(A3) $\phi_K(t) = 0$, for $|t| \geq 1$.

(A4) The characteristic function ϕ_ϵ of ϵ satisfies $\phi_\epsilon(t) \neq 0$, for any t.

(A5) $|\phi_\epsilon(t)||t|^{-\beta_0} exp(|t|^\beta/\gamma) \geq d_0$ (as $t \to \infty$) for some positive constants β, γ, d_0 and a constant β_0.

(A6) $|t^l \phi_K(th_n)/\phi_\epsilon(t)|$ is integrable on $(-\infty, \infty)$, $l = 0, 1$.

(A7) $|\phi_\epsilon(t)t^\beta| \geq d_0$ as $t \to \infty$, for some positive constants d_0 and β.

(A8) $\int_{-\infty}^{\infty} |\phi_K(t)t^{\beta+l}|dt < \infty$ and $\int_{-\infty}^{\infty} |\phi_K(t)t^{\beta+l}|^2 dt < \infty$, for some positive constant β and $l = 0, 1$.

The assumptions **(A1)**,**(A2)** and **(A3)** imply that K is a second-order kernel function.

For convenience, consider the following class of priors

$$\begin{aligned}\mathcal{F}_B &= \{G : G \text{ is a prior on } \Omega \text{ such that } \sup_x |f_X(x)| \leq B \\ &\qquad \text{with } f_X \text{ is given by (2.3)}\} \qquad (3.6)\end{aligned}$$

for some finite positive constant B.

Theorem 3.1. Let $f_{X|\theta}$ be given by (1.2) with $a = -\infty$. Let $G \in \mathcal{F}_B$, where $\mathcal{F}_B$ is given by (3.6). Further, suppose that the distributions G and F_ϵ are such that f_X given by (2.3) is twice differentiable on $(-\infty, \infty)$, $\int_\Omega \theta^2 dG(\theta) < \infty$,

$\int_\Omega \int_{-\infty}^{\infty} |\theta| f_{X|\theta}(y-x) dF_\epsilon(x) dG(\theta) < \infty$ uniformly in y, and $E\left[\int_{-\infty}^{\infty} \left(\frac{u^{(1)}(Y-x)}{u(Y-x)}\right)^2 dF_\epsilon(x)\right]$ $< \infty$. Furthermore, assume that the condtions (**A1**) to (**A6**) hold. Then, for the bandwidth $h_n = O((\log n)^{-1/\beta})$ and the sequence $\Delta_n = o((\log n)^{-1/\beta})$ (see (3.4)), we have

$$\lim_{n\to\infty} R(\delta_n, G) = R(G), \tag{3.7}$$

where $R(\delta_n, G)$ is the Bayes risk of the EB estimator δ_n defined by (3.3) with $a = -\infty$, and $R(G)$ is the minimum Bayes risk.

Theorem 3.2. Assume that the hypotheses of Theorem 3.1 hold now with the conditions (**A1**) to (**A6**) replaced by the conditions (**A1**) to (**A4**),(**A6**),(**A7**) and (**A8**). Then, for the bandwidth $h_n = O(n^{-1/(\beta+5)})$ and $\Delta_n = o(n^{-1/(\beta+5)})$, we have

$$\lim_{n\to\infty} R(\delta_n, G) = R(G). \tag{3.8}$$

The distributions normal and Cauchy satisfy the assumption (**A5**) above, whereas gamma and double exponential satisfy (**A7**). A kernel satisfying (**A1**), (**A2**), (**A3**) and (**A8**) can be easily constructed; see, e.g., Fan(1991a,b;1992). We now revisit Example 2.1 and investigate the validity of assumptions made in theorems above for this example.

Example 2.1 (continued). Let the error distribution $F_\epsilon = Normal(0,1)$. Then ϕ_ϵ satisfies (**A4**) and (**A5**) with $\beta = 2$. Since $f_{X|\theta}(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}$, we have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|\theta}(x) dG(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-\theta)^2/2} dG(\theta)$$

for any prior on $\Omega = (-\infty, \infty)$. Hence, by Theorem 2.9 of Lehmann(1959), one obtains

$$f_X^{(1)}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -(x-\theta) e^{-(x-\theta)^2/2} dG(\theta)$$

and

$$f_X^{(2)}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\theta)^2 e^{-(x-\theta)^2/2} dG(\theta) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-\theta)^2/2} dG(\theta).$$

Then

$$\sup_x |f_X^{(2)}(x)| \le B, \quad \text{where } B = \sqrt{2/\pi}; \text{ for any } G.$$

Also, it is easy to show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\theta| f_{X|\theta}(y-x) dF_\epsilon(x) dG(\theta) < \infty$ uniformly in y. Denote

$$\rho = E\left[\int_{-\infty}^{\infty} \left(\frac{u^{(1)}(Y-x)}{u(Y-x)}\right)^2 dF_\epsilon(x)\right].$$

Then

$$\rho = E\left[\int_{-\infty}^{\infty} (Y-x)^2 dF_\epsilon(x)\right] \le 2EY^2 + \int_{-\infty}^{\infty} x^2 dF_\epsilon(x), \text{ where}$$

$$\begin{aligned}
EY^2 &= \int_{-\infty}^{\infty} y^2 f_Y(y) dy \\
&= \int_{-\infty}^{\infty} y^2 \int_{-\infty}^{\infty} f_X(y-x) dF_\epsilon(x) dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} y^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(y-x-\theta)^2/2} e^{-x^2/2} dx dy dG(\theta) \\
&\le C_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 e^{-(y-\theta)^2/4} dy dG(\theta) \\
&\le C_2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-\theta)^2 e^{-(y-\theta)^2/4} dy dG(\theta) C_3 \int_{-\infty}^{\infty} \theta^2 dG(\theta) \\
&\le C_2 + C_3 \int_{-\infty}^{\infty} \theta^2 dG(\theta)
\end{aligned}$$

where C_1, C_2 and C_3 are some finite positive constants. Thus, $\rho < \infty$ if $\int_{-\infty}^{\infty} \theta^2 dG(\theta) < \infty$. That is, if G is such that $\int_{-\infty}^{\infty} \theta^2 dG(\theta) \le B$, then the assumptions of Theorem 3.1 are satisfied. Similarly, for $f_{X|\theta}(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}$, it is easy to show that the assumptions of Theorem 3.2 are satisfied if $\int_{-\infty}^{\infty} \theta^2 dG(\theta) \le B$ and $F_\epsilon = Gamma(1,p)$ distribution for some constant $p > 1$.

4. Proofs

In this section we prove Theorem 3.1 above. The proof of Theorems 3.2 is similar. First, we state two lemmas useful in proving these theorems. For proofs of the lemmas, see, e.g., Fan(1990a,b; 1992).

Lemma 4.1. Under the assumptions of **(A1)** to **(A6)** and with the choice $h_n = O((\log n)^{-1/\beta})$ of the bandwidth, one has

$$\sup_{-\infty<x<\infty} \sup_{G\in\mathcal{F}_B} E|f_n^{(l)}(x) - f_X^{(l)}(x)|^2 \leq Const. \times (\log n)^{-2(2-l)/\beta} \tag{4.1}$$

for $l = 0, 1$, where $f_n^{(l)}(x), f_X(x)$ and $\mathcal{F}_B$ are given by (3.1),(2.3) and (3.6), respectively.

Lemma 4.2. Under the assumptions of **(A1)** to **(A4)**,**(A6)**,**(A7)** and **(A8)**, and with the choice $h_n = O(n^{-1/(\beta+5)})$ of the bandwidth, one has

$$\sup_{-\infty<x<\infty} \sup_{G\in\mathcal{F}_B} E|f_n^{(l)}(x) - f_X^{(l)}(x)|^2 \leq Const. \times n^{-2(2-l)/(\beta+5)} \tag{4.2}$$

for $l = 0, 1$, where$f_n^{(l)}(x), f_X(x)$ and $\mathcal{F}_B$ are given by (3.1),(2.3) and (3.6), respectively.

Proof of Theorem 3.1. From (2.2) and (3.3) with $a = -\infty$ and by (3.5), we obtain

$$0 \leq R(\delta_n, G) - R(G) = \int_{-\infty}^{\infty} E(\delta_n(y) - \delta_G(y))^2 f_Y(y)dy, \tag{4.3}$$

where E denotes the expectation w.r.t. $Y_1, Y_2, ..., Y_n$ and θ. From (2.2) and (3.3) together with the $C_r - inequality$ (Loève (1963), p. 157) yield

$E(\delta_n(y) - \delta_G(y))^2$

$$\leq 2E\left[\frac{\int_{-\infty}^{\infty}(f_n^{(1)}(y-x) - f_X^{(1)}(y-x))dF_\epsilon(x)}{\hat{f}_n(y)}\right.$$

$$
- \left. \frac{\int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)}(f_n(y-x) - f_X(y-x))dF_\epsilon(x)}{\hat{f}_n(y)} \right]^2
$$

$$
+2E\left[(\frac{f_Y(y) - \hat{f}_n(y)}{\hat{f}_n(y)})^2 \right] \delta_G^2(y)
$$

$$
\leq \; 4E\left\{ \frac{\left[\int_{-\infty}^{\infty}(f_n^{(1)}(y-x) - f_X^{(1)}(y-x))dF_\epsilon(x)\right]^2}{\hat{f}_n(y)^2} \right\}
$$

$$
+4E\left\{ \frac{\left[\int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)}(f_n(y-x) - f_X(y-x))dF_\epsilon(x)\right]^2}{\hat{f}_n(y)^2} \right\}
$$

$$
+2\delta_G^2(y)E\left(\frac{f_Y(y) - \hat{f}_n(y)}{\hat{f}_n(y)}\right)^2. \tag{4.4}
$$

Since $\hat{f}_n(y) \geq \Delta_n$ (see (3.4)) and by the moment inequality , one obtains

$$
\text{the } 1^{st}\text{term of RHS of (4.4)} \leq 4\Delta_n^{-2}E\int_{-\infty}^{\infty}(f_n^{(1)}(y-x) - f_X^{(1)}(y-x))^2 dF_\epsilon(x)
$$

$$
= 4\Delta_n^{-2}\int_{-\infty}^{\infty} E(f_n^{(1)}(y-x) - f_X^{(1)}(y-x))^2 dF_\epsilon(x) \tag{4.5}
$$

Similarly,

$$
\text{the } 2^{nd}\text{term of RHS of (4.4)} \;\leq\; 4\Delta_n^{-2}\int_{-\infty}^{\infty}\left(\frac{u^{(1)}(y-x)}{u(y-x)}\right)^2
$$

$$
E(f_n^{(1)}(y-x) - f_X^{(1)}(y-x))^2 dF_\epsilon(x). \tag{4.6}
$$

Now consider

$$
\begin{aligned}
E\left(\frac{f_Y(y) - \hat{f}_n(y)}{\hat{f}_n(y)}\right)^2 &= E\left(\frac{f_Y(y) - \hat{f}_n(y)}{\hat{f}_n(y)}\right)^2 I[f_Y(y) \geq \Delta_n] \\
&\quad +E\left(\frac{f_Y(y) - \hat{f}_n(y)}{\hat{f}_n(y)}\right)^2 I[f_Y(y) < \Delta_n] \\
&= J_{1,n} + J_{2,n},
\end{aligned} \tag{4.7}
$$

where $J_{1,n}$ and $J_{2,n}$ are the first and second terms of RHS of (4.7). By definition (3.4) of $\hat{f}_n(y)$, we note that

$$
J_{1,n} \;\leq\; \Delta_n^{-2}E(\hat{f}_n(y) - f_Y(y))^2 I[f_Y(y) \geq \Delta_n] \tag{4.8}
$$

$$
\begin{aligned}
&= \Delta_n^{-2} E\left\{(\hat{f}_n(y) - f_Y(y))^2 I[|\int_{-\infty}^{\infty} f_n(y-x)dF_\epsilon(x)| \geq \Delta_n]\right. \\
&\quad \left. +(\hat{f}_n(y) - f_Y(y))^2 I[|\int_{-\infty}^{\infty} f_n(y-x)dF_\epsilon(x)| < \Delta_n]\right\} I[f_Y(y) \geq \Delta_n] \\
&= \Delta_n^{-2} E\left\{\left(\int_{-\infty}^{\infty} f_n(y-x)dF_\epsilon(x) - f_Y(y)\right)^2 I[|\int_{-\infty}^{\infty} f_n(y-x)dF_\epsilon(x)| \geq \Delta_n]\right. \\
&\quad \left. +(\Delta_n - f_Y(y))^2 I[|\int_{-\infty}^{\infty} f_n(y-x)dF_\epsilon(x)| < \Delta_n]\right\} I[f_Y(y) \geq \Delta_n] \\
&\leq \Delta_n^{-2} E\left\{(\int_{-\infty}^{\infty} f_n(y-x)dF_\epsilon(x) - f_Y(y))^2 I[|\int_{-\infty}^{\infty} f_n(y-x)dF_\epsilon(x)| \geq \Delta_n]\right. \\
&\quad \left. +(\int_{-\infty}^{\infty} f_n(y-x)dF_\epsilon(x) - f_Y(y))^2 I[|\int_{-\infty}^{\infty} f_n(y-x)dF_\epsilon(x)| < \Delta_n]\right\} \\
&\quad I(f_Y \geq \Delta_n) \\
&= \Delta_n^{-2} E\left[\int_{-\infty}^{\infty} (f_n(y-x) - f_X(y-x))dF_\epsilon(x)\right]^2 I[f_Y \geq \Delta_n] \\
&\leq \Delta_n^{-2}\{\int_{-\infty}^{\infty} E(f_n(y-x) - f_X(y-x))^2 dF_\epsilon(x)\} I[f_Y \geq \Delta_n]
\end{aligned}
$$

by the moment inequality. By combining (4.4),(4.5),(4.6),(4.7),(4.8) and using the fact that $J_{2,n} \leq 4I(f_Y(y) < \Delta_n)$, we obtain

$$E(\delta_n(y) - \delta_G(y))^2$$

$$
\begin{aligned}
&\leq 4\Delta_n^{-2}\left\{\int_{-\infty}^{\infty} E(f_n^{(1)}(y-x) - f_X^{(1)}(y-x))^2 dF_\epsilon(x)\right. \\
&\quad + \int_{-\infty}^{\infty}\left(\frac{u^{(1)}(y-x)}{u(y-x)}\right)^2 E(f_n(y-x) - f_X(y-x))^2 dF_\epsilon(x) \\
&\quad \left. +\delta_G^2(y) I[f_Y(y) \geq \Delta_n] \int_{-\infty}^{\infty} E(f_n(y-x) - f_X(y-x))^2 dF_\epsilon(x)\right\} \\
&\quad +4\delta_G^2(y) I[f_Y(y) < \Delta_n] \\
&\leq C_4\Delta_n^{-2}\left\{(\log n)^{-2/\beta}\int_{-\infty}^{\infty} dF_\epsilon(x)\right. \\
&\quad +(\log n)^{-4/\beta}\int_{-\infty}^{\infty}\left(\frac{u^{(1)}(y-x)}{u(y-x)}\right)^2 dF_\epsilon(x) \\
&\quad \left. +(\log n)^{-4/\beta}\delta_G^2(y) I[f_Y(y) \geq \Delta_n]\right\} \\
&\quad +4\delta_G^2(y) I[f_Y(y) < \Delta_n], \qquad (4.9)
\end{aligned}
$$

where the last inequality is obtained using Lemma 4.1 and C_4 is a constant independent of n and y. Now from (4.9), we see that

$$E(\delta_n(y) - \delta_G(y))^2 \leq M(y),$$

where $M(y) = C_5 + C_6 \int_{-\infty}^{\infty} \left(\frac{u^{(1)}(y-x)}{u(y-x)}\right)^2 dF_\epsilon(x) + C_7\delta_G^2(y)$, with C_5, C_6 and C_7 being constants independent of n and y. Observe that $\int_{-\infty}^{\infty} M(y) f_Y(y)dy < \infty$ by the assumptions of Theorem 3.1. Also, from (4.9) with the choice of $\Delta_n = o((logn)^{-1/\beta})$, we see that $\lim_{n\to\infty} E(\delta_n(y) - \delta_G(y))^2 = 0$ for each fixed y. Then, by an application of the dominated convergence theorem, we obtain $\lim_{n\to\infty} \int_{-\infty}^{\infty} E(\delta_n(y) - \delta_G(y))^2 f_Y(y)dy = 0$, under the assumptions of Theorem 3.1. The result (3.7) now follows in view of (4.3).

Proof of Theorem 3.2. The proof of (3.8) is verbatim the same as that of Theorem 3.1, except now that $\Delta_n = o(n^{-1/(\beta+5)})$ and Lemma 4.2 are used instead of $\Delta_n = o((logn)^{-1/\beta})$ and Lemma 4.1.

5. Simulation Results

To study the convergence of the regret $R(\delta_n, G) - R(G)$ of the proposed estimator (3.3), we have conducted simulation studies, and some of the results are reported here. Specifically, the following two cases of the results are presented here.

Case I: We take $f_{X|\theta}(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}$, prior $g(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2}, \; -\infty < \theta < \infty$, and the error distribution F_ϵ as the standard normal distribution. Further, we assume that the bandwidth $h_n = \sqrt{2}(\log n)^{-1/2}$ and the sequence Δ_n in (3.4) as $\Delta_n = \sqrt{3}(\log n)^{-1}$.

Case II: Here, we take $f_{X|\theta}(x) = \frac{1}{\sqrt{2\pi}}e^{-(x-\theta)^2/2}$, $g(\theta) = \frac{1}{\sqrt{2\pi}}e^{-\theta^2/2}$ and the error distribution F_ϵ as $Gamma(1,1)$ dsitribution. Also, we assume $h_n = n^{-1/6}$ and $\Delta_n = n^{-1/4}$.

For both cases, we used a second-order kernel,

$$K(x) = \frac{48\cos x}{\pi x^4}\left(1 - \frac{15}{x^2}\right) - \frac{144\sin x}{\pi x^5}\left(2 - \frac{5}{x^2}\right), \quad -\infty < x < \infty.$$

The Fourier transform of the above kernel is $\phi_K(t) = (1-t^2)^3$ for $|t| \leq 1$ and $\phi_K(t) = 0$ otherwise. Then the deconvolution kernel density estimator (3.2) is with the following kernel

$$K_{nl}(x) = \frac{(-1)^l}{\pi}\int_0^1 t^l(\cos tx)^{1-l}(\sin tx)^l(1-t^2)^3\exp\left(\frac{t^2}{2h_n^2}\right)dt, \ l = 0, 1.$$

Under the above specifications, we calculated the regret $R(\delta_n, G) - R(G)$ (see (3.5)) of our estimator for n ranging from 10 to 300. The integrals in the estimator (3.3) are evaluated by numerical integration. For the two cases considered, our simulation results are exhibited in Figure 1. The figure represents the behaviour of the regret $R(\delta_n, G) - R(G)$ as n ranges from 10 to 300.

Figure 1 about here

We find that the regret decreases to zero faster in Case II than in Case I. That is, the rate of convergence of the regret appears faster in Case II compared to Case I, see Section 6 for more elaboration on this issue. A number of the other variations of the cases I and II were also considered. Again, the patterns were very similar to those of Figure 1. Overall, the simulation results indicate that the nature of convergence is fairly satisfactory for moderate values of n. For small values of n ($n < 30$), the convergence is slightly conservative.

6. Concluding Remarks

In this paper we have introduced both Bayes and empirical Bayes problems with contaminated data (errors in variables) and obtained Bayes and empirical Bayes estimators under the squared error loss. In the latter case, it is shown that the proposed empirical Bayes estimator is asymptotically optimal. For these results, we only assume that the prior G has first few absolute moments finite. In the empirical Bayes estimation problem considered here, we can obtain rates of convergence for the regret $R(\delta_n, G) - R(G)$. However, the techniques involved in obtaining such rate results are markedly different from the approach used here. As the referee correctly guessed, the rate of convergence of the regret is extremely slow when the characteristic function of the error variable satisfies the condition (**A5**). However, when ϕ_ϵ, the characteristic function of ϵ, satisfies the condition (**A7**) then the rate of convergence of the regret is rather comparable when compared with the direct (or pure) data case, see Zhang and Karunamuni (1995) for more details. Our results revealed that the presence of errors in observation affect the performance of empirical Bayes estimators. In applications, thus, one needs extra care in identifying the error distribution if errors are likely to occur in observation.

A great deal of work has been done recently with measurement error models (errors in variables) in the nonparametric front; see, e.g., the work of Stefanski and Carroll(1990) and Fan(1991a,b; 1992) and the references therein. However, to the best of our knowledge, no work is reported in the literature on either Bayesian or empirical Bayesian settings with errors in variables. We believe that the present work will fill in this gap a little, and clearly further research is needed.

References

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors-an empirical Bayes approach. *J. Amer. Statist. Assoc.,* 68, 117-130.

Fan, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.,* 19, 1257-1272.

Fan, J. (1991b). Global behavior of deconvolution kernel estimators. *Statistica Sinica,* 1, 541-551.

Fan, J. (1992). Deconvolution with supersmooth distributions. *Canad. J. Statist.,* 20, 155-169.

Hannan, J. and Macky, D.W. (1971). Empirical Bayes squared-error loss estimation of unbounded functionals in exponential families. *Report No. RM- 290,* Dept. Statistics and Probablity, Michigan State Univ.

Lehmann (1959). *Testing Statistical Hypothesis.* Wiley, New York.

Lin, P.E. (1975). Rates of convergence in empirical Bayes estimation problems.Continuous case. *Ann. Statist.,* 3, 155-164.

Loève, M. (1963). *Probalility Theory* I, 4^{th} ed., Springer-Verlag, New York.

Maritz, J.S. and Lwin, T. (1989). *Empirical Bayes methods.* 2^{nd} edition. Chapman and Hall, London.

Robbins, H. (1956). An empirical Bayes approach to statistica. *Proc. Third Berkeley Symp. Math. Statist. Prob. I,* University of California Press, 157-163.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.,* 35, 1-20.

Singh, R.S. (1976). Empirical Bayes estimation with convergence rates in noncontinuous Lebesgue -exponential families. *Ann. Statist.,* 5, 394-400.

Singh, R.S. (1979). Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate. *Ann. Statist.*, 7, 890-902.

Singh, R.S. and Wei, L. (1992). Empirical Bayes with rates and best rates of convergence in $u(x)C(\theta)exp(-x/\theta)$-family: Estimation case. *Ann. Inst. Statist. Math.*, 44, 435-449.

Stefanski, L.A. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, 21, 169-184.

Van Houwelingen, J.C. and Stijnen, Th. (1983). Monotone empirical Bayes estimators for the continuous one-parameter exponential family. *Statist. Neerlandica*, 37, 29-43.

Yu, Benito Ong (1970). Rates of convergence in empirical Bayes two-action and estimation problems and in extended sequence-compound estimation problems. *Ph.D. Thesis.*, Michigan State University, East Lansing, Michigan.

Zhang, C.H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, 18, 806-830.

Zhang, S. and Karunamuni, R.J. (1995). Empirical Bayes estimation for the continuous one-parameter exponential family with errors in variables. *Statistics & Decisions*, to appear.

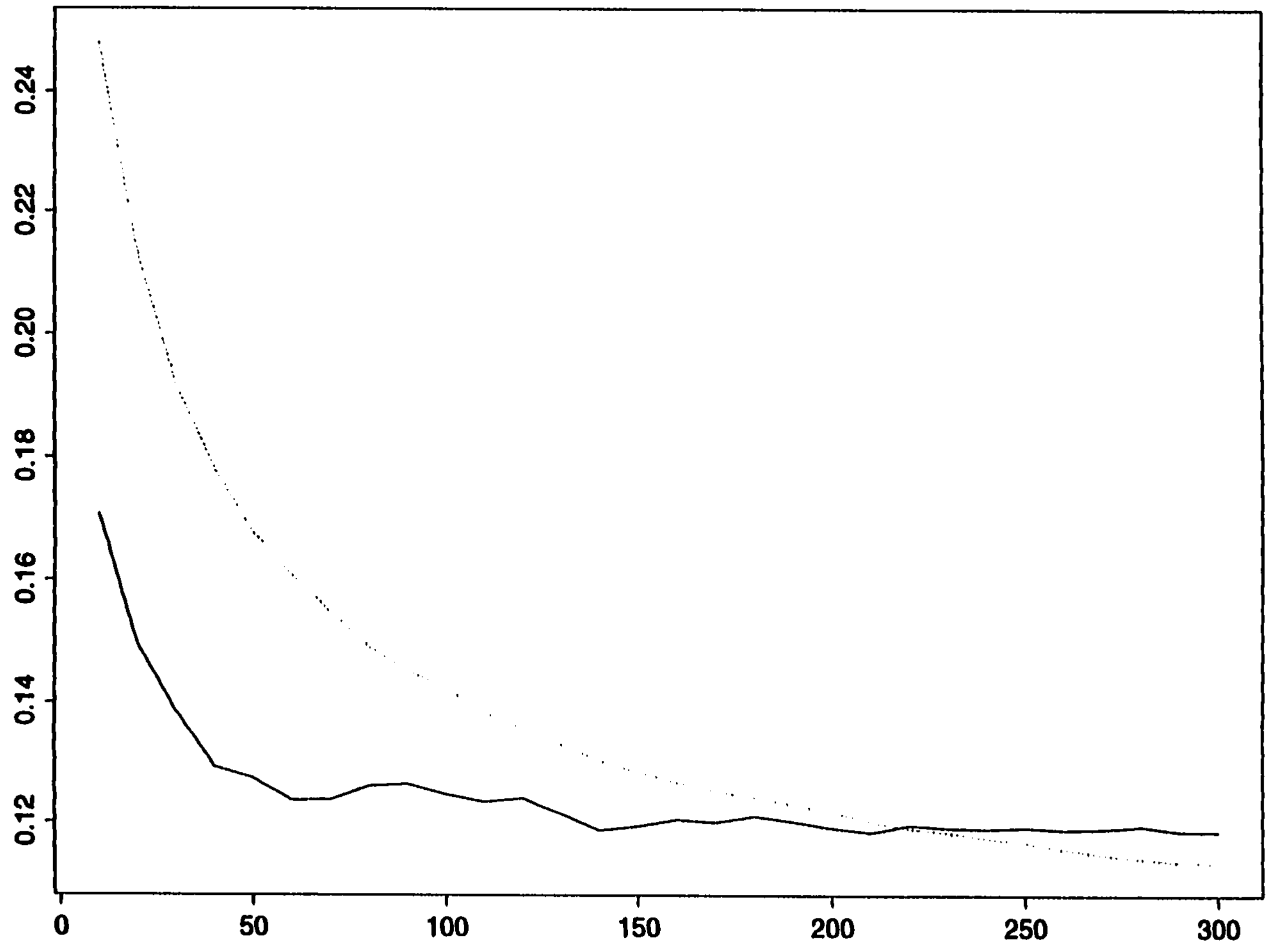


Figure 1. Behaviors of the regret with respect to n (solid line-Case I, dotted line-Case II)

Chapter 2

Empirical Bayes Estimation for the Continuous One-Parameter Exponential Family with Errors in Variables

1. Introduction

The empirical Bayes approach of Robbins [14, 15] is applicable to statistical decision problems when one is experienced with an independent sequence of Bayes decision problems each having similar structure. The statistical similarity in these decision problems includes the assumption of an unknown prior distribution G on the parameter space involved. Robbins argued that much can be gained by using the empirical Bayes approach which uses the data available in the first n decision problems in the $(n+1)^{st}$ decision problem. Since Robbins's initiation of this idea, many papers evolved on developing empirical Bayes procedures and their asymptotic properties as the number of problems, n, approaches infinity; see, e.g., the monograph of Maritz and Lwin [13] and the review article of Susarla [20]. Most of these empirical Bayes methods have treated situations in

---

A version of this chapter has been accepted for publication. Shunpu Zhang and Rohana J. Karunamuni. *Statistics and Decisions.* Vol. 15, 1997

which the observed data is uncontaminated.

Work on empirical Bayes problems in which the observed data is contaminated (errors in variables) first appeared in Zhang and Karunamuni [24]. They discuss both Bayes and empirical Bayes problems of squared error loss estimation and exhibit an asymptotically optimal empirical Bayes estimator for the continuous one-parameter exponential family. In this paper, we study an improved empirical Bayes estimator and investigate the rates of convergence of the corresponding regret (excess risk).

Section 2 describes the estimation problem in the one-parameter exponential family and its empirical Bayes analogue with errors in variables. It also constructs the proposed empirical Bayes estimator, which is the subject of our investigation. Section 3 discusses the performance of our empirical Bayes estimator. In particular, uniform rates of convergence results of the corresponding regret are exhibited for two types of error distributions (here, the uniformity is over a class of prior distributions, $\mathcal{G}$). Compared to the 'pure' data case, a slow rate of convergence is observed. However, it appears that this is an inherent expense that one has to pay when dealing with contaminated data. Proofs of the main results are given in Section 3. Section 4 discusses possible adjustments to improve the slow rate of convergence of some results in Section 2.

There has been a great deal of work done on empirical Bayes problems, in particular on squared error loss estimation problem for exponential families, in the case when the data is not contaminated; see, for example, the work of Yu [22], Hannan and Macky [10], Lin [12], Efron and Morris [6], Singh [16, 17], Van Houwelingen and Stijnen [21], Singh and Wei [18] and Datta [3, 4, 5], among others.

2. The Empirical Bayes Estimation Problem with Errors in Variables

Let (X, θ) be a random vector, where θ has a prior distribution G and, given θ, X has a density $f_{X|\theta}(x)$ with respect to Lebesgue measure on the real line. The pair (X, θ) is unobservable. Instead we observe Y, where $Y = X + \epsilon$ with ϵ denoting the random error. We assume that ϵ is independent of (X, θ) and has a known distribution F_ϵ. This is the situation one generally encounters in analyzing contaminated (errors in variables) data due to the measurement error or due to the nature of the environment. In this paper, we assume that $f_{X|\theta}(x)$ is the one-parameter exponential family, i.e.,

$$f_{X|\theta}(x) = u(x)c(\theta)e^{\theta x}, -\infty < x < \infty, \tag{2.1}$$

where $u(x) > 0$, $c(\theta) = [\int e^{\theta x}u(x)dx]^{-1}$ and $\Omega = \{\theta : c(\theta) > 0\}$, the natural parameter space. Consider the Bayes statistical decision problem for squared error loss estimation of θ using Y. If $\int_\Omega |\theta| \int_{-\infty}^{\infty} f_{X|\theta}(y-x)dF_\epsilon(x)dG(\theta) < \infty$ uniformly in y, the Bayes decision rule can be shown to be (see Zhang and Karunamuni [24])

$$\delta_G(y) = \frac{\int_{-\infty}^{\infty} f_X^{(1)}(y-x)dF_\epsilon(x) - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} f_X(y-x)dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_X(y-x)dF_\epsilon(x)}, \tag{2.2}$$

where

$$f_X(x) = \int_\Omega c(\theta)u(x)e^{\theta x}dG(\theta) \tag{2.3}$$

and $u^{(1)}$ and $f_X^{(1)}$ denote the first derivatives of u and f_X, respectively. The Bayes estimate δ_G minimizies the risk among all estimates. Its Bayes risk w.r.t. G is denoted by $R(G)$, the Bayes envelope value of the problem. If G is known, then we can use δ_G and attain $R(G)$. (It is assumed that G is such that $R(G) < \infty$.) This paper considers the case that G exists but it is completely unknown.

Suppose now that the above decision problem occurs $(n+1)$-times leading to the random vectors $(\theta_i, X_i, Y_i, \epsilon_i), i = 1, 2, ..., n+1$, where the each pair (θ_i, X_i) has the same probability structure as (θ, X) given above, $Y_i = X_i + \epsilon_i$, and the X_i's and the ϵ_i's are independent. The random vectors $\{(\theta_i, X_i, \epsilon_i)\}_{i=1}^{n+1}$ are unobservable and $\{Y_1, ..., Y_{n+1}\}$ is the only available observable data at the $(n+1)^{st}$ problem. In this setup, a generalization of the empirical Bayes approach of Robbins [14, 15] is applicable wherein one constructs estimates of the form $\delta_n(y) = \delta_n(y; Y_1, ..., Y_n)$ with $Y_{n+1} = y$ to estimate θ_{n+1}. Since $E(\theta_{n+1}|Y_1, ..., Y_{n+1}) = E(\theta_{n+1}|Y_{n+1})$, it follows that δ_G defined by (2.2) continues to be Bayes in the empirical Bayes problem. This motivates the use of the excess risk (regret) $R(\delta_n, G) - R(G)$ as a measure of goodness of the estimator δ_n.

DEFINITION 2.1. A sequence of estimates $\{\delta_n\}$ is said to be asymptotically optimal w.r.t. G with order a_n if

$$0 \le R(\delta_n, G) - R(G) \le c_1 a_n,$$

where c_1 is a positive constant (independent of n but may depend on G) and $\{a_n\}$ is a sequence of positive numbers such that $a_n \to 0$ as $n \to \infty$.

DEFINITION 2.2. A sequence of estimates $\{\delta_n\}$ is said to be asymptotically optimal *uniformly over a class of priors* $\mathcal{G}$ with order b_n if

$$0 \le \sup_{G \in \mathcal{G}} (R(\delta_n, G) - R(G)) \le c_2 b_n,$$

where c_2 is a positive constant (independent of n but may depend on $\mathcal{G}$) and $\{b_n\}$ is a sequence of posotive numbers such that $b_n \to 0$ as $n \to \infty$.

We shall find the sequence $\{b_n\}$ in Definition 2.2 for the following empirical Bayes estimator δ_n. Our δ_n is motivated by the fact that in view of the right hand side of (2.2), we need good estimators of f_X and $f_X^{(1)}$ based on $Y_1, ..., Y_n$. We use following kernel estimators of $f_X^{(l)}(l = 0, 1; f_X^{(0)} = f_X)$.

Let $\phi_\epsilon(t)$ denote the characteristic function of the error variable ϵ. Let $\hat{\phi}_n$ denote the empirical characterisric function defined by $\hat{\phi}_n(t) = \frac{1}{n}\sum_{j=1}^{n} exp(itY_j)$. For a nice symmetric (about 0) kernel K, let ϕ_K be its Fourier transform with $\phi_K(0)$=1. If the function $|t^l\phi_K(th_n)/\phi_\epsilon(t)|$ is integrable, define the kernel estimator

$$f_n^{(l)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} exp(-itx)(-it)^l\phi_K(th_n)\frac{\hat{\phi}_n(t)}{\phi_\epsilon(t)}dt \tag{2.4}$$

as the estimator of $f_X^{(l)}(x)$, where h_n is the bandwidth ($h_n \to 0$ as $n \to \infty$). This type of estimators are proposed by Stefanski and Carroll [19] and exclusively studied by Carroll and Hall [2] and Fan [7, 8 ,9]. Define

$$K_{nl}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} exp(-itx)\frac{(-it)^l\phi_K(t)}{\phi_\epsilon(t/h_n)}dt.$$

Note that (2.4) can be rewritten as a kernel type of estimate

$$f_n^{(l)}(x) = \frac{1}{nh_n^{l+1}}\sum_{j=1}^{n} K_{nl}(\frac{x - Y_j}{h_n}), l = 0, 1. \tag{2.5}$$

The fact that ϕ_K is real-valued and even implies that K_{nl} is real-valued, and so is $f_n^{(l)}(x)$; see Stefanski and Carroll [19] for more details. In view of (2.2), we can estimate δ_G through f_X and $f_X^{(1)}$, where f_X is given by (2.3). We define

$$\delta_n(y) = \left[\frac{\int_{-\infty}^{\infty} f_n^{(1)}(y - x)dF_\epsilon(x) - \int_{-\infty}^{\infty}\frac{u^{(1)}(y-x)}{u(y-x)}f_n(y - x)dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_n(y - x)dF_\epsilon(x)}\right]_{h_n^{-1}} \tag{2.6}$$

as our estimate of δ_G, where for $c > 0$, $[b]_c$ is $-c$, b or c according as $b < -c$, $|b| < c$, or $b > c$, and $f_n^{(l)}$ $(l = 0, 1; f_n^{(0)} = f_n)$ are given by (2.5). The preceeding idea of truncation is due to Singh [17].

Let $R(\delta_n, G)$ denote the Bayes risk of (2.6) w.r.t. G. Restricting to G to those with finite Bayes risk, then the excess risk $R(\delta_n, G) - R(G)$ satisfies (by an application of Lemma 2.1 of Singh [17])

$$0 \leq R(\delta_n, G) - R(G) = E(\delta_n - \delta_G)^2, \tag{2.7}$$

where $R(G)$ is the Bayes envelope value and δ_G is given by (2.2). Rates of convergence of the right hand side of (2.7) are given in the theorems below. First, we state some basic assumptions on the kernel K and the error variable ϵ. More assumptions will be given in the theorems.

(A1) The kernel K is a symmetric function about 0 on $(-\infty, \infty)$ and is of order k. That is, K satisfies $\int_{-\infty}^{\infty} K(y)dy = 1$, $\int_{-\infty}^{\infty} y^j K(y)dy = 0$ for $j = 1, ..., k-1$, and $\int_{-\infty}^{\infty} y^k K(y)dy \neq 0$.

(A2) $|\phi_\epsilon(t)| > 0$ for all t.

We also need to impose some smoothness conditions on the unknown densities f_X's given by (2.3). Define

$$\mathcal{F}_{B,k} = \{G : G \text{ is a prior on } \Omega \text{ such that } \sup_x |f_X^{(k)}(x)| \leq B\} \tag{2.8}$$

for some finite constant $B > 0$ and an integer $k \geq 1$.

THEOREM 2.1. *For any $G \in \mathcal{F}_{B,k}$, suppose that the distributions G and F_ϵ (the distribution of ϵ) satisfy $\int_\Omega \int_{-\infty}^{\infty} |\theta| f_{X|\theta}(y-x)dF_\epsilon(x)dG(\theta) < \infty$ uniformly in y. Further, suppose that* **(A1)** *and* **(A2)** *hold and the following conditions are satisfied:*

(B1) $\phi_K(t) = 0$ *for* $|t| \geq 1$*;*

(B2) $|\phi_\epsilon(t)||t|^{-\beta_0} exp(|t|^\beta/\gamma) \geq d_0$ *(as $t \to \infty$) for some positive constants* β, γ, d_0

and a constant β_0;

(B3) *for some δ $(\frac{1}{k} < \delta < \frac{1}{2})$ and $\xi > 0$*

$$\sup_{G\in\mathcal{F}_{B,k}} E\left[\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(Y-x)}{u(Y-x)}\right| dF_\epsilon(x)\right]^{\frac{2\delta}{(1-2\delta)}} < \infty, \quad \sup_{G\in\mathcal{F}_{B,k}} E|Y|^{\frac{2\delta(1+\xi)}{(1-2\delta)}} < \infty,$$

$$\sup_{G\in\mathcal{F}_{B,k}} E\left\{|Y|^{\frac{2\delta(1+\xi)}{(1-2\delta)}}\left[\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(Y-x)}{u(Y-x)}\right| dF_\epsilon(x)\right]^{\frac{2\delta}{(1-2\delta)}}\right\} < \infty$$

and

(B4) *for some $t' > 1$*

$$\sup_{G\in\mathcal{F}_{B,k}} E_G|\theta|^{2t'}\left(E_G|\theta|^{\frac{2(k\delta-1)}{(t'-1)}}\right)^{t'-1} < \infty.$$

Then, by choosing the bandwidth $h_n = (4/\gamma)^{1/\beta}(\log n)^{-1/\beta}$, we have

$$\sup_{G\in\mathcal{F}_{B,k}} [R(\delta_n, G) - R(G)] = O\left((\log n)^{-2(k\delta-1)/\beta}\right), \tag{2.9}$$

where δ_n is given by (2.6).

THEOREM 2.2. *For any $G \in \mathcal{F}_{B,k}$, suppose that the distributions G and F_ϵ (the distribution of ϵ) satisfy $\int_\Omega \int_{-\infty}^{\infty} |\theta| f_{X|\theta}(y-x)dF_\epsilon(x)dG(\theta) < \infty$ uniformly in y. Further, suppose that* **(A1)**, **(A2)**, **(B3)** *and* **(B4)** *hold and the following conditions are satisfied:*

(C1) *$\int_{-\infty}^{\infty} |\phi_K(t)t^{\beta+l}|dt < \infty$ and $\int_{-\infty}^{\infty} |\phi_K(t)t^{\beta+l}|^2 dt < \infty$, for some constants $\beta \geq 0$ and $l = 0, 1$, and*

(C2) *$|\phi_\epsilon(t)t^\beta| \geq d_0$ as $t \to \infty$, for some constants $d_0 > 0$ and $\beta \geq 0$.*

Then, by choosing the bandwidth $h_n = O\left(n^{-1/(2(k+\beta)+1)}\right)$, we have

$$\sup_{G\in\mathcal{F}_{B,k}} [R(\delta_n, G) - R(G)] = O\left(n^{-2(k\delta-1)/(2(k+\beta)+1)}\right). \tag{2.10}$$

REMARK 2.1. The examples of error distributions satisfying **(B2)** are normal, mixture normal and Cauchy, and the examples of distributions satisfying **(C2)** include gamma, double exponential and symmetric gamma distributions. These two types of distributions are called 'supersmooth' and 'ordinary smooth' distributions, respectively (Fan [7, 8, 9]).

REMARK 2.2. The rates of convergence in Theorem 2.1 is extremely slow compared to that of Theorem 2.2. This fact can be attributed to the smoothness of the distribution of the error variable ϵ. From the proofs below, it will be clear that the rates of convergence of the regret $R(\delta_n, G) - R(G)$ are obtained via the rates of convergence results of $(Ef_n^{(l)}(x) - f_X^{(l)}(x))$ and $Var f_n^{(l)}(x), l = 0, 1$ for each x, where $f_n^{(l)}(x)$ is given by (2.5). Estimation of f_X (the marginal density of X, see (2.3)) and its derivatives based on Y, knowing F_ϵ, is a deconvolution problem; see, e.g., Stafanski and Carroll [19] and Fan [7, 8, 9]. It is well-known now in the nonparametric deconvolution literature that the smoother the error distribution F_ϵ is, the harder the deconvolution will be; see Fan [7] for further details on this point. We believe that the deconvolution is rather inevitable in the EB estimation problem we discussed above when dealing with the contaminated data.

REMARK 2.3. In practice, the conditions in Theorems 2.1 and 2.2 are easy to verify, see the examples given below. The rate of convergence of Theorem 2.2 is rather compatible (but slightly slower) with the rate obtained in Singh [17] for the direct (or pure) data case. From a practical point of view this means that the errors which have an ordinary smooth distribution cause a little damage to the performance of the proposed EB estimator. But, the errors which have a supersmooth distribution will have a considerable effect on the performance of the EB estimator.

REMARK 2.4. A kernel which satisfies conditions **(A1)** and **(B1)** can be constructed as follows: First, choose a real-valued symmetric function ϕ_K with the support $[-1, 1]$ and satisfying $\phi_K = 1 + 0(t^k)$. For example, for $k = 2$, take

$$\phi_K(t) = \begin{cases} (1-t^2)^3 & \text{if } -1 \leq t \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

Then, the Fourier inverse K of ϕ_K is a k^{th} order kernel function since $\phi_K(0) = 1$ entails $\int_{-\infty}^{\infty} K(t)dt = 1$ and

$$\int_{-\infty}^{\infty} t^l K(t)dt = \frac{\phi_K^{(l)}(0)}{i^l} = 0, \text{ for } l = 1, ..., k-1.$$

For example, the Fourier inverse of (2.11) is

$$K(t) = \frac{48\cos t}{\pi t^4}\left(1 - \frac{15}{t^2}\right) - \frac{144\sin t}{\pi t^5}\left(2 - \frac{5}{t^2}\right), \quad -\infty < t < \infty. \tag{2.12}$$

REMARK 2.5. Assumption **(A1)** simply says that $K(\cdot)$ is a k^{th} order kernel. Assumption **(A2)** ensures that the true density $f_X(x)$is idetifiable from the model $Y = X + \epsilon$. Assumption **(B1)** is only required for Theorem 2.1, where the tail of the characteristic function of the error distribution ϵ converges to 0 at an exponential rate. In Theorem 2.2, this assumption is relaxed to **(C1)**. **(B3)** and **(B4)** are technical assumptions for achieving the rates of convergence of the regret in the theorems.

It is worthy to note that **(C2)** includes the model where only $100p\%$ ($0 \leq p \leq 1$) of the data are obtaibed with error and the remaining data are error-free. That is

$$Y = X + \epsilon, \tag{2.13}$$

with $P(\epsilon = 0) = 1 - p$ and $P(\epsilon = \epsilon^*) = p$, where ϵ^* is the error variable with distribution F_{ϵ^*} and the characteristic function $\phi_{\epsilon^*}(t)$. Denote the characteristic function of ϵ^* by ϕ_{ϵ^*}. Then

$$\phi_\epsilon(t) = (1-p) + p\phi_{\epsilon^*}(t). \tag{2.14}$$

If $Re(\phi_{\epsilon^*}) \geq 0$, then $|\phi_\epsilon(t)| \geq 1-p$. Here '*Re*' means the real-part of ϕ_{ϵ^*}. Thus, ϵ satisfies **(C2)** with $\beta = 0$ no matter what the ϵ^* is, supersmooth or ordinary smooth.

COROLLARY 2.1. *For any $G \in \mathcal{F}_{B,k}$, suppose that the distributions G and F_ϵ satisfy $\int_\Omega \int_{-\infty}^{\infty} |\theta| f_{X|\theta}(y-x)dF_\epsilon(x)dG(\theta) < \infty$ uniformly in y. Further, suppose that* **(A1)**, **(B3)**, **(B4)** *hold and the following conditions are satisfied:*

(D1) $\int_{-\infty}^{\infty} |\phi_K(t)t^l|dt < \infty$ *and* $\int_{-\infty}^{\infty} |\phi_K(t)t^l|^2 dt < \infty$, *for* $l = 0, 1$.

(D2) $Re(\phi_{\epsilon^*}(t)) \geq 0$, *for all* t.

Then, by choosing the bandwidth $h_n = O\left(n^{-1/(2k+1)}\right)$, *we have*

$$\sup_{G \in \mathcal{F}_{B,k}} [R(\delta_n^*, G) - R(G)] = O\left(n^{-2(k\delta-1)/(2k+1)}\right). \tag{2.15}$$

Corollary 2.1 is a direct consequence of Theorem 2.2. The proof will be omitted.

When the error distribution is 'supersmooth', the rate of convergence in Theorem 2.1 is too slow to be practical. Since the very common normal distribution is supersmooth, it is natural to find some ways by which the rates of convergence can be improved. However, Corollary 2.1 exhibits that the rate is not much affected if the data are partly contaminated even if the error ϵ is supersmooth. We now propose another model, under which the rate of convergence can be as good as that we obtained from the uncontaminated data.

Assume that all the data were contaminated but the noise level can be controlled. That is

$$Y = X + \epsilon, \tag{2.16}$$

where $\epsilon = \sigma_0\tilde{\epsilon}$, σ_0 parametrizes the noise level. Then

$$\phi_\epsilon(t) = \phi_{\tilde{\epsilon}}(\sigma_0 t). \tag{2.17}$$

With the corresponding EB estimator defined by (2.6), we have

THEOREM 2.3. *Let $\sigma_0 = O(n^{-1/(2k+1)})$, for some positive constant c. For any $G \in \mathcal{F}_{B,k}$, suppose that the distributions G and F_ϵ satisfy $\int_\Omega \int_{-\infty}^{\infty} |\theta| f_{X|\theta}(y - x) dF_\epsilon(x) dG(\theta) < \infty$ uniformly in y. Further, suppose that* **(A1)**, **(A2)** *with ϵ is replaced by $\tilde{\epsilon}$,* **(B3)** *and* **(B4)** *hold and the following condition is satisfied.*
(E1) $\int_{-\infty}^{\infty} |\phi_K(t)t^l|^2 dt < \infty$.
Then, by choosing the bandwidth $h_n = O\left(n^{-1/(2k+1)}\right)$, we have

$$\sup_{G \in \mathcal{F}_{B,k}} [R(\tilde{\delta}_n, G) - R(G)] = O\left(n^{-2(k\delta-1)/(2k+1)}\right). \tag{2.18}$$

EXAMPLE 1. Consider the exponential family in (2.1) with $u(x) = e^{-x^2/2}$ and $c(\theta) = (2\pi)^{-1/2} e^{-\theta^2/2}$. Then $f_{X|\theta}(x) = (2\pi)^{-1/2} e^{-(x-\theta)^2/2}$, $-\infty < x < \infty$. Also, the natural parameter space $\Omega = (-\infty, \infty)$. The marginal density (2.3) of X is now

$$f_X(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-(x-\theta)^2/2} dG(\theta). \tag{2.19}$$

Case (1): Suppose that the error distribution F_ϵ is $N(0,1)$. Then ϕ_ϵ satisfies **(B2)** with $\beta = 2$. From (2.19), by repeated differentiation under the integral sign (by Theorem 2.9 of Lehmann [11]), we obtain

$$f_X^{(k)}(x) = (-1)^k \int_{-\infty}^{\infty} H_k(x - \theta) f_{X|\theta}(x) dG(\theta)$$

for any G, where H_k is the k^{th} Hermite polynomial. Thus, $\sup_x |f_X^{(k)}(x)| \leq (2\pi)^{-1/2} \sum_{j=0}^{k} |a_j|$, for any G, where a_j is the j^{th} coefficient of the k^{th} Hermite polynomial. Therefore, $G \in \mathcal{F}_{B,k}$, where $B = (2\pi)^{-1/2} \sum_{j=0}^{k} |a_j|$. Also,

$\int_\Omega \int_{-\infty}^{\infty} |\theta| f_{X|\theta}(y-x) dF_\epsilon(x) dG(\theta) \le \int_{-\infty}^{\infty} |\theta| dG(\theta) < \infty$ if $E_G|\theta| < \infty$. Since $\frac{u^{(1)}(y-x)}{u(y-x)} = -(y-x)$, we have

$$\int_{-\infty}^{\infty} \left| \frac{u^{(1)}(y-x)}{u(y-x)} \right| dF_\epsilon(x) \le |y| + \int_{-\infty}^{\infty} |x| dF_\epsilon(x) \le |y| + C_1$$

for some finite constant C_1. By the C_r-inequality,

$$\left(\int_{-\infty}^{\infty} \left| \frac{u^{(1)}(y-x)}{u(y-x)} \right| dF_\epsilon(x) \right)^{2\delta/(1-2\delta)} \le C_\delta |y|^{2\delta/(1-2\delta)} + C_2$$

for some finite constants C_δ and C_2. Thus, to verify **(B3)** , it is enough to prove $\sup_{G\in\mathcal{F}_{B,k}} E|Y|^{2\delta(1+\xi)/(1-2\delta)} < \infty$ for some $k^{-1} < \delta < 1/2$ and $\xi > 0$. Now, the marginal density of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(y-x) dF_\epsilon(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (2\pi)^{-1/2} e^{-(y-x-\theta)^2/2} dG(\theta) dF_\epsilon(x).$$

Then, with $t = 2\delta(2+\xi)/(1-2\delta)$, by routine algebra we have

$$\begin{aligned} E|Y|^t &= (2\pi)^{-1/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |y|^t e^{-(y-x-\theta)^2/2} dG(\theta) dF_\epsilon(x) dy \\ &= (2\pi)^{-1/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |y|^t e^{-(y-\theta)^2/4} \int_{-\infty}^{\infty} e^{-(x-\frac{1}{2}(y-\theta))^2/2} dx dy dG(\theta) \\ &\le C_3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |y|^t e^{-(y-\theta)^2/4} dy dG(\theta) \\ &\le C_4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |y-\theta|^t e^{-(y-\theta)^2/4} dy dG(\theta) \\ &\quad + C_5 \int_{-\infty}^{\infty} |\theta|^t \int_{-\infty}^{\infty} e^{-(y-\theta)^2/4} dy dG(\theta) \\ &\le C_6 + C_5 \int_{-\infty}^{\infty} |\theta|^t dG(\theta) < \infty, \end{aligned}$$

where C_3, C_4, C_5 and C_6 are finite positive constants. Therefore, $\sup_{G\in\mathcal{F}_{B,k}} E|Y|^t < \infty$ if $\sup_{G\in\mathcal{F}_{B,k}} E_G|\theta|^t < \infty$. Let $t' = k\delta (> 1)$, then **(B4)** is equivalent to $\sup_{G\in\mathcal{F}_{B,k}} E_G|\theta|^{2k\delta} < \infty$. Combining above results, we see that $\sup_{G\in\mathcal{F}_{B,k}} E_G|\theta|^{2k\delta \vee \frac{2\delta(2+\xi)}{(1-2\delta)}} < \infty$ (for $1/k < \delta < 1/2, \xi > 0$) is sufficient for

(B3) and **(B4)**. If $1/k < \delta < 1/2 - 1/k$, then the assumption needed is $\sup_{G\in\mathcal{F}_{B,k}} E_G|\theta|^{2k\delta} < \infty$.

Case (2): Suppose the error distribution F_ϵ is $Gamma(1,p)$ with density

$$f_\epsilon(x) = \frac{1}{\Gamma(p)}x^{p-1}e^{-x}, x > 0, p > 1.$$

Then ϕ_ϵ satisfies **(C2)** with $\beta = p$. The Bayes estimator $\delta_G(y)$ defined by (2.2) now reduces to

$$\delta_G(y) = \frac{\int_0^\infty f_X^{(1)}(y-x)\frac{1}{\Gamma(p)}x^{p-1}e^{-x}dx + \int_0^\infty (y-x)f_X(y-x)\frac{1}{\Gamma(p)}x^{p-1}e^{-x}dx}{\int_0^\infty f_X(y-x)\frac{1}{\Gamma(p)}x^{p-1}e^{-x}dx},$$

where $f_X(x)$ is given by (2.19). The EB estimator δ_n defined by (2.6) now takes the form

$$\delta_n(y) = \left[\frac{\int_0^\infty f_n^{(1)}(y-x)\frac{1}{\Gamma(p)}x^{p-1}e^{-x}dx + \int_0^\infty (y-x)f_n(y-x)\frac{1}{\Gamma(p)}x^{p-1}e^{-x}dx}{\int_0^\infty f_n(y-x)\frac{1}{\Gamma(p)}x^{p-1}e^{-x}dx}\right]_{h_n^{-1}}.$$

Similar to Case (1) above, we can show that for $1/k < \delta < 1/2$, $\sup_{G\in\mathcal{F}_{B,k}} E_G|\theta|^{2k\delta\vee\frac{2\delta(2+\xi)}{(1-2\delta)}} < \infty$ is enough for **(B3)** and **(B4)**. For $1/k < \delta < 1/2 - 1/k$, the assumption becomes $\sup_{G\in\mathcal{F}_{B,k}} E_G|\theta|^{2k\delta} < \infty$.

EXAMPLE 2. Consider the exponential family in (2.1) with $u(x) = 2e^{-x^2} + e^{-2x^2}$ and $c(\theta) = \frac{1}{(\pi)^{1/2}(1+2e^{\theta^2/8})e^{\theta^2/8}}$. Then

$$f_X(x) = (\pi)^{-1/2}e^{-x^2}(2+e^{-x^2})\int_{-\infty}^{\infty}\frac{e^{\theta x)}}{(1+2e^{\theta^2/8})e^{\theta^2/8}}dG(\theta). \tag{2.20}$$

Since $\frac{u^{(1)}(y-x)}{u(y-x)} = -4(y-x)$, similar to the proof in Example 1, we can prove that **(B3)** and **(B4)** are satisfied if $\sup_{G\in\mathcal{F}_{B,k}} E_G|\theta|^{2k\delta\vee\frac{2\delta(2+\xi)}{(1-2\delta)}} < \infty$, for $1/k < \delta < 1/2, \xi > 0$ and $\epsilon \sim N(0,1)$ or $\epsilon \sim Gamma(1,p)$, $p > 1$. Details are omitted here.

3. Proofs

In this section we prove Theorem 2.1 above. The proofs of Theorems 2.2 and Theorem 2.3 are similar. First we state five lemmas useful in proving these theorems. For proofs of Lemmas 3.1 and 3.2, see Theorems 3.1 and 3.2 of Fan [7]. Also, see Fan [8, 9].

LEMMA 3.1. Under the assumptions of **(A1)**, **(A2)**, **(B1)** and **(B2)** and with the choice $h_n = (4/\gamma)^{\frac{1}{\beta}}(\log n)^{-1/\beta}$ of the bandwidth, we have

$$\sup_{-\infty<x<\infty} \sup_{G\in\mathcal{F}_{B,k}} E|f_n^{(l)}(x) - f_X^{(l)}(x)| \leq Const. \times (\log n)^{-(k-l)/\beta} \tag{3.1}$$

for $l = 0, 1$, where $f_n^{(l)}(x), f_X(x)$ and $\mathcal{F}_{B,k}$ are given by (2.4), (2.3) and (2.8), respectively.

LEMMA 3.2. Under the assumptions of **(A1)**, **(A2)**, **(C1)** and **(C2)** and with the choice $h_n = O(n^{-1/(2(k+\beta)+1)})$ of the bandwidth, we have

$$\sup_{-\infty<x<\infty} \sup_{G\in\mathcal{F}_{B,k}} E|f_n^{(l)}(x) - f_X^{(l)}(x)| \leq Const. \times n^{-(k-l)/(2(k+\beta)+1)} \tag{3.2}$$

for $l = 0, 1$, where$f_n^{(l)}(x), f_X(x)$ and $\mathcal{F}_{B,k}$ are given by (2.4), (2.3) and (2.8), respectively.

The following lemma is a version of Singh-Datta inequality, see Datta [3].

LEMMA 3.3. For every pair (Y, Y') of random variables and for real numbers $y \neq 0, y', 0 < L < \infty$ and $0 < \gamma \leq 2$

$$E\left(\left|\frac{Y'}{Y} - \frac{y'}{y}\right| \wedge L\right)^\gamma \leq 2|y|^{-\gamma}\left\{E|Y' - y'|^\gamma + \left(\left|\frac{y'}{y}\right| + L\right)^\gamma E|Y - y|^\gamma\right\}. \tag{3.3}$$

LEMMA 3.4. Let $\delta_n(y)$ and $\delta_G(y)$ be defined by (2.6) and (2.2), respectively. Then for any $0 < t < 1$, we have

$$E(\delta_n(y) - \delta_G(y))^2$$

$$\le 2^{3-t} h_n^{t-2} f_Y^{-t}(y) \Big\{ C_t \Big(\int_{-\infty}^{\infty} E|f_n^{(1)}(y-x) - f_X^{(1)}(y-x)| dF_\epsilon(x) \Big)^t$$
$$+ C_t \Big(\int_{-\infty}^{\infty} \Big| \frac{u^{(1)}(y-x)}{u(y-x)} \Big| E|f_n(y-x) - f_X(y-x)| dF_\epsilon(x) \Big)^t$$
$$+ 3^t h_n^{-t} \Big(\int_{-\infty}^{\infty} E|f_n(y-x) - f_X(y-x)| dF_\epsilon(x) \Big)^t \Big\} I(|\delta_G(y)| \le h_n^{-1})$$
$$+ 2|\delta_G(y)|^2 I(|\delta_G(y)| > h_n^{-1}), \tag{3.4}$$

where C_t is a finite positive constant independent of n, y and G.

PROOF OF LEMMA 3.4: Write

$$\begin{aligned} E(\delta_n(y) - \delta_G(y))^2 &= E(\delta_n(y) - \delta_G(y))^2 I(|\delta_G(y)| \le h_n^{-1}) \\ &\quad + E(\delta_n(y) - \delta_G(y))^2 I(|\delta_G(y)| > h_n^{-1}) \\ &= I_1 + I_2. \end{aligned}$$

Then, by straightforward simplifications we obtain, for $0 < t < 1$,

$$\begin{aligned} I_1 &= E|\delta_n(y) - [\delta_G(y)]_{h_n^{-1}}|^2 I(|\delta_G(y)| \le h_n^{-1}) \\ &= E\Bigg| \left[\frac{\int_{-\infty}^{\infty} f_n^{(1)}(y-x) dF_\epsilon(x) - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} f_n(y-x) dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_n(y-x) dF_\epsilon(x)} \right]_{h_n^{-1}} \\ &\quad - \left[\frac{\int_{-\infty}^{\infty} f_X^{(1)}(y-x) dF_\epsilon(x) - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} f_X(y-x) dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_X(y-x) dF_\epsilon(x)} \right]_{h_n^{-1}} \Bigg|^2 \\ &\quad I(|\delta_G(y)| \le h_n^{-1}) \\ &\le E\Bigg\{ \Bigg| \frac{\int_{-\infty}^{\infty} f_n^{(1)}(y-x) dF_\epsilon(x) - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} f_n(y-x) dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_n(y-x) dF_\epsilon(x)} \\ &\quad - \frac{\int_{-\infty}^{\infty} f_X^{(1)}(y-x) dF_\epsilon(x) - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} f_X(y-x) dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_X(y-x) dF_\epsilon(x)} \Bigg| \wedge 2h_n^{-1} \Bigg\}^2 \\ &\quad I(|\delta_G(y)| \le h_n^{-1}) \\ &\le \left(2h_n^{-1}\right)^{2-t} E\Bigg\{ \Bigg| \frac{\int_{-\infty}^{\infty} f_n^{(1)}(y-x) dF_\epsilon(x) - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} f_n(y-x) dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_n(y-x) dF_\epsilon(x)} \end{aligned}$$

$$- \left. \frac{\int_{-\infty}^{\infty} f_X^{(1)}(y-x)dF_\epsilon(x) - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} f_X(y-x)dF_\epsilon(x)}{\int_{-\infty}^{\infty} f_X(y-x)dF_\epsilon(x)} \right| \wedge 2h_n^{-1} \Bigg\}^t$$

$$I(|\delta_G(y)| \le h_n^{-1})$$

$$= \left(2h_n^{-1}\right)^{2-t} \Delta(y),$$

where $\Delta(y)$ denotes the expectation term. By an application of Lemma 3.3 followed by the Hölder inequality we have

$$\begin{aligned}
\Delta(y) \le\ & 2f_Y^{-t}(y) \Bigg\{ E \Bigg| \int_{-\infty}^{\infty} \left(f_n^{(1)}(y-x) - f_X^{(1)}(y-x) \right) dF_\epsilon(x) \\
& - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} \left(f_n(y-x) - f_X(y-x) \right) dF_\epsilon(x) \Bigg|^t \\
& + (|\delta_G(y)| + 2h_n^{-1})^t E \left| \int_{-\infty}^{\infty} \left(f_n(y-x) - f_X(y-x) \right) dF_\epsilon(x) \right|^t \Bigg\} \\
& I(|\delta_G(y)| \le h_n^{-1}) \\
\le\ & 2f_Y^{-t}(y) \Bigg\{ \Bigg[E \Bigg| \int_{-\infty}^{\infty} \left(f_n^{(1)}(y-x) - f_X^{(1)}(y-x) \right) dF_\epsilon(x) \\
& - \int_{-\infty}^{\infty} \frac{u^{(1)}(y-x)}{u(y-x)} \left(f_n(y-x) - f_X(y-x) \right) dF_\epsilon(x) \Bigg| \Bigg]^t \\
& + (|\delta_G(y)| + 2h_n^{-1})^t \left[E \left| \int_{-\infty}^{\infty} \left(f_n(y-x) - f_X(y-x) \right) dF_\epsilon(x) \right| \right]^t \Bigg\} \\
& I(|\delta_G(y)| \le h_n^{-1}) \\
\le\ & 2f_Y^{-t}(y) \Bigg\{ \Bigg(E \int_{-\infty}^{\infty} \left| f_n^{(1)}(y-x) - f_X^{(1)}(y-x) \right| dF_\epsilon(x) \\
& + \int_{-\infty}^{\infty} \left| \frac{u^{(1)}(y-x)}{u(y-x)} \right| |f_n(y-x) - f_X(y-x)| \, dF_\epsilon(x) \Bigg)^t \\
& + (|\delta_G(y)| + 2h_n^{-1})^t \left(E \int_{-\infty}^{\infty} |f_n(y-x) - f_X(y-x)| \, dF_\epsilon(x) \right)^t \Bigg\} \\
& I(|\delta_G(y)| \le h_n^{-1}) \\
\le\ & 2f_Y^{-t}(y) \Bigg\{ C_t \left(E \int_{-\infty}^{\infty} \left| f_n^{(1)}(y-x) - f_X^{(1)}(y-x) \right| dF_\epsilon(x) \right)^t \\
& + C_t \left(\left| \frac{u^{(1)}(y-x)}{u(y-x)} \right| E \int_{-\infty}^{\infty} |f_n(y-x) - f_X(y-x)| \, dF_\epsilon(x) \right)^t \\
& + (|\delta_G(y)| + 2h_n^{-1})^t \left(E \int_{-\infty}^{\infty} |f_n(y-x) - f_X(y-x)| \, dF_\epsilon(x) \right)^t \Bigg\}
\end{aligned}$$

$I(|\delta_G(y)| \le h_n^{-1})$,

the last inequality follows from the $C_r - inequality$, where C_t is the coefficient from the $C_r - inequality$. Thus, by Fubini's theorem we obtain

$$\begin{aligned} I_1 &\le 2^{3-t} h_n^{t-2} f_Y^{-t}(y) \Big\{ C_t \left(\int_{-\infty}^{\infty} E|f_n^{(1)}(y-x) - f_X^{(1)}(y-x)| dF_\epsilon(x) \right)^t \\ &+ C_t \left(\int_{-\infty}^{\infty} \left| \frac{u^{(1)}(y-x)}{u(y-x)} \right| E|f_n(y-x) - f_X(y-x)| dF_\epsilon(x) \right)^t \\ &+ 3^t h_n^{-t} \left(\int_{-\infty}^{\infty} E|f_n(y-x) - f_X(y-x)| dF_\epsilon(x) \right)^t \Big\} \\ & I(|\delta_G(y)| \le h_n^{-1}). \end{aligned} \quad (3.5)$$

Obviously,

$$I_2 \le 4|\delta_G(y)|^2 I(|\delta_G(y)| > h_n^{-1}). \quad (3.6)$$

Combining (3.5) and (3.6) completes the proof.

LEMMA 3.5. For model (2.16), under the assumptions **(A1)**, **(A2)** with ϵ is replaced by $\tilde{\epsilon}$, and **(E1)** $\int_{-\infty}^{\infty} |\phi_K(t) t^l|^2 dt < \infty$, $l = 0, 1$. With the choice of the bandwidth $h_n = O\left(n^{-1/(2k+1)}\right)$, we have

$$\sup_{-\infty < x < \infty} \sup_{G \in \mathcal{F}_{B,k}} E|f_n^{(l)}(x) - f_X^{(l)}(x)| \le Const. \times n^{-(k-l)/(2k+1)} \quad (3.7)$$

for $l = 0, 1$, where$f_n^{(l)}(x), f_X(x)$ and $\mathcal{F}_{B,k}$ are given by (2.4), (2.3) and (2.8), respectively.

PROOF: Write

$$\begin{aligned} E|f_n^{(l)}(x) - f_X^{(l)}(x)| &\le |Ef_n^{(l)}(x) - f_X^{(l)}(x)| + (Var f_n^{(l)}(x))^{1/2} \\ &= J_1 + J_2. \end{aligned}$$

One can easily prove that $J_1 = O(h_n^{k-l})$, for all x and $G \in \mathcal{F}_{B,k}$, see Fan [7]. Further, note that $\sup_{-\infty<y<\infty} \sup_{G\in\mathcal{F}_{B,k}} f_Y(y) < C$ for some finite positive constant C, by Lemma 1 of Bickel and Ritov [1]. Since $\phi_\epsilon(t/h_n) = \phi_{\tilde\epsilon}(\sigma_0 t/h_n) = \phi_{\tilde\epsilon}(t)$ when $\sigma_0 = O(h_n) = cn^{-1/(2k+1)}$, for some positive constant c. Then

$$\int_{-\infty}^{\infty} \frac{|\phi_K(t)t^l|^2}{|\phi_\epsilon(t/h_n)|^2} dt = \int_{-\infty}^{\infty} \frac{|\phi_K(t)t^l|^2}{|\phi_{\tilde\epsilon}(ct)|^2} dt.$$

Therefore

$$\begin{aligned}
Var f_n^{(l)}(x) &\leq \frac{1}{nh_n^{2+2l}} E\left|K_{nl}(\frac{x-Y_1}{h_n})\right|^2 \\
&= \frac{1}{nh_n^{1+2l}} \int_{-\infty}^{\infty} |K_{nl}(y)|^2 f_Y(x-h_n y)dy \\
&\leq \sup_{-\infty<y<\infty} f_Y(y) \frac{1}{nh_n^{1+2l}} \int_{-\infty}^{\infty} |K_{nl}(y)|^2 dy \\
&\leq \frac{C}{2\pi} \frac{1}{nh_n^{1+2l}} \int_{-\infty}^{\infty} \frac{|\phi_K(t)t^l|^2}{|\phi_\epsilon(t/h_n)|^2} dt \\
&\leq \frac{C}{2\pi} \frac{1}{nh_n^{1+2l}} \int_{-\infty}^{\infty} \frac{|\phi_K(t)t^l|^2}{|\phi_{\tilde\epsilon}(ct)|^2} dt \\
&\leq \frac{C}{2\pi \min|\phi_{\tilde\epsilon}(t)|^2} \int_{-\infty}^{\infty} |\phi_K(t)t^l|^2 dt \\
&\leq C\frac{1}{nh_n^{1+2l}},
\end{aligned}$$

under assumptions **(A2)** with ϵ replaced by $\tilde\epsilon$ and **(E1)**, where C is a finite positive constant independent of n, x and G. In different locations, it may take different values. Then, $J_2 = \left(C\frac{1}{nh_n^{1+2l}}\right)^{1/2}$, for all x and $G \in \mathcal{F}_{B,k}$. Combining J_1 and J_2 and choosing $h_n = O\left(n^{-1/(2k+1)}\right)$ completes the proof.

PROOF of THEOREM 2.1: By Lemma 3.1 and 3.4 with $t = 2\delta(<1)$ in Lemma 3.4 and the bandwidth $h_n = (4/\gamma)^{\frac{1}{\beta}}(\log n)^{-1/\beta}$, one obtains

$$\begin{aligned}
&E(\delta_n(y) - \delta_G(y))^2 \\
&\leq 2^{3-2\delta}(\log n)^{2(1-\delta)/\beta} f_Y^{-2\delta}(y) \left\{ C_1(\log n)^{-2\delta(k-1)/\beta} \left(\int_{-\infty}^{\infty} dF_\epsilon(x)\right)^{2\delta}\right.
\end{aligned}$$

$$
\begin{aligned}
&+ C_2(\log n)^{-2k\delta/\beta}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right| dF_\epsilon(x)\right)^{2\delta} \\
&+ C_3(\log n)^{-2\delta(k-1)/\beta}\left(\int_{-\infty}^{\infty} dF_\epsilon(x)\right)^{2\delta}\Bigg\} I(|\delta_G(y)| \le h_n^{-1} \\
&+ 4|\delta_G(y)|^2 I(|\delta_G(y)| > h_n^{-1}) \\
&\le C_0(\log n)^{-2(k\delta-1)/\beta} f_Y^{-2\delta}(y) \\
&+ \left\{C_1 + C_2(\log n)^{-2\delta/\beta}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right| dF_\epsilon(x)\right)^{2\delta}\right\} I(|\delta_G(y)| \le h_n^{-1}) \\
&+ 4|\delta_G(y)|^2 I(|\delta_G(y)| > h_n^{-1}),
\end{aligned}
\tag{3.8}
$$

where C_0, C_1 and C_2 are finite positive constants independent of n, y and G.

Note that $R(G) < \infty$ is guaranteed by $E\theta^2 < \infty$. By (2.7) and (3.7), we have

$$
\begin{aligned}
&R(\delta_n, G) - R(G) \\
&\quad = E(\delta_n(Y) - \delta_G(Y))^2 \\
&\quad \le C_0(\log n)^{-2(k\delta-1)/\beta}\Bigg\{C_1 \int_{-\infty}^{\infty} f_Y^{1-2\delta}(y)dy \\
&\quad +C_2(\log n)^{-2\delta/\beta}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right| dF_\epsilon(x)\right)^{2\delta} f_Y^{1-2\delta}(y)dy\Bigg\} \\
&\quad + 4E|\delta_G(Y)|^2 I(|\delta_G(Y)| > h_n^{-1}).
\end{aligned}
$$

But

$$
E|\delta_G(Y)|^2 I(|\delta_G(Y)| > h_n^{-1}) \le \left\{E|\delta_G(Y)|^{2t'}\right\}^{1/t'} \left\{EI(|\delta_G(Y)| > h_n^{-1})\right\}^{(t'-1)/t'},
$$

for any $t' > 1$. By Markov inequality, we get

$$
\begin{aligned}
&E|\delta_G(Y)|^2 I(|\delta_G(Y)| > h_n^{-1}) \\
&\quad \le \left\{E|\delta_G(Y)|^{2t'}\right\}^{1/t'} \left\{E|\delta_G(Y)|^{2(k\delta-1)t'/(t'-1)}\right\}^{(t'-1)/t'} h_n^{2(k\delta-1)} \\
&\quad \le h_n^{2(k\delta-1)} \left\{E|\theta|^{2t'}\right\}^{1/t'} \left\{E|\theta|^{2(k\delta-1)t'/(t'-1)}\right\}^{(t'-1)/t'},
\end{aligned}
$$

for $\delta > 1/k$. Thus,

$$
\begin{aligned}
&R(\delta_n, G) - R(G) \\
&\quad \le C_0(\log n)^{-2(k\delta-1)/\beta}\left\{C_1\int_{-\infty}^{\infty} f_Y^{1-2\delta}(y)dy\right. \\
&\quad \left. + C_2(\log n)^{-2\delta/\beta}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right| dF_\epsilon(x)\right)^{2\delta} f_Y^{1-2\delta}(y)dy\right\} \\
&\quad + 4(\log n)^{-2(k\delta-1)/\beta}\left\{E|\theta|^{2t'}\right\}^{1/t'}\left\{E|\theta|^{2(k\delta-1)t'/(t'-1)}\right\}^{(t'-1)/t'}.
\end{aligned}
$$

So, if

(i) $\sup_{G\in\mathcal{F}_{B,k}} \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right| dF_\epsilon(x)\right)^{2\delta} f_Y^{1-2\delta}(y)dy < \infty$,

(ii) $\sup_{G\in\mathcal{F}_{B,k}} \int_{-\infty}^{\infty} f_Y^{1-2\delta}(y)dy < \infty$, and

(iii) $\sup_{G\in\mathcal{F}_{B,k}} E|\theta|^{2t'\vee\left(\frac{2(k\delta-1)t'}{t'-1}\right)} < \infty$, for some $t' > 1$,

we have

$$
\sup_{G\in\mathcal{F}_{B,k}} [R(\delta_n, G) - R(G)] = O\left((\log n)^{-2(k\delta-1)/\beta}\right). \tag{3.9}
$$

For simplicity of the assumptions, we give sufficient conditions for (i) and (ii) to hold. For any $\xi > 0$, using the Hölder inequality followed by the $C_r - inequality$,

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right| dF_\epsilon(x)\right)^{2\delta} f_Y^{1-2\delta}(y)dy \\
&\quad = \int_{-\infty}^{\infty}(1+|y|)^{-2\delta(1+\xi)} \\
&\quad (1+|y|)^{2\delta(1+\xi)}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right| dF_\epsilon(x)\right)^{2\delta} f_Y^{1-2\delta}(y)dy \\
&\quad \le \left(\int_{-\infty}^{\infty}(1+|y|)^{-(1+\xi)}dy\right)^{2\delta} \\
&\quad \left(\int_{-\infty}^{\infty}(1+|y|)^{\frac{2\delta(1+\xi)}{1-2\delta}}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right| dF_\epsilon(x)\right)^{\frac{2\delta}{1-2\delta}} f_Y(y)dy\right)^{1-2\delta} \\
&\quad \le C_3\left\{\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right| dF_\epsilon(x)\right)^{\frac{2\delta}{1-2\delta}} f_Y(y)dy\right.
\end{aligned}
$$

$$+\int_{-\infty}^{\infty}|y|^{\frac{2\delta(1+\xi)}{1-2\delta}}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right|dF_\epsilon(x)\right)^{\frac{2\delta}{1-2\delta}}f_Y(y)dy\Bigg\}^{1-2\delta}$$

$$=C_3\left\{E\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(Y-x)}{u(Y-x)}\right|dF_\epsilon(x)\right)^{\frac{2\delta}{1-2\delta}}\right.$$

$$\left.+E|Y|^{\frac{2\delta(1+\xi)}{1-2\delta}}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(Y-x)}{u(Y-x)}\right|dF_\epsilon(x)\right)^{\frac{2\delta}{1-2\delta}}\right\}^{1-2\delta},$$

where C_3 is a finite positive constant independent of n, y and G. Then,

$$\sup_{G\in\mathcal{F}_{B,k}}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(y-x)}{u(y-x)}\right|dF_\epsilon(x)\right)^{2\delta}f_Y^{1-2\delta}(y)dy<\infty,$$

if

$$\sup_{G\in\mathcal{F}_{B,k}}E\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(Y-x)}{u(Y-x)}\right|dF_\epsilon(x)\right)^{\frac{2\delta}{1-2\delta}}<\infty$$

and

$$\sup_{G\in\mathcal{F}_{B,k}}E|Y|^{\frac{2\delta(1+\xi)}{1-2\delta}}\left(\int_{-\infty}^{\infty}\left|\frac{u^{(1)}(Y-x)}{u(Y-x)}\right|dF_\epsilon(x)\right)^{\frac{2\delta}{1-2\delta}}<\infty.$$

Similarly, we can show $\sup_{G\in\mathcal{F}_{B,k}}E|Y|^{\frac{2\delta(1+\xi)}{1-2\delta}}<\infty$, for some $\xi>0$, is sufficient for (ii). Further, notice that the RHS of (3.8) is independent of G. Hence the result.

PROOFS OF THEOREM 2.2 and THEOREM 2.3: The proofs are the same as that of Theorem 2.1 except that now we use Lemma 3.2 in the proof of Theorem 2.2 and Lemma 3.5 in the proof of THEOREM 2.3, instead of Lemma 3.1.

References

[1] Bickel, P.J. and Ritov, Y.: Estimating integrated squared density derivatives: Sharp best order convergence estimates. *Sankhya*, Ser. A 50, 381-393 (1988).

[2] Carroll, R.J. and Hall, P.: Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* 83, 1184-1186 (1988).

[3] Datta, S.: Asymptotic optimality of Bayes compound estimators in compact exponential families. *Ann. Statist.* 19, 354-365 (1991).

[4] Datta, S.: Nonparametric empirical Bayes estimation with $O(n^{-1/2})$ rate of a truncation parameter. *Statistics and Decisions,* 9, 45-61 (1991).

[5] Datta, S.: Empirical Bayes estimation in a threshold model. *Sankhya,* Ser. A, 54, 106-117 (1994).

[6] Efron, B. and Morris, C.: Stein's estimation rule and its competitors-an empirical Bayes approach. *J. Amer. Statist. Assoc.* 68, 117-130 (1973).

[7] Fan, J.: On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* 19, 1257-1272 (1991).

[8] Fan, J.: Global behavior of deconvolution kernel estimators. *Statistica Sinica* 1, 541-551 (1991).

[9] Fan, J.: Deconvolution with supersmooth distributions. *Canad. J. Statist.* 20, 155-169 (1992).

[10] Hannan, J. and Macky, D.W.: Empirical Bayes squared-error loss estimation of unbounded functionals in exponential families. *Report No. RM- 290,* Dept. Statistics and Probablity, Michigan State University. (1971).

[11] Lehmann, E.L.: *Testing Statistical Hypothesis.* Wiley, New York (1959).

[12] Lin, P.E.: Rates of convergence in empirical Bayes estimation problems. Continuous case. *Ann. Statist.* 3, 155-164 (1975).

[13] Maritz, J.S. and Lwin, T.: *Empirical Bayes methods.* 2^{nd} edition. Chapman and Hall, London (1989).

[14] Robbins, H.: An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob. I,* University of California Press, 157-163 (1956).

[15] Robbins, H.: The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* 35, 1-20 (1964).

[16] Singh, R.S.: Empirical Bayes estimation with convergence rates in non-continuous Lebesgue-exponential families. *Ann. Statist.* 5, 394-400 (1976).

[17] Singh, R.S.: Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate. *Ann. Statist.* 7, 890-902 (1979).

[18] Singh, R.S. and Wei, L.: Empirical Bayes with rates and best rates of convergence in $u(x)C(\theta)\exp(-x/\theta)$-family: Estimation case. *Ann. Inst. Statist. Math.* 44, 435-449 (1992).

[19] Stefanski, L.A. and Carroll, R.J.: Deconvoluting kernel density estimators. *Statistics*, 21, 169-184 (1990).

[20] Susarla, V.: Empirical Bayes theory,. *Encyclopedia of Statistical Sciences* (eds. S. U. Kotz and N. L. Johnson), Vol. 2, 440-503, Wily, New York, (1982).

[21] Van Houwelingen, J.C. and Stijnen, Th.: Monotone empirical Bayes estimators for the continuous one-parameter exponential family. *Statist. Neerlandica,* 37, 29-43 (1983).

[22] Yu, Benito Ong.: Rates of convergence in empirical Bayes two-action and estimation problems and in extended sequence-compound estimation problems. *Ph.D. Thesis.* Michigan State University, East Lansing, Michigan (1970).

[23] Zhang, C.H.: Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* 18, 806-830 (1990).

[24] Zhang, S. and Karunamuni, R.J.: Bayes and empirical Bayes estimation with errors in variables. *Statist. Prob. Letters,* 33, 23-34 (1997).

Chapter 3

Empirical Bayes Two-Action Problem for the Continuous One-Parameter Exponential Family with Errors in Variables

1. Introduction

Let (X, θ) be a random vector, where θ has a prior distribution G, given θ, X has a density $f_{X|\theta}(x)$ with respect to Lebesgue measure on the real line. The pair (X, θ) is not observable. Instead, we observe only Y, where $Y = X + \varepsilon$ with ε denoting the random error. We assume that ε is independent of (X, θ) and has a known distribution F_ε on $(-\infty, \infty)$. This is the situation when one encounters in analyzing contaminated data (errors in variables) due to measurement error or due to nature of the environment. In this paper, we assume that $f_{X|\theta}(x)$ has the following form:

$$f_{X|\theta}(x) = u(x)C(\theta)e^{\theta x}, -\infty \leq a < x < b \leq \infty, \tag{1.1}$$

where $u > 0$ on (a, b) and $C(\theta) = (\int_a^b u(x)e^{x\theta}dx)^{-1}$. Let Ω denote the natural parameter space, $\Omega = \{\theta : C(\theta) > 0\}$. We study the testing problem, $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$, based on an observation Y with the loss function $L(\theta, 0)$

Results of this chapter were presented as an invited talk in the international meeting "Statistical Research for the 21st Century" held in November, 1996 in Montrèal in honor of Professor C. R. Rao. A version of this chapter has been submitted for the proceedings of this meeting, which is to be published as a special issue of Journal of Statistical Planning and Inference.

$= \max\{\theta - \theta_0, 0\}$ for accepting H_0, and $L(\theta, 1) = \max\{\theta_0 - \theta, 0\}$ for accepting H_1. If $\delta(y) = Pr\{\text{accepting} H_0 | Y = y\}$ denote a randomized decision rule for the preceding testing problem, then its Bayes risk with respect to (w.r.t.) G is given by

$$R(\delta, G) = \int_{-\infty}^{\infty} \alpha_G(y)\delta(y)dy + \int_{\Omega} L(\theta, 0)dG(\theta), \tag{1.2}$$

where

$$\alpha_G(y) = \int_{\Omega} \theta f_{Y|\theta}(y)dG(\theta) - \theta_0 f_Y(y) \tag{1.3}$$

with

$$f_Y(y) = \int_{\Omega} f_{Y|\theta}(y)dG(\theta) \tag{1.4}$$

and $f_{Y|\theta}(y)$ denotes density of Y given θ, i.e., $f_{Y|\theta}(y) = \int f_{X|\theta}(y - x)dF_\epsilon(x)$. Then a Bayes rule w.r.t. G (i.e., a minimizer of (1.2)) for the present testing problem is given by

$$\delta_G(y) = \begin{cases} 1 & \text{if } \alpha_G(y) \leq 0 \\ 0 & \text{elsewhere,} \end{cases} \tag{1.5}$$

where α_G is given by (1.3). Let $R(G)$ denote the Bayes risk of δ_G w.r.t. G :

$$R(G) = R(\delta_G, G) = \inf_{\delta'} R(\delta', G). \tag{1.6}$$

The quantity $R(G)$ is called the Bayes risk envelope value of the problem.

When the prior G is not completely known, then the Bayes rule (1.5) is not available. The empirical Bayes (EB) approach considers the situation where G is unknown but information is available from the past experiences. Suppose now that the above decision problem occurs $(n+1)$-times leading to the random vectors $(\theta_i, X_i, Y_i, \epsilon_i)$, $i = 1, 2, \ldots, n+1$, where the each pair (X_i, θ_i) has the same probability structure as (X, θ) given above, $Y_i = X_i + \varepsilon_i$, and the X_i's and the ε_i's are independent. The random vectors $\{(\theta_i, X_i, \varepsilon_i)\}_{i=1}^{n+1}$ are unobservable and $\{Y_1, \ldots, Y_{n+1}\}$ are the only available observable data at the $(n+1)^{\text{st}}$-problem. In this set up, a generalization of the EB approach of Robbins (1956, 1964) is applicable wherein one constructs testing rules of the form $\delta_n(y) = \delta_n(y; Y_1, \ldots, Y_n)$ with $Y_{n+1} = y$ to make a decision about θ_{n+1}. Since $E(\theta_{n+1}|Y_1, \ldots, Y_{n+1}) = E(\theta_{n+1}|Y_{n+1})$, δ_G defined by (1.5) continues to be Bayes in the EB problem. This motivates the use of the regret (excess risk) as a measure of goodness of an EB rule δ_n.

Definition 1.1. A sequence of EB testing rules $\{\delta_n\}$ is said to be asymptotically optimal w.r.t. G with order a_n if

$$0 \leq R(\delta_n, G) - R(G) \leq c_1 a_n,$$

where c_1 is a positive constant (independent of n but may depend on G) and $\{a_n\}$ is a sequence of postiive numbers such that $a_n \to 0$ as $n \to \infty$,

Definition 1.2. A sequence of EB testing rules $\{\delta_n\}$ is said to be asymptotically optimal uniformly over a class of priors $\mathcal{G}$ with order b_n if

$$0 \leq \sup_{G \in \mathcal{G}} (R(\delta_n, G) - R(G)) \leq c_2 b_n,$$

where c_2 is a positive constant (independent of n but may depend on $\mathcal{G}$) and $\{b_n\}$ is a sequence of positive numbers such that $b_n \to 0$ as $n \to \infty$.

Work on EB problems in which the observed data are contaminated first appeared in Zhang and Karunamuni (1997a, 1997b). They discuss the EB estimation problem with squared error loss. There has been a great deal of work done on the linear loss two-action problem for the continuous one-parameter exponential family in the case of uncontaminated data, that is of the 'pure' data case. A notable work includes Samuel (1963), Yu (1970), Johns and Van Ryzin (1972), Van Houwelingen (1976), Stijnen (1985), Karunamuni and Yang (1995) and Karunamuni (1996), among others.

The proposed EB testing rule is given in Section 2 and its asymptotic properties are considered in Section 3. The proofs are deferred to Section 4.

2. The Proposed Empirical Bayes Rule

A natural estimator of the Bayes rule δ_G defined by (1.5) can be obtained by estimating α_G defined by (1.3) based on the past observed data $Y_1, \ldots, Y_n$. Note that

$$\int_\Omega \theta f_{Y|\theta}(y)dG(\theta) = \int_\Omega \theta \int_{-\infty}^{\infty} f_{X|\theta}(y-x)dF_\varepsilon(x)dG(\theta).$$

Thus, if $\int_\Omega |\theta| \int f_{X|\theta}(y-x)dF_\varepsilon(x)dG(\theta) < \infty$ then by Fubini's theorem

$$\begin{aligned}\int_\Omega \theta f_{Y|\theta}(y)dG(\theta) &= \int_{-\infty}^{\infty}\int_\Omega \theta f_{X|\theta}(y-x)dG(\theta)dF_\varepsilon(x)\\ &= \int_{-\infty}^{\infty} f_X^{(1)}(y-x)dF_\varepsilon(x) - \int_{-\infty}^{\infty}\frac{u^{(1)}(y-x)}{u(y-x)}f_X(y-x)dF_\varepsilon(x),\end{aligned}$$

where $f_X(x) = \int_\Omega f_{X|\theta}(x)dG(\theta)$ and $f_X^{(1)}$ denotes the first derivative of f_X. Then α_G can be written as [see (1.3)]

$$\begin{aligned}\alpha_G(y) &= \int_{-\infty}^{\infty} f_X^{(1)}(y-x)dF_\varepsilon(x) - \int_{-\infty}^{\infty}\frac{u^{(1)}(y-x)}{u(y-x)}f_X(y-x)dF_\varepsilon(x)\\ &\quad -\theta_0\int_{-\infty}^{\infty} f_X(y-x)dF_\varepsilon(x)\\ &= \int_{-\infty}^{\infty} f_X^{(1)}(y-x)dF_\varepsilon(x) - \int_{-\infty}^{\infty}\nu(y-x)f_X(y-x)dF_\varepsilon(x), \qquad (2.1)\end{aligned}$$

where $\nu(t) = u^{(1)}(t)/u(t) + \theta_0$. It is natural to estimate $\alpha_G(y)$ via estimating f_X and $f_X^{(1)}$ in (2.1). Let $\phi_\varepsilon(t)$ denote the characteristic function of the error variable ε. Let $\hat{\phi}_n$ denote the empirical characteristic function defined by $\hat{\phi}_n(t) = n^{-1}\sum_{j=1}^n \exp(itY_j)$. For a nice symmetric (about 0) kernel K, let ϕ_K be its Fourier transform with $\phi_K(0) = 1$. Then a kernel estimator of $f_X^{(\ell)}$, the ℓ^{th} derivative of f_X ($\ell = 0, 1$), is given by

$$f_n^{(\ell)}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)(-it)^\ell \phi_K(th_n) \frac{\hat{\phi}_n(t)}{\phi_\varepsilon(t)} dt, \tag{2.2}$$

where h_n is the bandwidth ($h_n \to 0$ as $n \to \infty$). [Here we assumed that the function $|t^\ell \phi_K(th_n)/\phi_\varepsilon(t)|$ is integrable on $(-\infty, \infty)$.] This type of estimators are proposed by Stefanski and Carroll (1990) and exclusively studied by Carroll and Hall (1988) and Fan (1991a, 1991b, 1992). Define

$$K_{n\ell}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)(-it)^\ell \frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)} dt. \tag{2.3}$$

Then (2.2) can be written as a usual kernel type estimator:

$$f_n^{(\ell)}(x) = \frac{1}{nh_n^{\ell+1}} \sum_{j=1}^{n} K_{n\ell}\left(\frac{x - Y_j}{h_n}\right) \tag{2.4}$$

for $\ell = 0, 1$. The fact that the function ϕ_K is real-valued and even implies that $K_{n\ell}$ defined by (2.3) is a real-valued function, and so is $f_n^{(\ell)}$ defined by (2.4); see Stefanski and Carroll (1990) for more details. In view of (2.1), an estimator of α_G is defined by

$$\alpha_n(y) = \int_{-\infty}^{\infty} f_n^{(\ell)}(y - x) dF_\varepsilon(x) - \int_{-\infty}^{\infty} v(y - x) f_n(y - x) dF_\varepsilon(x), \tag{2.5}$$

where $f_n^{(\ell)}$ ($\ell = 0, 1$) is given by (2.4). Our proposed EB testing rule is now defined by

$$\delta_n(y) = \begin{cases} 1 & \text{if } \alpha_n(y) \le 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.6}$$

By (1.2), the uncondiitonal Bayes risk of (2.6) is given by

$$R(\delta_n, G) = E \int_{-\infty}^{\infty} \alpha_G(y) \delta_n(y) dy + \int_\Omega L(\theta, 0) dG(\theta). \tag{2.7}$$

3. Asymptotic Properties of δ_n

In this section, we investigate asymptotic properties of $R(\delta_n, G) - R(G)$, the regret of δ_n, where $R(G)$ is given by (1.6). First we state some assumptions on the kernel K [see (2.2)] and the error variable ε :

(A1) The kernel K is a symmetric function about 0 on $(-\infty, \infty)$ and is of order r, where r is some positive integer. That is, K satisfies $\int_{-\infty}^{\infty} K(y)dy = 1$, $\int_{-\infty}^{\infty} y^j K(y)dy = 0$ for $j = 1, \ldots, (r-1)$ and $\int_{-\infty}^{\infty} y^r K(y)dy \neq 0$.

(A2) The characteristic function of ε satisfies $|\phi_\varepsilon(t)| > 0$ for all t.

(A3) $\int_{-\infty}^{\infty} |t^\ell \phi_K(t)/\phi_\varepsilon(t)|dt < \infty$ for $\ell = 0, 1$.

We also require that the distributions G and F_ε (the distribution function of ε) satisfy

(A4) $\int_\Omega \int_{-\infty}^{\infty} |\theta| f_{X|\theta}(y - x)dF_\varepsilon(x)dG(\theta) < \infty$ uniformly in y, and

(A5) for some $0 < \delta < 1$ and positive constants B_1 and B_2 (independent of G),

$$\int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta}dy \leq B_1 \text{ and } \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta}\Big(\int_{-\infty}^{\infty} |\nu(y-x)|dF_\varepsilon(x)\Big)^{\delta} dy \leq B_2,$$

where α_G is given by (2.1).

Now define a class of prior distributions by

$$\begin{aligned} \mathcal{F}_{B_0,r} &= \{G : G \text{ is a prior on } \Omega \text{ s.t. } \sup_x |f_X^{(r)}(x)| \leq B_0, \\ &\quad \text{and (A.4) and (A.5) are satisfied}\} \qquad (3.1) \end{aligned}$$

for some constant $B_0 > 0$, where r is as used in (A1) and $f_X^{(r)}$ is the r^{th} derivative of f_X.

Theorem 3.1. For some integer $r \geq 1$ and constants $0 < \delta < 1$, $B_i > 0$ $(i = 0, 1, 2)$, let $\mathcal{F}_{B_0,r}$ be defined by (3.1). Further suppose that K and F_ε are such that (A1) to (A5) hold, and the following conditions are satisfied:

(B1) $\phi_K(t) = 0$ for $|t| \geq 1$.

(B2) $|\phi_\varepsilon(t)|\ |t|^{-\beta_0}\ \exp(|t|^\beta/\gamma) \geq \gamma_0$ as $|t| \to \infty$ for some positive constants β, γ and γ_0 are a constant β_0.

Then, by choosing the bandwidth $h_n = O((\log\, n)^{-1/\beta})$, we obtain

$$\sup_{G \in \mathcal{F}_{B_0,r}} (R(\delta_n, G) - R(G)) = O((\log\, n)^{-\delta(r-1)/\beta}). \tag{3.2}$$

Theorem 3.2. For some integer $r \geq 1$ and constants $0 < \delta < 1$, $B_i > 0$ $(i = 0, 1, 2)$, let $\mathcal{F}_{B_0,r}$ be defined by (3.1). Further suppose that K and F_ε are such that (A1) to (A5) hold, and the following conditions are satisfied:

(C1) $\int_{-\infty}^{\infty} |\phi_K(t)|\ |t|^{\beta+1} dt < \infty$ and $\int_{-\infty}^{\infty} |\phi_K(t) t^{\beta+1}|^2 dt < \infty$.

(C2) $|\phi_\varepsilon(t) t^\beta| \geq \gamma_0$ as $|t| \to \infty$ for some constants $\gamma_0 > 0$ and $\beta \geq 0$.

Then, by choosing the bandwidth $h_n = O(n^{-\frac{1}{2(r+\beta)+1}})$, we obtain

$$\sup G \in \mathcal{F}_{B_0,r}(R(\delta_n, G) - R(G)) = O(n^{-\frac{\delta(r-1)}{2(r+\beta)+1}}). \tag{3.3}$$

Remark 3.1. The examples of error distributions satisfying (B2) are normal, mixture normal and Cauchy, and the examples of distributions satisfying (C2) include gamma, double exponential and symmetric gamma distributions. These two types of distributions are called 'supersmooth' and 'ordinary smooth' distributions, respectively (Fan (1991a, 1991b, 1992)).

Remark 3.2. Theorem 3.1 indicates that the rate of convergence of δ_n is extremely slow for very common error distributions such as normal. Fan (1991a, 1991b) showed that when estimating $f_X^{(1)}$ based on the observations $Y_1, \ldots, Y_n$ with supersmooth errors, the optimal rate of convergence is of the order $O((\log n)^{-(r-1)/\beta})$. For the pure data case, Karunamuni (1996) showed that the optimal rate of convergence of monotone EB testing rules for the present problem is the same as that of estimators of $f_X^{(1)}$. In view of these facts and Johns and Van Ryzin (1972), it is clear that the rates of convergence in Theorems 3.1 and 3.2 are in the best possible forms. Furthermore, the errors which are distributed according to a supersmooth distribution will make a considerable impact on the performance of the proposed EB rule and need to be concerned in application.

Remark 3.3. In practice, the conditions of Theorems 3.1 and 3.2 are easy to verify. Also, the convergence rate of Theorem 3.2 is compatible with the rates that have been obtained in the literature for the pure data case of non-monotone EB rules, see Johns and Van Ryzin (1972). This means that the errors with an ordinary smooth distribution cause only a little damage to the performance of the proposed EB testing rule.

Remark 3.4. Kernel functions satisfying (A1) are easily available in the kernel density literature; see, e.g. Singh (1977). Assumption (A2) ensures that the true densities f_X are identifiable from the model $Y = X + \varepsilon$. Kernels satisyfing (A1), (B1) and (C1) can be constructed by the method used in Fan (1992).

In some practical situations not all the observations are contaminated, but partially contaminated. Assume that only $100p\%$ $(0 < p < 1)$ of the data are measured with error and the remaining data are error free. Then, we have the model $Y = X + \varepsilon$ with $P(\varepsilon = 0) = 1 - p$ and $P(\varepsilon = \varepsilon^*) = p$, where ε^* is an error variable with distribution F_{ε^*} and the characteristic function ϕ_{ε^*}. Then the characteristic function of ε is given by $\phi_\varepsilon(t) = (1-p) + p\phi_{\varepsilon^*}(t)$. If $\mathrm{Re}(\phi_{\varepsilon^*}) \geq 0$, then $|\phi_\varepsilon(t)| \geq 1 - p$, where 'Re' means the real-part. Thus, ε satisfies (C2) with $\beta = 0$ no matter what the distribution of ε^* is, i.e., supersmooth or ordinary smooth. Corollary 3.1 below is a direct consequence of Theorem 3.2, and, therefore, its proof will be omitted.

Corollary 3.1. For some integer $r \geq 1$ and constants $0 < \delta < 1$, $B_i > 0$ $(i = 0, 1, 2)$, let $\mathcal{F}_{B_0,r}$ be defined by (3.1). Further suppose that K and F_ε are such that (A1) to (A5) hold, and the following conditions are satisfied:

(D1) $\int_{-\infty}^{\infty} |\phi_K(t)t|dt < \infty$ and $\int_{-\infty}^{\infty} |\phi_K(t)t|^2 dt < \infty$.

(D2) $\mathrm{Re}(\phi_{\varepsilon^*}(t)) \geq 0$ for all t.

Then, by choosing the bandwidth $h_n = O(n^{-\frac{1}{2r+1}})$, we obtain

$$\sup_{G \in \mathcal{F}_{B_0,r}} (R(\delta_n, G) - R(G)) = O\Big(n^{-\frac{\delta(r-1)}{2r+1}}\Big). \qquad (3.4)$$

Remark 3.5. The rate in (3.4) is exactly the same rate of Johns and Van Ryzin

(1972) for the pure data case. Their rate is the best available rate for non-monotone EB testing rules for the pure data. Note that the proposed EB rule of the present paper is also non-monotone. Therefore, the situation is as good as the pure data case when the errors are partially contaminated even if the errors are supersmooth. Van Houwelingen (1976) and Karunamuni (1996) have shown that montone EB testing rules exhibit faster rate of convergence compared to non-monotone ones. However, investigation of such rules for the present setup is beyond the scope of the present paper.

We now exhibit another model under which the rate of convergence can also be as good as that of the uncontaminated data case, while all the data are contaminated with supersmooth errors.

Assume that all the data are contaminated but the error level can be controlled. That is, $Y = X + \varepsilon$ with $\varepsilon = \sigma_0 \tilde{\varepsilon}$, where σ_0 parametrizes the error level. This model has been proposed by Fan (1992). Then, $\phi_\varepsilon(t) = \phi_{\tilde{\varepsilon}}(\sigma_0 t)$.

Theorem 3.3. For integer $r \geq 1$ and constants $0 < \delta < 1$, B_i $(i = 0, 1, 2)$, let $\mathcal{F}_{B_0,r}$ be defined by (3.1). Also, let $\sigma_0 = O(n^{-\frac{1}{(2r+1)}})$. Further suppose that K and F_ε are such that (A1) to (A5) hold with ε replaced by $\tilde{\varepsilon}$, and that (D1) holds. Then, by choosing the bandwidth $h_n = O(n^{-\frac{1}{2r+1}})$, we obtain

$$\sup_{G \in \mathcal{F}_{B_0,r}} (R(\delta_n, G) - R(G)) = O(n^{-\frac{\delta(r-1)}{2r+1}}). \tag{3.5}$$

Example 1. Consider the exponential family in (1.1) with $u(x) = e^{-x^2/2} I_{(-\infty,\infty)}(x)$ and $C(\theta) = (2\pi)^{-1/2} e^{-\theta^2/2} I_{(-\infty,\infty)}(\theta)$. Then, $f_{X|\theta}(x) = (2\pi)^{-1/2} e^{-(x-\theta)^2/2}$, $-\infty < x < \infty$, and the natural parameter space $\Omega = (-\infty, \infty)$. The marginal density of

X is

$$f_X(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-(x-\theta)^2/2} dG(\theta). \tag{3.6}$$

Case (i): Suppose that the error distribution F_ϵ is $N(0,1)$, the standard normal distribution. Then ϕ_ϵ satisfies (B2) of Theorem 3.1 with $\beta = 2$. By repeated differentiation under the integral sign (by Theorem 2.9 of Lehmann (1986)), the k^{th} derivative of f_X is given by

$$f_X^{(k)}(x) = (-1)^k \int_{-\infty}^{\infty} H_k(x-\theta) f_{X|\theta}(x) dG(\theta), \tag{3.7}$$

for any G, where H_k is the k^{th} Hermite polynomial. Thus, $\sup_x |f_X^{(k)}(x)| \leq (2\pi)^{-1/2} \sum_{j=0}^{k} |a_j|$ for any G, where a_j is the j^{th} coefficient of the k^{th} Hermite polynomial. Therefore, $G \in \mathcal{F}_{B_0,r}$ if $B_0 = (2\pi)^{-1/2} \sum_{j=0}^{r} |a_j|$, provided (A4) and (A5) are satisfied. Note that

$$\int_\Omega \int_{-\infty}^{\infty} |\theta| f_{X|\theta}(y-x) dF_\epsilon(x) dG(\theta) \leq \int_{-\infty}^{\infty} |\theta| dG(\theta) < \infty,$$

if $E_G|\theta| < \infty$. Thus, (A4) is satisfied if $E_G|\theta| < \infty$. Note that $\nu(y-x) = -(y-x)+\theta_0$. Therefore, $|\nu(y-x)| \leq |y-x|+|\theta_0|$ and $\int |y-x| dF_\epsilon(x) \leq |y| + \int |x| dF_\epsilon(x)$. Now by the C_r-inequality,

$$\begin{aligned}
&\int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} \Big(\int_{-\infty}^{\infty} |\nu(y-x)| dF_\epsilon(x) \Big)^\delta dy \\
&\quad \leq \int_{-\infty}^{\infty} |\alpha(y)|^{1-\delta} \Big(\int_{-\infty}^{\infty} |y-x| dF_\epsilon(x) + |\theta_0| \Big)^\delta dy \\
&\quad \leq C_1 \int_{-\infty}^{\infty} |\alpha(y)|^{1-\delta} \Big(\int_{-\infty}^{\infty} |y-x| dF_\epsilon(x) \Big)^\delta dy + C_2 |\theta_0|^\delta \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} dy \\
&\quad \leq C_1 \int_{-\infty}^{\infty} |y|^\delta |\alpha_G(y)|^{1-\delta} dy + C_2 \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} dy, \qquad (3.8)
\end{aligned}$$

where C_1 and C_2 are some positive constants (independent of G). Now, for any $\xi > 0$, by Hölder inequality,

$$\int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} dy$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} (1+|y|)^{-(1+\xi)\delta}(1+|y|)^{(1+\xi)\delta}|\alpha_G(y)|^{1-\delta}dy \\
&\le \Big(\int_{-\infty}^{\infty}(1+|y|)^{-(1+\xi)}dy\Big)^{\delta}\Big(\int_{-\infty}^{\infty}(1+|y|)^{\delta(1+\xi)/(1-\delta)}|\alpha_G(y)|dy\Big)^{1-\delta} \\
&\le C_3\Big\{\int_{-\infty}^{\infty}(1+|y|)^{\delta(1+\xi)/(1-\delta)}|\alpha_G(y)|dy\Big\}^{1-\delta} \\
&\le C_4\Big\{\int_{-\infty}^{\infty}|\alpha_G(y)|dy + \int_{-\infty}^{\infty}|y|^{\delta(1+\xi)/(1-\delta)}|\alpha_G(y)|dy\Big\}^{1-\delta}, \qquad (3.9)
\end{aligned}
$$

where C_3 and C_4 are some positive constants (independent of G). Similary, we can show that

$$
\begin{aligned}
\int_{-\infty}^{\infty}|y|^{\delta}|\alpha_G(y)|^{1-\delta}dy \;\le\; C_5\Big\{&\int_{-\infty}^{\infty}|y|^{\delta/(1-\delta)}|\alpha_G(y)|dy \\
&+\int_{-\infty}^{\infty}|y|^{\delta(2+\xi)/(1-\delta)}|\alpha_G(y)|dy\Big\}^{1-\delta} \qquad (3.10)
\end{aligned}
$$

for some constant $C_5 > 0$. In view of (3.8), (3.9) and (3.10), we see that (A5) is satisfied if

$$
\int_{-\infty}^{\infty}|y|^{\delta(2+\xi)/(1-\delta)}|\alpha_G(y)|dy \le C_6 \qquad (3.11)
$$

for some constant $C_6 > 0$ (independent of G). Since $F_\varepsilon \equiv N(0,1)$ and

$$
\begin{aligned}
|\alpha_G(y)| &= \Big|\int_{-\infty}^{\infty} f_X^{(1)}(y-x)dF_\varepsilon(x) - \int_{-\infty}^{\infty}\nu(y-x)f_X(y-x)dF_\varepsilon(x)\Big| \\
&\le \int_{-\infty}^{\infty}\int_{\Omega}(2\pi)^{-1/2}(|y-x-\theta|+|\theta_0|+|y-x|)e^{-(y-x-\theta)^2/2}dG(\theta)dF_\varepsilon(x) \\
&\le C_7\int_{-\infty}^{\infty}\int_{\Omega}\{e^{-(y-x-\theta)^2/4}+e^{-(y-x-\theta)^2/2}+|\theta|e^{-(y-x-\theta)^2/2}\}e^{-x^2/2}dG(\theta)dx \\
&\le C_8\Big\{\int_{\Omega}e^{-(y-\theta)^2/6}\int_{-\infty}^{\infty}e^{-(\sqrt{3}x-(y-\theta)/\sqrt{3})^2/4}dx\, dG(\theta) \\
&+\int_{\Omega}e^{-(y-\theta)^2/4}\int_{-\infty}^{\infty}e^{-(\sqrt{2}x-(y-\theta)/\sqrt{2})^2/2}dx\, dG(\theta) \\
&+\int_{\Omega}|\theta|e^{-(y-\theta)^2/4}\int_{-\infty}^{\infty}e^{-(\sqrt{2}x-(y-\theta)^2/\sqrt{2})^2/2}dx\, dG(\theta) \\
&\le C_9\Big\{\int_{\Omega}e^{-(y-\theta)^2/6}dG(\theta)+\int_{\Omega}|\theta|e^{-(y-\theta)^2/4}dG(\theta) \\
&+\int_{\Omega}e^{-(y-\theta)^2/4}dG(\theta)\Big\} \qquad (3.12)
\end{aligned}
$$

for some constants $C_i > 0$ ($i = 7, 8, 9$) (independent of G). Thus, from (3.11) and (3.12),

$$\begin{aligned}\int_{-\infty}^{\infty} & |y|^{\delta(2+\xi)/(1-\delta)}|\alpha_G(y)|dy \\ \leq & \; C_{10}\Big\{\int_\Omega \int_{-\infty}^{\infty} |y-\theta|^{\delta(2+\xi)/(1-\delta)} e^{-(y-\theta)^2/6} dy\, dG(\theta) \\ & + \int_\Omega |\theta|^{\delta(2+\xi)/(1-\delta)} \int_{-\infty}^{\infty} e^{-(y-\theta)^2/6} dy\, dG(\theta) \\ & + \int_\Omega |\theta| \int_{-\infty}^{\infty} |y-\theta|^{\delta(2+\xi)/(1-\delta)} e^{-(y-\theta)^2/4} dy\, dG(\theta) \\ & + \int_\Omega |\theta|^{\delta(2+\xi)/(1-\delta)+1} \int_{-\infty}^{\infty} e^{-(y-\theta)^2/4} dy\, dG(\theta) \\ & + \int_\Omega \int_{-\infty}^{\infty} |y-\theta|^{\delta(1+\xi)/(1-\delta)} e^{-(y-\theta)^2/4} dy\, dG(\theta) \\ & + \int_\Omega |\theta|^{\delta(2+\xi)/(1-\delta)} \int_{-\infty}^{\infty} e^{-(y-\theta)^2/4} dy\, dG(\theta)\Big\} \qquad (3.13)\end{aligned}$$

for some constant $C_{10} > 0$ (independent of G). Hence, from (3.13) it is clear that (3.11) holds if $E_G|\theta|^{\delta(2+\xi)/(1-\delta)+1} \leq C_{11}$ for some constant $C_{11} > 0$ (independent of G). In other words, the assumption (A5) reduces to a single moment condition of the G's and $G \in \mathcal{F}_{B_0,r}$ if $E_G|\theta|^{\delta(2+\xi)/(1-\delta)+1} \leq C_{11}$.

Case (ii): Suppose the error distribution F_ε is gamma $(1,p)$ with density $f_\varepsilon(x) = (\Gamma(p))^{-1}x^{p-1}e^{-x}$, $x > 0$, $p > 1$. Then ϕ_ε satisifes (C2) of Theorem 3.2 with $\beta = p$. Similar to case (i) above, we can again show that (A5) reduces a single moment condition $E_G|\theta|^{\delta(2+\xi)/(1-\delta)+1} \leq C_{12}$ for some constant $C_{12} > 0$ (independent of G), $0 < \delta < 1$ and $\xi > 0$.

Example 2. Consider the exponential family in (1.1) with $u(x) = (2e^{-x^2} + e^{-2x^2})I_{(-\infty,\infty)}(x)$ and $C(\theta) = \pi^{-1/2}(\sqrt{2}/2+2e^{\theta^2/8})^{-1}e^{-\theta^2/8}I_{(-\infty,\infty)}(\theta)$. Then, $f_{X|\theta}(x)$ is a bimodel density and

$$f_X(x) = \pi^{-1/2}e^{-x^2}(2+e^{-x^2})\int_{-\infty}^{\infty} e^{\theta x}(\sqrt{2}/2+2e^{\theta^2/8})^{-1}e^{-\theta^2/8}dG(\theta).$$

Since $u^{(1)}(y-x)/u(y-x) = -4(y-x)(e^{-(y-x)^2}+e^{-2(y-x)^2})/\{2e^{-(y-x)^2}+e^{-2(y-x)^2}\}$, it is obvious that $|v(y-x)| \leq |u^{(1)}(y-x)/u(y-x)+\theta_0| \leq 4|y-x|+|\theta_0|$. Following similar steps as in Example 1, we can again show that (A4) is satisfied if $E_G|\theta| < \infty$ and (A5) is satisfied if $E_G|\theta|^{\delta(2+\xi)/(1-\delta)+1} \leq C_{13}$ for some constant $C_{13} > 0,\ 0 < \delta < 1$ and $\xi > 0$.

4. Proofs

In this section, we shall prove Theorems 3.1, 3.2 and 3.3. Only Theorem 3.1 is proved in details, since the proofs of Theorems 3.2 and 3.3 are similar. First we shall state a few lemmas. Lemmas 4.1 and 4.2 are due to Fan (1991a, 1991b). The proof of Lemma 4.4 can be found in Johns and Van Ryzin (1972).

Lemma 4.1. Under the assumptions (A1), (A2), (B1) and (B2), and with the choice $h_n = O((\log n)^{-1/\beta})$ of the bandwidth, we have

$$\sup_x \sup_{G\in\mathcal{G}_{B,r}} E|f_n^{(\ell)}(x) - f_X^{(\ell)}(x)| \leq \text{Const. } \times (\log n)^{-(r-\ell)/\beta} \tag{4.1}$$

for $\ell = 0, 1$, where $f_n^{(\ell)}$ is given by (2.2), and

$$\mathcal{G}_{B,r} = \{G : G \text{ is a prior on } \Omega \text{ such that } \sup_x |f_X^{(r)}(x)| \leq B\} \tag{4.2}$$

for some constant $B > 0$.

Lemma 4.2. Under the assumptions of (A1), (A2), (C1) and (C2) and with the choice $h_n = O(n^{-1/(2(r+\beta)+1)})$ of the bandwidth, we have

$$\sup_x \sup_{G\in\mathcal{G}_{B,r}} E|f_n^{(\ell)}(x) - f_X^{(\ell)}(x)| \leq \text{Const. } \times n^{-(r-\ell)/(2(r+\beta)+1)} \tag{4.3}$$

for $\ell = 0, 1$, where $f_n^{(\ell)}$ and $\mathcal{G}_{B,r}$ are given by (2.2) and (4.2), respectively.

Lemma 4.3. Consider the model $Y = X + \varepsilon$, where $\varepsilon = \sigma_0\tilde{\varepsilon}$ with $\sigma_0 = O(n^{-1/(2r+1)})$. Then under the assumptions (A1) and (A2) with ε replaced by $\tilde{\varepsilon}$ and (D1), with the choice of the bandwidth $h_n = O(n^{-1/(2r+1)})$, we have

$$\sup_x \sup_{G\in\mathcal{G}_{B,r}} E|f_n^{(\ell)}(x) - f_X^{(\ell)}(x)| \leq \text{Const.} \times n^{-(r-\ell)/(2r+1)}, \tag{4.4}$$

for $\ell = 0, 1$, where $f_n^{(\ell)}$ and $\mathcal{G}_{B,r}$ are given by (2.2) and (4.2), respectively.

Lemma 4.4. Let $R(\delta_n, G)$ and $R(G)$ be defined by (2.7) and (1.6), respectively. Then

$$0 \leq R(\delta_n, G) - R(G) \leq \int_{-\infty}^{\infty} |\alpha_G(y)| Pr\{|\alpha_n(y) - \alpha_G(y)| \geq |\alpha_G(y)|\} dy, \tag{4.5}$$

where α_G and α_n are given by (2.1) and (2.5), respectively.

Proof of Theorem 3.1. By Lemma 4.4 and by the Markov inequality, for $0 < \delta < 1$,

$$\begin{aligned} R(\delta_n, G) - R(G) &\leq \int_{-\infty}^{\infty} |\alpha_G(y)| Pr\{|\alpha_n(y) - \alpha_G(y)| \geq |\alpha_G(y)|\} dy \\ &\leq \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} E|\alpha_n(y) - \alpha_G(y)|^{\delta} dy. \end{aligned} \tag{4.6}$$

By applying the C_r-inequality followed by Lyapunov's inequality and using Fubini's theorem, we obtain

$$\begin{aligned} E|\alpha_n(y) - \alpha_G(y)|^{\delta} &\leq C_{14}\Big\{E\Big|\int_{-\infty}^{\infty} (f_n^{(1)}(y-x) - f_X^{(1)}(y-x)) dF_\varepsilon(x)\Big|^{\delta} \\ &\quad + E\Big|\int_{-\infty}^{\infty} \nu(y-x)(f_n(y-x) - f_X(y-x)) dF_\varepsilon(x)\Big|^{\delta}\Big\} \\ &\leq C_{14}\Big(\int_{-\infty}^{\infty} E|f_n^{(1)}(y-x) - f_X^{(1)}(y-x)| dF_\varepsilon(x)\Big)^{\delta} \\ &\quad + C_{14}\Big(\int_{-\infty}^{\infty} |\nu(y-x)| E|f_n(y-x) - f_X(y-x)| dF_\varepsilon(x)\Big)^{\delta} \end{aligned} \tag{4.7}$$

for some constant $C_{14} > 0$ (independent of G). From Lemma 4.1, (4.6) and (4.7), for some constant $C_{15} > 0$ (independent of G), we have

$$\begin{aligned}
&\sup_{G\in\mathcal{F}_{B,r}} (R(\delta_n, G) - R(G)) \le \sup_{G\in\mathcal{F}_{B,r}} \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} E|\alpha_n(y) - \alpha_G(y)|^{\delta} dy \\
&\le \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} \sup_{G\in\mathcal{F}_{B,r}} E|\alpha_n(y) - \alpha_G(y)|^{\delta} dy \\
&\le C_{14} \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} \Big(\int_{-\infty}^{\infty} \sup_{G\in\mathcal{F}_{B,r}} E|f_n^{(1)}(y-x) - f_X^{(1)}(y-x)| dF_\varepsilon(x) \Big)^{\delta} dy \\
&+C_{14} \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} \Big(\int_{-\infty}^{\infty} |\nu(y-x)| \sup_{G\in\mathcal{F}_{B,\nabla}} E|f_n(y-x) - f_X(y-x)| dF_\varepsilon(x) \Big)^{\delta} dy \\
&\le C_{15} (\log n)^{-\delta(r-1)/\beta} \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} dy \\
&+C_{15} (\log n)^{-\delta r/\beta} \int_{-\infty}^{\infty} |\alpha_G(y)|^{1-\delta} \Big(\int_{-\infty}^{\infty} |\nu(y-x)| dF_\varepsilon(x) \Big)^{\delta} dy \\
&= O((\log n)^{-\delta(r-1)/\beta})
\end{aligned}$$

by the assumptions of Theorems 3.1. This completes the proof.

Proof of Theorems 3.2 and 3.3. The proofs are similar to that of Theorem 3.1 above, except, that we use Lemma 4.2 and Lemma 4.3 in the place of Lemma 4.1 of the proofs of Theorem 3.2 and Theorem 3.3, respectively.

REFERENCES

Carroll, R.J. and Hall, P. (1986). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184-1186.

Fan, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.

Fan, J. (1991b). Global behavior of deconvolution kernel estimators. *Statistica Sinica* **1**, 541-551.

Fan, J. (1992). Deconvolution with supersmooth distributions. *Canad. J. Statist.* **20**, 155-169.

Johns, M.V., Jr. and Van Ryzin, J.R. (1972). Convergence rates for empirical Bayes two-action problems. II. Continuous Case. *Ann. Math. Statist.* **43**, 934-947.

Karunamuni, R.J. and Yang, H. (1995). On convergence rates of monotone empirical Bayes tests for the continuous one-parameter exponential family. *Statist. Decisions* **13**, 181-192.

Karunamuni, R.J. (1996). Optimal rates of convergence of empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **24**, 212-231.

Lehmann, E. (1986). *Testing Statistical Hypotheses,* 2^{nd} ed. Wiley, New York.

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob. I*, 157-163, University of California Press, Berkeley.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1-20.

Samuel, E. (1963). An empirical Bayes approach to the testing of certain parameteric hypotheses. *Ann. Math. Statist.* **34**, 1370-1385.

Singh, R.S. (1977). Applications of estimators to a density and its derivatives to certain statistical problems. *J. Roy. Statist. Soc. Ser. B*, **39**, 357-363.

Stefanski, L.A. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, **21**, 169-184.

Stijnen, Th. (1985). On the asymptotic behavior of empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **13**, 403-412.

Van Houwelingen, J.C. (1976). Monotone empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **4**, 981-989.

Yu, B.O. (1970). Rates of convergence in empirical Bayes two-action and estimation problems and in extended sequence-compound estimation problems. *Ph.D. Dissertation*, Dept. Statistics and Probability, Michigan State University.

Zhang, S. and Karunamuni, R.J. (1997a). Bayes and empirical Bayes estimation with errors in variables. *Statist. Prob. Letters*, **33**, 23-34.

Zhang, S. and Karunamuni, R.J. (1997b). Empirical Bayes estimation for the continuous one-parameter exponential family with errors in variables. *Statistics & Decisions*, to appear.

Chapter 4

On Kernel Density Estimation Near Endpoints with Application to Line Transect Sampling

1. Introduction

An important problem in population biology is estimation of the density of objects, such as animals or plants, in a study area. A convenient method of doing so is the *line transect sampling* method. An observer moves along the transect line and records from each object he/she detects either the perpendicular distance x from the line or the radial distance r and the angle θ ($x = r\sin\theta$). In practice, several transects would usually be selected, but sampling techniques will not be discussed here; the reader is referred to Buckland et al. (1993) for experimental guidelines. The present work deals with perpendicular distance models for ungrouped data. The basic assumptions that the models of this paper rely on are (Buckland (1985)):

(i) Objects are randomly distributed.

(ii) Objects on the transect line are seen with probability 1.

(iii) Any movement of objects before detection is slower relative to the speed of the observer and is independent of the observer.

(iv) Perpendicular distances are recorded without error.

(v) Sightings are independent events.

A version of this paper has been submitted for publication.

(vi) No object is recorded more than once.

The density D of the objects may be estimated from the equation $\hat{D} = n\hat{f}(0)/2L$, where $\hat{f}(0)$ is an estimate of $f(0)$, L is the length of the transect line and n is the number of objects sighted (see, e.g., Buckland et al. (1993)). A number of parametric models have been proposed for f, and there is extensive literature for use of the maximum likelihood technique as the criterion for estimation of $f(0)$; see, e.g., Gates, Marshal and Olson (1968), Burnham and Anderson (1976), Pollock (1978), Ramsey (1979), Quinn and Gallucci (1980), Burnham, Anderson and Laake (1980) and Buckland (1985). For additional references the reader is referred to the recent monograph by Buckland et al. (1993). Bayesian estimation of $f(0)$ is investigated in Karunamuni and Quinn (1995). It is a standard practice that if there are theoretical reasons for supposing that the detection density has a given parametric form, then parametric modelling may be carried out. Otherwise, robust or nonparametric methods such as Fourier series, splines, kernel methods or polynomials might be preferred. A nonparametric Fourier series estimator of $f(0)$ is given by Burnham et al. (1980) and a Hermite polynomial estimator is due to Buckland (1985). Kernel estimates of $f(0)$ and D are investigated by Buckland (1992) and Chen (1996).

Nonparametric estimation of $f(0)$ is particularly difficult due to *boundary effects* that occur in nonparametric curve estimation problems. Such effects are well known to occur in nonparametric density estimation when the support of the density has a finite endpoint (0 is an endpoint in the present situation). They are a major problem both for application and asymptotic theory. Theoretically, the rate of convergence at boundary points is slower than that at the interior points. See Gasser and Müller (1979), Hall (1981), Rice (1984), Schuster (1985), Gasser et al. (1985), Cline and Hart (1991) and Marron and Ruppert (1994), among others, for further discussions on this topic. In the literature, on the problem of estimating density at the boundary points by the kernel method, a number of solutions (adjustments and modifications) have been proposed:

(i) The reflection method (Schuster (1985), Cline and Hart (1991) and Silverman (1986)). [This method is specially designed for the case $f^{(1)}(0) = 0$, which is the shape criterion of Burnham et al. (1980, p. 47). Kernel estimates of $f(0)$ of Buckland (1992) and Chen (1996) are established under the preceding assumption.]

(ii) The use of "boundary kernels" (Gasser and Müller (1979), Gasser et al. (1985) and Müller (1991)).

(iii) The transformation technique (Marron and Ruppert (1994)).

The purpose of this paper is to study the local polynomial fitting method. Fan (1992) and Fan and Gijbels (1992) introduced the local polynomial smoothing technique in the context of nonparametric regression function estimation. They showed that their method of smoothing has very high asymptotic efficiency among all possible (nonparametric) smoothers, including those produced by kernel, orthogonal series and spline method in estimation of a regression function. Further, they have observed that their method removes boundary effects as well as other disadvantages mainly in estimation problem of regression function.

An attempt to estimate $f(0)$ without boundary effects and a careful investigation of the local polynomial smoothing technique in the context of density estimation at boundary points form the basis of this paper. In the next section we derive a density estimator via the local polynomial fitting method. To implement the techniques developed for the regression problem by Fan (1992) and Fan and Gijbels (1992), we generate pseudo variables $Y_1, ..., Y_n$ (response variables) based on the observed sample $X_1, ..., X_n$ (explanatory variables). (Recall that in the regression problem one observes two variables, the response variable Y and the explanatory variable X.) We observe that our local polynomial fitting estimator also possesses somewhat similar nice properties exhibited by the correspoding regression estimator of Fan (1992) and Fan and Gijbels (1992). That is, the kernel functions derived via the local polynomial fitting method automatically adapt for boundary effects at the boundary region, see Section 3. Another interesting feature that we notice is that, in the interior region, the preceding mentioned kernels coincide with (up to a normalizing constant) those of the "optimal kernels" derived by Gasser et al. (1985). Furthermore, we argue that the local polynomial fitting method provides an intuitive explanation for the "boundary kernel" proposal of Gasser and Müller (1979) and Gasser et al. (1985). In Section 4 we obtain an optimal kernel in order to estimate the density at the endpoints, as a solution of a variational problem. Section 5 presents numerical comparisons between various boundary kernel methods. Section 6 discusses the estimation problem of $f(x)$ at the boundary region under the shape criterion. Finally, in Section 7 an illustration of the techniques and its capabilities are displayed using two examples from line transect sampling.

2. Local Polynomial Density Estimation

Assume that $f(x)$ is a probability density function with support $[0, a]$, $a \leq \infty$. Our goal is to estimate $f(x)$. To derive an estimator of f via the local polynomial fitting method, we first assume that we have the pseudo data $Y_1 = f(X_1)$, $Y_2 = f(X_2)$, ..., $Y_n = f(X_n)$ corresponding to the observed data X_1, X_2, ..., X_n. It is obvious that

$$f(x) = E[f(X)|X = x].$$

Now we view $Y_1, Y_2, ..., Y_n$ as response variables and $X_1, X_2, ..., X_n$ as explanatory variables. Since (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) are independent and identically distributed (i.i.d.) random vectors, we can use the local polynomial smoother to fit $f(x)$, the regression function in this case. That is, minimize

$$\sum_{i=1}^{n}\left[Y_i - \sum_{j=0}^{p} b_j(x)(x-X_j)^j\right]^2 K\left(\frac{x-X_i}{h}\right) \qquad (2.1)$$

with respect to (w.r.t.) b_j, where K is a non-negative kernel function with support [-1,1] and $h = h_n$ ($h_n \to 0$ as $n \to \infty$) is the smoothing parameter. Let $\hat{b}_j(x)$ ($j = 0, 1, ..., p$) denote the solution of the least square problem (2.1). Then $\hat{f}^{(v)}(x) = (-1)^v v! \hat{b}_v(x)$ is an estimator of $f^{(v)}(x)$, the v^{th} derivative of $f(x)$, $v = 0, 1, ..., p$. Denoting

$$X = \begin{pmatrix} 1 & (x-X_1) & \dots & (x-X_1)^p \\ 1 & (x-X_2) & \dots & (x-X_2)^p \\ \vdots & \vdots & & \vdots \\ 1 & (x-X_n) & \dots & (x-X_n)^p \end{pmatrix}$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \hat{b}(x) = \begin{pmatrix} \hat{b}_0(x) \\ \hat{b}_1(x) \\ \vdots \\ \hat{b}_p(x) \end{pmatrix}$$

and W =diag $\left(K\left(\frac{x-X_i}{h}\right)\right)$, the $n \times n$ diagonal matrix of weights, the solution to (2.1) is given by

$$\begin{aligned} \hat{b}(x) &= (X^T W X)^{-1} X^T W Y \\ &= \begin{pmatrix} S_{n,0}(x) & S_{n,1}(x) & \dots & S_{n,p}(x) \\ S_{n,1}(x) & S_{n,2}(x) & \dots & S_{n,p+1}(x) \\ \vdots & \vdots & & \vdots \\ S_{n,p}(x) & S_{n,p+1}(x) & \dots & S_{n,2p}(x) \end{pmatrix}^{-1} \begin{pmatrix} T_{n,0}(x) \\ T_{n,1}(x) \\ \vdots \\ T_{n,p}(x) \end{pmatrix} \\ &= S_n^{-1} T_n \qquad (2.2) \end{aligned}$$

where

$$S_{n,j}(x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)(x - X_i)^j,\ j = 0, 1, ..., 2p$$

and

$$T_{n,j}(x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)(x - X_i)^j Y_i,\ j = 0, 1, ..., p.$$

Hence

$$\hat{b}_v(x) = e_v^T \hat{b}(x) = \sum_{i=1}^{n} W_v^n\left(\frac{x - X_i}{h}\right) Y_i, \tag{2.3}$$

where $e_v = (0, 0, ..., 0, 1, 0, ..., 0)^T$ with 1 at the $(v+1)^{st}$ position and $W_v^n(t) = e_v^T S_n^{-1}(1, th, ..., (th)^p)^T K(t)$.

The inverse of the matrix S_n in (2.2) exists when the kernel function K is nonnegative and the samples $X_1, ..., X_n$ contain at least p distinct X_i with nonnegative weights. The following lemma shows that (2.3) can be represented as a kernel type estimator which only depends on X_i, $i = 1, 2, \ldots n$.

Lemma 2.1. Assume that the density function $f(x)$ is continuous and positive on $[0, a]$, and satisfies $\sup_x |f^{(i)}(x)| < \infty$, $i = 1, 2$. Then $\hat{b}_v(x) = \tilde{b}_v(x)(1 + o_P(1))$, where

$$\tilde{b}_v(x) = \frac{1}{nh^{v+1}} \sum_{i=1}^{n} K_v^*\left(\frac{x - X_i}{h}\right) \tag{2.4}$$

with

$$K_v^*(t) = e_v^T S^{-1}(1, t, ..., t^p)^T K(t), \tag{2.5}$$

and the matrix $S = (s_{j,l})$, where $s_{j,l} = \int t^{j+l} K(t) dt$, $0 \le j,\ l \le p$.

The proof of Lemma 2.1 is given in the Appendix. Lemma 2.1 shows that one can use $(-1)^v v! \tilde{b}_v(x)$ as an estimator of $f^{(v)}(x)$. A similar estimator has been derived by Fan et al. (1993) in the context of nonparametric regression function estimation. The kernel function $K_v^*(t)$ defined by (2.5) is called an equivalent kernel, and it is of order $(v, p+1)$ up to normalizing constants; see Gasser et al. (1985) and Fan et al. (1993). Using an equivalent kernel, the asymptotic mean squared error of a kernel estimator of the type (2.4) in the interior region depends on

$$\left|\int t^{p+1} K_v^*(t)dt\right|^{2v+1} \left(\int K_v^{*2}(t)dt\right)^{p+1-v}. \tag{2.6}$$

"Optimal" polynomial kernels of orders (0, 2), (0, 4), (1, 3), (1, 5), (0, 6) and (2, 4) have been derived by Gasser et al. (1985) in the sense of minimizing (2.6) among all "minimal" kernels (Gasser et al. (1985)). When the kernel K in (2.5) is chosen as $K(t) = \frac{3}{4}(1-t^2)I_{[-1,1]}$, our equivalent kernels K_v^* defined by (2.5) coincides with optimal kernels of Gasser et al. (1985) up to a normalizing constant. Also, it is easy to show that if we choose K as $\frac{1}{2}I_{[-1,1]}$, i.e., the minimum variance kernel, then the induced kernels (2.5) also satisfy the minimum variance property as in the regression case (Fan et al. (1993), Gasser et al. (1985) and Granovsky and Müller (1989)). In this paper, our main interest is to study the performance of the kernel estimator $\tilde{f}^{(v)}(x) = (-1)^v v! \tilde{b}_v(x)$ at the boundary region including the endpoints , where $\tilde{b}_v(x)$ is given by (2.4).

3. Boundary Kernels

3.1. Equivalent Boundary Kernels

Assume that a constant bandwidth h is used in the interior $I = \{x : h \le x \le a-h\}$. Denote $B_L = \{x : 0 \le x \le h\}$ and $B_R = \{x : a-h \le x \le a\}$ as left and right boundary regions, respectively. Also, define

$$\bar{K}_{v,c}(t) = \begin{cases} K_{v,1}(t), & x \in I \\ K_{v,c}(t), & x \in B_L \\ K^{\dagger}_{v,\frac{a-x}{h}}(t), & x \in B_R, \end{cases} \tag{3.1}$$

where $c = x/h$, $K_{v,c}$ is a boundary kernel with support $(-1, c]$, $0 \leq c \leq 1$, see Definition 3.1 below, and $K^{\dagger}_{v,c}$ has support $[-c, 1)$ with $K_{v,c}(t) = K^{\dagger}_{v,c}(-t)$. Boundary kernels have been investigated by Gasser and Müller (1979), Gasser et al. (1985), Müller (1988, 1991) and Müller and Wang (1994).

Definition 3.1 A boundary kernel $K_{v,c}$ is said to be of order (v, k) if

$$\int_{-1}^{c} K_{v,c}(t)t^j dt = \begin{cases} 0, & j = 0, ..., v-1, v+1, ..., k-1 \\ 1, & j = v \\ B_{v,k}(c), & j = k, \end{cases} \tag{3.2}$$

where $B_{v,k}(c) < \infty$. In the interior region I, $B_{v,k}(c) = B_{v,k}(1)$ and we require that $B_{v,k}(1) \neq 0$. In the boundary regions B_L and B_R, it is not always possible that $B_{v,k}(c) \neq 0$ for all $0 \leq c < 1$. But, it will be seen below that at least we can require $B_{v,k}(0) \neq 0$. We now derive (equivalent) boundary kernels by the polynomial fitting method of the previous section. In this paper, we only consider the left boundary point, and, by symmetry, the right boundary point can also be treated similarily.

Consider a sequence of points $x = ch$ $(0 \leq c \leq 1)$ in B_L. Define the moments of K now by $s_{j,c} = \int_{-1}^{c} u^j K(u)du$. The equivalent kernel $K^*_{v,c}$ defined by (2.5) now leads to an equivalent boundary kernel

$$K^*_{v,c}(t) = e_v^T S_c^{-1}(1, t, ..., t^p)^T K(t), \tag{3.3}$$

where $S_c = (s_{j+l,c})_{0 \leq j,l \leq p}$ and $S_c^{-1} = \left(s^{-1}_{j+l,c}\right)_{0 \leq j,l \leq p}$. Thus, the local polynomial fitting method adapts automatically for boundary effects. Fan et al. (1993) have derived equivalent boundary kernels in the regression context. It is easy to show that $K^*_{v,c}$ defined by (3.3) satisfies

$$\int_{-1}^{c} K^*_{v,c}(t)t^j dt = \begin{cases} 0, & j = 0, ..., v-1, v+1, ..., p \\ 1, & j = v, \end{cases}$$

i.e., $K^*_{v,c}$ is of order $(v, p+1)$. Since the optimality of the equivalent kernels induced from the Epanechnikov kernel in the interior region, it is our interest to investigate the properties of the equivalent boundary kernels at the boundary. The analytic representation of equivalent boundary kernels $K^*_{v,c}$for the left boundary (defined by (3.3)) of orders (0, 2), (0, 4), (1, 3), (1, 5), (2, 4) and (0, 6), induced from the Epanechnikov kernel $K(t) = \frac{3}{4}(1-t^2)I_{[-1,1]}$ are presented below, which will be used to obtain the bandwidth variation function in Section 5.

order (0, 2):

$$K^*_{0,c}(t) = \frac{12(1-t^2)I_{[-1,1]}}{(1+c)^4(19-18c+3c^2)}[8-16c+24c^2-12c^3+(15-30c+15c^2)t]$$

order (0, 4):

$$\begin{aligned}K^*_{0,c}(t) = \frac{60(1-t^2)I_{[-1,1]}}{(1+c)^8(501-900c+510c^2-100c^3+5c^4)}&[4(38-456c+2964c^2\\ &-8057c^3+11970c^4-10395c^5+5320c^6-1635c^7+300c^8-25c^9)\\ &-35(c-1)^2(-37+370c-1119c^2+1220c^3-617c^4+150c^5-15c^6)t\\ &+56(52-624c+1956c^2-2873c^3+2220c^4-900c^5+180c^6-15c^7)t^2\\ &-210(c-1)^2(-9+90c-75c^2+20c^3-2c^4)t^3]\end{aligned}$$

order (1, 3):

$$\begin{aligned}K^*_{1,c}(t) = \frac{60(1-t^2)I_{[-1,1]}}{(1+c)^6(24-33c+12c^2-c^3)}&[-5(c-1)^2(-4+16c-12c^2+3c^3)\\ &+8(16-51c+66c^2-36c^3+6c^4)t-35(c-1)^2(c-4)t^2]\end{aligned}$$

order (1, 5):

$$\begin{aligned}K^*_{1,c}(t) = \frac{10(1-t^2)I_{[-1,1]}}{(1+c)^{10}(1334-2995c+2280c^2-730c^3+90c^4-3c^5)}&\\ [-1764(c-1)^2(-28+504c-3084c^2+8073c^3-10410c^4+7365c^5&\\ -3000c^6+705c^7-90c^8+5c^9)+1344(656-8500c+45360c^2&\end{aligned}$$

$$-121800c^3 + 188160c^4 - 178227c^5 + 106050c^6 - 39615c^7 + 9000c^8$$
$$-1120c^9 + 56c^{10})t - 5292(c-1)^2(-726 + 6243c - 18028c^2 + 22326c^3$$
$$-13910c^4 + 4535c^5 - 720c^6 + 40c^7)t^2 + 12096(496 - 4670c + 14840c^2$$
$$-23415c^3 + 20580c^4 - 10325c^5 + 2870c^6 + 2870c^6 - 400c^7 + 20c^8)t^3$$
$$-19404(c-1)^2(-158 + 1059c - 1204c^2 + 508c^3 - 90c^4 + 5c^5)t^4]$$

order (0, 6):

$$K^*_{0,c}(t) = \frac{(1+c)^{-12}(1-t^2)I_{[-1,1]}}{(14407 - 37926c + 37065c^2 - 16660c^3 + 3465c^4 - 294c^5 + 7c^6)}$$
$$[1344(422 - 12660c + 196230c^2 - 1410305c^3 + 5727270c^4 - 14245908c^5$$
$$+23083970c^6 - 25347420c^7 + 19288710c^8 - 10286290c^9 + 3849510c^{10}$$
$$-1000860c^{11} + 176890c^{12} - 20580c^{13} + 1470c^{14} - 49c^{15}) - 52920(c-1)^2$$
$$(-205 + 5740c - 57628c^2 + 266868c^3 - 652658c^4 + 918372c^5 - 792068c^6$$
$$+434588c^7 - 153855c^8 + 34888c^9 - 4900c^{10} + 392c^{11} - 14c^{12})t + 20160$$
$$(3124 - 93720c + 856428c^2 - 3869349c^3 + 10041510c^4 - 16313694c^5$$
$$+17418870c^6 - 12530553c^7 + 6122256c^8 - 2016875c^9 + 436170c^{10}$$
$$-58590c^{11} + 4410c^{12} - 147c^{13})t^2 - 194040(c-1)^2(-791 + 22148c$$
$$-141204c^2 + 387996c^3 - 546594c^4 + 437052c^5 - 207552c^6 + 58548c^7$$
$$-9399c^8 + 784c^9 - 28c^{10})t^3 + 221760(746 - 22380c + 156378c^2$$
$$-483009c^3 + 823290c^4 - 848823c^5 + 548688c^6 - 223188c^7 + 55860c^8$$
$$-8155c^9 + 630c^{10} - 21c^{11})t^4 - 1513512(c-1)^2(-43 + 1204c$$
$$-5764c^2 + 8428c^3 - 5560c^4 + 1876c^5 - 328c^6 + 28c^7 - c^8)t^5]$$

order (2, 4):

$$K^*_{0,c}(t) = \frac{60(1-t^2)I_{[-1,1]}}{(1+c)^8(501 - 900c + 510c^2 - 100c^3 + 5c^4)}$$
$$[56(52 - 624c + 1956c^2 - 2873c^3 + 2220c^4 - 900c^5 + 180c^6 - 15c^7)$$

$$-490(c-1)^2(-77+290c-279c^2+100c^3-10c^4)t+560(c-2)$$
$$(-92+290c-335c^2+150c^3-15c^4)t^2-4410(c-1)^2$$
$$(-17+10c-c^2)t^3]$$

For some important cases of c, the equivalent boundary kernels of order (0, 2) are shown in Figure 1.

Figure 1 about here

Remark: The equivalent boundary kernels of order (0, 2) are in fact the boundary kernels suggested by Gasser and Müller (1979) where they obtained them by multiplying the interior kernel by a linear function. The local polynomial fitting method offers an intuitive explanation for their arguments. The equivalent boundary kernels of order (0, 2) correct the boundary effects by increasing the order of the kernel by 1 at the boundary. In the context of regression problem, Fan et al. (1993) have indicated that increasing the order of the kernel by 1 at the boundary is more convincing than keeping the same order at the boundary as in the interior. The reason is that if we use the minimum variance kernel $K(t) = \frac{1}{2}I_{[-1,1]}$ in (3.3), then the order of equivalent boundary kernels will increase by 1 at the boundary while keeping the same minimum variance property. Since the equivalent boundary kernel is in fact a product of a linear function and an interior kernel, we can see that increasing the order by 1 at the boundary is not necessary. It will be clear in Section 4 that the optimal boundary kernel of order (0, 2) at $c = 0$ is in fact a polynomial of degree 2, not 3.

3.2. Bandwidth Variation Function

Recall that a constant bandwidth h is used in the interior region I. For x in B_L, write for the local bandwidth $h(x) = b(x|h)h$ for some function $b : [0,1] \to R^+$

with $b(1) = 1$. Then at $x = ch$ for the estimator (2.4) with K_v^* replaced $K_{v,c}^*$ given by (3.3), we obtain

$$\text{Bias}(\tilde{b}_v, x) = \frac{v!}{(p+1)!}\{b(c)h\}^{p+1-v} f^{(p+1)}(x) \int_{-1}^{c/b(c)} t^{p+1} K_{v,c/b(c)}^*(t)dt$$

and

$$\text{Var}(\tilde{b}_v, x) = \frac{(v!)^2}{n[hb(c)]^{2v+1}} f(x) \int_{-1}^{c/b(c)} (K_{v,c/b(c)}^*(t))^2 dt.$$

Write $B_{v,p+1}(c) = \int_{-1}^{c} t^{p+1} K_{v,c}^*(t)dt$ and $V_v(c) = \int_{-1}^{c} (K_{v,c}^*(t))^2 dt$. Then the leading terms of the asymptotic mean squared error (MSE) are

$$\begin{aligned} \text{MSE}(\tilde{b}_v, x) \ \sim\ & \left[\frac{v!}{(p+1)!}\right]^2 \left[f^{(p+1)}(0)\right]^2 \{hb(c)\}^{2(p+1-v)} \left[B_{v,p+1}\left(\frac{c}{b(c)}\right)\right]^2 \\ & + \frac{(v!)^2}{n[hb(c)]^{2v+1}} f(0) V_v\left(\frac{c}{b(c)}\right). \end{aligned} \tag{3.4}$$

Assume that the local bandwidth is chosen optimally at $x = h$, then it minimizes the preceding expression (3.4). Differentiating (3.4) w.r.t. h yields

$$h = \left[\frac{((p+1)!)^2(2v+1)f(0)V_v(1)}{2n(p+1-v)\left(f^{(p+1)}(0)\right)^2 [B_{v,p+1}]^2}\right]^{\frac{1}{2p+3}}. \tag{3.5}$$

Substituting h into (3.4), we obtain

$$\begin{aligned} \text{MSE}(\tilde{b}_v, x) \ \sim\ & \frac{(v!)^2 \left[f^{(p+1)}(0)\right]^{\frac{2(2v+1)}{2p+3}} [B_{v,p+1}(1)]^{\frac{2(2v+1)}{2p+3}} (2(p+1-v))^{\frac{2v+1}{2p+3}}}{f(0)^{-\frac{2(p+1-v)}{2p+3}} ((p+1)!)^{\frac{2(2v+1)}{2p+3}} [V_v(1)]^{\frac{2v+1}{2p+3}} (2v+1)^{\frac{2v+1}{2p+3}} n^{\frac{2(p+1-v)}{2p+3}}} \\ & \left\{\frac{1}{[b(c)]^{2v+1}} V_v\left(\frac{c}{b(c)}\right) + \frac{V_v(1)(2v+1)}{2(p+1-v)[B_{v,p+1}(1)]^2}\right. \\ & \left. [b(c)]^{2(p+1-v)} \left[B_{v,p+1}\left(\frac{c}{b(c)}\right)\right]^2\right\}. \end{aligned} \tag{3.6}$$

Then the optimal bandwidth variation function b is the solution of the variational problem of minimizing

$$\begin{aligned} & \frac{1}{(b(y))^{2v+1}} V_v\left(\frac{y}{b(y)}\right) + \frac{V_v(1)(2v+1)}{2[B_{v,p+1}(1)]^2(p+1-v)} \\ & \qquad (b(y))^{2(p+1-v)} \left[B_{v,p+1}\left(\frac{y}{b(y)}\right)\right]^2, \end{aligned} \tag{3.7}$$

under the requirement $b(y) \geq y$; compare with Müller (1991).

Since $B_{v,p+1}(t)$ has a zero for some $t \in (0,1)$, (3.7) won't yield solutions for which b is smooth in y, thus leading to jumps in the estimated curve. Instead, Müller (1991) suggests to assume that, in (3.7), $B_{v,p+1}\left(\frac{y}{b(y)}\right)$ is fixed at $B_{v,p+1}(1)$ and then try to minimize (3.7). The defect of this method is that $b(\cdot)$ is only optimal at $x = h$. Since it is the most important case to estimate the density at the endpoint, it is required that the bandwidth variation function is optimal at $x = 0$. Assuming $B_{v,p+1}(0) \neq 0$, we suggest using

$$|B_{v,p+1}(0)| + (B_{v,p+1}(1) - |B_{v,p+1}(0)|)\, y/b(y)$$

as an approximation to $B_{v,p+1}\left(\frac{y}{b(y)}\right)$ of (3.7). Unfortunately, the above two methods wouldn't yield an optimal b which is smooth in y. But, later we shall see that our method at least gives an optimal bandwidth variation in the sense of minimizing MSE at $x = 0$ and $x = h$. This issue will be further discussed in Section 5, i.e., in simulation studies. As Müller (1991) suggested, an obvious choice of b will be $b(y) = 1$ for all $0 \leq y \leq 1$, which means no bandwidth variation at all.

A natural question is whether the optimality can be carried into the boundary uniformly when we use the equivalent kernel $K^*_{v,c}$ (see (3.3)) induced by the Epanechnikov kernel. In the next section, we shall investigate optimal kernels for end-points estimation. We shall find a kernel of order (0, 2) and shall show that it is optimal in the sense of minimizing MSE at the (left) end-point. Surprisingly, it is not the equivalent kernel induced by the Epanechnikov kernel.

4. Optimal Boundary Kernels at the End-points

Here we investigate the most important case of order (0, 2) kernels for the left end-point 0. The right endpoint can be treated similarly. From (3.2), boundary kernels should satisfy the conditions

$$\begin{aligned}&\text{(i)}\ \int_{-\infty}^{0} K_{0,0}(t)dt = 1\\ &\text{(ii)}\ \int_{-\infty}^{0} tK_{0,0}(t)dt = 0.\end{aligned} \tag{4.1}$$

Now (3.6) reduces to

$$\frac{1}{b(0)}V_0(0) + \frac{V_0(1)}{4\,[B_{0,2}(1)]^2}(b(0))^4\,[B_{0,2}(0)]^2\,. \tag{4.2}$$

Minimizing (4.2) w.r.t. $b(0)$ under the restriction $b(0) \geq 0$, we obtain the minimizing value

$$b^*(0) = \left\{\frac{V_0(0)\,[B_{0,2}(1)]^2}{V_0(1)\,[B_{0,2}(0)]^2}\right\}^{\frac{1}{5}}. \tag{4.3}$$

Substituting $b^*(0)$ back in (4.2), one obtains

$$\frac{5}{4}\left\{\frac{[B_{0,2}(0)]^2\,[V_0(0)]^4}{[B_{0,2}(1)]^2}\right\}^{\frac{1}{5}}[V_0(1)]^{\frac{1}{5}}. \tag{4.4}$$

So the problem of minimizing MSE (at the left end-point) becomes the problem of finding a boundary kernel $K_{0,0}$ which minimizes $[B_{0,2}(0)]^2\,[V_0(0)]^4$. That is, the $K_{0,0}$ which minimizes

$$T\,(K_{0,0}) = \left(\int_{-\infty}^{0} t^2K_{0,0}(t)dt\right)^2\left(\int_{-\infty}^{0} K_{0,0}^2(t)dt\right)^4. \tag{4.5}$$

The functional $T\,(K_{0,0})$ is invariant for an "equivalence class" of kernels (see Gasser and Müller (1979)).

Lemma 4.1. For any boundary kernel $K_{0,0}(t)$ of order (0, 2) with exact one change of sign on $(-\infty, 0]$, we have $\int_{-\infty}^{0} t^2K_{0,0}(t)dt \neq 0$.

The proof of Lemma 4.1 is given in the Appendix.

By Lemma 1 of Gasser and Müller (1979), $\int_{-\infty}^{0} t^2 K_{0,0}(t)dt$ may be normalized without affecting the solution to (4.5). We take

$$\int_{-\infty}^{0} t^2 K_{0,0}(t)dt = -1. \tag{4.6}$$

Now, if ΔK represents a small deviation for an extremum subject to conditions (4.1) and (4.6), then the variation of

$$\begin{aligned}\int_{-\infty}^{0} K_{0,0}^2(t)dt + \lambda_1 \left[\int_{-\infty}^{0} K_{0,0}(t)dt - 1\right] &+ \lambda_2 \left[\int_{-\infty}^{0} K_{0,0}(t)tdt\right] \\ &+ \lambda_3 \left[\int_{-\infty}^{0} K_{0,0}(t)t^2dt + 1\right]\end{aligned}$$

should be zero, where λ_1, λ_2 and λ_3 are the Lagrange multipliers. Hence,

$$\begin{aligned}2\int_{-\infty}^{0} K_{0,0}(t)\Delta K(t)dt + \lambda_1 \int_{-\infty}^{0} \Delta K(t)dt &+ \lambda_2 \int_{-\infty}^{0} \Delta K(t)tdt \\ &+ \lambda_3 \int_{-\infty}^{0} \Delta K(t)t^2dt = 0.\end{aligned}$$

Therefore

$$2K_{0,0}(t) + \lambda_1 + \lambda_2 t + \lambda_3 t^2 = 0$$

and hence

$$K_{0,0}(t) = \frac{-\lambda_1 - \lambda_2 t - \lambda_3 t^2}{2}. \tag{4.7}$$

Thus, the calculus of variation yields polynomials of degree 2 as possible solutions to the minimizing problem (4.5). But it is not difficult to see that T in (4.5) can be made as small as possible with a suitable choice of γ, where $[\gamma, 0]$ is the support of $K_{0,0}(t)$. To get around this problem, we restrict our attention to the class of kernels with only one change of sign on their support. There are two reasons for choosing such a kernel: 1) A kernel with more than one change of sign on its support usually causes significantly larger variance, 2) Lemma 4.1 ensures that such a kernel satisfying $\int_{-\infty}^{0} t^2 K_{0,0}(t)dt \neq 0$, which further ensures the existence of the optimal bandwidth variation function at $x = 0$, see (3.8). For $K_{0,0}$ defined by (4.7), there are at most two roots such that $K_{0,0}$ takes value 0. Also (4.1) shows that $K_{0,0}$ has at least one root in its support. Naturally we can take the left most root as the left end point of the support, denote it by γ. Thus, $K_{0,0}(\gamma) = 0$ and it is essential that $K_{0,0}(\gamma) = 0$ in order to make the estimator smooth (see Müller (1991) and Müller and Wang (1994)). With the above specifications and under the conditions (4.1) and (4.6), the optimal kernel $K_{0,0}(t)$ given by (4.7) is

$$K_{0,0}(t) = \begin{cases} \frac{3\sqrt{30}}{5} + \frac{27}{5}t + \frac{9\sqrt{30}}{25}t^2, & -\sqrt{\frac{10}{3}} \le t \le 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.8}$$

Normalizing the support of $K_{0,0}(t)$ in (4.8) as $[-1, 0]$, $K_{0,0}(t)$ becomes

$$K_0(t) = \begin{cases} 6 + 18t + 12t^2, & -1 \le t \le 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4.9}$$

Starting from $K_0(t)$, naturally we expect the other boundary kernels for $0 \le c < 1$ are of order 2 with support $[-1, c]$. In fact, for a boundary kernel K_c, under (4.1) and another assumption that $K_c(-1) = 0$, $0 < c < 1$, we can show that

$$K_c(t) = \begin{cases} \frac{12}{(1+c)^4}(1+t)\left[t(1-2c) + \frac{3c^2-2c+1}{2}\right], & -1 \le t \le c \\ 0, & \text{otherwise.} \end{cases} \tag{4.10}$$

Note that $K_c(t)$ is a natural continuation of $K_0(t)$ (given by (4.9)) and the Epanechnikov's kernel $K_{opt}(t) = \frac{3}{4}(1 - t^2)I_{[-1,1]}$. That is, putting $c = 0$ and $c = 1$ in (4.10) one obtains K_0 and K_{opt}, respectively. Interestingly, Müller and Wang (1994) have derived the same boundary kernel using a different approach.

Theorem 4.1. Among all the end-point kernels of order (0, 2) with one change of sign on $[-1, 0]$, the kernel $K_0(t)$ defined by (4.9) is optimal in the sense of minimizing MSE at $x = 0$.

The proof of Theorem 4.1 is given in the Appendix. Although the kernel (4.9) is the optimal kernel at the left end-point, the optimality of K_c (given by (4.10)) at other boundary points remains unknown.

5. Simulations

In this section, we estimate the density f given by $f(x) = e^{-x}$, $x \geq 0$, at the left boundary region using order (0, 2) kernels. The two boundary kernels that are employed are (see (4.10) and (3.3))

$$K_c(t) = \frac{12}{(1+c)^4}(t+1)\left[t(1-2c) + \frac{3c^2-2c+1}{2}\right] I_{[-1,c]} \qquad (5.1)$$

and

$$\begin{aligned} K^*_{0,c}(t) &= \frac{12(1-t^2)}{(1+c)^4(3c^2-18c+19)} \\ &\quad \left\{8 - 16c + 24c^2 - 12c^3 + t(15 - 30c + 15c^2)\right\} I_{[-1,c]}. \end{aligned} \qquad (5.2)$$

For (5.1), we have

$$\begin{aligned} B(c) &= \int_{-1}^{c} K_c(t)t^2 dt = \frac{-1+6c-3c^2}{10}, \\ V(c) &= \int_{-1}^{c} (K_c(t))^2\, dt = \frac{12(2-3c+3c^2)}{5(1+c)^3}. \end{aligned} \qquad (5.3)$$

For (5.2), we have

$$\begin{aligned} B'(c) &= \int_{-1}^{c} K^*_{0,c}(t)t^2 dt = \frac{11-66c+81c^2-36c^3+6c^4}{5(-19+18c-3c^2)} \\ V'(c) &= \int_{-1}^{c} \left(K^*_{0,c}(t)\right)^2 dt = \frac{48(1184-3936c+6600c^2-6345c^3+3345c^4-891c^5+99c^6)}{35(1+c)^3(-19+18c-3c^2)^2}. \end{aligned} \qquad (5.4)$$

Figure 2 gives the values of $B(c)$, $V(c)$, $B'(c)$ and $V'(c)$ at the boundary region $(0, h)$. The x-axis of the figure represents values of c ranging from 0 to 1.

Figure 2 about here

From (3.6), MSE at $x = ch$, $0 \le c \le 1$, is

$$\mathrm{MSE}(x) = \frac{n^{-\frac{4}{5}}}{5^{\frac{2}{5}}(0.6)^{\frac{1}{5}}}\left\{\frac{1}{b(c)}V_0\left(\frac{c}{b(c)}\right) + \frac{0.6}{4(0.2)^2}b^4(c)B_{0,2}^2\left(\frac{c}{b(c)}\right)\right\}, \qquad (5.5)$$

where $V_0(y) = \int_{-1}^{y} K_{0,y}^2(t)dt$ and $B_{0,2} = \int_{-1}^{y} K_{0,y}(t)t^2dt$, with $b(c)$ is determined by minimizing

$$\frac{1}{b(c)}V_0\left(\frac{c}{b(c)}\right) + \frac{0.6}{4(0.2)^2}b^4(c)B_{0,2}^2\left(\frac{c}{b(c)}\right). \qquad (5.6)$$

For kernel (5.1), since $B_{0,2}(0) = -0.1$ and $B_{0,2}(1) = 0.2$, we approximate $B_{0,2}\left(\frac{c}{b(c)}\right)$ by $0.1\left(\frac{c}{b(c)} + 1\right)$. Using (5.3), the optimal bandwidth variation, $b_1(c)$ say, (one which minimizes (5.6) with the above $B_{0,2}$) is shown by the thin-bold line in Figure 3 (top). The optimal bandwidth variation when we fix $B_{0,2}(\cdot)$ at $B_{0,2}(1)$ is also calculated, $b_2(c)$ say. The latter is shown by the thick-bold line in Figure 3 (top). They both have discontinuity points, so they won't yield smooth estimators. We also compared their MSE values as well. It is clear to us that b_1 gives better results than b_2. Furthermore, b_1 in Figure 3 (top) shed some lights on how to choose an approximation to the optimal bandwidth variation function. We propose

Method 1: Choose constant bandwidth variation, i.e., $b_3(c) = 1$.

Method 2: Use the line which connects the points $(1, 1)$ and $(0, 2)$ as the bandwidth variation function, i.e., $b_4(c) = 2 - c$.

Method 3: Define

$$b_5(c) = \begin{cases} 4c^2 - 4c + 2, & c \le 0.5 \\ 1, & c > 0.5. \end{cases}$$

Functions b_3, b_4 and b_5 are also exhibited in Figure 3 (top) for comparison.

Figure 3 about here

Note that methods 2 and 3 give optimal bandwidth variation at $x = 0$ and $x = h$ in the sense of minimizing (5.6). That means b_4 and b_5 minimize (5.6) at $c = 0$ and $c = 1$.

For kernel (5.2), a similar analysis is done, but now with $b_4(c) = 1.86174 - 0.86174c$ and

$$b_5(c) = \begin{cases} 3.44696c^2 - 3.44696c + 1.86174, & c \leq 0.5 \\ 1, & c > 0.5. \end{cases}$$

Again, b_i ($i = 1, 2, 3, 4$ and 5) are displayed in Figure 3 (bottom).

In the following, all the MSE values are computed from (5.5). Figure 4 (top) exhibits MSE values (w.r.t. c, $0 \leq c \leq 1$) of estimating $f(x) = e^{-x}$, $x \geq 0$, by three different bandwidth variation choices (namely, b_3, b_4 and b_5) with kernel (5.1) in the boundary region $(0, h)$, where $h = \left(\frac{V_0(1)}{nB_{0,2}^2(1)}\right)^{\frac{1}{5}}$, the local optimal bandwidth (see (3.5)). With $n = 200$, $h = 0.5956789$. (In practice, h can be approximated by the cross-validation method, etc.)

Similarly, Figure 4 (bottom) displays MSE values (w.r.t. c, $0 \leq c \leq 1$) by b_3, b_4 and b_5 with kernel (5.2).

Figure 4 about here

Note that b_i's with kernel (5.1) are different from b_i's with kernel (5.2), although they are obtained using the same method of choosing the bandwidth variation. Both plots in Figure 4 (top and bottom) reveal that b_4 and b_5 are better than b_3 near the left endpoint. At the points near $x = h$ (or $c = 1$), b_3 (namely, no bandwidth variation) yields good results. Overall, b_4 and b_5 are good choices of bandwidth variation functions.

Figure 5 (top) compares MSE values (w.r.t. c, $0 \leq c \leq 1$) for kernels (5.1) and (5.2) with bandwidth variation function b_4. Figure 5 (bottom) compares MSE values (w.r.t. c, $0 \leq c \leq 1$) for kernels (5.1) and (5.2) with bandwidth variation function b_5. We find that at the points near $x = 0$, kernel (5.1) is better than kernel (5.2). This confirms Theorem 4.1, according to which kernel (5.1) is optimal at $x = 0$. At the points near $x = h$ (or $c = 1$), kernel (5.2) is better. So, we can conclude that linear polynomial smoothing method brings some optimality near $x = h$ but not to the left endpoint 0, whereas kernel (5.1) is optimal at the left endpoint. This optimality, however, doesn't prevail to the points near $x = h$.

Figure 5 about here

We also compared the performance of kernels (5.1), (5.2), Müller's kernel (referred to Müller (1991)) and the Epanechnikov kernel (without a boundary correction) for estimating the density $f(x) = e^{-x}$, $x \geq 0$. The bandwidth variation function that was employed in the simulation for the first three kernels was b_4, i.e., Method 2. For comparison, the theoretical optimal bandwidth is adopted at $x = h$. The number of replications is 100. The sample size was $n = 200$. Summary results are presented in Figure 6. The x-axis is for values of x from 0 to 2.

Figure 6 about here

Clearly, the Epanechnikov kernel didn't remove the boundary effect, since it was implemented without a boundary correction. The kernels (5.1) and (5.2) have good comparable performances in approximating the true density, $f(x) = e^{-x}$, $x \geq 0$. They both perform better than Müller's kernel. We also noticed the optimality property of kernel (5.1) for estimating the density at the endpoint $x = 0$.

6. The case of $f^{(1)}(0) = 0$

The order (0, 2) boundary kernels discussed in the previous sections are used to obtain the second order term and to cancel out the first order term in the bias expansion. But, when $f^{(1)}(0) = 0$, the first order term disappears automatically. Assume that the bandwidth h is used in the interior of the support of $f(x)$, where $f(x)$ is a density with support $[0, a]$, $a \leq \infty$. Then, the usual kernel estimator of f is

$$\tilde{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{6.1}$$

where K is a kernel of order (0, 2) with support $[-1, 1]$. The bias expansion of $\tilde{f}(x)$ at $x = ch$ is

$$\text{Bias}(\tilde{f}, x) = f(x)\left(\int_{-1}^{c} K(t)^2 dt - 1\right) + \frac{h^2}{2} f^{(2)}(x) \int_{-1}^{c} t^2 K(t) dt. \tag{6.2}$$

In (6.2), the first order term of h doesn't appear. But, obviously, $\tilde{f}(x)$ is not a consistent estimator of $f(x)$ at $x = ch$ for $0 \leq c < 1$. To get around this problem, a number of methods have been suggested in the literature. Two popular ones are the so-called cut-and-normalized kernel method (Gasser and Müller (1979)) and the reflection method (Schuster (1985) and Silverman (1986, p. 31)). The cut-and-normalized kernel method in fact uses an order (0, 1) kernel with support $[-1, c]$ in (6.1). The order (0, 1) kernel is referred to a kernel K_c which satisfies

$$\int_{-1}^{c} K_c(t) dt = 1, \tag{6.3}$$

with the condition that K_c is nonnegative. In our investigation, we have observed that the local polynomial fitting method discussed in Section 2 with $p = 0$ (i.e., the constant fit) yields a cut-and-normalized kernel. Define

$$K(t) = \frac{3}{4}(1-t^2)I_{[-1,1]}. \tag{6.4}$$

The constant fit equivalent kernel induced by (6.4) is (see (2.5))

$$K_c^*(t) = \frac{3}{2+3c-c^3}(1-t^2)I_{[-1,c]}. \tag{6.5}$$

When $c = 1$, (6.5) becomes (6.4), so $K_c^*(t)$ provides a natural continuation of $K(t)$, keeping the nonnegativity property intact.

We now compare the above constant fit method with the reflection method. First, let's recall the reflection method. Assume that we have a random sample $X_1, ..., X_n$ from an unknown density f. Then, the estimator of f from the reflection method is

$$f^{\dagger}(x) = \begin{cases} \tilde{f}(x) + \tilde{f}(-x) & x \in [0,h) \\ \tilde{f}(x) & x \in [h, a-h] \\ \tilde{f}(x) + \tilde{f}(2a-x) & x \in [a-h, a] \end{cases}, \tag{6.6}$$

where $\tilde{f}(x)$ is defined by (6.1) and K is a usual nonnegative order (0, 2) kernel function. For $x = ch$,

$$\begin{aligned} f^{\dagger}(x) &= \frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{x-X_i}{h}\right) + K\left(\frac{-x-X_i}{h}\right)\right] \\ &= \frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x-X_i}{h} - 2c\right)\right]. \end{aligned}$$

So, one can view $f^{\dagger}(x)$ as the usual kernel estimator with the kernel

$$K_c^{\dagger}(t) = K(t) + K(t-2c). \tag{6.7}$$

From the facts that $\int_{-1}^{c} K_c^{\dagger}(t)dt = 1$ and $K_c^{\dagger}(t) \geq 0$, the reflection method is in fact a boundary kernel method with the boundary kernel (6.7). To get a better understanding of the difference between the constant fit kernel (6.5) and the boundary kernel (6.7) derived from the refletion method, we plotted them for different values of c with K given by (6.4) in Figures 7 and 8. We see that (6.7) looks very much like the cut-and-normalized kernel (6.5) except that (6.7) is not smooth inside its support, while (6.5) is.

Figures 7 and 8 about here

It is generally believed that an unsmoothness of the kernel function will result in an unsmooth density estimator. Therefore, a natural question is " Is it necessary to use the reflection method at the risk of obtaining an unsmooth estimator?" One may argue that (6.7) can be smoothen if we use a smoother kernel K. Figure 9 below shows that the plot of the kernel from the reflection method with $K(t) = \frac{15}{16}(1-t^2)^2 I_{[-1,1]}$, which is smoother than (6.4). Now the kernel (6.7) is a smooth function. However, a smoother kernel will increase the variance of the estimator. So, if our purpose is to minimize MSE, the smoother-kernel method is not preferred.

Figure 9 about here

A comparison between the performance of the kernel (6.5) and (6.7) was done as well. As in Section 5, here again we face the problem of choosing the bandwidth variation function. For a kernel K_c which satisfies (6.3) and $\tilde{f}$ defined by (6.1) with K replaced by K_c at $x = ch$, the asymptotic bias and variance at $x = ch$ are

$$\text{Bias}(\tilde{f}, x) = \frac{h^2}{2} f^{(2)}(0) \int_{-1}^{c} t^2 K_c(t)dt, \tag{6.8}$$

and

$$\text{Var}(\tilde{f}, x) = \frac{f(0)}{nh} \int_{-1}^{c} K_c^2(t)dt. \tag{6.9}$$

Therefore,

$$\text{MSE}(\tilde{f}, x) = \frac{h^4}{4}[f^{(2)}(0)]^2 \left[\int_{-1}^{c} t^2 K_c(t)dt\right]^2 + \frac{f(0)}{nh}\int_{-1}^{c} K_c^2(t)dt. \qquad (6.10)$$

Assume that h is chosen by (3.5), then

$$h = \left[\frac{f(0)\int_{-1}^{1} K^2(t)dt}{[f^{(2)}(0)]^2 \left(\int_{-1}^{1} t^2 K(t)dt\right)^2}\right]^{\frac{1}{5}} n^{-\frac{1}{5}}. \qquad (6.11)$$

Using a similar discussion as in Subsection 3.2, we obtain the optimal bandwidth variation function as

$$b(c) = \left[\frac{\int_{-1}^{c} K_c^2(t)dt}{\int_{-1}^{1} K^2(t)dt}\right]^{-\frac{1}{5}} \left[\frac{\int_{-1}^{1} t^2 K(t)dt}{\int_{-1}^{c} t^2 K_c(t)dt}\right]^{-\frac{1}{5}}. \qquad (6.12)$$

Here, $b(c)$ makes sense for every $0 \leq c \leq 1$, since $\int_{-1}^{c} t^2 K_c(t)dt > 0$ for $0 \leq c \leq 1$.

For $K(t)$ and $K_c^*(t)$ defined by (6.4) and (6.5),

$$\int_{-1}^{c} [K_c^*(t)]^2 dt = \frac{3(8 - 9c + 3c^2)}{20 - 15c^2 + 5c^3}$$

and

$$\int_{-1}^{c} t^2 K_c^*(t)dt = \frac{-2 + 4c - 6c^2 + 3c^3}{5(c-2)},$$

then, the optimal bandwidth variation function is

$$b^*(c) = \left[\frac{3(8 - 9c + 3c^2)}{(c+1)(-2 + 4c - 6c^2 + 3c^3)^2}\right]^{\frac{1}{5}}. \qquad (6.13)$$

For $K_c^\dagger(t)$ defined by (6.7) with K given by (6.4),

$$\int_{-1}^{c} [K_c^\dagger(t)]^2 dt = 1.2 - 3c^2 + 3c^3 - 0.6c^5$$

and

$$\int_{-1}^{c} t^2 K_c^\dagger(t)dt = 0.2 - 0.75c + 2c^2 - 1.5c^3 + 0.25c^5,$$

then, the optimal bandwidth variation function is

$$b^{\dagger}(c) = \left[\frac{1.2 - 3c^2 + 3c^3 - 0.6c^5}{15(0.2 - 0.75c + 2c^2 - 1.5c^3 + 0.25c^5)^2}\right]^{\frac{1}{5}}. \tag{6.14}$$

Figure 10 displays b^* and $b^{\dagger}$ for $0 \le c \le 1$. Note that $b^*(0) = b^{\dagger}(0)$. In fact, we can also see that $K_0^*(t) = K_0^{\dagger}(t)$. Thus, the constant fit method is equivalent to the reflection method at $x = 0$ when the same interior kernel is used.

Figure 10 about here

Figure 11 gives the MSE values of the constant fit method and the reflection method with K defined by (6.4). The estimated density is $f(x) = \frac{1}{2}(x+1)e^{-x}$, $x \ge 0$. The bandwidth variation functions $b^*(c)$ and $b^{\dagger}(c)$ given by (6.13) and (6.14), respectively, are employed. For the purpose of comparison, MSE values are calculated from (6.10) and the theoretical optimal bandwidth at $x = h$ is employed. In practice, this bandwidth can be approximated by the cross-validation method or by the plug-in method. From Figure 11, it is clear that the reflection method performs slightly better than the cut-and-normalized kernel method at about one-third of the boundary region near the endpoint. At the other two-third of the boundary region near $x = h$, the reflection method is obviously inferior to the cut-and-normalized kernel method. We also computed the IMSE (integrated MSE) values of the two estimators at the boundary region. The IMSE value from the reflection method is 0.004424794, whereas the IMSE value from the cut-and-normalized kernel method is 0.004385766. All indications are that the cut-and-normalized kernel method (or the constant fit method) marginally outperforms the reflection method. Another advantage of the cut-and-normalized kernel method is that it needs less computation than the reflection method.

Figure 11 about here

In conclusion, we recommend the use of the cut-and-normalized kernel method with a bandwidth variation function in practice. If our interest is only on the estimation of the density at the end-points, then there is no difference between the two methods, taking the bandwidth variation into consideration. Chen (1996) discussed the problem of estimating animal abundance using the kernel method. In his paper, he estimated $f(0)$ by the reflection method, under the shape criterion $f^{(1)}(0) = 0$. Though his simulation results are good, they can be improved further by using a bandwidth variation function.

7. Examples

In this section, we are mainly interested in the cases where the data are peaked at $x = 0$. In estimating the animal abundance, many methods have been suggested. The Fourier series method is one of the most popular methods that is frequently used in practice. We now compare our boundary kernel method with the Fourier series method. Assume that the true model is the truncated exponential model of Crain et al. (1979), where the density of the detected right angle distance X is

$$f(x) = \frac{\lambda exp(-\lambda x)}{1 - exp(-\lambda w)}, \; 0 \leq x \leq w \tag{7.1}$$

with w being the distance from the travel path beyond which no object can ever be detected. In simulations, we used the boundary kernel (5.1), which is optimal at $x = 0$. The bandwidth variation function b_4 is employed at $x = 0$. The optimal bandwidth at $x = h$ is approximated by using the reference density method (see Silverman (1986) and Chen (1996)). We chose the exponential density $f(x) =$

$\lambda\exp(-\lambda x)$, $x \geq 0$, as our reference density. Since

$$h = \left[\frac{15f(0)}{f^{(2)}(0)^2}\right]^{1/5} n^{-1/5} = 15^{1/5}\lambda^{-1}n^{-1/5},$$

and since the expectation of the random variable X with density $\lambda\exp(-\lambda x)$, $x \geq 0$, is λ^{-1}, it is natural to use $\hat{h} = 15^{1/5}E(X)n^{-1/5}$ as the estimator of h. For the Fourier series method, we computed the estimates for terms ranging from one to six. In simulations, We generated 100 replications of the samples from the model (7.1) with $w = 100$. For each sample, the sample size was chosen as 100. Table 1 summarizes the simulation results: $\hat{f}_{F,i}(0)$ denotes the Fourier series estimate of $f(0)$ when the estimator contains i number of terms, $i = 1, ..., 6$. The values in the parentheses are the corresponding MSE values. The boundary kernel estimate of $f(0)$ is denoted by $\hat{f}(0)$ in the table. Entries in the table are averages over 100 replications.

Note that the values of $\hat{f}(0)$ are very close to the true values of $f(0)$ consistently for all values of λ. Furthermore, Table 1 indicates that $\hat{f}(0)$ is a better estimate than $\hat{f}_{F,i}(0)$ for all $i = 1, 2, ..., 6$ and for all λ considered. For example, when λ=0.15, the true value of $f(0)$ is 0.15. The estimates $\hat{f}_{F,6}(0)$ and $\hat{f}(0)$ are 0.08953989 and 0.1408818, respectively, and the correspoding MSE values are 0.003675134 and 0.0005483571, respectively.

We also computed the estimates without taking the bandwidth variation into consideration. These values are not reported here, and they are not better than those values of the Fourier series method. Therefore, it appears that use of a bandwidth variation function is very important to guarantee the good performance of the kernel estimator.

Table 1. Comparison between the boundary kernel method
and the Fourier Series method

	λ=0.05	λ=0.10	λ=0.15
f(0)	0.05033918	0.1000045	0.15
$\hat{f}_{F,1}(0)$	0.02445792 (0.0006705535)	0.02809392 (0.0051713006)	0.02915134 (0.014604443)
$\hat{f}_{F,2}(0)$	0.03207089 (0.0003377232)	0.04221390 (0.0033414883)	0.04613698 (0.0.010788065)
$\hat{f}_{F,3}(0)$	0.03643933 (0.0002027705)	0.05257510 (0.0022551197)	0.06043987 (0.008023476)
$\hat{f}_{F,4}(0)$	0.03918505 (0.0001410504)	0.06012749 (0.0016017385)	0.07217877 (0.006062488)
$\hat{f}_{F,5}(0)$	0.04103015 (0.0001094010)	0.06576078 (0.0011924119)	0.08173865 (0.004671730)
$\hat{f}_{F,6}(0)$	0.04236107 (0.00009284404)	0.07010664 (0.0009232401)	0.08953989 (0.003675134)
$\hat{f}(0)$	0.04751006 (0.00006672282)	0.09171048 (0.0002812265)	0.1408818 (0.0005493571)

As a second example, we calculated the estimate for Ruffed grouse data (Gates (1979)), which shows a peak at $x = 0$ and has been discussed in Buckland (1985). The boundary kernel method can not be applied directly since the data are grouped. To get around this problem, we first fitted a parametric model for the data (this idea was suggested to us by Professor S. T. Buckland). The range of this data is from 0 to 30, which is grouped into six equal-length intervals. Since this data shows a peak at $x = 0$, it is reasonable to fit this data by the following exponential model

$$f(x) = \lambda exp(-\lambda x), \; x \geq 0. \tag{7.2}$$

A natural estimator of λ is

$$\hat{\lambda} = \frac{n}{2.5n_1 + 7.5n_2 + 12.5n_3 + 17.5n_4 + 22.5n_5 + 27.5n_6},$$

where n_i $(i = 1, ..., 6)$ are the number of observations in each group, $n = \sum_{i=1}^{6} n_i$.

Next, we generated random samples according to the density (7.2) with λ replaced by $\hat{\lambda}$ for each group interval, keeping the number of observations unchanged. we then used the boundary kernel method to find an estimate using these pseudo data. Our estimate of $f(0)$ is $\hat{f}(0)$=0.1229497 with the standard error = 0.001194368. The preceding values are averages of 100 replications. The above estimate $\hat{f}(0)$ is slightly higher than the value reported in Buckland (1985). We believe that our estimate is more accurate than Buckland's, since the polynomial method implemented in his paper is mainly efficient when the data exhibit a shoulder at $x = 0$. Also, a small standard error indicates that our method of dealing with the grouped data is convincing.

From the above simulation results, one can see that the boundary kernel method is very promissing in estimating the density when the density exhibites a peak at $x = 0$ and has an exponential-type density. In addition, it is obvious that the boundary kernel method can be used for any type of data, whatever the true model is. However, the choice of bandwidth is also rather crucial for better results. The method of choosing the optimal bandwidth by using a reference density may be sensitive to the choice of the reference density. A wise choice of h is the use of the cross-validation method, though the computation may be burdensome. In the cases where one has some information about the true density, the reference density method may be a good choice.

REFERENCES

Buckland, S. T. (1985). Perpendicular distance models for line transect sampling. *Biometrics* 41, 177-195.

Buckland, S. T. (1992). Fitting density functions with polynomials. *Applied Statistics* 41, 63-76.

Buckland, S. T., Anderson, D. R., Burnham, K. P. and Laake, J. L. (1993). *Distance Sampling.* London: Chapman and Hall.

Burnham, K. P. and Anderson, D. R. (1976). Mathematical models for nonparametric inferences from line transect data. *Biometrics* 32, 325-336.

Burnham, K. P., Anderson, D. R. and Lakke, J. L. (1980). Estimation of density for line transect sampling of biological populations. *Wildlife Monographs* 72.

Chen, S. X. (1996). A kernel estimate for the density of a biological population by using line transect sampling. *Applied Statistics* 45, 135-150.

Cline, D. B. H. and Hart, J. D. (1991). Kernel estimation of densities of discontinuous derivatives. *Statistics* 22, 69-84.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87, 998-1004.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics* 20, 2008-2036.

Fan, J., Gasser, T., Gijbels, I., Brockman, M. and Joachim, E. (1993). Local polynomial fitting: A standard for nonparametric regression. Unpublished manuscript.

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.

Gasser, T., Müller, H. G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society,* Series B. 47, 238-256.

Gates, C. E., Marshall, W. H. and Olson, D. P. (1968). Line transect method of estimating grouse population densities. *Biometrics* 24, 135-145.

Gates, C. E. (1979). Line transect and related issues. In *Sampling Biological Populations*, R. M. Cormack, G. P. Patil, and D. S. Robson (eds), 71-154. Fairland, Maryland: International Cooperative Publishing House.

Granovsky, B. L. and Müller, H. G. (1989). On the optimality of a class of polynomial kernel functions. *Statistics and Decisions* 7, 301-312.

Hall, P. (1981). On trigonometric series estimates of densities. *The Annals of Statistics* 9, 683-685.

Karunamuni, R. J. and Quinn, T. J. II (1995). Bayesian estimation of animal abundance for line transect sampling. *Biometrics* 51, 1325-1337.

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

Müller, H. G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* 78, 521-530.

Müller, H. G. and Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* 50, 61-76.

Pollock, K. H. (1978). A family of density estimators for line transect sampling. *Biometrics* 34, 475-478.

Quinn, T. J. II and Gallucci, V. F. (1980). Parametric models for line-transect estimators of abundance. *Ecology* 61, 293-302.

Ramsey, F. L. (1979). Parametric models for line transect surveys. *Biometrika* 66, 505-512.

Rice, J. (1984). Boundary modifications for kernel regression. *Communications in Statistics- Theory and Methods* 13, 890-900.

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

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

Stromberg, K. R. (1981). *An Introduction to Classical Real Analysis.* Wadsworth, California.

APPENDIX: Proofs

Proof of Lemma 2.1. First note that (see (2.2))

$$\begin{aligned} S_{n,j}(x) &= \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)(x-X_i)^j \\ &= ES_{n,j}(x) + o_P\left(\sqrt{Var[S_{n,j}(x)]}\right). \qquad (A.1) \end{aligned}$$

Assume the support of K is [-1,1], by a change of variables followed by an application of Taylor formula,

$$\begin{aligned} \mathrm{E}S_{n,j}(x) &= n\int K\left(\frac{x-u}{h}\right)(x-u)^j f(u)du \\ &= nh\int_{-1}^{1} K(t)(th)^j f(x-th)dt \\ &= nh^{j+1}\int_{-1}^{1} K(t)t^j \left\{f(x) - thf^{(1)}(x) + \frac{t^2h^2}{2}f^{(2)}(x^*)\right\} dt \\ &= nh^{j+1} f(x)\int_{-1}^{1} K(t)t^j dt(1+o(1)). \qquad (A.2) \\ \mathrm{Var}[S_{n,j}(x)] &\leq nEK^2\left(\frac{x-X_1}{h}\right)(x-X_1)^{2j} \\ &= n\int K^2\left(\frac{x-u}{h}\right)(x-u)^{2j} f(u)du \\ &= nh^{2j+1}\int_{-1}^{1} K^2(t)t^{2j} f(x-th)dt \\ &= O\left(nh^{2j+1}\right). \qquad (A.3) \end{aligned}$$

By combining (A.1), (A.2) and (A.3), we obtain

$$S_{n,j}(x) = nh^{j+1} f(x)\int_{-1}^{1} K(t)t^j dt\,(1+o_P(1)). \qquad (A.4)$$

Now $\hat{b}_v(x)$ defined by (2.3) can be rewritten as

$$\begin{aligned}
\hat{b}_v(x) &= \frac{1}{nf(x)}\sum_{i=1}^{n} e_v^T \\
&\quad \begin{pmatrix} h\int_{-1}^{1}K(t)dt, & h^2\int_{-1}^{1}K(t)tdt, & \dots & h^{p+1}\int_{-1}^{1}K(t)t^pdt \\ h^2\int_{-1}^{1}K(t)tdt, & h^3\int_{-1}^{1}K(t)t^2dt, & \dots & h^{p+2}\int_{-1}^{1}K(t)t^{p+1}dt \\ \vdots & \vdots & & \vdots \\ h^{p+1}\int_{-1}^{1}K(t)t^pdt, & h^{p+2}\int_{-1}^{1}K(t)t^{p+1}dt, & \dots & h^{2p+1}\int_{-1}^{1}K(t)t^{2p}dt \end{pmatrix}^{-1} \\
&\quad \cdot(1, x-X_i, \dots, (x-X_i)^p)^T K\left(\frac{x-X_i}{h}\right) Y_i(1+o_P(1)) \\
&= \frac{1}{nf(x)}\sum_{i=1}^{n} e_v^T \mathrm{Diag}\left(h^{-1}, h^{-2}, \dots, h^{-(p+1)}\right) \\
&\quad \cdot \begin{pmatrix} \int_{-1}^{1}K(t)dt, & \int_{-1}^{1}K(t)tdt, & \dots & \int_{-1}^{1}K(t)t^pdt \\ \int_{-1}^{1}K(t)tdt, & \int_{-1}^{1}K(t)t^2dt, & \dots & \int_{-1}^{1}K(t)t^{p+1}dt \\ \vdots & \vdots & & \vdots \\ \int_{-1}^{1}K(t)t^pdt, & \int_{-1}^{1}K(t)t^{p+1}dt, & \dots & \int_{-1}^{1}K(t)t^{2p}dt \end{pmatrix}^{-1} \\
&\quad \cdot\mathrm{Diag}\left(1, h^{-1}, \dots, h^{-p}\right)(1, x-X_i, \dots, (x-X_i)^p)^T \\
&\quad K\left(\frac{x-X_i}{h}\right) f(X_i)(1+o_P(1)) \\
&= \frac{1}{nh^{v+1}f(x)}\sum_{i=1}^{n} e_v^T S^{-1}\left(1, \frac{x-X_i}{h}, \dots, \left(\frac{x-X_i}{h}\right)^p\right) \\
&\quad K\left(\frac{x-X_i}{h}\right) f(X_i)(1+o_P(1)) \qquad \text{(A.5)}
\end{aligned}$$

Since K has support on $[-1,1]$, it is enough to consider the points such that $x-h \le X_i \le x+h$. But, for $x-h \le X_i \le x+h$, we have $f(X_i) = f(x)(1+o(1))$ under the assumptions of the lemma. Combining the preceding result together with (A.5) completes the proof.

Proof of Theorem 4.1. Define $p_0(t) = 6(1+3t+2t^2)$. Suppose $\tilde{K}(t) = K_0(t) +$

$\Delta K_0(t)$ is also a kernel of order (0, 2) satisfying conditions (4.1) and

$$\int_{-\infty}^{0} \tilde{K}(t)t^2 dt = \int_{-\infty}^{0} K_0(t)t^2 dt = -0.1, \tag{A.6}$$

compare with (4.6). Then by (4.1) and (A.6), we obtain

$$\begin{aligned}
\int_{-\infty}^{0} \tilde{K}^2(t)dt &= \int_{-\infty}^{0} \left[K_0(t) + \Delta K_0(t)\right]^2 dt \\
&\quad -12\left\{\int_{-\infty}^{0} \left[K_0(t) + \Delta K_0(t)\right] dt - 1\right\} \\
&\quad -36\left\{\int_{-\infty}^{0} \left[K_0(t) + \Delta K_0(t)\right] t dt\right\} \\
&\quad -24\left\{\int_{-\infty}^{0} \left[K_0(t) + \Delta K_0(t)\right] t^2 dt + \frac{1}{10}\right\} \\
&= \int_{-\infty}^{0} K_0^2(t)dt + \int_{-\infty}^{0} \left[2K_0(t) - 12 - 36t - 24t^2\right] \Delta K_0(t)dt \\
&\quad + \int_{-\infty}^{0} \left(\Delta K_0(t)\right)^2 dt \\
&\geq \int_{-\infty}^{0} K_0^2(t)dt + 2\int_{-\infty}^{0} \left[K_0(t) - p_0(t)\right] \Delta K_0(t)dt,
\end{aligned} \tag{A.7}$$

since $\int_{-\infty}^{0} (\Delta K_0(t))^2 dt \geq 0$. So, to prove $\int_{-\infty}^{0} \tilde{K}^2(t) \geq \int_{-\infty}^{0} K_0^2(t)dt$ it is enough to show that

$$\int_{-\infty}^{0} \left[K_0(t) - p_0(t)\right] \Delta K_0(t)dt = -\int_{-\infty}^{-1} p_0(t)\Delta K_0(t)dt \geq 0, \tag{A.8}$$

since $K_0(t) = p_0(t)I_{[-1,0]}$. In other words, $\int_{-\infty}^{-1} p_0(t)\Delta K_0(t)dt \leq 0$. We claim that the only change of sign of $\tilde{K}(t)$ in $(-\infty, 0]$ occurs at some point $\phi < 0$, goes from $-$ to $+$. Suppose the contrary is true. Then, by the second mean value theorem for integrals (see, e.g., Stromberg (1981, p.238)), there exist two points ξ_1 and ξ_2 such that $-\infty < \xi_1 < \xi_2 < 0$ and

$$0 = \int_{-\infty}^{0} \tilde{K}(t)tdt = \xi_1 \int_{-\infty}^{\phi} \tilde{K}(t)dt + \xi_2 \int_{\phi}^{0} \tilde{K}(t)dt. \tag{A.9}$$

But, RHS of (A.9)$< \xi_2 \int_{-\infty}^{0} \tilde{K}(t)dt < 0$, and this is a contradiction with the left hand side of (A.9). Thus, our claim is true.

Now to prove (A.8), consider the following two cases:

Case (a): $\phi \in [-1, 0]$.

Case (b): $\phi \in (-\infty, -1)$.

If $\phi \in [-1, 0]$, then $\int_{-\infty}^{-1} p_0(t)\Delta K_0(t)dt \leq 0$ is immediate, since $p_0(t) > 0$ for $t < -1$, and $\Delta K_0(t) = \tilde{K}(t) \leq 0$ for $t < -1$. Now consider the case $\phi \in (-\infty, -1)$. Then, again by the second mean value theorem for integrals,

$$\begin{aligned}\int_{-\infty}^{-1} p_0(t)\Delta K_0(t)dt &= \int_{-\infty}^{\phi} p_0(t)\Delta K_0(t)dt + \int_{\phi}^{-1} p_0(t)\Delta K_0(t)dt \\ &= \frac{p_0(\omega_1)}{\omega_1}\int_{-\infty}^{\phi} \Delta K_0(t)tdt + \frac{p_0(\omega_2)}{\omega_2}\int_{\phi}^{-1} \Delta K_0(t)tdt \qquad \text{(A.10)}\end{aligned}$$

for some ω_1 and ω_2 such that $-\infty < \omega_1 < \phi < \omega_2 < -1$. Observe that $p_0(t) > 0$ for $t < -1$ and that $\frac{p_0(t)}{t}$ is a monotonely increasing function for $t < -1$. Then, by the fact that $\int_{-\infty}^{\phi} \Delta K_0(t)tdt \geq 0$, we obtain from (A.10),

$$\int_{-\infty}^{-1} p_0(t)\Delta K_0(t)dt \leq \frac{p_0(\omega_2)}{\omega_2}\int_{-\infty}^{-1} \Delta K_0(t)tdt.$$

Thus, it is enough to show that $\int_{-\infty}^{-1} \Delta K_0(t)tdt \geq 0$. But, this follows in view of the following equation:

$$0 = \int_{-\infty}^{0} \tilde{K}(t)tdt = \int_{-1}^{0} \tilde{K}(t)tdt + \int_{-\infty}^{-1} \Delta K_0(t)(t)tdt.$$

This completes the proof.

Proof of Lemma 4.1. From the proof of Theorem 4.1 above, we noticed that the only change of sign of $K_{0,0}(t)$ in $(-\infty, 0]$, at some point $\phi < 0$, goes from $-$ to $+$. Thus, by the second mean value theorem for integrals, we obtain

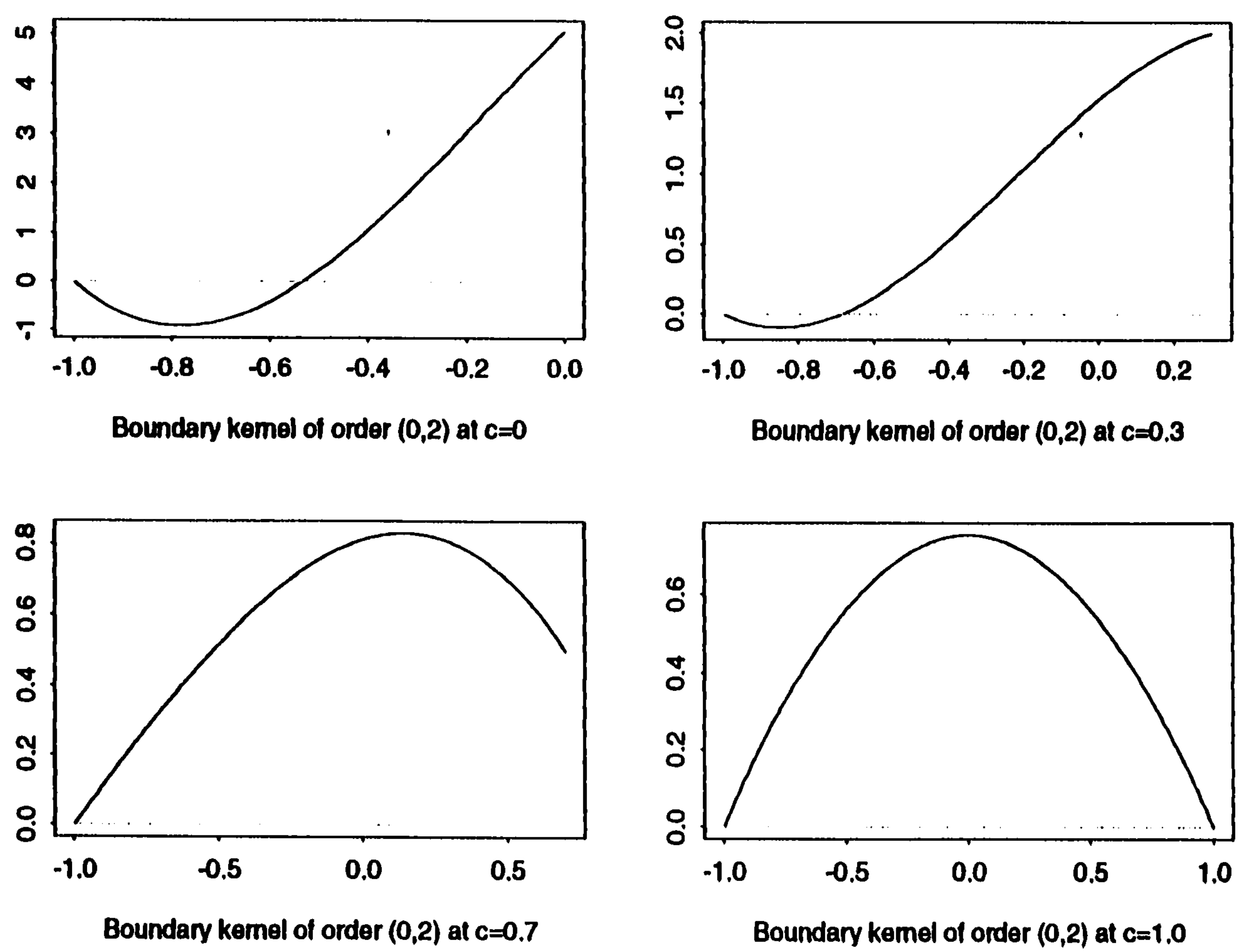
$$\begin{aligned}\int_{-\infty}^{0} K_{0,0}(t)t^2dt &= \int_{-\infty}^{\phi} K_{0,0}(t)t^2dt + \int_{\phi}^{0} K_{0,0}(t)t^2dt \\ &= v_1\int_{-\infty}^{\phi} K_{0,0}(t)tdt + v_2\int_{\phi}^{0} K_{0,0}(t)tdt \qquad \text{(A.11)}\end{aligned}$$

for some v_1 and v_2 such that $-\infty < v_1 < v_2 < 0$. But $\int_{-\infty}^{\phi} K_{0,0}(t)tdt > 0$ and $\int_{\phi}^{0} K_{0,0}(t)tdt < 0$. Thus, from (A.11),

$$\begin{aligned}\int_{-\infty}^{0} K_{0,0}(t)t^2dt &< v_2\left\{\int_{-\infty}^{\phi} K_{0,0}(t)tdt + \int_{\phi}^{0} K_{0,0}(t)tdt\right\} \\ &= v_2\int_{-\infty}^{0} K_{0,0}(t)tdt \\ &= 0.\end{aligned}$$

This completes the proof.

Figure 1. Equivalent boundary kernels of order (0,2).



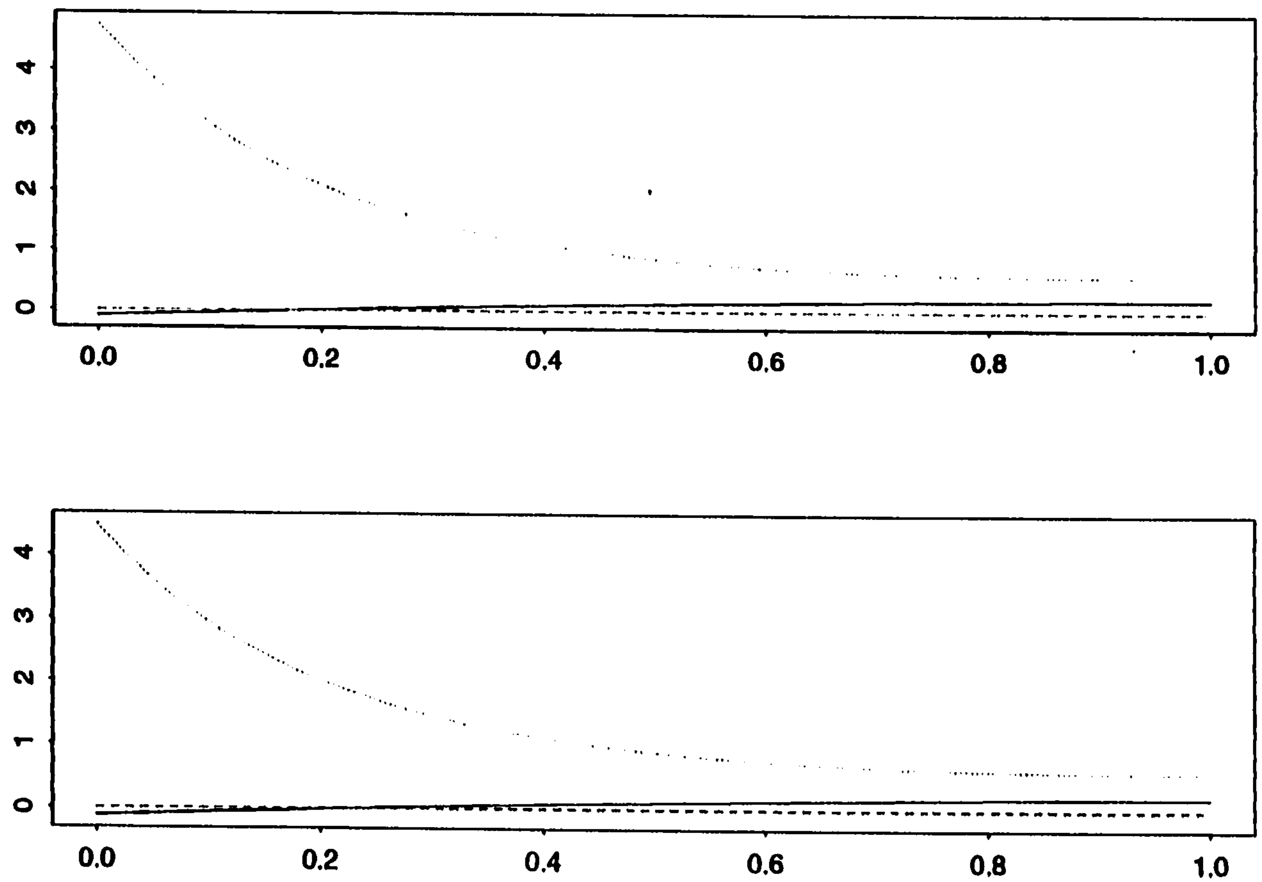


Figure 2. Top: Values of B(c) (bold curve) and V(c) (dotted curve);
Bottom: Values of B'(c) (bold line) and V'(c) (dotted curve). The dashed line is x=0.

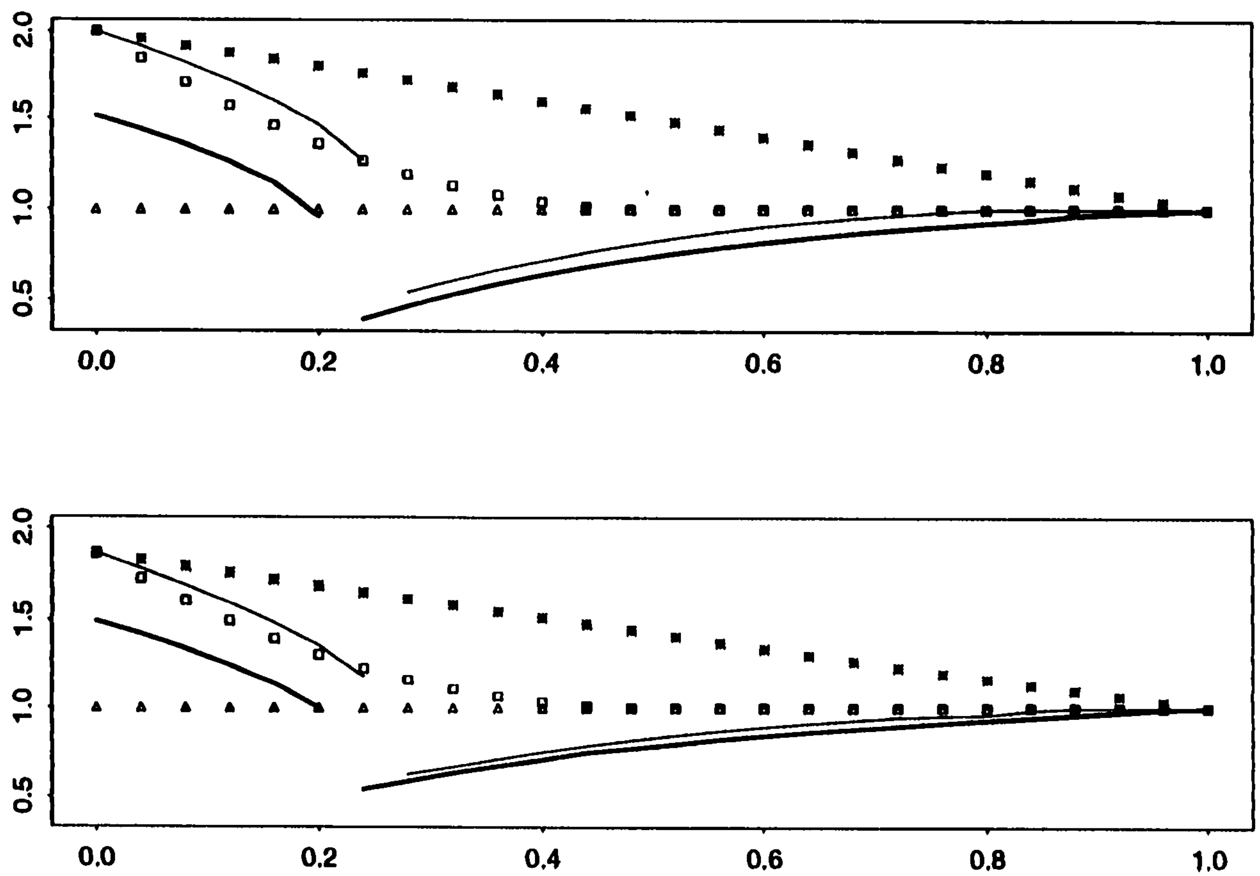


Figure 3. Bandwidth Variation functions for kernels (5.1) (top) and (5.2) (bottom).
b1=thin-bold curve; b2=thick-bold curve; b3=triangle line; b4=star line, b5=square curve, The x-axis is c values.

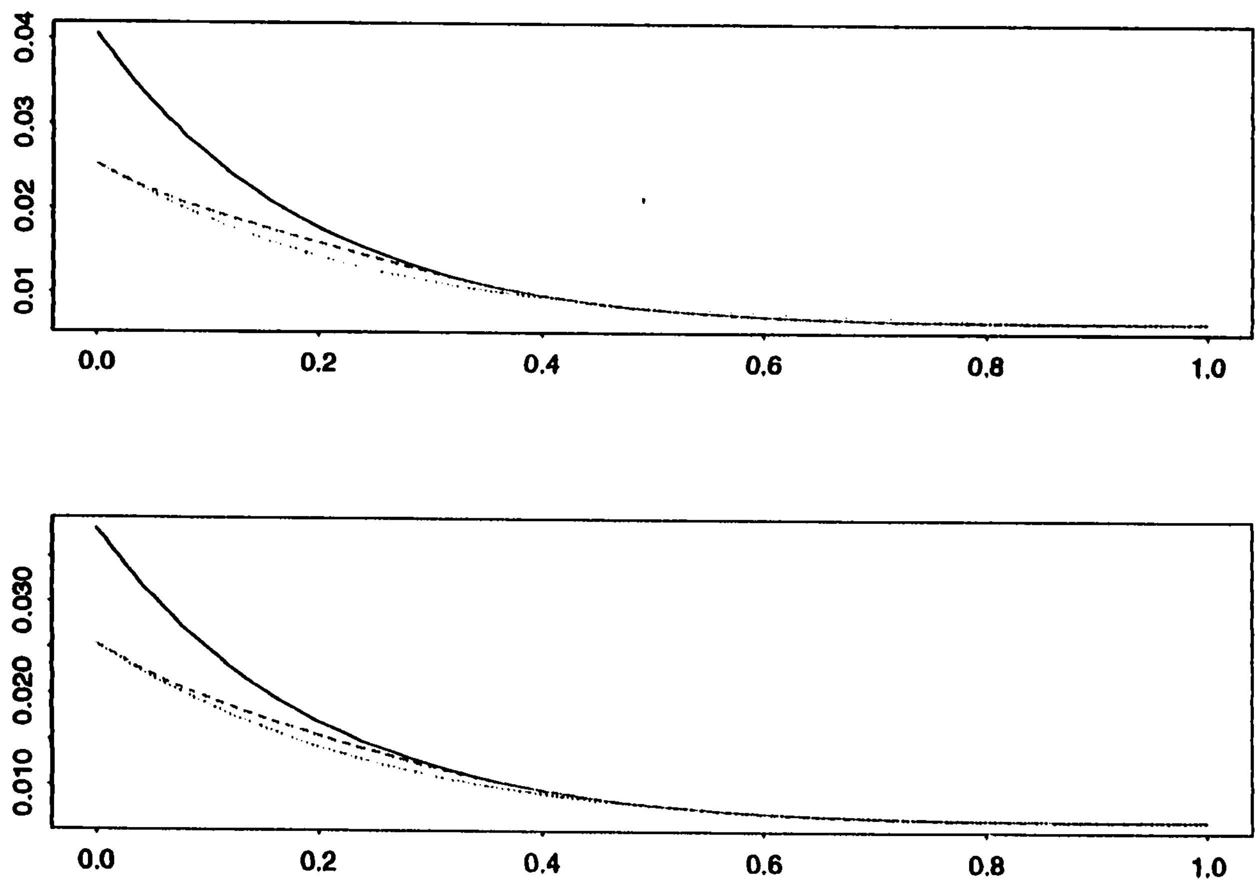


Figure 4. MSE values for kernels (5.1) (top) and (5.2) (bottom) with bandwidth variations b3 (bold curve), b4 (dotted curve) and b5 (dashed curve).

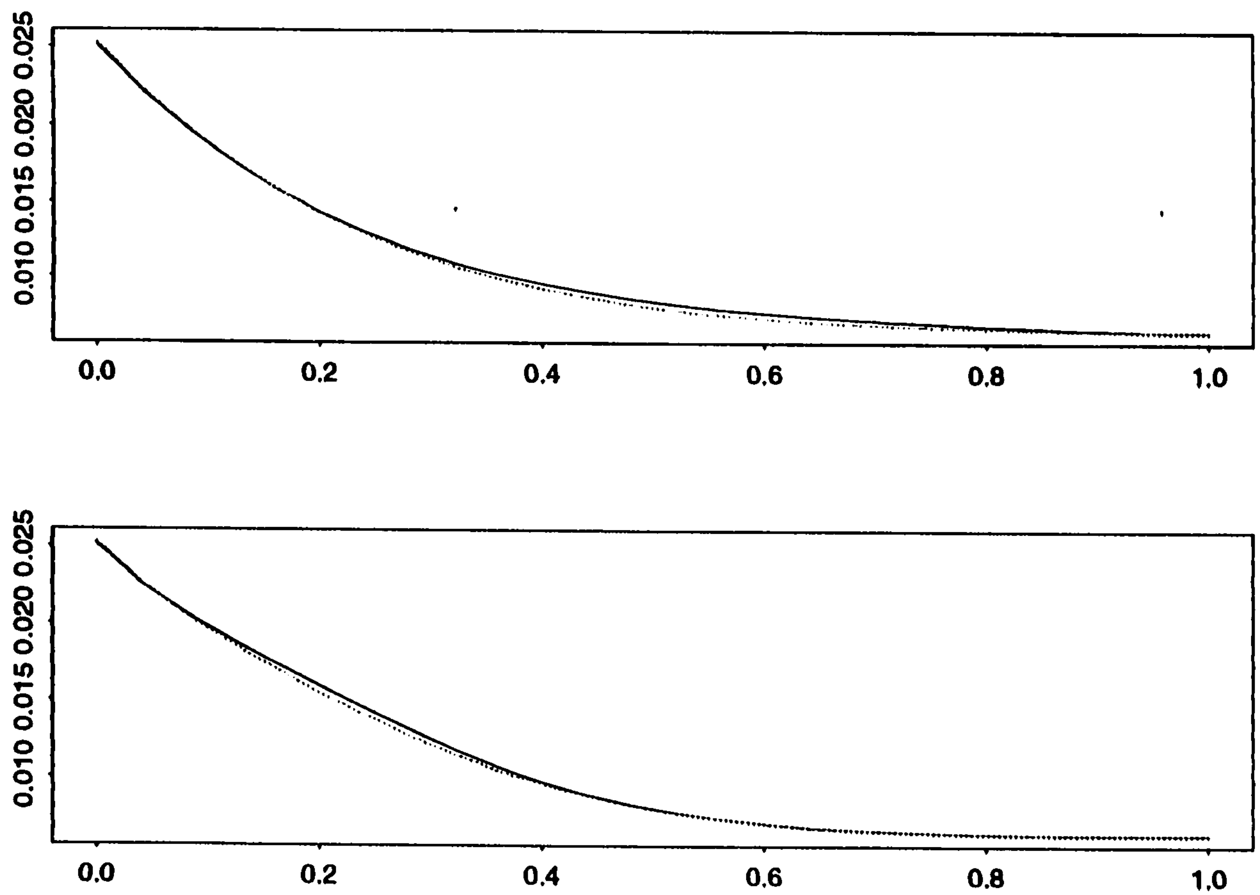


Figure 5. Top: MSE values for kernels (5.1) (dotted curve) and (5.2) (bold curve) with bandwidth variation b4.
Bottom: MSE values for kernels (5.1) (dotted curve) and (5.2) (bold curve) with bandwidth variation b5.

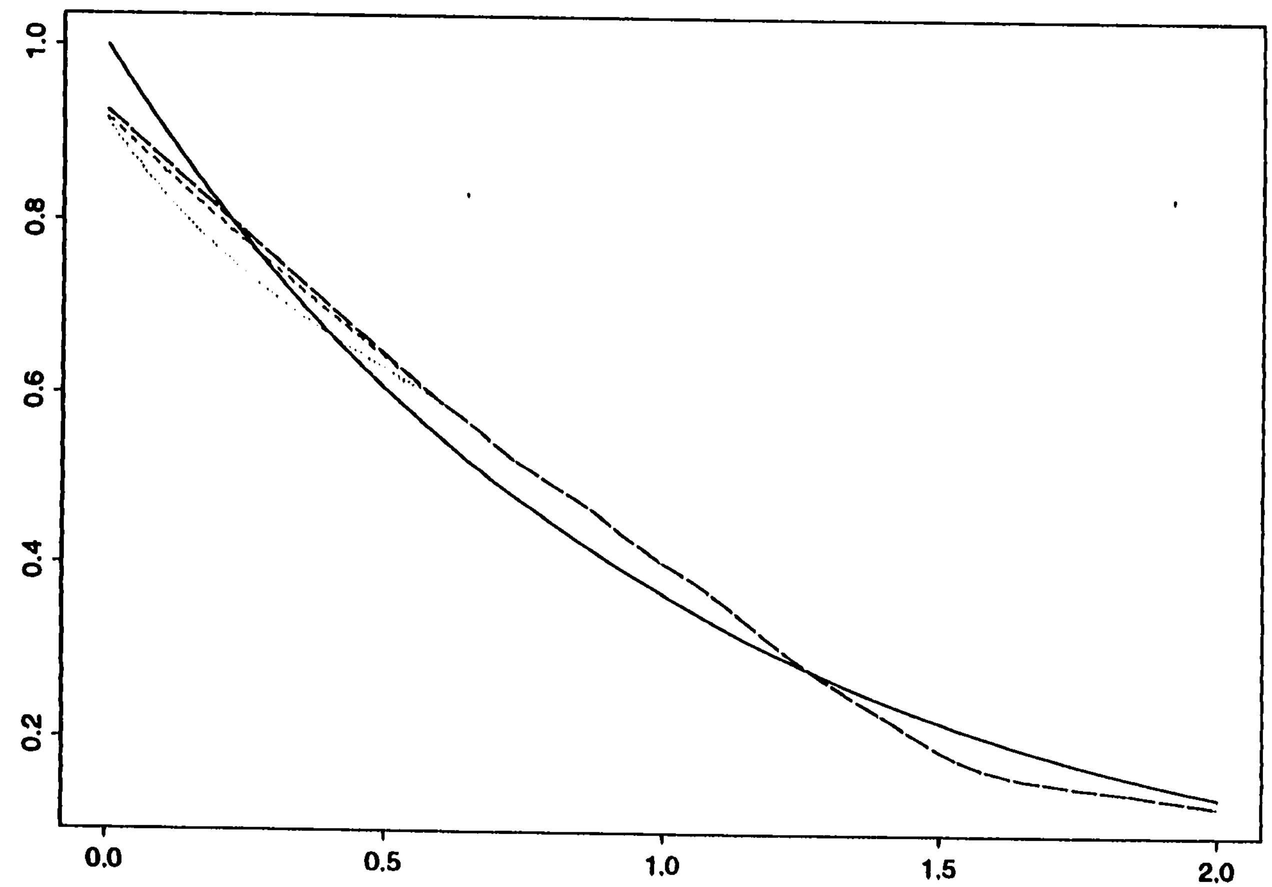


Figure 6. Plots of $f(x)=\exp(-x)$, $x>0$ (bold curve) estimated by Epanechnikov kernel (thin-light-dashed curve); kernel (5.1) (thick-bold-dashed curve); kernel (5.2) (thin-bold-dashed curve); and by Muller's kernel (dotted curve).

Figure 7. Cut-and-normalized boundary kernels of order (0,2).

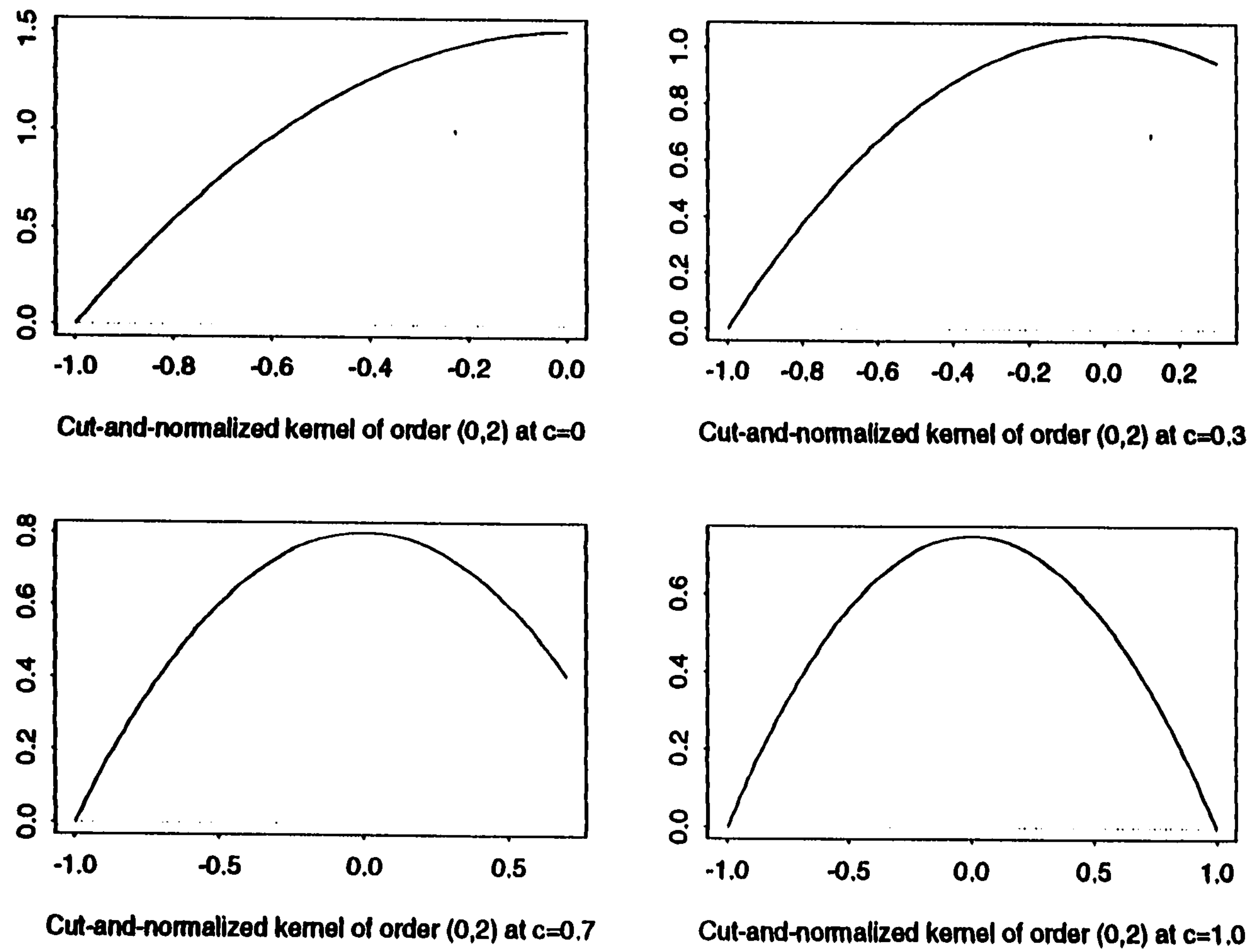


Figure 8. Reflection boundary kernels of order (0,2).

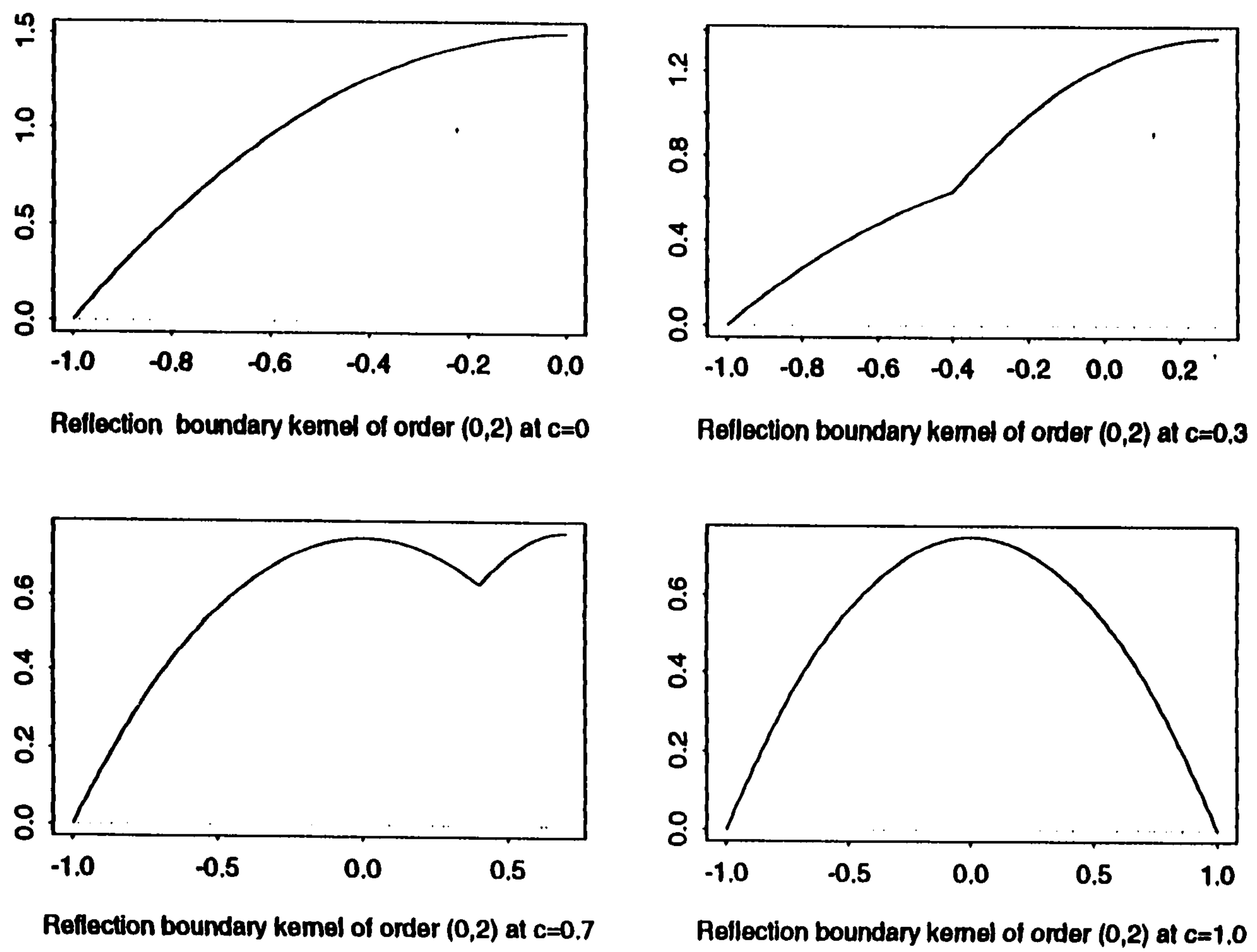
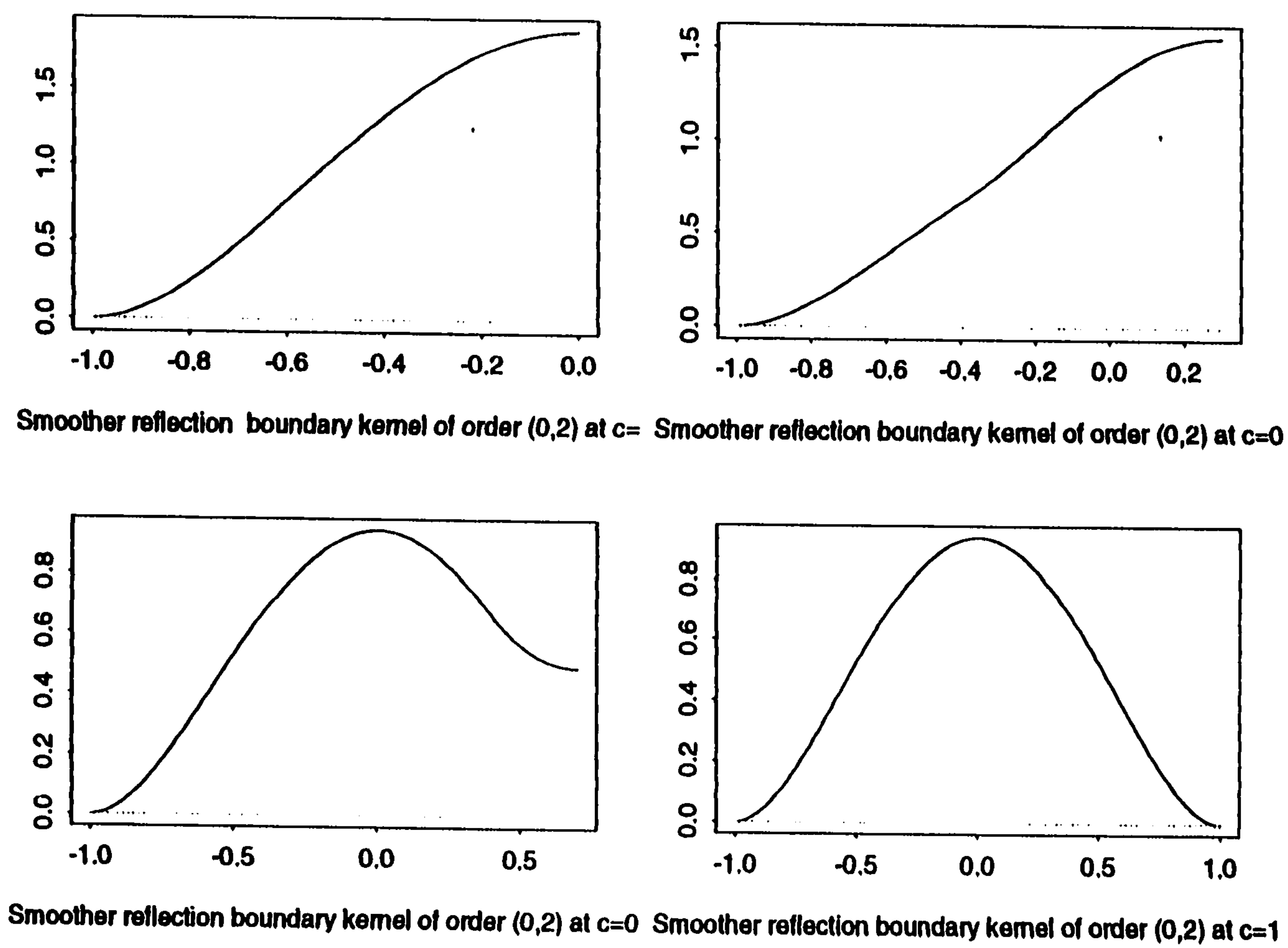


Figure 9. Smoother reflection boundary kernels of order (0,2).



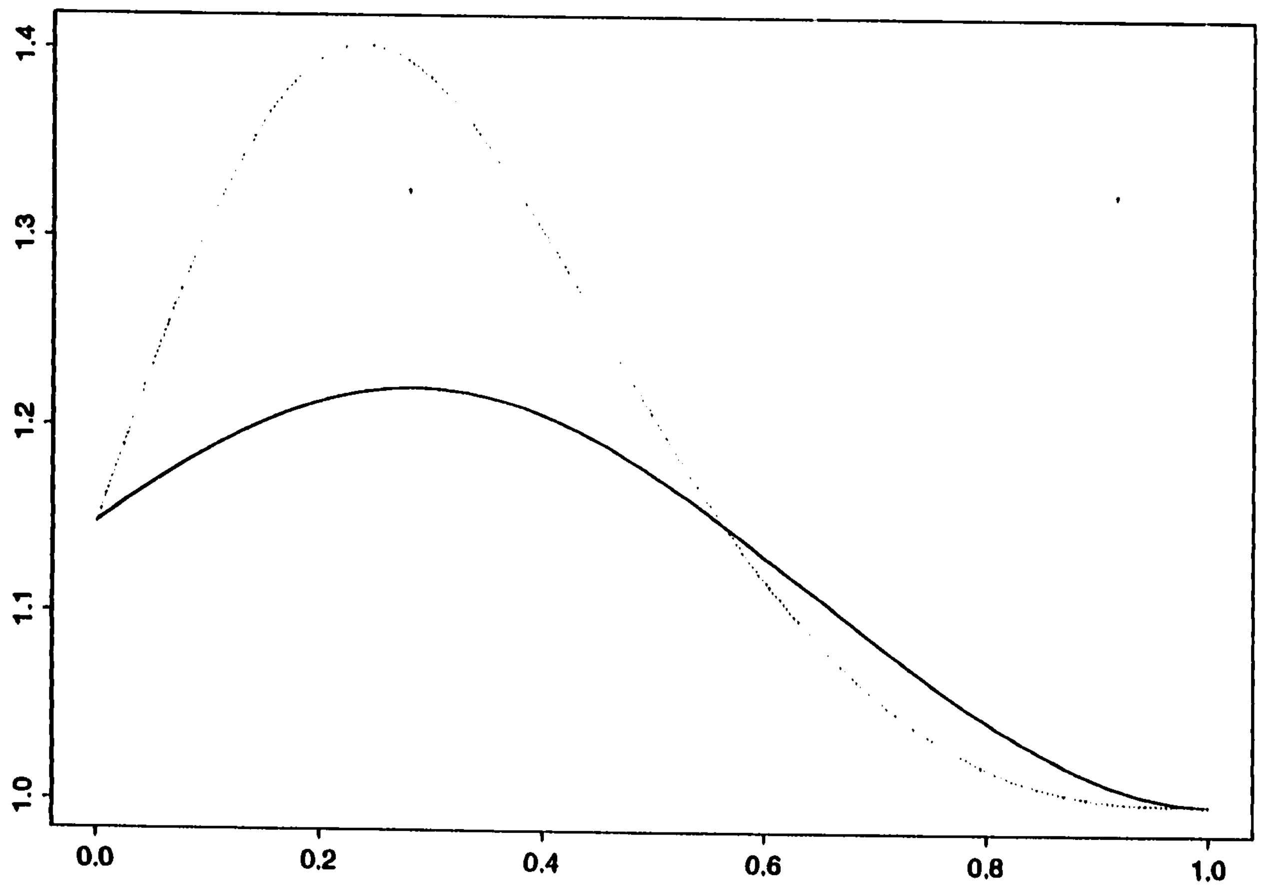


Figure 10. The optimal bandwidth variation function for cut-and-normalized kernel method-bold curve, and the optimal bandwidth variation function for the reflection method-dotted curve.

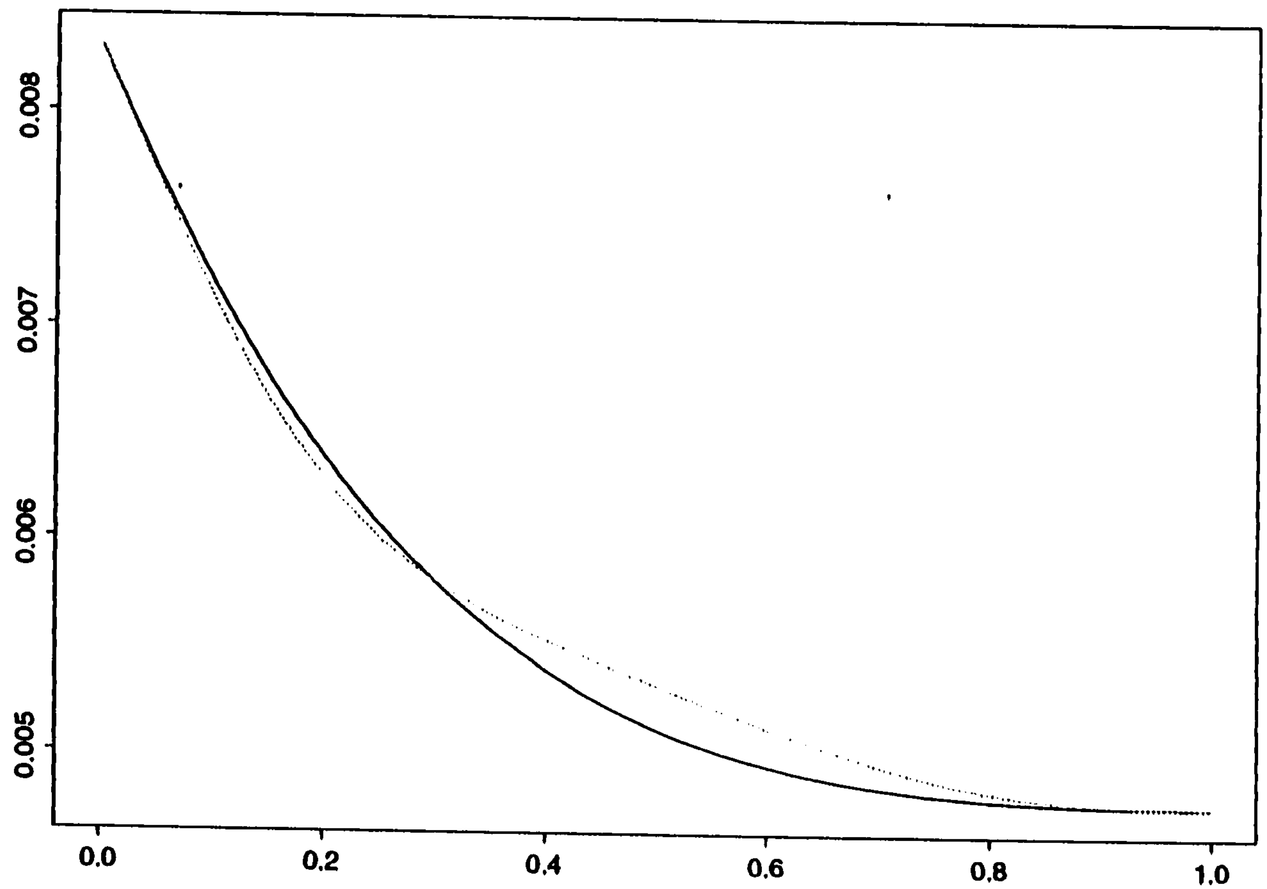


Figure 11. The pointwise MSE values from (6.5)-bold curve, with bandwidth variation (6.13); and the pointwise MSE values from (6.7)-dotted curve, with bandwidth variation (6.14).

Chapter 5

On Nonparametric Density Estimation at the Boundary

1. Introduction

Boundary effects are well known to occur in nonparametric density estimation when the support of the density has finite endpoints. Specifically, suppose that f_X is a probability density function with support $[0,1]$. Let $X_1, \ldots, X_n$ be a random sample from f_X. Then, the conventional kernel estimate of f_X is

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

where K is a kernel function of order $(0, 2)$ with support $[-1, 1]$ (i.e., $\int_{-1}^{1} K(t)dt = 0$, $\int_{-1}^{1} tK(t)dt = 0$ and $\int_{-1}^{1} t^2K(t)dt \neq 0$), and h is the bandwidth ($h \to 0$ as $n \to \infty$). Assuming that $f_X^{(2)}$, the second derivative of f_X, is continuous in a neighborhood of the left endpoint 0, we have for $x = ch$, $c \in [0,1]$,

$$\begin{aligned} Ef_n(x) &= f_X(x)\int_{-1}^{c} K(t)dt - hf_X^{(1)}(x)\int_{-1}^{c} tK(t)dt + \frac{h^2}{2}f_X^{(2)}(x)\int_{-1}^{c} t^2K(t)dt \\ &\quad + o(h^2). \end{aligned} \tag{1.2}$$

When $c = 1$, (1.2) gives the usual interior bias expansion. When $c \in [0, 1)$, $f_n(x)$ is not a consistent estimator of $f_X(x)$, since $\int_{-1}^{c} K(t)dt \neq 1$. This is known as the boundary effect at the left boundary. A similar phenominon occurs at the right boundary. To remove these boundary effects, a variety of methods have been

A version of this chapter has been submitted for publication.

developed during the past two decades. See, for instance, Gasser and Müller (1979), Rice (1984), Gasser et al. (1985), Schuster (1985), Silverman (1986), Cline and Hart (1991), Karunamuni and Mehra (1991), Müller (1991), Hall and Wehrly (1991), Jones (1993), Marron and Ruppert (1994) and Cowling and Hall (1996). In the context of the nonparametric regression estimation, Fan (1993) and Fan et al. (1993) revived the local polynomial fitting method. It has been shown that this method can automatically adapt to boundary effects. Zhang and Karunamuni (1995) extended this method to the case of density estimstion. They showed that the local polynomial fitting method in density estimation yields a class of boundary kernels. The idea of using boundary kernels to remove boundary effects first appeared in Gasser and Müller (1979). Since then, a variety of boundary kernels have been suggested. Most studied boundary kernels are the so-called smooth optimum kernels (Müller (1991)). The problem with these kernels is that they are not intuitive, since they put zero weights on the estimated point. Müller and Wang (1994) noticed this defect and derived another class of boundary kernels of order (0, 2). Later, Zhang and Karunamuni (1995) derived the same kernels by the local polynomial fitting method and showed that the kernel was in fact optimal among a class of boundary kernels in the sense of minimizing the mean squared error (MSE). The purpose of the present paper is to extend the results of Zhang and Karunamuni (1995) to the general case and to propose a new method to remove boundary effects. During the preparation of this paper, we learned that a more general result of Zhang and Karunamuni (1995), similar to the present paper, has been proved in a technical report of Cheng at al (1995), where they not only proved the optimality of the above stated order (0, 2) boundary kernel, but also obtained the optimal kernels of other orders as we have done in this paper.

However, there are number of notable differences between the results of our paper and the paper of Cheng et al. (1995). Here, we discuss the following important issues but not in theirs: (1) We propose a new and intuitive way to remove boundary effects by replacing the unwanted terms in the bias expansion by their estimators. Interestingly, our method offers a new way to construct boundary kernels. (2) We show that the class of boundary kernels derived from the local polynomial fitting method is a special case of ours. (3) We give an explicit geometric characterization of the class of kernels, which the optimality is based on, and that was obscured in Cheng et al. (1995). Also, an easy way to construct the optimal endpoint kernel is proposed. (An endpoint kernel is referred to as a kernel that is used to estimate a density at an endpoint of its support.) (4) We also discuss the important problem of choosing the optimal bandwidth when estimating a density at the boundary and provide an easy and general method to choose the bandwidth at the boundary.

Section 2 of this paper describes the method we propose to remove boundary effects. Section 3 gives some examples of the boundary kernels derived from the method of Section 2. Section 4 discusses the problem of how to choose the bandwidth variation function at the boundary region. In Section 5, the optimal endpoint kernels amomg a class of so-called *minimal* kernels are obtained. Some numerical results are given in Section 6.

2. Density estimation without boundary effects

From (1.2), it is clear that the term $-hf_X^{(1)}(x)\int_{-1}^{c} tK(t)dt$ is the main source of the boundary effect. Without this term, $f_n(x)/\int_{-1}^{c} K(t)dt$ will be a consistent estimator of $f(x)$. In fact, (1.2) implies that

$$E\left\{\frac{f_n(x)}{\int_{-1}^{c} K(t)dt}+\frac{\int_{-1}^{c} tK(t)dt}{\int_{-1}^{c} K(t)dt}hf_X^{(1)}(x)\right\} = f_X(x)+\frac{h^2}{2}\frac{f_X^{(2)}(x)\int_{-1}^{c} t^2K(t)dt}{\int_{-1}^{c} K(t)dt} + o(h^2). \qquad (2.1)$$

From (2.1), we note that a linear combination $\alpha f_n(x) + \beta h f_n^{(1)}(x)$ would serve as an estimator of $f_X(x)$, where α and β need to be determined and

$$f_n^{(1)}(x) = \frac{1}{nh^2} \sum_{i=1}^{n} K_1 \left(\frac{x - X_i}{h} \right) \tag{2.2}$$

is an estimate of $f_X^{(1)}(x)$ with K_1 being a kernel of order (1, k), $k \geq 2$, with support [-1, 1] (see (2.11) for the definition of order of kernels).

We now extend this idea to more general case. That is, consider the estimation problem of $f_X^{(v)}$, the v^{th} derivative of f_X, $v \geq 0$ (when $v = 0$, $f_X^{(0)} = f_X$). Assume that K_v is a kernel of order (v, k), $v \leq k - 1$, with support [-1, 1]. Then, the usual kernel estimate of $f_X^{(v)}(x)$ is

$$f_n^{(v)}(x) = \frac{1}{nh^{v+1}} \sum_{i=1}^{n} K_v \left(\frac{x - X_i}{h} \right). \tag{2.3}$$

Assume that $f_X^{(k)}$ is continuous in a neighborhood of $x = 0$. Then, for $x = ch$, $c \in$[0, 1], we have

$$\begin{aligned} Ef_n^{(v)}(x) &= \frac{1}{h^{v+1}} \int K_v \left(\frac{x - y}{h} \right) f_X(y) dy \\ &= \frac{1}{h^{v+1}} \{ f_X(x) \int_{-1}^{c} K_v(t) dt - h f_X^{(1)}(x) \int_{-1}^{c} t K_v(t) dt +, ..., \\ & \frac{(-h)^k}{k!} f_X^{(k)}(x) \int_{-1}^{c} t^k K_v(t) dt + o(h^k) \}. \end{aligned} \tag{2.4}$$

When $c < 1$, $f_n^{(v)}(x)$ is not consistent for $f_X^{(v)}(x)$. In order to remove this boundary effect, we suppose that a sequence of polynomials $K_0(t)$, $K_1(t)$, ..., $K_{k-1}(t)$ is available, where $K_i(t)$ is a kernel of order (i, k_i), $k_i \geq k$, i=0, 1, ..., $k - 1$. Then, $f_n^{(v)}(x)$ is a kernel estimator of the form (2.3), $v = 0$, 1, ..., $k - 1$. By repeating a calculation similar to (2.4), we obtain

$$\begin{aligned} Ef_n^{(0)}(x) &= f_X(x) \int_{-1}^{c} K_0(t) dt - h f_X^{(1)}(x) \int_{-1}^{c} t K_0(t) dt \\ & + \frac{(-h)^k}{k!} f_X^{(k)}(x) \int_{-1}^{c} t^k K_0(t) dt + o(h^k) \end{aligned}$$

$$
\begin{aligned}
Ef_n^{(1)}(x) &= \frac{1}{h}\{f_X(x)\int_{-1}^{c} K_1(t)dt - hf_X^{(1)}(x)\int_{-1}^{c} tK_1(t)dt \\
&\quad + \frac{(-h)^k}{k!} f_X^{(k)}(x)\int_{-1}^{c} t^k K_1(t)dt + o(h^k)\}
\end{aligned}
$$

$$\cdots\cdots$$

$$
\begin{aligned}
Ef_n^{(k-1)}(x) &= \frac{1}{h^{k-1}}\{f_X(x)\int_{-1}^{c} K_{k-1}(t)dt - hf_X^{(1)}(x)\int_{-1}^{c} tK_{k-1}(t)dt \\
&\quad + \frac{(-h)^k}{k!} f_X^{(k)}(x)\int_{-1}^{c} t^k K_{k-1}(t)dt + o(h^k)\}.
\end{aligned}
$$

Thus, the following linear combination (of $f_n^{(0)}(x)$, ..., $f_n^{(k-1)}(x)$) is defined as our estimator of $f_X^{(v)}(x)$:

$$
\hat{f}_n^{(v)}(x) = \sum_{i=0}^{k-1} a_i h^i f_n^{(i)}(x), \tag{2.5}
$$

where coefficients a_i ($i = 0$, ..., $k-1$) are to be determined. We now show that the estimator in (2.5) can be written in the usual form (2.3). Simple algebra yields

$$
\begin{aligned}
E\hat{f}_n^{(v)}(x) &= f_X(x)\left[\int_{-1}^{c}\sum_{i=0}^{k-1} a_i K_i(t)dt\right] - hf_X^{(1)}(x)\left[\int_{-1}^{c}\sum_{i=0}^{k-1} a_i tK_i(t)dt\right] \\
&\quad + \ldots + \frac{(-h)^k}{k!} f_X^{(k)}(x)\left[\int_{-1}^{c}\sum_{i=0}^{k-1} a_i t^k K_i(t)dt\right] + o(h^k).
\end{aligned}
$$

To remove the boundary effect, i.e., to cancel out the terms $f_X^{(i)}$, $0 \le i \le k-1$, $i \ne v$, we need

$$
\begin{cases}
a_0 \int_{-1}^{c} K_0(t)dt + \cdots + a_{k-1}\int_{-1}^{c} K_{k-1}(t)dt = 0 \\
a_0 \int_{-1}^{c} tK_0(t)dt + \cdots + a_{k-1}\int_{-1}^{c} tK_{k-1}(t)dt = 0 \\
\qquad\vdots \\
a_0 \int_{-1}^{c} t^v K_0(t)dt + \cdots + a_{k-1}\int_{-1}^{c} t^v K_{k-1}(t)dt = \frac{(-1)^v v!}{h^v} \\
\qquad\vdots \\
a_0 \int_{-1}^{c} t^{k-1} K_0(t)dt + \cdots + a_{k-1}\int_{-1}^{c} t^{k-1} K_{k-1}(t)dt = 0.
\end{cases}
\tag{2.6}
$$

Denote

$$S = \begin{pmatrix} \int_{-1}^{c} K_0(t)dt & \int_{-1}^{c} K_1(t)dt & \dots & \int_{-1}^{c} K_{k-1}(t)dt \\ \int_{-1}^{c} tK_0(t)dt & \int_{-1}^{c} tK_1(t)dt & \dots & \int_{-1}^{c} tK_{k-1}(t)dt \\ \vdots & \vdots & & \vdots \\ \int_{-1}^{c} t^{k-1}K_0(t)dt & \int_{-1}^{c} t^{k-1}K_1(t)dt & \dots & \int_{-1}^{c} t^{k-1}K_{k-1}(t)dt \end{pmatrix}^T.$$

Then, (2.6) can be rewritten as

$$S^T \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ (-1)^v v!/h^v \\ \vdots \\ 0 \end{pmatrix}. \tag{2.7}$$

LEMMA 2.1. The matrix S is regular.

The proof of Lemma 2.1 is given in the Appendix. From Lemma 2.1 and (2.7), we obtain

$$(a_0, a_1, ..., a_{k-1}) = (0, ..., (-1)^v v!/h^v, ..., 0)S^{-1}. \tag{2.8}$$

Now, from (2.5) and (2.8), one has

$$\begin{aligned} \hat{f}_n^{(v)}(x) &= (a_0, a_1, ..., a_{k-1}) \begin{pmatrix} f_n^{(0)}(x) \\ hf_n^{(1)}(x) \\ \dots \\ h^{k-1}f_n^{(k-1)}(x) \end{pmatrix} \qquad (2.9) \\ &= \frac{1}{nh^{v+1}} \sum_{i=1}^{n} (0, ..., (-1)^v v!, ..., 0)S^{-1} \begin{pmatrix} K_0\left(\frac{x-X_i}{h}\right) \\ K_1\left(\frac{x-X_i}{h}\right) \\ \vdots \\ K_{k-1}\left(\frac{x-X_i}{h}\right) \end{pmatrix} \\ &= \frac{1}{nh^{v+1}} \sum_{i=1}^{n} K_{v,c}\left(\frac{x-X_i}{h}\right), \end{aligned}$$

where

$$K_{v,c}(t) = (0, ..., (-1)^v v!, ..., 0)S^{-1} \begin{pmatrix} K_0(t) \\ K_1(t) \\ \vdots \\ K_{k-1}(t) \end{pmatrix}. \tag{2.10}$$

DEFINITION 2.1. A function $K_{v,c}(\cdot)$ with support $[-1, c]$, $b > 0$, $0 \leq c \leq 1$, is said to be a kernel of order $(v,\ k)$ if $K_{v,c}$ satisfies

$$\int_{-1}^{c} K_{v,c}(t)t^j dt = \begin{cases} 0 & j = 0,\ ...,\ v-1,\ v+1,\ ...,\ k-1 \\ (-1)^v v! & j = v \\ \neq 0 & j = k. \end{cases} \tag{2.11}$$

When $0 \leq c < 1$, $K_{v,c}$ is called a boundary kernel. When $c = 1$, $K_{v,1}$ is the usual interior kernel of order $(v,\ k)$.

LEMMA 2.2. The kernel $K_{v,c}$ defined by (2.10) is a kernel of order $(v,\ k)$ for $0 \leq c \leq 1$. When $c = 1$, $K_{v,c} \equiv K_v$, where K_v is a $(v,\ k)^{th}$ order interior kernel.

Proof. Write

$$\left(\int_{-1}^{c} K_{v,c}(t)dt, \int_{-1}^{c} tK_{v,c}(t)dt, ..., \int_{-1}^{c} t^{k-1}K_{v,c}(t)dt\right) = (0, ..., (-1)^v v!, ..., 0)S^{-1}$$

$$\begin{pmatrix} \int_{-1}^{c} K_0(t)dt & \int_{-1}^{c} K_1(t)dt & \cdots & \int_{-1}^{c} K_{k-1}(t)dt \\ \int_{-1}^{c} tK_0(t)dt & \int_{-1}^{c} tK_1(t)dt & \cdots & \int_{-1}^{c} tK_{k-1}(t)dt \\ \vdots & \vdots & & \vdots \\ \int_{-1}^{c} t^{k-1}K_0(t)dt & \int_{-1}^{c} t^{k-1}K_1(t)dt & \cdots & \int_{-1}^{c} t^{k-1}K_{k-1}(t)dt \end{pmatrix}^T$$

$$= (0, ..., (-1)^v v!, ..., 0).$$

Now $K_{v,1}(t) = K_v(t)$ is obvious from (2.10).

One important application of Lemma 2.2 is that the estimator $\hat{f}_n^{(v)}$ defined by (2.9) will automatically adapt to boundary effects. Further, (2.10) offers a way to construct boundary kernels. Examples of such kernels are given in the next section. Since boundary kernels $K_{v,c}$ defined by (2.10) are formed by taking a linear combination of different order kernels, we shall call them "*Combined boundary kernels.*"

3. Combined boundary kernels

The purpose of this section is to study more about boundary kernels defined by (2.10) and give a few examples of combined boundary kernels.

Assume that functions K_v $(0 \le v \le k-1)$ in (2.10) are polynomials of degree k with support [-1, 1]; that is, $K_v(t) = \sum_{i=0}^{k} \lambda_{i,v} t^i$. Since by assumption K_v is of order $(v,\ k)$, K_v must satisfy

$$\int_{-1}^{1} K_v(t) t^j dt = \begin{cases} 0 & j = 0,\ \ldots,\ v-1,\ v+1,\ \ldots,\ k-1 \\ (-1)^v v! & j = v. \end{cases} \tag{3.1}$$

Now (3.1) together with $K_v(-1) = 0$ lead to a system of $k-1$ linear equations for the coefficients $\lambda_{0,v}, \ldots, \lambda_{k,v}$:

$$\begin{pmatrix} 2 & 0 & \frac{2}{3} & \ldots & \frac{1-(-1)^{k+1}}{k+1} \\ 0 & \frac{2}{3} & 0 & \ldots & \frac{1-(-1)^{k+2}}{k+2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & -1 & 1 & \ldots & (-1)^k \end{pmatrix} \begin{pmatrix} \lambda_{0,v} \\ \lambda_{1,v} \\ \vdots \\ \lambda_{k,v} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ (-1)^v v! \\ \vdots \\ 0 \end{pmatrix}. \tag{3.2}$$

Denote

$$S^+ = \begin{pmatrix} 2 & 0 & \frac{2}{3} & \ldots & \frac{1-(-1)^{k+1}}{k+1} \\ 0 & \frac{2}{3} & 0 & \ldots & \frac{1-(-1)^{k+2}}{k+2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & -1 & 1 & \ldots & (-1)^k \end{pmatrix}^T_{(k+1)\times(k+1)}.$$

Then, it is easy to show that S^+ is regular. Thus, from (3.2),

$$(\lambda_{0,v}, \ldots, \lambda_{k,v}) = (0, \ldots, (-1)^v v!, 0, \ldots, 0)(S^+)^{-1},$$

and hence

$$\begin{pmatrix} K_0(t) \\ K_1(t) \\ \vdots \\ K_{k-1}(t) \end{pmatrix} = V(S^+)^{-1} \begin{pmatrix} 1 \\ t \\ \vdots \\ t^k \end{pmatrix}, \tag{3.3}$$

where

$$V = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & (-1)^{k-1}(k-1)! & 0 \end{pmatrix}_{k\times(k+1)}.$$

LEMMA 3.1. With $K_0, \ldots, K_{k-1}$ given by (3.3), we have

$$\begin{aligned} K_{v,c}(t) &= (0, \ldots, (-1)^v v!, 0, \ldots, 0) \\ &\quad \begin{pmatrix} c+1 & \frac{c^2-1}{2} & \ldots & 1 \\ \frac{c^2-1}{2} & \frac{c^3-1}{3} & \ldots & -1 \\ \vdots & \vdots & & \vdots \\ \frac{c^{k+1}-(-1)^{k+1}}{k+1} & \frac{c^{k+2}-(-1)^{k+2}}{k+2} & \ldots & (-1)^k \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ t \\ \vdots \\ t^k \end{pmatrix} \\ &= (0, \ldots, (-1)^v v!, 0, \ldots, 0)(S^*)^{-1} \begin{pmatrix} 1 \\ t \\ \vdots \\ t^k \end{pmatrix}. \end{aligned} \tag{3.4}$$

The proof of Lemma 3.1 is given in the Appendix. The regularity of S^+ can be proved along the same lines of Theorem 5 of Gasser and Müller (1979). Lemma

3.1 implies that $K_{v,c}$ is a polynomial of degree k satisfying

$$\int_{-1}^{c} K_{v,c}(t)t^j dt = \begin{cases} 0 & j = 0,\ \ldots,\ v-1,\ v+1,\ \ldots,\ k-1 \\ (-1)^v v! & j = v. \end{cases} \tag{3.5}$$

and $K_{v,c}(-1) = 0$, i.e., $K_{v,c}$ is of order $(v,\ k)$.

We now give a few examples of combined kernels with $K_0,\ \ldots,\ K_{k-1}$ given by (3.3).

EXAMPLE 1. *Order (0, 2).* Let $K_0(t) = \frac{3}{4}(1-t^2)$ and $K_1(t) = \frac{3}{4}(3t^2+2t-1)$. Then, the combined boundary kernel of order (0, 2) is

$$K_{0,c}(t) = \frac{12(1+t)}{(1+c)^4}[(1-2c)t + \frac{3c^2-2c+1}{2}], \tag{3.6}$$

see Figure 1. At $c = 0$, (3.5) becomes $K_{0,0}(t) = 12t^2 + 18t + 6$. We call it an endpoint kernel.

Figure 1 about here

EXAMPLE 2. *Order (0, 4).* Let $K_0(t) = \frac{15}{32}(3 - 10t^2 + 7t^4)$, $K_1(t) = \frac{15}{32}(3 - 20t - 30t^2 + 28t^3 + 35t^4)$, $K_2(t) = \frac{15}{16}(-21 + 84t + 210t^2 - 140t^3 - 245t^4)$. Then, the combined boundary kernel of order (0, 4) is

$$\begin{aligned} K_{0,c}(t) = \ & \frac{5}{4(1+c)^8}[3(69 - 828c + 2292c^2 - 2676c^3 + 1525c^4 - 400c^5 + 50c^6) \\ & (3 - 10t^2 + 7t^4) - 5(c-1)^2(31 - 310c + 333c^2 - 100c^3 + 10c^4) \\ & (3 - 20t - 30t^2 + 28t^3 + 35t^4) + 35(c-1)^2 \\ & (25 - 250c + 255c^2 - 76c^3 + 10c^4)(-1 + 6t^2 - 5t^4) \\ & -5(c-1)^2(7 - 70c + 69c^2 - 20c^3 + 2c^4) \\ & (-21 + 84t + 210t^2 - 140t^3 - 245t^4)] \end{aligned} \tag{3.7}$$

see Figure 2. The corresponding endpoint kernel is $K_{0,0}(t) = 20(1 + 10t + 30t^2 + 35t^3 + 14t^4)$.

Figure 2 about here

EXAMPLE 3. *Order (1, 3)*. Let $K_0(t) = \frac{3}{8}(3 + 3t - 5t^2 - 5t^3)$, $K_1(t) = \frac{15}{4}(t^3 - t)$, and $K_2(t) = \frac{15}{4}(5t^3 + 3t^2 - 3t - 1)$. Then, the combined boundary kernel of order (1, 3) is

$$\begin{aligned} K_{1,c}(t) &= \frac{30}{(1+c)^6}[-3(3-2c)(c-1)^2(3+3t-5t^2-5t^3) \\ &\quad +4(16-39c+30c^2-5c^3)(t^3-t) \\ &\quad -5(5-2c)(c-1)^2(5t^3+3t^2-3t-1)], \end{aligned} \tag{3.8}$$

see Figure 3. The corresponding endpoint kernel is $K_{1,0}(t) = -60(1 + 8t + 15t^2 + 8t^3)$.

Figure 3 about here

The endpoint kernel described in Example 1 above possesses an important property- it is optimal in the sense of minimizing the mean squared error in the class of all kernels of order (0, 2) with exact one change of sign in their support (Zhang and Karunamuni (1995)). In Section 5, we shall prove that the endpoint kernels in Example 2 and 3 are also optimal.

Zhang and Karunamuni (1995) constructed a class of boundary kernels for estimating density functions based on the local polynomial fitting method developed by Fan (1992) and Fan et al. (1993) in the context of nonparametric regression function estimation. See, Cheng et al. (1995) for a different way to use the local polynomial fitting method in estimation of a density function. The boundary kernels of order (v, k) derived in Zhang and Karunamuni (1995) are of the form

$$K^*_{v,c}(t) = (0, \ \ldots, \ (-1)^v v!, \ 0 \ , \ldots, \ 0)M^{-1}(1, \ t, \ \ldots, \ t^{k-1})^T K(t), \qquad (3.9)$$

where K is a nonnegative weight function and $M = (m_{j,l})$ is a matrix with $m_{j,l} = \int_{-1}^{c} t^{j+l} K(t)dt$, $0 \leq j, \ l \leq k-1$. When $c = 1$, $K^*_{v,1}$ is the interior kernel of order (v, k). If K is chosen as $\frac{3}{4}(1-t^2)I_{[-1, \ 1]}$, $K^*_{v,1}$ is the optimal kernel of order (v, k) among the class of minimal kernels. See Gasser et al. (1985) and Zhang and Karunamuni (1995) for more details. The next result shows that (3.9) is a special case of (2.10).

THEOREM 3.1. If $K_i(t)$, $(i = 0, \ \ldots, \ k-1)$ in (2.10) are given by $K^*_{i,1}(t)$ $(i = 0, \ \ldots, \ k-1)$, where $K^*_{i,1}$ $(i = 0, \ \ldots, \ k-1)$ are defined by (3.9), then $K_{v,c}(t) = K^*_{v,c}$ for $0 \leq v \leq k-1$ and $0 \leq c \leq 1$, where $K_{v,c}$ is given by (2.10).

4. Boundary adaptive bandwidth

Although the boundary kernels derived in the previous section automatically adapt to boundary effects, the choice of the bandwidth at the boundary regions plays an important role in determining the performance of the estimator; see Müller (1991) and Zhang and Karunamuni (1995). In this section, we discuss various bandwidth variation functions suitable for applications.

For any boundary kernel $K^+_{v,c} \in L_2 = \{f : \ \int f^2 dx < \infty\}$ of order (v, k), write $B_{v,k}(c) = \int_{-\infty}^{c} K^+_{v,c}(t)t^k dt$ and $V_{v,k}(c) = \int_{-\infty}^{c} \left(K^+_{v,c}(t)\right)^2 dt$. For $x = ch$, $0 \leq c \leq 1$, assume that a local bandwidth $h(x) = b(x/h)h$ is used, where h will be specified later in this section and b: $[0, \ 1] \to R^+$ is a bandwidth variation function such that $b(1) = 1$. Then, the leading terms of the asymptotic MSE of

$f_n^{+(v)}(x) = \frac{1}{nh^{v+1}} \sum_{i=1}^n K_{v,c}^+ \left(\frac{x-X_i}{h}\right)$ at $x = ch$ is

$$\text{MSE}(f_n^{+(v)}, x) = \frac{1}{k!^2}[f^{(k)}(0)]^2[hb(c)]^{2(k-v)} \left[B_{v,k}\left(\frac{c}{b(c)}\right)\right]^2 + \frac{f(0)}{n[hb(c)]^{2v+1}} V_{v,k}\left(\frac{c}{b(c)}\right)(1 + o(1)). \tag{4.1}$$

Assume that the optimal bandwidth (say h_1) is chosen at $x = h_1$. Then, h_1 minimizes (4.1) for $c = 1$. Differentiating the expression (4.1) w.r.t. h, we obtain

$$h_1 = \left\{\frac{(k!)^2(2v+1)f(0)V_{v,k}(1)}{2n(k-v)f^{(k)}(0)[B_{v,k}(1)]^2}\right\}^{\frac{1}{2k+1}}. \tag{4.2}$$

The corresponding MSE for $h = h_1$ is

$$\begin{aligned}\text{MSE}(f_n^{+(v)}, x) \quad = \quad & \frac{[2(k-v)]^{\frac{2v+1}{2k+1}} f(0)^{\frac{2(k-v)}{2k+1}} [f^{(k)}(0)B_{v,k}(1)]^{\frac{2(2v+1)}{2k+1}}}{(k!)^{\frac{2(2v+1)}{2k+1}}[(2v+1)V_{v,k}(1)]^{\frac{2v+1}{2k+1}}} \left\{\frac{V_{v,k}\left(\frac{c}{b(c)}\right)}{b(c)^{2v+1}} + \right. \\ & \left. \frac{(2v+1)V_{v,k}(1)}{2(k-v)[B_{v,k}(1)]^2} b(c)^{2(k-v)} \left[B_{v,k}\left(\frac{c}{b(c)}\right)\right]^2\right\} n^{-\frac{2(k-v)}{2k+1}}. \end{aligned} \tag{4.3}$$

Then, the optimal bandwidth variation function $b(c)$ is the solution of the variational problem of minimizing

$$\frac{V_{v,k}\left(\frac{c}{b(c)}\right)}{b(c)^{2v+1}} + \frac{(2v+1)V_{v,k}(1)}{2(k-v)[B_{v,k}(1)]^2} b(c)^{2(k-v)} \left[B_{v,k}\left(\frac{c}{b(c)}\right)\right]^2, \tag{4.4}$$

under the requirement that $b(y) \geq y$. Generally, the minimization problem (4.4) does not yield a solution which is a smooth function in y. A variety of methods have been proposed to remedy this situation and to obtain suboptimal solutions. For more details, see Müller (1991) and Zhang and Karunanuni (1995).

Assuming that $B_{v,k}(0) \neq 0$ it is natural to require that the bandwidth is chosen optimally at $x = 0$. Let h_0 denote the optimal value of h at $x = 0$. Then

$$h_0 = \left\{\frac{(k!)^2(2v+1)f(0)V_{v,k}(0)}{2n(k-v)f^{(k)}(0)[B_{v,k}(0)]^2}\right\}^{\frac{1}{2k+1}}.$$

The preceding expression together with (4.2) give

$$\frac{h_0}{h_1} = \left\{ \frac{[B_{v,k}(1)]^2 V_{v,k}(0)}{[B_{v,k}(0)]^2 V_{v,k}(1)} \right\}^{\frac{1}{2k+1}}.$$

Then, a suboptimal choice of $b(c)$ would be

$$b_1(c) = 1 - (c-1)\left(\left\{ \frac{[B_{v,k}(1)]^2 V_{v,k}(0)}{[B_{v,k}(0)]^2 V_{v,k}(1)} \right\}^{\frac{1}{2k+1}} - 1 \right). \tag{4.5}$$

Another reasonable choice of $b(c)$ can be obtained as follows: Substitute $V_{v,k}\left(\frac{c}{b(c)}\right)$ and $B_{v,k}\left(\frac{c}{b(c)}\right)$ in (4.4) by $V_{v,k}(y)$ and $B_{v,k}(1) - (\mathrm{Sign}(B_{v,k}(1))|B_{v,k}(0)| - B_{v,k}(1))(y-1)$, respectively, $0 \le y \le 1$. Then, minimize

$$\begin{aligned} &\frac{V_{v,k}(y)}{b(c)^{2v+1}} + \frac{(2v+1)V_{v,k}(1)}{2(k-v)[B_{v,k}(1)]^2} b(y)^{2(k-v)} [B_{v,k}(1) \\ &\quad -(\mathrm{Sign}(B_{v,k}(1))|B_{v,k}(0)| - B_{v,k}(1))(y-1)]^2 \end{aligned} \tag{4.6}$$

w.r.t. $b(y)$. The resulting minimizer is

$$b_2(c) = \left\{ \frac{[B_{v,k}(1)]^2 V_{v,k}(c)}{V_{v,k}(1)\,[B_{v,k}(1) - (\mathrm{Sign}(B_{v,k}(1))|B_{v,k}(0)| - B_{v,k}(1))(y-1)]^2} \right\}^{\frac{1}{2k+1}} . \tag{4.7}$$

Note that (4.6) is the optimal choice of $b(\cdot)$ at $x = 0$ and 1.

5. The optimal endpoint kernel

In this section, we shall restrict our attention to estimation of a density at the left end-point, i.e., $f(0)$. The right endpoint can be treated similarly.

With the optimal choice of the bandwidth at $x = 0$, i.e., h_0, from (4.3), we obtain

$$\begin{aligned} \mathrm{MSE}(f_n^{+(v)}, 0) &= \frac{(2k+1) f(0)^{\frac{2(k-v)}{2k+1}} [f^{(k)}(0) B_{v,k}(0)]^{\frac{2(2v+1)}{2k+1}} [V_{v,k}(0)]^{\frac{2(k-v)}{2k+1}}}{(k!)^{\frac{2(2v+1)}{2k+1}} (2v+1)^{\frac{2v+1}{2k+1}} [2(k-v)]^{\frac{2(k-v)}{2k+1}}} \\ &\quad \times n^{-\frac{2(k-v)}{2k+1}}. \end{aligned} \tag{5.1}$$

For definition of $f_n^{+(v)}$, see circa (4.1). Thus, to minimize (5.1) it is enough to minimize

$$T_0(K_{v,0}) = |B_{v,k}(0)|^{2v+1}(V_{v,k}(0))^{k-v}.$$

For any two functions f, $g \in L_2([a,\ b])$, denote the inner product of f, g by

$$< f, g >= \int_a^b f(t)g(t)dt, \quad -\infty \leq a \leq b \leq \infty.$$

Then,

$$T_0(K_{v,0}) = |< K_{v,0}, t^k >|^{2v+1} < K_{v,0}, K_{v,0} >^{k-v}. \tag{5.2}$$

Then, we have the following lemma.

LEMMA 5.1. Define a mapping H_λ: $L_2 \to L_2$ such that $H_\lambda f(\cdot) = (\lambda^{v+1})^{-1} f(\cdot/\lambda)$. Then, $T_0(K_{v,0}) = T_0(H_\lambda K_{v,0})$ for all $\lambda > 0$. (That is, T_0 is invariant under scale transforms.)

The proof of Lemma 5.1 is immediate from (5.2). Also, see Granovosky and Müller (1989). Lemma 5.1 implies that the support of $K_{v,0}$ or $B_{v,k}(0)$ can be normalized without affecting the solution. This makes it possible to choose an optimal kernel for endpoint estimation. Since $B_{v,k}(0)$ may be zero, the problem of minimizing (5.2) could be degenerate. To get around this difficulty, we restrict our attention to the following special class of kernels.

DEFINITION 5.1. An endpoint kernel of order $(v,\ k)$ is said to be *minimal* if it has $(k-1)$ changes of sign on its support. We shall denote this class of kernels by $\mathcal{N}_{k-1}$.

LEMMA 5.2. An endpoint kernel of order $(v,\ k)$ has at least $(k-1)$ changes of sign on its support.

The proof of Lemma 5.2 is given in the Appendix. Lemma 5.2 explains that a minimal endpoint kernel of order $(v,\ k)$ has to be of degree $(k-1)$ polynomial with at least $(k-1)$ real roots on its support. If the support is defined by the outmost roots of the polynomial, then, it has to be of degree at least k with k real roots on its support.

LEMMA 5.3. (i) For the minimal endpoint kernel $K^{+}_{v,0}$ of order $(v,\ k)$ with support $[-\tau,\ 0]$, $\int_{-\tau}^{0} K^{+}_{v,0}(t)t^k dt \neq 0$.
(ii) For the minimal endpoint kernel with its support defined by its outmost root, its coefficients are all positive (or negative) according to v is even (or odd), respectively.

The proof of Lemma 5.3 is given in the Appendix as well.

THEOREM 5.1. Assume that $K^{+}_{v,0}(t)$ is a polynomial of degree k satisfying

$$\int_{-1}^{0} K^{+}_{v,0}(t)t^j dt = \begin{cases} 0 & j = 0,\ \ldots,\ v-1,\ v+1,\ \ldots,\ k-1 \\ (-1)^v v! & j = v \end{cases} \tag{5.3}$$

and

$$K^{+}_{v,0}(-1) = 0. \tag{5.4}$$

Then, $K^{+}_{v,0}(t)$ is the optimal kernel in the sense that it minimizes $T_0(K_{v,0})$ w.r.t. the class of minimal endpoint kernels.

Cheng et al. (1995) showed that the kernel (3.9) with $K(t) = (1+t)I_{[-1,\ 0]}(t)$ is optimal in the sense of minimizing MSE among all nonnegative K. To prove

Theorem 5.1, it is enough to prove

(i) $K^{+}_{v,0}(t)$ defined by (5.3) and (5.4) is the same as $K^{*}_{v,0}$ defined by (3.9) with $K(t) = (1-t)I_{[-1,\ 0]}$.

(ii) $\mathcal{N}_{k-1} = \{K^{*}_{v,0}(t),$ for all $K \geq 0\}$.

The proof of Theorem 5.1 is given in the Appendix. Note that (3.4) provides the solution to the equations (5.3) and (5.4). The endpoint kernels defined by (3.7) and (3.8) are optimal.

6. Numerical results

In this section, we plan to demonstrate the importance of choosing a bandwidth variation function at the boundary by using the kernels defined by (3.7) and (3.8), which are of orders (0, 4) and (1, 3), respectively, and are optimal at the endpoint. The case when the kernel is of order (0, 2) has been discussed by Zhang and Karunamuni (1995). We assume that $f_X(x) = e^{-x}I_{(0,\ \infty)}(x)$ throughout the simulation. For simplicity of comparison, we shall use the true density values whenever they appear.

For the kernel defined by (3.7), we obtain (see Section 4)

$$\begin{aligned} B_{0,4}(c) &= \frac{-1+20c-60c^2+40c^3-5c^4}{126}. \\ V_{0,4}(c) &= \frac{80(2-21c+117c^2-205c^3+165c^4-45c^5+5c^6)}{9(1+c)^7}. \\ h_1 &= \left[\frac{24^2 5/4}{8(1/21)^2 n}\right]^{\frac{1}{9}} = 3.243132n^{-\frac{1}{9}} \end{aligned}$$

$$\begin{aligned}
\text{MSE}(f_n^n, x = ch_1) &= \frac{8^{\frac{1}{9}}(\frac{1}{21})^{\frac{2}{9}}}{24^{-\frac{2}{9}}(\frac{5}{4})^{\frac{1}{9}}} \\
&\quad \times \left\{ \frac{V_{0,4}\left(\frac{c}{b(c)}\right)}{b(c)} + \frac{5(21)^2}{32} b(c)^8 [B_{0,4}\left(\frac{c}{b(c)}\right)]^2 \right\} n^{-\frac{8}{9}}. \qquad (6.1) \\
b_1(c) &= 2 - c. \\
b_2(c) &= \left\{ \frac{256(2 - 21c + 117c^2 - 205c^3 + 165c^4 - 45c^5 + 5c^6)}{(1+5c)^2(1+c)^7} \right\}^{\frac{1}{9}}.
\end{aligned}$$

In order to compare b_1, b_2 with the constant bandwidth variation function ($b(\cdot) \equiv 1$), we calculated their corresponding MSE values from (6.1) for $0 \le c \le 1$. The MSE values are plotted in Figure 4 below. Both plots show that the bandwidth variation functions significantly decrease MSE values, while the performance of b_1 and b_2 is similar. The same phenomenon was observed for order (0, 2) boundary kernels in Zhang and Karunamuni (1995).

Figure 4 about here

A similar comparison was done with the order (1, 3) boundary kernel defined by (3.8). In this case,

$$\begin{aligned}
B_{1,3}(c) &= \frac{3(1 - 4c + 2c^2}{7}. \\
V_{1,3}(c) &= \frac{240(8 - 17c + 11c^2)}{7(1+c)^5}. \\
h_1 &= 315^{\frac{1}{7}} n^{-\frac{1}{7}}. \\
\text{MSE}(f_n^n, x = ch_1) &= \frac{4^{\frac{3}{7}}\left(\frac{3}{7}\right)^{\frac{6}{7}}}{3^{\frac{3}{7}} 6^{\frac{6}{7}} \left(\frac{15}{7}\right)^{\frac{3}{7}}} \\
&\quad \times \left\{ \frac{V_{1,3}\left(\frac{c}{b(c)}\right)}{b(c)^3} + \frac{315}{36} b(c)^4 [B_{1,3}\left(\frac{c}{b(c)}\right)]^2 \right\} n^{-\frac{4}{7}}. \qquad (6.2)
\end{aligned}$$

$$
\begin{aligned}
b_1(c) &= 2 - c. \\
b_2(c) &= \left\{ \frac{16(8 - 17c + 11c^2)}{(1+c)^5} \right\}^{\frac{1}{7}}.
\end{aligned}
$$

Also, the MSE values are calculated from (6.2) w.r.t. b_1, b_2 and $b(\cdot) \equiv 1$ and were plotted in Figure 5. Again, our conclusion is the same as that of the previous one.

Figure 5 about here

Figures 4 and 5 demonstrate that the use of a bandwidth variation function at the boundary is very important and that improve the performance of the estimator greatly. Since the simplicity and the better performance of b_1, we suggest to use b_1 as the bandwidth variation function in practice.

Appendix: Proofs

Proof of Lemma 2.1. Assume that the highest degree among K_0, K_1, ..., K_{k-1} is l, then

$$K_i(t) = \sum_{j=0}^{l} \beta_{ij} t^j, \ i = 0, \ ..., \ k-1.$$

If S is singular, there exist α_0, ..., α_{k-1}, (not all equal to zero) such that

$$\sum_{i=0}^{k-1} \alpha_i \int_{-1}^{c} t^m K_i(t) dt = \sum_{i=0}^{k-1} \alpha_i \int_{-1}^{c} \sum_{j=0}^{l} \beta_{ij} t^{m+j} dt = 0, \ m = 0, \ 1, \ ..., \ k-1. \quad \text{(A.1)}$$

By simple algebra, (A.1) can be written as

$$
\begin{pmatrix}
c+1 & \frac{c^2-1}{2} & \cdots & \frac{c^{l+1}-(-1)^{l+1}}{l+1} \\
\frac{c^2-1}{2} & \frac{c^3-1}{3} & \cdots & \frac{c^{l+2}-(-1)^{l+2}}{l+2} \\
\vdots & \vdots & & \vdots \\
\frac{c^{k+1}-(-1)^{k+1}}{k+1} & \frac{c^{k+2}-(-1)^{k+2}}{k+2} & \cdots & \frac{c^{l+k}-(-1)^{l+k}}{l+k}
\end{pmatrix} \times
$$

$$\begin{pmatrix} \beta_{00} & \beta_{10} & \cdots & \beta_{(k-1)0} \\ \beta_{01} & \beta_{11} & \cdots & \beta_{(k-1)1} \\ \vdots & \vdots & & \vdots \\ \beta_{0l} & \beta_{1l} & \cdots & \beta_{(k-1)l} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{k-1} \end{pmatrix} = \underline{0}. \tag{A.2}$$

Denote

$$C = \begin{pmatrix} c+1 & \frac{c^2-1}{2} & \cdots & \frac{c^{l+1}-(-1)^{l+1}}{l+1} \\ \frac{c^2-1}{2} & \frac{c^3-1}{3} & \cdots & \frac{c^{l+2}-(-1)^{l+2}}{l+2} \\ \vdots & \vdots & & \vdots \\ \frac{c^{k+1}-(-1)^{k+1}}{k+1} & \frac{c^{k+2}-(-1)^{k+2}}{k+2} & \cdots & \frac{c^{l+k}-(-1)^{l+k}}{l+k} \end{pmatrix}$$

and

$$B = \begin{pmatrix} \beta_{00} & \beta_{10} & \cdots & \beta_{(k-1)0} \\ \beta_{01} & \beta_{11} & \cdots & \beta_{(k-1)1} \\ \vdots & \vdots & & \vdots \\ \beta_{0l} & \beta_{1l} & \cdots & \beta_{(k-1)l} \end{pmatrix}.$$

Then, (A.2) becomes

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{k-1} \end{pmatrix} = \underline{0}. \tag{A.3}$$

We first claim that B is column independent. Otherwise, there exist $\beta_0, \ldots, \beta_{k-1}$ (not all equal to 0) such that $\sum_{i=0}^{k-1} \beta_i K_i(t) = 0$. Write $\beta_{i_*} = \min\{\beta_i : \beta_i \neq 0, i = 0, \ldots, k-1\}$. Then

$$\beta_{i_*} K_{i_*}(t) = - \sum_{i=0,\ i \neq i_*}^{k-1} \beta_i K_i(t)$$

and

$$0 \neq \int_{-1}^{1} \beta_{i_*} t^{i_*} K_{i_*}(t) dt = - \sum_{i=0,\ i \neq i_*}^{k-1} \beta_i \int_{-1}^{1} t^{i_*} K_i(t) = 0.$$

The fact that B is column independent implies that $l + 1 \geq k$. This further implies that C is row independent (also, see Gasser and Müller (1979)). Since B is column indenpendent and C is row independent, there exist two matrices Q, P with $|Q| = \pm 1$, $|P| = \pm 1$ such that

$$CQ = (C_1, \underline{0}),\ PB = (B_1^T, \underline{0})^T,$$

where C_1 and B_1 are regular metrices. Then,

$$CB = CQQ^{-1}P^{-1}PB = (C_1, \underline{0})Q^{-1}P^{-1}(B_1^T, \underline{0})^T.$$

Denote $Q^{-1}P^{-1} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$. Then we have

$$CB = (C_1, \underline{0}) \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} (B_1^T, \underline{0})^T = C_1 M_{11} B_1,$$

where M_{11} is regular. So, $rank(CB) = rank(C_1 M_{11} B_1) = k$. Therefore, CB is regular and the only solution to (A.3) is $\underline{0}$. This proves that S is regular.

Proof of Lemma 3.1. From (2.10) and (3.3),

$$K_{v,c}(t) = (0,\ \ldots,\ (-1)^v v!,\ 0\ ,\ldots,\ 0)S^{-1}V(S^+)^{-1} \begin{pmatrix} 1 \\ t \\ \vdots \\ t^k \end{pmatrix}.$$

Note that $(0,\ \ldots,\ (-1)^v v!,\ 0\ ,\ldots,\ 0)$ is a $1 \times k$ vector, S is a $k \times k$ matrix, V is $k \times (k+1)$ matrix and S^+ is a $(k+1) \times (k+1)$ matrix. To prove (3.4), it is enough to prove

$$(0,\ \ldots,\ (-1)^v v!,\ 0\ ,\ldots,\ 0)S^{-1}V(S^+)^{-1} = (0,\ \ldots,\ (-1)^v v!,\ 0\ ,\ldots,\ 0)(S^*)^{-1}.$$

This is equivalent to prove that

$$(0,\ ...,\ (-1)^v v!,\ 0\ ,...,\ 0)\begin{pmatrix} S^{-1} & \underline{0} \\ \Delta & 1 \end{pmatrix}\begin{pmatrix} V \\ V_1 \end{pmatrix}(S^+)^{-1}$$
$$= (0,\ ...,\ (-1)^v v!,\ 0\ ,...,\ 0)(S^*)^{-1},$$

where $(0,\ ...,\ (-1)^v v!,\ 0\ ,...,\ 0)$ is a $1 \times (k+1)$ vector with $(-1)^v v!$ at its v^{th} place, Δ is a $1 \times k$ vector to be determined later and $V_1 = (0,\ 0,\ ...,\ 0, 1)_{1\times(k+1)}$. So, it is enough to prove

$$S^+ \begin{pmatrix} V \\ V_1 \end{pmatrix}^{-1} \begin{pmatrix} S & \underline{0} \\ C & 1 \end{pmatrix} = S^*, \tag{A.4}$$

C is corresponding to Δ of the above expression. Note that

$$S = \begin{pmatrix} \int_{-1}^{c} K_0(t)dt & \int_{-1}^{c} K_1(t)dt & \dots & \int_{-1}^{c} K_{k-1}(t)dt \\ \int_{-1}^{c} tK_0(t)dt & \int_{-1}^{c} tK_1(t)dt & \dots & \int_{-1}^{c} tK_{k-1}(t)dt \\ \vdots & \vdots & & \vdots \\ \int_{-1}^{c} t^{k-1}K_0(t)dt & \int_{-1}^{c} t^{k-1}K_1(t)dt & \dots & \int_{-1}^{c} t^{k-1}K_{k-1}(t)dt \end{pmatrix}^T .$$

Denote $(S^+)^{-1} = \left(s_{jl}^+\right)_{0\le j,\ l\le k}$, then, by (3.3), $S = (s_{ij})_{0\le i,\ j\le k}$, where $s_{ij} = (-1)^i i! \int_{-1}^{c}(s_{i0}^+ + s_{i1}^+ t + ... + s_{ik}^+ t^j)dt$. Simple calculation leads to

$$\begin{aligned} S \;=\; & \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & (-1)^{k-1}(k-1)! \end{pmatrix} \begin{pmatrix} s_{00}^+ & s_{01}^+ & \dots & s_{0k}^+ \\ s_{10}^+ & s_{11}^+ & \dots & s_{1k}^+ \\ \vdots & \vdots & & \vdots \\ s_{(k-1)0}^+ & s_{(k-1)1}^+ & \dots & s_{(k-1)k}^+ \end{pmatrix} \\ & \begin{pmatrix} c+1 & \frac{c^2-1}{2} & \dots & \frac{c^k-(-1)^k}{k} \\ \frac{c^2-1}{2} & \frac{c^3-1}{3} & \dots & \frac{c^{k+1}-(-1)^{k+1}}{k+1} \\ \vdots & \vdots & & \vdots \\ \frac{c^{k+1}-(-1)^{k+1}}{k+1} & \frac{c^{k+2}-(-1)^{k+2}}{k+2} & \dots & \frac{c^{2k}-(-1)^{2k}}{2k} \end{pmatrix} \\ =\; & V^* V_1^* V_2. \end{aligned}$$

So

$$\begin{pmatrix} S & \underline{0} \\ C & 1 \end{pmatrix} = \begin{pmatrix} V^* & \underline{0} \\ C & 1 \end{pmatrix} \begin{pmatrix} V_1^* V_2 & 0 \\ C & 1 \end{pmatrix} = \begin{pmatrix} V \\ V_1 \end{pmatrix} \begin{pmatrix} V_1^* V_2 & \underline{0} \\ \underline{0} & 1 \end{pmatrix}$$

Consequently,

$$\text{LHS of (A.4)} = S^+ \begin{pmatrix} V_1^* V_2 & \underline{0} \\ \underline{0} & 1 \end{pmatrix}. \tag{A.5}$$

Since the solution to the following equation

$$\begin{pmatrix} 2 & 0 & \frac{2}{3} & \cdots & \frac{1-(-1)^{k+1}}{k+1} \\ 0 & \frac{2}{3} & 0 & \cdots & \frac{1-(-1)^{k+2}}{k+2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & -1 & 1 & \cdots & (-1)^k \end{pmatrix} \begin{pmatrix} \lambda_0 \\ \lambda_1 \\ \vdots \\ \lambda_k \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ (-1)^k k! \end{pmatrix}$$

is $(\lambda_0, \lambda_1, ..., \lambda_k) = (0, 0, ..., (-1)^k k!)(S^+)^{-1}$, it follows that

$$(s^+_{k,0}, s^+_{k,1}, ..., s^+_{k,k})(1, -1, ..., (-1)^k)^T = 1. \tag{A.6}$$

Define

$$C = (s^+_{k,0}, s^+_{k,1}, ..., s^+_{k,k})V_2 \tag{A.7}$$

Then, by (A.6), (A.7) and the fact that $K_v(-1) = 0$, $v = 0, ..., k-1$, we have

$$\begin{aligned} \begin{pmatrix} V_1^* V_2 & \underline{0} \\ \underline{0} & 1 \end{pmatrix} &= \begin{pmatrix} V_1^* \\ s^+_{k0}, s^+_{k1}, ..., s^+_{kk} \end{pmatrix} \begin{pmatrix} (V_2)^T \\ 1, -1, ..., (-1)^k \end{pmatrix}^T \\ &= (S^+)^{-1} S^*. \end{aligned} \tag{A.8}$$

This together with (A.5) completes the proof.

Proof of Theorem 3.1. From (3.9)

$$\begin{pmatrix} K_0(t) \\ K_1(t) \\ \vdots \\ K_{k-1}(t) \end{pmatrix} = \mathrm{Diag}(1, \ , -1, \ 2!, \ ..., \ (-1)^k(k-1)!)M_1^{-1} \begin{pmatrix} 1 \\ t \\ \vdots \\ t^{k-1} \end{pmatrix} K(t)$$

$$= DM_1^{-1} \begin{pmatrix} 1 \\ t \\ \vdots \\ t^{k-1} \end{pmatrix} K(t),$$

where $M_1 = (m'_{j,l})_{0 \le j, \ l \le k-1}$ is a matrix with $m'_{j,l} = \int_{-1}^{1} t^{j+l} K(t)dt$. Then,

$$K_{v,c}(t) = (0, \ ..., \ (-1)^v v!, \ ..., \ 0)S^{-1}DM_1^{-1} \begin{pmatrix} 1 \\ t \\ \vdots \\ t^{k-1} \end{pmatrix} K(t).$$

Since

$$K^*_{v,c}(t) = (0, \ ..., \ (-1)^v v!, \ ..., \ 0)M^{-1} \begin{pmatrix} 1 \\ t \\ \vdots \\ t^{k-1} \end{pmatrix} K(t),$$

it is enough to prove $S^{-1}DM_1^{-1} = M^{-1}$ or equivalently $M = M_1 D^{-1} S$. This can be proved along the same lines of proof of Lemma 3.1.

Proof of Lemma 5.2. We shall only prove the case when $v = 0$. Assume the support of the endpoint kernel $K^+_{0,0}(t)$ is [-τ, 0] for some $\tau > 0$. Decompose [-τ, 0] into j subintervals I_1, I_2, ..., I_j defined by the points where a change of sign occurs, and by $-\tau$ and 0. Then, the following matrix is regular.

$$M = \begin{pmatrix} \int_{I_1} tK_{0,0}^{+}(t)dt & \dots & \int_{I_j} tK_{0,0}^{+}(t)dt \\ \vdots & \vdots & \vdots \\ \int_{I_1} t^j K_{0,0}^{+}(t)dt & \dots & \int_{I_j} t^j K_{0,0}^{+}(t)dt \end{pmatrix}.$$

Otherwise, there exist $\{\alpha_1, ..., \alpha_j\} \neq \{0, ..., 0\}$ such that

$$\int_{I_l} \sum_{i=1}^{j} \alpha_i t^i K_{00}^{+}(t)dt = 0, \ l = 1, \ 2, \ ..., \ j.$$

Since $K_{00}^{+}(t)$ does not change its sign in I_l, $\sum_{i=1}^{j} \alpha_i t^i$ has a root within I_l. In total, this gives j roots not equal to 0 and one root at 0. We conclude that $\{\alpha_1, ..., \alpha_j\} = \{0, ..., 0\}$.

Now we prove that $j \geq k$. If $j < k$, then the matrix M would map the vector $\{1, ..., 1\}$ to 0 by the moment conditions. This is a contradiction to the regularity of M.

Proof of (i) of Lemma 5.3. By Lemma 5.2, if $\int_{-\tau}^{0} K_{v,0}^{+}(t)t^k dt = 0$, then there are at least k changes of sign on $[-\tau,0]$. This is a contradiction to the assumption that $K_{v,0}(t)$ is minimal.

Proof of (ii) of Lemma 5.3. Lemma 5.2 implies that all the roots of $K_{v,0}^{+}(t)$ are negative. This fact together with $\int_{-\tau}^{0} t^v K_{v,0}^{+}(t)dt = (-1)^v v!$ completes the proof.

Proof of (i) of Theorem 5.1. It is obvious that $K_{v,0}^{+}$ and $K_{v,0}^{*}$ satisfy (5.3) and (5.4), since they are both $(v, k)^{th}$ order endpoint kernels. Note that a polynomial of degree k is uniquely determined by (5.3) and (5.4). Hence the result.

Proof of (ii) of Theorem 5.1. For any boundary kernel function $K_{v,0} \in \mathcal{N}_{k-1}$, by definition of $\mathcal{N}_{k-1}$, $K_{v,0}$ has $k-1$ changes of sign on its support. Assume that $K_{v,0}$ has m roots outside its support and on its boundaries. Denote them by t_1, ..., t_m. Define $K(t) = d(x-t_1)...(x-t_m)$, where d is a normalizing constant, which makes $K(t) \geq 0$ and $\int_{-1}^{1} K(t)dt = 1$. By (3.9), $K_{v,0}^*$ is uniquely determined by $K(t)$. This means that $K_{v,0}$ can be written in the form of (3.9).

On the other hand, assume that $K_{v,0} \in \{K_{v,0}^*(t),$ for all $K \geq 0\}$. From Lemma 5.3, $K_{v,0}$ has at least $k-1$ changes of sign on its support. Since $K(t) \geq 0$, this means $(0, ..., (-1)^v v!, 0, ..., 0)M^{-1}(1, t, ..., t^{k-1})^T$ has at least $k-1$ changes of sign on the support of $K(t)$. Since $(0, ..., (-1)^v v!, 0, ..., 0)M^{-1}(1, t, ..., t^{k-1})^T$ is at most a polynomial of degree $k-1$, it has to have exactly $k-1$ changes of sign on the support of $K(t)$. Therefore, $K_{v,0} \in \mathcal{N}_{k-1}$.

References

Cheng, M. Y., Fan, J., and Marron, J. S. (1995), "On Automatic Boundary Corrections," Technical report.

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

Cowling, A., and Hall, P. (1996), "On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation," *Journal of the Royal Statiscal Society*, Series B, 58, 551-563.

Fan, J. (1992), "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998-1004.

Fan, J., Gasser, T., Gijbels, I., Brockman, M., and Joachim, E. (1993), "Local Polynomial Fitting: A Standard for Nonparametric Regression," Unpublished manuscript.

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

Gasser, T., Müller, H. G., and Mammitzsch, V. (1985), "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society,* Series B, 47, 238-256.

Jones, M. C. (1993), "Simple boundary correction for kernel density estimation" *Statist. Comput.,* 3, 135-146.

Karunamuni, R. J., and Mehra, K. L. (1991), "Optimal Convergence Properties of Kernel Density Estimators without Differentiability Conditions," *Ann. Inst. Statist. Math.,* 43, 327-346.

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

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

Müller, H. G., and Wang, J. L. (1994), "Hazard Rate Estimation Under Random Censoring with Varying Kernels and Bandwidths," *Biometrics,* 50, 61-76.

Rice, J. (1984), "Boundary Modifications for Kernel Regression," *Communications in Statistics- Theory and Methods,* 13, 890-900.

Schuster, E. F. (1985), "Incorporating Support Constraints into Nonparametric Estimators of Densities," *Communications in Statistics- Theory and Methods,* 14, 1123-1136.

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

Zhang, S., and Karunamuni, R. J. (1995)," On Kernel Density Estimation Near Endpoints with an Application to Line Transect Sampling," Technical Report, Department of Mathematical Sciences, Statistics Center, Univ. of Alberta.

Figure 1. Boundary kernels of order (0,2).

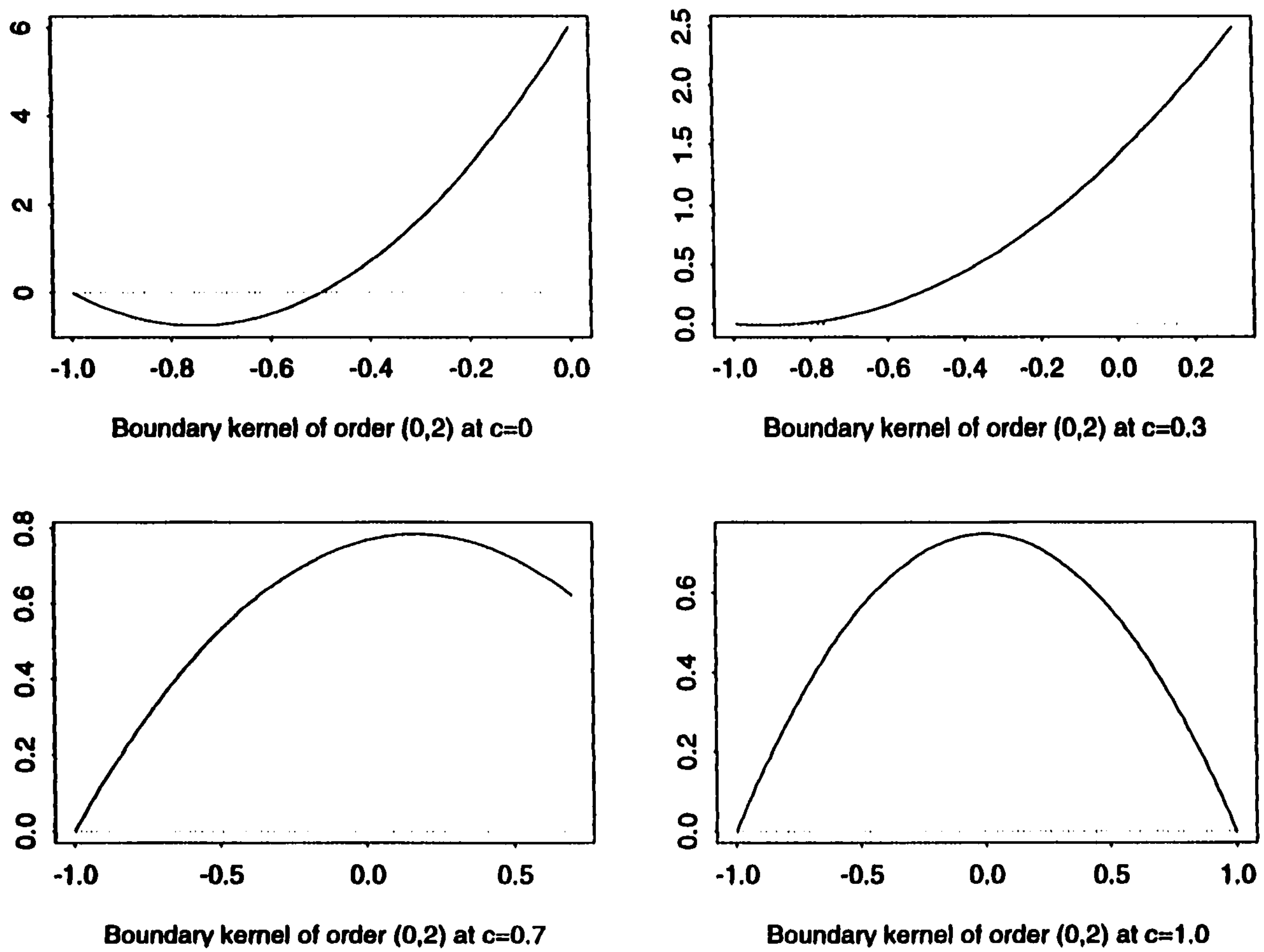


Figure 2. Boundary kernels of order (0,4).

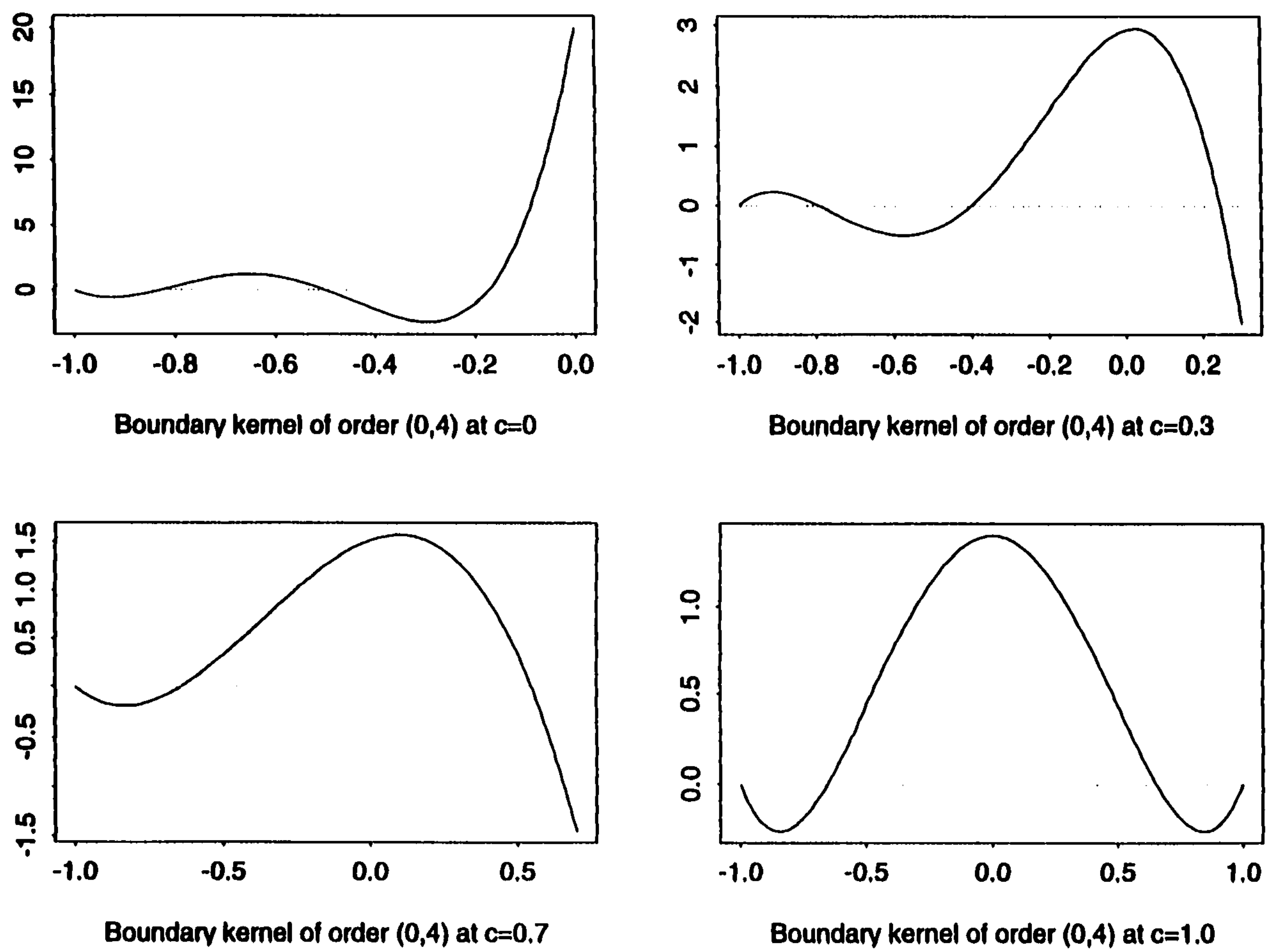
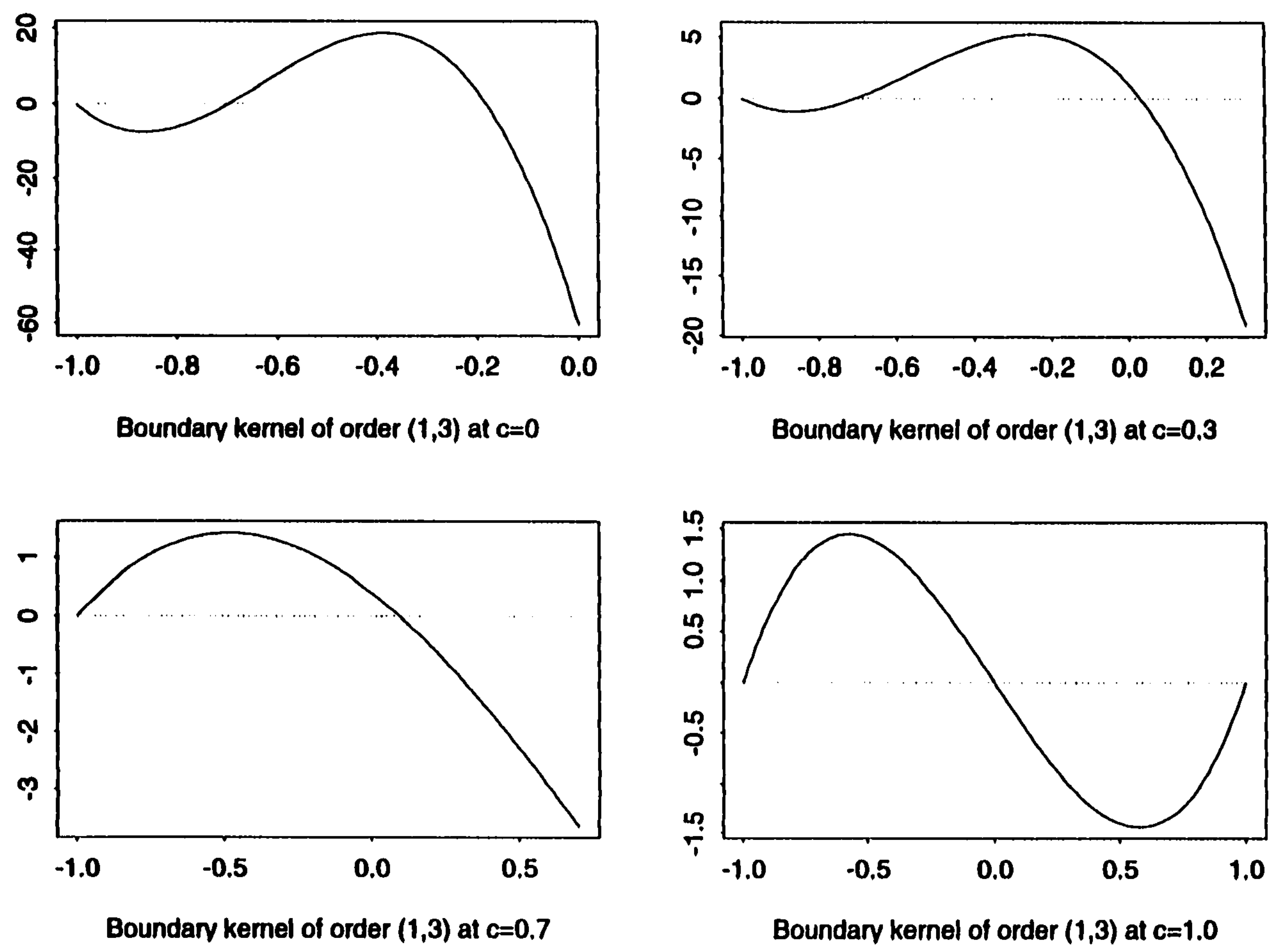


Figure 3. Boundary kernels of order (1,3), optimal at 0.



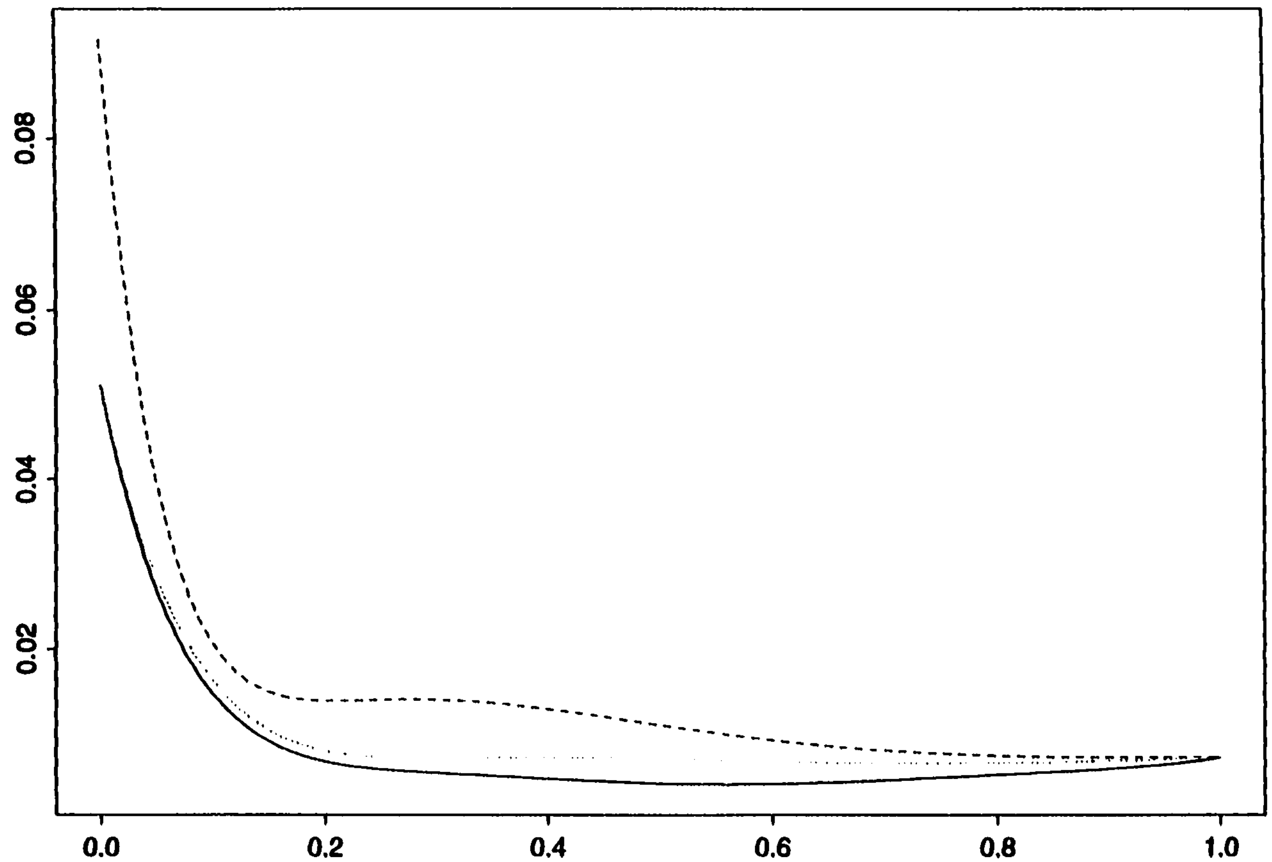


Figure 4. MSE from the endpoint kernel of order (0,4), with b1 (Solid Line), b2 (Dotted Line) and b=1 (Dashed Line)

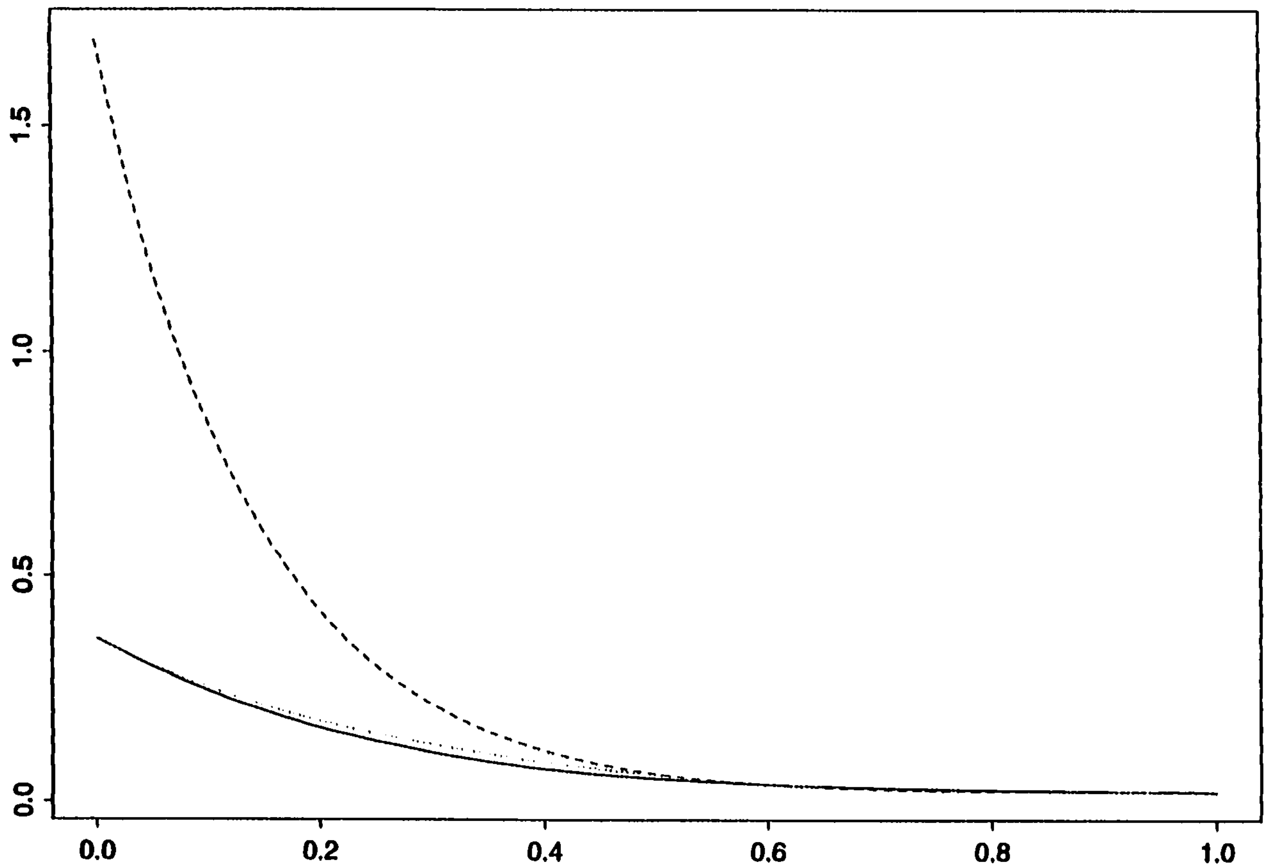


Figure 5. MSE from optimal kernel of order (1,3) with b1 (Solid Line), b2 (Dotted Line) and b=1 (Dashed Line)

Chapter Six

An Improved Estimator of the Density Function at the Boundary

1. Introduction

Let f denote a probability density function with support $[0, \infty)$, and consider nonparametric estimation of f based on a random sample X_1, ..., X_n from f. Then the conventional kernel estimator of f at x 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 a non-negative symmetric kernel function with support $[-1, 1]$, and h is the bandwidth ($h \to 0$, as $n \to \infty$). For $x = ch$, $0 \leq c < 1$, the estimate $f_n(x)$ is not a consistent estimate of $f(x)$. This is known as the boundary effect.

There has been an extensive literature on how to correct this boundary effect. Some well-known methods are summarized below:

(i) The reflection method (Schuster (1985), Cline and Hart (1991) and Silverman (1986)). This method is specially designed for the case $f^{(1)}(0) = 0$, where $f^{(1)}$ denotes the first derivative of f.

(ii) The boundary kernel method (Gasser and Müller (1979), Gasser et al. (1985), Müller (1991), Jones (1993), Cheng et al. (1995) and Zhang and Karunamuni (1995, 1996)). This method is more general than the reflection method in the sense that it can adapt to any shape of densities. However, a drawback of this method is that the estimates might be negative. To correct this deficiency,

A version of this chapter has been submitted for publication.

some remedies have been proposed, see Jones (1993).

(iii) The transformation method (Wand, Marron and Ruppert (1993) and Marron and Ruppert (1994)).

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

The boundary kernel related methods usually focus on getting the bias as one wants it, the price for that being an increase in variance. It has been gradually gotten through to researchers that this variance inflation is important, reflecting a real practical phenomenon, and so such methods do in fact allow for improvement (Cowling and Hall (1996)). Approaches involving only kernel modifications without regard to f are always associated with larger variance. To make the all important reductions in variance, one has to do something involving functions of f near the boundary (as done here). Whereas ordinary reflection has a bad bias but has low variance (in Jones (1993), among others), f-dependent generalizations of such a method are well worth exploring to see if one can improve the bias but hold on to the low variance. It seems the answer is yes! Cowling and Hall (1996)) already have one version of this, but it is not clear from their paper whether the variance is kept in check for all f, it seems not from their Table 2. The present work can be viewed as an improvement of a particularly sensible class of methods which can keep variance down. Our method is a combination of methods of pseudo-data, transformation and reflection. Furthermore, the proposed estimator is non-negative everywhere. In simulations, we show that this idea produces smaller mean squared error (MSE) values compared to other methods for almost all shapes of densities.

2. The Methodology

As in Cowling and Hall (1996), we also need to generate some pseudo data. We aim to generate the pseudo data beyond the left endpoint of the support of the density f such that the pseudo data offers a natural extension of the density f outside its support locally. Our method of generating pseudo data combines the transformation and reflection methods, consisting of the following three steps:

(1) Transform the original data X_1, ..., X_n to $g(X_1)$, ..., $g(X_n)$ while keeping the original data, where g is a non-negative, continuous and monotonically increasing function from $[0, \infty)$ to $[0, \infty)$.

(2) Reflect $g(X_1)$, ..., $g(X_n)$ around the origin so we have $-g(X_1)$, ..., $-g(X_n)$.

(3) Based on the enlarged data sample $-g(X_1)$, ..., $-g(X_n)$, X_1, ..., X_n, the new estimator is defined as

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x+g(X_i)}{h}\right)\right\}, \tag{2.1}$$

where K is a usual kernel function as in (1.1).

In order to define the transformation g, we first need to obtain explicit forms of the bias and variance expressions of the estimator (2.1). Under certain conditions on g and f, it is easy to show that (see Lemma 1 of Appendix) for $x = ch$, $0 \le c \le 1$,

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

and

$$\mathrm{Var}\hat{f}_n(x) = \frac{f(0)}{nh}[\int_{-1}^{1} K(t)^2dt + 2\int_{-1}^{c} K(t)K(2c-t)dt](1+o(1)). \tag{2.3}$$

[Here $f^{(i)}$ and $g^{(i)}$ denote i^{th} derivatives of f and g, respectively, with $f^{(0)} = f$ and $g^{(0)} = g$.] A simple consequence of (2.3) is

$$\mathrm{Var}\hat{f}_n(x) \le \frac{2f(0)}{nh}\int_{-1}^{1} K(t)^2dt. \tag{2.4}$$

The primary goal of our transformation g is to eliminate the first order term in the bias expression (2.2). So, it is enough to let

$$g^{(2)}(0) = \frac{2f^{(1)}(0)}{f(0)}. \tag{2.5}$$

Combining (2.5) with the assumptions in Lemma 1 of Appendix, g should satisfy

(i) $g^{-1}(0) = 0$,

(ii) $g^{(1)}(0) = 1$,

(iii) $g^{(2)}(0) = \frac{2f^{(1)}(0)}{f(0)}$,

(iv) g is monotonically increasing.

Functions satisfying conditions (i)-(iv) are easy to construct. Based on extensive simulations, we find that the following transformation adapts to various shapes of densities well:

$$g(x) = x + dx^2 + Ad^2x^3, \tag{2.6}$$

where

$$d = f^{(1)}(0)/f(0) \text{ and } 3A > 1. \tag{2.7}$$

For g defined by (2.6) and for $x = ch$, $0 \leq c \leq 1$, the bias term (2.2) can be written as

$$\begin{aligned} \mathrm{E}\hat{f}_n(x) - f(x) \;=\; & \frac{h^2}{2}\Big\{f^{(2)}(0)[\int_{-1}^{c} t^2K(t)dt + \int_{c}^{1}(t-c)^2K(t)dt] \\ & -6(A-1)\frac{[f^{(1)}(0)]^2}{f(0)}\int_{c}^{1}(t-c)^2K(t)dt\Big\} + o(h^2). \end{aligned} \tag{2.8}$$

An interesting case of (2.8) is when $A = 1$. Then (2.8) becomes

$$\mathrm{E}\hat{f}_n(x) - f(x) = \frac{h^2}{2}f^{(2)}(0)[\int_{-1}^{c} t^2K(t)dt + \int_{c}^{1}(t-c)^2K(t)dt] + o(h^2)$$

and

$$|\mathrm{E}\hat{f}_n(x) - f(x)| \leq \frac{h^2}{2}|f^{(2)}(0)|\int_{-1}^{1} t^2K(t)dt] + o(h^2).$$

The preceding inequality shows that the boundary bias is even less than the interior bias while (2.4) shows that boundary variance is at most twice of the interior. So, we can expect the MSE behavior of the estimator (2.1) at the boundary points to be similar to that of the interior points.

2.1. Estimation of g

The transformation g defined by (2.6) is not available in practice, since d defined by (2.7) is unknown. A natural estimator of d, and hence g, can be obtained by directly substituting corresponding kernel estimators of $f(0)$ and $f^{(1)}(0)$ in (2.7). However, the resulting estimator suffers a great variability, which in turn affects the performance of our estimator (2.1). A better estimator of d is obtained as follows. Note that d can be written as $d = \frac{d}{dx} \log f(x) \mid_{x=0}$. So, d can be estimated by

$$d_n = \frac{\log f_n(h) - \log f_n(0)}{h}, \tag{2.9}$$

where

$$f_n(h) = f_n^*(h) + \frac{1}{n^2} \tag{2.10}$$

$$f_n(0) = \max\left(f_n^*(0), \frac{1}{n^2}\right) \tag{2.11}$$

with

$$\begin{aligned} f_n^*(h) &= \tfrac{1}{nh} \textstyle\sum_{i=1}^n K\left(\tfrac{h-X_i}{h}\right) \\ f_n^*(0) &= \tfrac{1}{nh_0} \textstyle\sum_{i=1}^n K_{(0)}\left(\tfrac{-X_i}{h_0}\right), \end{aligned} \tag{2.12}$$

where K is a usual kernel, $K_{(0)}$ is a so-called end-point kernel satistying

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

where

$$h_0 = b(0)h$$

and

$$b(0) = \{[\int_{-1}^{1} t^2K(t)dt]^2 \int_{-1}^{0} K_0(t)^2dt / [\int_{-1}^{0} t^2K_0(t)dt]^2 / \int_{-1}^{1} K(t)^2dt\}^{\frac{1}{5}}.$$

The rationale for choosing h_0 is that the local optimal bandwidth for estimating $f(0)$ is $b(0)$ times the local optimal bandwidth for estimating $f(h)$ (see Zhang and Karunamuni (1995)). Furthermore, the factor $1/n^2$ in (2.10) and (2.11) is employed to make $f_n^*(h)$ and $f_n^*(0)$ bounded away from 0. As we shall prove later, it really does not affect the statistical properties of $f_n^*(h)$ and $f_n^*(0)$. It is shown in Lemma 2 of Appendix that

$$\begin{aligned} E|f_n^*(h) - f(h)|^k &= O(h^{2k}) \\ E|f_n^*(0) - f(0)|^k &= O(h^{2k}), \end{aligned} \tag{2.13}$$

for any integer $k \geq 2$, provided $f^{(2)}$ is continuous near 0. A direct consequence of (2.13) is the following results (see Lemmas 3 and 4 of Appendix)

$$\begin{aligned} E|f_n(h) - f(h)|^k &= O(h^{2k}) \\ E|f_n(0) - f(0)|^k &= O(h^{2k}) \end{aligned} \tag{2.14}$$

and

$$E|d_n - d|^k = O(h^k), \tag{2.15}$$

for any integer $k \geq 2$, where d_n, $f_n(h)$ and $f_n(0)$ are given by (2.9), (2.10) and (2.11), respectively.

We therefore define

$$g_n(x) = x + d_nx^2 + Ad_n^2x^3 \tag{2.16}$$

as our estimator of $g(x)$, where d_n is defined by (2.9)

2.2. The Proposed Estimator

Based on g_n defined by (2.16), our proposed new estimator of f corresponding to (2.1) is defined as

$$\hat{f}_{new}(x) = \frac{1}{nh}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x+g_n(X_i)}{h}\right)\right\}. \tag{2.17}$$

It is easy to see that for $x \geq h$, (2.17) reduces to

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

which is the usual interior kernel estimator. So, (2.17) is a natural boundary continuation of the usual estimator. Also note that the only data that need to be transformed are those within $4Ah/(4A-1)$ of the boundary. It is also important to remark here that the estimator (2.17) is non-negative, a property shared with other reflection estimators (see Jones and Foster (1996)) and the transformation-based estimator of Marron and Ruppert (1995)) but not most boundary kernel approaches. The properties of the bias and variance of (2.17) are discussed in the following theorem.

Theorem 1. Assume that $f(x) > 0$ for $x = 0$, h, and $f^{(2)}$ is continuous in a neighborhood of 0. Then, for $x = ch$, $0 \leq c \leq 1$, we have

$$|\mathrm{E}\hat{f}_{new}(x) - f(x)| = O(h^2) \tag{2.18}$$

and

$$\mathrm{Var}\hat{f}_{new}(x) = O\left(\frac{1}{nh}\right), \tag{2.19}$$

where $\hat{f}_{new}$ is given by (2.17).

The proof of Theorem 1 is deferred to the Appendix.

3. Simulations

To compare the performance of our proposed estimator $\hat{f}_{new}$ defined by (2.17) with other existing estimators, extensive simulations were carried out. The other well-known estimators used in the comparison are the boundary kernel estimator, the pseudo-data method estimator and the non-negative boundary correction method estimator, defined as follows.

The boundary kernel estimator is defined as (see Zhang and Karunamuni (1995))

$$\hat{f}_B(x) = \frac{1}{nh_c}\sum_{i=1}^{n} K_{(c)}\left(\frac{x - X_i}{h_c}\right), \tag{3.1}$$

where $c = \min\left\{\frac{x}{h}, 1\right\}$, $K_{(c)}$ is a boundary kernel satisfying $K_{(1)}(t) = K(t)$, $h_c = b(c)h$ with $b(c) = 1 - (c-1)(b(0) - 1)$ and $b(0)$ as defined in Section 2 (see circa (2.12)). In simulations, we employed the following kernel and boundary kernel

$$K(t) = \frac{3}{4}(1 - t^2)I_{[-1,1]}$$

and

$$K_{(c)}(t) = \frac{12}{(1+c)^4}(1+t)\left[(1-2c)t + \frac{3c^2 - 2c + 1}{2}\right] I_{[-1,c]}, \tag{3.2}$$

respectively, where I_A denotes the indicator function on the set A.

Zhang and Karunamuni (1995) have shown that $K_{(0)}$ defined by (3.2) at $c = 0$ is the optimal end-point kernel under some restraints other from those in Müller (1991), i.e. $K_{(0)}(t)$ minimizes MSE when estimating $f(0)$. Note that for the above K and $K_{(c)}$, $b(c) = 2 - c$. Therefore $h_c = (2-c)h$. When $c = 0$, h_0 $(= 2h)$ is the optimal bandwidth for estimating $f(0)$. Zhang and Karunamuni (1995) demonstrated that the use of h_0 reduces the MSE values to about half of that when the fixed bandwidth h is used. The factor $b(c)(= 2 - c)$ is called the bandwidth variation function (see Müller (1991)).

The pseudo-data method estimator is defined as (see Cowling and Hall (1996))

$$\hat{f}_{CH}(x) = \frac{1}{nh}\left\{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) + \sum_{i=1}^{m} K\left(\frac{x - X_{(-i)}}{h}\right)\right\}, \tag{3.3}$$

where $X_{(-i)} = -5X_{(i/3)} - 4X_{(2i/3)} + \frac{10}{3}X_{(i)}$, $i = 1, 2, ..., n$, $X_{(i)}$ is the i^{th}-order statistic of $X_1, ..., X_n$, and m is an integer such that $nh < m < n$. In our simulations, we used $m = n^{\frac{4}{5}+\frac{1}{10}}$. The above rule of generating $X_{(-i)}$ is called the best three point rule by Cowling and Hall (1996).

The non-negative boundary correction method estimator is defined as (see Jones and Foster (1996))

$$\hat{f}_{JF}(x) = \bar{f}(x)\exp\left\{\frac{\hat{f}(x)}{\bar{f}(x)} - 1\right\}, \tag{3.4}$$

where

$$\bar{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K_c\left(\frac{x - X_i}{h}\right),$$

K_c is the cut-and-normalized kernel (see Gasser and Müller (1979)) and

$$\hat{f}(x) = (nh)^{-1}\sum_{i=1}^{n} K_{(c)}\left(\frac{x - X_{(-i)}}{h}\right),$$

where $K_{(c)}$ is a boundary kernel. Note that, except the pseudo-data method, the other three methods of estimation locally modify the estimator only in the boundary region. For the interior points $x \geq h$, they yield the same estimator.

The optimal global bandwidth is defined by (see Silverman (1986, pp 39-40))

$$h = \left\{\frac{\int_{-1}^{1} K(t)^2 dt}{[\int_{-1}^{1} t^2 K(t)dt]^2 \int [f^{(2)}(x)]^2 dx}\right\}^{\frac{1}{5}} n^{-\frac{1}{5}} \tag{3.5}$$

was implemented as the bandwidth throughout simulations. The reason for using (3.5) as the bandwidth is that the comparisons based on the optimal bandwidth are more convincing than comparisons based on approximated bandwidths which might - because of the quality or otherwise of the bandwidth selection method -

might be misleading. Also, a global rather than local bandwidth choice is made because this is much the more likely to be used in applications.

Our simulations consist of two parts. In the first part, we calculated the bias, variance and MSE values of the four estimates when estimationg $f(0)$. The sample size $n = 200$. Thirteen different shapes of densities were considered. For densities (1) to (9), $f(0) \neq 0$. The summary results are reported in Table 1. The values in Table 1 represent the averages over 500 repetitions.

Table 1 about here

In the second part of the simulation, we examined the behavior of each estimator inside the boundary region. This was done by plotting ten typical realizations of each estimator. Figures 1 to 5 represent these behaviors for five different densities inside the boundary region as well as for a part of the interior region. Figure 5 represents density estimates over the full domain $[0, 1]$.

Figures 1 to 5 about here

Discussion: By close examination of Table 1 and Figures 1 to 5, it is apparent that for densities (1) to (4) in Table 1, which satisfy $f^{(1)}(0) < 0$, the estimator $\hat{f}_{new}$ is by some way the best among the four considered. Also, $\hat{f}_{CH}$ is better than $\hat{f}_B$ for densities (1) to (3) and $\hat{f}_{JF}$ appears to be the least favorable one in all these cases - its MSE values are more than three times those of $\hat{f}_{new}$. Figure 1, which is concerned with density (2) shows clearly the variability associated with $\hat{f}_{JF}$ and also points out an unattractive performance of $\hat{f}_{CH}$ up to one bandwidth away from the boundary. For densities (5) to (9), which satisfy $f^{(1)}(0) \geq 0$, $\hat{f}_B$ is the best one among the four. But $\hat{f}_{new}$ takes a fairly close second place to $\hat{f}_B$, except in one case (i.e., for (5)). For densities (7) to (9), which satisfy $f^{(1)}(0) > 0$,

$\hat{f}_{CH}$ shows the largest MSE values among the four estimators. It also looks worst on Figures 2 and 3 which are concerned with densities (6) and (9). The other three estimators being roughly comparable in this case, although $\hat{f}_{JF}$ also suffers somewhat in comparison with $\hat{f}_{new}$ and $\hat{f}_B$. Though $\hat{f}_B$ has a better performance over other three for densities (5) to (9), it tends to take negative values when $f(0)$ is close to 0.

For densities (10) to (12), which satisfy $f(0) = 0$, $\hat{f}_{JF}$ is the best one among the four, whereas $\hat{f}_{new}$ again takes second place except in (10), in which case $\hat{f}_B$ has roughly the same performance as $\hat{f}_{JF}$, and $\hat{f}_{CH}$ matches $\hat{f}_{new}$, the two being not so good. Note that $\hat{f}_{JF}$ needs a little modification to insure $\bar{f}(x)$ in (3.4) is not equal to 0. Also, $\hat{f}_B$ has a marginal edge over $\hat{f}_{CH}$ for densities (10) and (11), but $\hat{f}_{CH}$ shows a better performance over $\hat{f}_B$ for density (12). Moreover, $\hat{f}_B$ has rather unappealing feature that it tends to take negative values when $f(0) = 0$ (the proportion of negativity of $\hat{f}_B$ are 5%, 85% and 100%, respectively, for densities (10), (11) and (12)). This is perhaps the major feature of Figure 4.

Finally, we discuss the performance of the estimators at $x = 0$ for the density (13). Since the density (13) has a pole at $x = 0$, the true value of $f(0)$ does not exist. However, $f(0)$ can be approximated by the four methods above. For this density, the bandwidth is subjectively chosen as $h = 0.18$ to ensure that the density estimator is smooth in the interior region. Table 1 shows that $\hat{f}_{new}$ has the smallest variance among the four estimators. $\hat{f}_{CH}$ has the smallest average value among the four; $\hat{f}_{JF}$ gives the largest average value and the largest variance as well. From Figure 5, one can see that $\hat{f}_{new}$, $\hat{f}_B$, and $\hat{f}_{JF}$ suggest the existence of a pole at $x = 0$, and $\hat{f}_{CH}$ completely fails in this case. It appears that $\hat{f}_{JF}$ seems the best one for this model, closely followed by $\hat{f}_B$.

It is worthwhile to mention that the use of the bandwidth variation function $b(c) = 2 - c$ for $\hat{f}_B$. It has been observed that the MSE values of $\hat{f}_B$ would be almost doubled without the use of the bandwidth variation function. If $\hat{f}_B$ is employed without $b(c)$, then clearly the estimator $\hat{f}_{new}$ becomes the best one for densities (1) to (9), while the performance of $\hat{f}_B$ would be similar to that of $\hat{f}_{JF}$, or even worse than $\hat{f}_{JF}$ in some cases.

In conclusion, we see that overall $\hat{f}_{new}$ is the best choice among the four competitors considered. It steadily outperforms $\hat{f}_{CH}$ for densities (1) to (9) (which only beats it once). Indeed, the performance of $\hat{f}_{CH}$ is very disappointing and this estimator cannot be recommended for use. $\hat{f}_{new}$ overwhelmingly defeats $\hat{f}_{JF}$ for densities (1) to (4), and remains a bit better elsewhere except when $f(0) = 0$. It has the edge on $\hat{f}_B$ (with bandwidth variation function) too for densities (1) to (4), loses out only a little when $f^{(1)}(0) \geq 0$, and any losing out when $f(0)$ is close to 0 is outweighted by the negativity of $\hat{f}_B$.

Appendix: Proofs

Lemma 1. Assume that $f^{(2)}(\cdot)$ and $g^{(3)}(\cdot)$ exist and continuous. Further, assume that $g^{-1}(0) = 0$, $g^{(1)}(0) = 1$, where g^{-1} is the inverse function of g, $f^{(i)}$ and $g^{(i)}$ are the i^{th} derivatives of f and g, respectively, $i \geq 0$. ($f^{(0)} = f$, $g^{(0)} = g$). Then for $x = ch$, $0 \leq c \leq 1$, we have

$$\begin{aligned}
\mathrm{E}\hat{f}_n(x) - f(x) &= h\int_{-1}^{c}(t-c)K(t)[2f^{(1)}(0) - g^{(2)}(0)f(0)] \\
&\quad + \frac{h^2}{2}f^{(2)}(0)[\int_{-1}^{c} t^2K(t)dt + \int_{-1}^{c}(t-c)^2K(t)dt] \\
&\quad - \frac{h^2}{2}\int_{-1}^{c}(t-c)^2K(t)dt\{g^{(3)}(0)f(0) \\
&\quad + 3g^{(2)}(0)[f^{(1)}(0) - g^{(2)}(0)f(0)]\} + o(h^2) \qquad \text{(A.1)}
\end{aligned}$$

and

$$\mathrm{Var}\hat{f}_n(x) = \frac{f(0)}{nh}[\int_{-1}^{1} K(t)^2dt + 2\int_{-1}^{c} K(t)K(2c-t)dt](1+o(1)). \qquad \text{(A.2)}$$

Proof: Note that

$$\begin{aligned}
\mathrm{E}\hat{f}_n(x) &= \frac{1}{h}\left[EK\left(\frac{x-X_1}{h}\right)+K\left(\frac{x+g(X_1)}{h}\right)\right] \\
&= f(x)\int_{-1}^{c}K(t)dt - hf^{(1)}(x)\int_{-1}^{c}tK(t)dt + \frac{f^{(2)}(x)}{2}h^2\int_{-1}^{c}t^2K(t)dt \\
&\quad +\frac{1}{h}EK\left(\frac{x+g(X_1)}{h}\right)+o(h^2) \qquad \text{(A.3)}
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{h}EK\left(\frac{x+g(X_1)}{h}\right) \\
&= \frac{1}{h}\int_0^\infty K\left(\frac{x+g(y)}{h}\right)f(y)dy \\
&= \int_c^1 K(t)\frac{f(g^{-1}(-x+ht))}{g^{(1)}(g^{-1}(-x+ht))}dt \\
&= \int_c^1 K(t)\frac{f(g^{-1}((t-c)h))}{g^{(1)}(g^{-1}((t-c)h))}dt \\
&= \int_c^1 K(t)\left\{\frac{f(g^{-1}(0))}{g^{(1)}(g^{-1}(0))}\right. \\
&+(t-c)h\frac{g^{(1)}(g^{-1}(0))f^{(1)}(g^{-1}(0))-g^{(2)}(g^{-1}(0))f(g^{-1}(0))}{[g^{(1)}(g^{-1}(0))]^3} \\
&+\frac{h^2}{2}(t-c)^2\left[\frac{g^{(1)}(g^{-1}(0))f^{(2)}(g^{-1}(0))-g^{(3)}(g^{-1}(0))f(g^{-1}(0))}{[g^{(1)}(g^{-1}(0))]^4}\right. \\
&\left.\left.-\frac{3g^{(2)}(g^{-1}(0))[g^{(1)}(g^{-1}(0))f^{(1)}(g^{-1}(0))-g^{(2)}(g^{-1}(0))f(g^{-1}(0))]}{[g^{(1)}(g^{-1}(0))]^5}\right]\right\} \\
&dt+o(h^2) \\
&= \int_c^1 K(t)dtf(0)+h\int_c^1(t-c)K(t)dt[f^{(1)}(0)-g^{(2)}(0)f(0)] \\
&+\frac{h^2}{2}\int_c^1(t-c)^2K(t)dt\{f^{(2)}(0)-g^{(3)}(0)f(0)-3g^{(2)}(0) \\
&[f^{(1)}(0)-g^{(2)}(0)f(0)]\}+o(h^2). \qquad \text{(A.4)}
\end{aligned}$$

Combine (A.3) and (A.4) to obtain

$$\mathrm{E}\hat{f}_n(x) = f(x)\int_{-1}^{c}K(t)dt + f(0)\int_c^1 K(t)dt$$

$$
\begin{aligned}
&+h\{-f^{(1)}(x)\int_{-1}^{c} tK(t)dt + \int_{c}^{1}(t-c)K(t)dt[f^{(1)}(0) - g^{(2)}(0)f(0)]\} \\
&+\frac{h^2}{2}f^{(2)}(x)\int_{-1}^{c} t^2K(t)dt + \frac{h^2}{2}\int_{c}^{1}(t-c)^2K(t)dt\{f^{(2)}(0) - g^{(3)}(0)f(0) \\
&-3g^{(2)}(0)[f^{(1)}(0) - g^{(2)}(0)f(0)]\} + o(h^2). \qquad \text{(A.5)}
\end{aligned}
$$

By the existence and continuity of $f^{(2)}(\cdot)$, for $x = ch$, we have

$$
f(0) = f(x) - chf^{(1)}(x) + \tfrac{c^2h^2}{2}f^{(2)}(x) + o(h^2)
$$

$$
f^{(1)}(x) = f^{(1)}(0) + chf^{(2)}(0) + o(h) \qquad \text{(A.6)}
$$

$$
f^{(2)}(x) = f^{(2)}(0) + o(1).
$$

By substituting (A.6) into (A.5), we obtain

$$
\begin{aligned}
\mathrm{E}\hat{f}_n(x) \quad = \quad & f(x) + h\int_{c}^{1}(t-c)K(t)dt[2f^{(1)}(0) - g^{(2)}(0)f(0)] \\
&+\frac{h^2}{2}f^{(2)}(0)[\int_{c}^{1} t^2K(t)dt + \int_{-1}^{c}(t-c)^2K(t)] \\
&-\frac{h^2}{2}\int_{c}^{1}(t-c)^2K(t)\{g^{(3)}(0)f(0) \\
&+3g^{(2)}(0)[f^{(1)}(0) - g^{(2)}(0)f(0)]\} + o(h^2).
\end{aligned}
$$

This completes the proof of (A.1). To prove (A.2), note that

$$
\begin{aligned}
\mathrm{Var}\hat{f}_n(x) \quad &= \quad \frac{1}{n^2h^2}\mathrm{Var}\left\{\sum_{i=1}^{n}\left[K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x+g(X_i)}{h}\right)\right]\right\} \\
&= \quad \frac{1}{n^2h^2}\mathrm{E}\left\{\sum_{i=1}^{n}\left\{\left[K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x+g(X_i)}{h}\right)\right]\right.\right. \\
&\quad \left.\left. -\mathrm{E}\left[K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x+g(X_i)}{h}\right)\right]\right\}\right\}^2 \\
&= \quad \frac{1}{n^2h^2}\mathrm{E}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x+g(X_i)}{h}\right)\right. \\
&\quad \left. -\mathrm{E}\left[K\left(\frac{h-X_i}{h}\right) + K\left(\frac{x+g(X_i)}{h}\right)\right]\right\}^2
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{nh^2}\mathrm{E}\left\{K\left(\frac{x-X_1}{h}\right)+K\left(\frac{x+g(X_1)}{h}\right)\right.\\
&\qquad \left.-\mathrm{E}\left[K\left(\frac{x-X_1}{h}\right)+K\left(\frac{x+g(X_1)}{h}\right)\right]\right\}^2\\
&= \frac{1}{nh^2}\mathrm{E}\left[K\left(\frac{x-X_1}{h}\right)+K\left(\frac{x+g(X_1)}{h}\right)\right]^2\\
&\qquad -\frac{1}{nh^2}\left\{\mathrm{E}\left[K\left(\frac{x-X_1}{h}\right)+K\left(\frac{x+g(X_1)}{h}\right)\right]\right\}^2\\
&= I_1+I_2,
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 &= \frac{1}{nh^2}\int\left[K\left(\frac{x-y}{h}\right)+K\left(\frac{x+g(y)}{h}\right)\right]^2 f(y)dy\\
&= \frac{1}{nh^2}\left[\int K\left(\frac{x-y}{h}\right)^2 f(y)dy+\int K\left(\frac{x+g(y)}{h}\right)^2 f(y)dy\right]\\
&\qquad +\frac{2}{nh^2}\int K\left(\frac{x-y}{h}\right)K\left(\frac{x+g(y)}{h}\right)f(y)dy\\
&= I_{11}+I_{12},
\end{aligned}
$$

with

$$
\begin{aligned}
I_{11} &= \frac{1}{nh^2}\left[h\int_{-1}^{c}K(t)^2 f(x-ht)dt+h\int_{c}^{1}K(t)^2\frac{f(g^{-1}(x-ht))}{g^{(1)}(g^{-1}(x-ht))}dt\right]\\
&= \frac{1}{nh}\left[\int_{-1}^{c}K(t)^2 f((c-t)h)dt+\int_{-1}^{c}K(t)^2\frac{f(g^{-1}((c-t)h))}{g^{(1)}(g^{-1}((c-t)h))}dt\right]\\
&= \frac{f(0)}{nh}\int_{-1}^{1}K(t)^2dt(1+o(1)).
\end{aligned}
$$

and

$$
\begin{aligned}
I_{12} &= \frac{2}{nh}\int_{-1}^{c}K(t)K\left(\frac{x+g(x-ht)}{h}\right)f(x-ht)dt\\
&= \frac{2}{nh}\int_{-1}^{c}K(t)K\left(\frac{x+g((c-t)h)}{h}\right)f((c-t)h)dt.
\end{aligned}
$$

Since $g^{(2)}(\cdot)$ is continuous in a neighborhood of 0, by a Taylor's expansion of order 2, we see that for some $0<\delta<1$,

$$
g((c-t)h) = g(0)+(c-t)hg^{(1)}(0)+\frac{(c-t)^2}{2}g^{(2)}(\delta(c-t)h)
$$

$$
\begin{aligned}
&= (c-t)h + \frac{(c-t)^2h^2}{2} g^{(2)}(\delta(c-t)h) \\
&= (c-t)h + O(h^2) \qquad \text{(A.7)}
\end{aligned}
$$

Substituting $g((c-t)h)$ defined by (A.7) into I_{12}, we have

$$
\begin{aligned}
I_{12} &= \frac{2}{nh}\int_{-1}^{c} K(t)K((2c-t)+O(h))f((c-t)h)dt \\
&= \frac{2f(0)}{nh}\int_{-1}^{c} K(t)K(2c-t)dt(1+o(1)).
\end{aligned}
$$

Now combine I_{11} and I_{12} to obtain

$$
I_1 = \frac{f(0)}{nh}\left[\int_{-1}^{1} K(t)^2dt + 2\int_{-1}^{c} K(t)K(2c-t)dt\right](1+o(1)).
$$

Similarly we can prove that

$$
\begin{aligned}
I_2 &= -\frac{1}{nh^2}\left\{\mathrm{E}\left[K\left(\frac{x-X_1}{h}\right)+K\left(\frac{x+g(X_1)}{h}\right)\right]\right\}^2 \\
&= O\left(\frac{1}{n}\right) = o\left(\frac{1}{nh}\right).
\end{aligned}
$$

Therefore,

$$
\mathrm{Var}\hat{f}_n(x) = \frac{f(0)}{nh}[\int_{-1}^{1} K(t)^2dt + 2\int_{-1}^{c} K(t)K(2c-t)dt](1+o(1)).
$$

This completes the proof.

Lemma 2. Let $f_n^*(h)$ and $f_n^*(0)$ be defined by (2.12). Suppose that $f^{(2)}(\cdot)$ is continuous near 0. Then

$$
\mathrm{E}|f_n^*(h) - f(h)|^k = O(h^{2k}) \qquad \text{(A.8)}
$$

$$
\mathrm{E}|f_n^*(0) - f(0)|^k = O(h^{2k}), \qquad \text{(A.9)}
$$

for any integer $k \geq 2$.

Proof: By the C_r-inequality,

$$\begin{aligned}
\mathrm{E}|f_n^*(h) - f(h)|^k &= \mathrm{E}|(f_n^*(h) - \mathrm{E}f_n^*(h)) + (\mathrm{E}f_n^*(h) - f(h))|^k \\
&\leq C_k\{\mathrm{E}|f_n^*(h) - \mathrm{E}f_n^*(h)|^k + |\mathrm{E}f_n^*(h) - f(h)|^k\} \\
&= C_k(\bar{I}_1 + \bar{I}_2), \qquad (\mathrm{A.10})
\end{aligned}$$

where C_k is a constant as a result of applying the C_r-inequality (Loève (1963), p. 157).

Since $\bar{I}_1 = \frac{1}{n^k h^k}\mathrm{E}\left|\sum_{i=1}^n \left[K\left(\frac{h-X_i}{h}\right) - \mathrm{E}K\left(\frac{h-X_i}{h}\right)\right]\right|^k$, applying the C_r-inequality repeatedly in $\bar{I}_1$, we obtain

$$\begin{aligned}
\bar{I}_1 &\leq \frac{C}{n^k h^k}\sum_{i=1}^n \mathrm{E}\left|K\left(\frac{h-X_i}{h}\right) - \mathrm{E}K\left(\frac{h-X_i}{h}\right)\right|^k \\
&= \frac{C}{n^{k-1}h^k}\mathrm{E}\left|K\left(\frac{h-X_1}{h}\right) - \mathrm{E}K\left(\frac{h-X_1}{h}\right)\right|^k \\
&\leq \frac{C}{n^{k-1}h^k}\left[C_k\mathrm{E}\left|K\left(\frac{h-X_1}{h}\right)\right|^k + C_k\left|\mathrm{E}K\left(\frac{h-X_1}{h}\right)\right|^k\right],
\end{aligned}$$

for some constant $C > 0$. Since

$$\begin{aligned}
\mathrm{E}\left|K\left(\frac{h-X_1}{h}\right)\right|^k &= \int K\left(\frac{h-y}{h}\right)^k f(y)dy \\
&= h\int_{-1}^1 K(t)^k f((1-t)h)dt \\
&= O(h)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}K\left(\frac{h-X_1}{h}\right) &= \int K\left(\frac{h-y}{h}\right) f(y)dy \\
&= O(h),
\end{aligned}$$

we have

$$\bar{I}_1 = O\left(\frac{1}{n^{k-1}h^{k-1}}\right) = O(h^{2k}). \qquad (\mathrm{A.11})$$

Also, since it is well-known that $Ef_n^*(h) - f(h) = \frac{h^2 f^{(2)}(0)}{2} \int_{-1}^{1} K(t)^2 dt + o(h^2)$, we have

$$\begin{aligned} \bar{I}_2 &= |Ef_n^*(h) - f(h)|^k \\ &= O(h^{2k}). \qquad \text{(A.12)} \end{aligned}$$

(A.8) is now proved by combining (A.10), (A.11) and (A.12). Similarly, we can prove (A.9).

Lemma 3. Let $f_n(h)$ and $f_n(0)$ be defined by (2.10) and (2.11). Suppose that $f^{(2)}(\cdot)$ is continuous in a neighborhood of 0. Then

$$E|f_n(h) - f(h)| = O(h^{2k}) \qquad \text{(A.13)}$$

$$E|f_n(0) - f(0)| = O(h^{2k}), \qquad \text{(A.14)}$$

for any integer $k \geq 2$.

Proof: The proof of (A.13) is obvious. We only prove (A.14). Note that

$$\begin{aligned} E|f_n(0) - f(0)|^k &= E|f_n(0) - f(0)|^k I_{[f_n^*(0) \geq \frac{1}{n^2}]} + E|f_n(0) - f(0)|^k I_{[f_n^*(0) < \frac{1}{n^2}]} \\ &= E|f_n^*(0) - f(0)|^k I_{[f_n^*(0) \geq \frac{1}{n^2}]} + E|\frac{1}{n^2} - f(0)|^k I_{[f_n^*(0) < \frac{1}{n^2}]} \\ &\leq E|f_n^*(0) - f(0)|^k + E|\frac{1}{n^2} - f_n^*(0) + f_n^*(0) - f(0)|^k I_{[0 < f_n^*(0) < \frac{1}{n^2}]} \\ &+ E|\frac{1}{n^2} - f(0)|^k I_{[f_n^*(0) \leq 0]} \\ &\leq E|f_n^*(0) - f(0)|^k + C_k E|\frac{1}{n^2} - f_n^*(0)|^k I_{[0 < f_n^*(0) < \frac{1}{n^2}]} \\ &+ C_k E|f_n^*(0) - f(0)|^k I_{[0 < f_n^*(0) < \frac{1}{n^2}]} + C_k \left(\frac{1}{n^2}\right)^k + C_k |f(0)|^k E I_{[f_n^*(0) \leq 0]} \\ &\leq (C_k + 1) E|f_n^*(0) - f(0)|^k + 2C_k \left(\frac{1}{n^2}\right)^k + C_k E|f_n^*(0) - f(0)|^k I_{[f_n^*(0) \leq 0]} \\ &\leq (2C_k + 1) E|f_n^*(0) - f(0)|^k + 2C_k \left(\frac{1}{n^2}\right)^k \\ &= O(h^{2k}), \end{aligned}$$

where C_k is a constant from applying the C_r-inequality. Hence the result.

Lemma 4. Let d_n be defined by (2.9). Assume that $f(x) > 0$ for $x = 0,\ h$ and $f^{(2)}(\cdot)$ is continuous near $x = 0$. Then

$$\mathrm{E}|d_n - d|^k = O(h^k), \tag{A.15}$$

for any integer $k \geq 2$.

Proof: Consider

$$\begin{aligned} \mathrm{E}|d_n - d|^k &\leq C_k \left\{ \mathrm{E}\left| d_n - \frac{\log f(h) - \log f(0)}{h} \right|^k + \left| \frac{\log f(h) - \log f(0)}{h} - d \right|^k \right\} \\ &= C_k[J_1 + J_2], \end{aligned} \tag{A.16}$$

where C_k is a constant again from applying the C_r-inequality. Note that

$$\begin{aligned} J_1 &= \mathrm{E}\left| \frac{\log f_n(h) - \log f_n(0)}{h} - \frac{\log f(h) - \log f(0)}{h} \right|^k \\ &\leq \frac{C_k}{h^k}\left[\mathrm{E}|\log f_n(h) - \log f(h)|^k + \mathrm{E}|\log f_n(0) - \log f(0)|^k \right]. \end{aligned}$$

By applying Taylor's expansion of the function $\log(\cdot)$ and by (A.13) and (A.14), for some $0 < \delta < 1$, we have for $x = 0,\ h$,

$$\begin{aligned} \mathrm{E}|\log f_n(x) - \log f(x)|^k &= \mathrm{E}\left| \frac{f_n(x) - f(x)}{f(x) + \delta(f_n(x) - f(x))} \right|^k \\ &\leq \frac{1}{((1-\delta)f(x))^k} \mathrm{E}|f_n(x) - f(x)|^k \\ &= O(h^{2k}). \end{aligned}$$

Therefore

$$J_1 = O(h^k). \tag{A.17}$$

Also, by Taylor's expansion of $\log f(\cdot)$, we have

$$\begin{aligned} J_2 &= \left| h \frac{f^{(2)}(0)f(0) - [f^{(1)}(0)]^2}{f(0)^2} + o(h) \right|^k \\ &= O(h^k). \end{aligned} \tag{A.18}$$

By combining (A.16), (A.17) and (A.18) complete the proof.

The proof of Theorem 1 is somewhat tedious because of the dependency of g_n on X_1, ..., X_n. The idea of our proof is as follows. First, for every X_i, we construct a transform $g_{ni}(x)$ which only depends on X_1, ..., X_{i-1}, X_{i+1}, ..., X_n. Using these g_{ni}'s, we obtain the pseudo data $-g_{n1}(X_1)$, ..., $-g_{nn}(X_n)$. Define

$$\bar{f}_{new}(x) = \frac{1}{nh} \sum_{i=1}^{n} \left\{ K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + g_{ni}(X_i)}{h}\right) \right\}. \tag{A.19}$$

It is obvious that g_{ni}, $i = 1, ..., n$ are very close to g_n. The only difference between g_{ni} and g_n is that g_{ni} depends on $n-1$ observations while g_n depends on n observations. Corresponding to d_n, we define

$$d_{ni} = \frac{\log f_{ni}(h) - \log f_{ni}(0)}{h}, \tag{A.20}$$

where

$$f_{ni}(h) = f^*_{ni}(h) + \frac{1}{n^2} \tag{A.21}$$

and

$$f_{ni}(0) = \max\left(f^*_{ni}(0), \frac{1}{n^2}\right) \tag{A.22}$$

with

$$f^*_{ni}(h) = \frac{1}{nh} \sum_{l=1, l \neq i}^{n} K\left(\frac{h - X_i}{h}\right), \; i = 1, \; ..., \; n. \tag{A.23}$$

$$f^*_{ni}(0) = \frac{1}{nh_0} \sum_{l=1, l \neq i}^{n} K_0\left(\frac{-X_l}{h_0}\right), \; i = 1, \; ..., \; n. \tag{A.24}$$

The definition of h, h_0 K and $K_{(0)}$ are the same as before. It can be seen that d_{ni} is independent of X_i. We state following lemmas without proofs. They are similar to Lemmas 2, 3 and 4.

Lemma 5. Let $f^*_{ni}(h)$ and $f^*_{ni}(0)$ be defined by (A.23) and (A.24), respectively. Suppose that $f^{(2)}(\cdot)$ is continuous in a neighborhood of 0. Then

$$\mathrm{E}|f^*_{ni}(h) - f(h)|^k = O(h^{2k}),\ i = 1,\ \ldots,\ n,$$

and

$$\mathrm{E}|f^*_{ni}(0) - f(0)|^k = O(h^{2k}),\ i = 1,\ \ldots,\ n,$$

for any integer $k \geq 2$.

Lemma 6. Let $f_{ni}(h)$ and $f_{ni}(0)$ be defined by (A.21) and (A.22), respectively. Suppose that $f^{(2)}(\cdot)$ is continuous in a neighborhood of 0. Then

$$\mathrm{E}|f_{ni}(h) - f(h)|^k = O(h^{2k}),\ i = 1,\ \ldots,\ n, \tag{A.25}$$

and

$$\mathrm{E}|f_{ni}(0) - f(0)|^k = O(h^{2k}),\ i = 1,\ \ldots,\ n, \tag{A.26}$$

for any integer $k \geq 2$.

Lemma 7. Let d_{ni} be defined by (A.20). Assume that $f(x) > 0$ for $x = 0$, h and $f^{(2)}(\cdot)$ is continuous near 0. Then

$$\mathrm{E}|d_{ni} - d_n|^k \ = \ O\left(\frac{1}{n^k h^{2k-1}}\right),\ i = 1,\ \ldots,\ n, \tag{A.27}$$

and

$$\mathrm{E}|d_{ni} - d|^k = O(h^k), \; i = 1, \; ..., \; n, \tag{A.28}$$

for any integer $k \geq 2$.

Proof of Theorem 1. For $x = ch$, $0 \leq c \leq 1$, we have

$$\begin{aligned} \mathrm{E}\hat{f}_{new}(x) - f(x) &= (\mathrm{E}\hat{f}_{new}(x) - \mathrm{E}\bar{f}_{new}(x)) + (\mathrm{E}\bar{f}_{new}(x) - \mathrm{E}\hat{f}_n(x)) \\ &\quad +(\mathrm{E}\hat{f}_n(x) - f(x)) \\ &= K_1 + K_2 + K_3, \end{aligned}$$

where $K_1 = \mathrm{E}\hat{f}_{new}(x) - \mathrm{E}\bar{f}_{new}(x)$, $K_2 = \mathrm{E}\bar{f}_{new}(x) - \mathrm{E}\hat{f}_n(x)$ and $K_3 = \mathrm{E}\hat{f}_n(x) - f(x)$. From applying Taylor's expansion, note that

$$\begin{aligned} |K_1| &= \frac{1}{nh}\left|\sum_{i=1}^{n} \mathrm{E}\left\{K\left(\frac{x+g_n(X_i)}{h}\right) - K\left(\frac{x+g_{ni}(X_i)}{h}\right)\right\}\right| \\ &\leq \frac{1}{nh}\sum_{i=1}^{n}\left|\mathrm{E}\left\{K\left(\frac{x+g_n(X_i)}{h}\right) - K\left(\frac{x+g_{ni}(X_i)}{h}\right)\right\}\right| \\ &= \frac{1}{nh}\sum_{i=1}^{n}\left|\mathrm{E}\left\{K^{(1)}\left(\frac{x+g_{ni}(X_i)}{h} + \delta\frac{g_n(X_i)-g_{ni}(X_i)}{h}\right)\frac{g_n(X_i)-g_{ni}(X_i)}{h}\right\}\right|, \end{aligned}$$

where $0 < \delta < 1$ is a constant.

Since for any d, $x \geq 0$ and $A > \frac{1}{3}$,

$$\begin{aligned} g(x) &= x + dx^2 + Ad^2x^3 \\ &= x(1 + dx + Ad^2x^2) \\ &= x\left[(\sqrt{A}dx + \frac{1}{2\sqrt{A}})^2 + \frac{4A-1}{4A}\right] \\ &\geq \frac{4A-1}{4A}x, \end{aligned}$$

it is easy to see that $g(x) \geq h$ for $x \geq ph$, where $p = \frac{4A}{4A-1}$. Hence

$$\begin{aligned} |K_1| &\leq \frac{1}{nh}\sum_{i=1}^{n}\left|\mathrm{E}\left\{\frac{g_n(X_i)-g_{ni}(X_i)}{h}K^{(1)}\left(\frac{x+(1-\delta)g_{ni}(X_i)+\delta g_n(X_i)}{h}\right)\right.\right. \\ &\quad \left.\left. I_{[0\leq X_i \leq ph]}\right\}\right| \\ &\leq \frac{C}{nh^2}\sum_{i=1}^{n}|\mathrm{E}(g_n(X_i) - g_{ni}(X_i))I_{[0\leq X_i\leq ph]}|, \end{aligned}$$

where C is a constant (in different positions, it may take different values). It is obvious that

$$\begin{aligned} \mathrm{E}|g_n(X_i) - g_{ni}(X_i))|I_{[0\le X_i\le ph]} &= \mathrm{E}|(d_n - d_{ni})X_i^2|I_{[0\le X_i\le ph]} \\ &\quad + A(d_n^2 - d_{ni}^2)X_i^3 I_{[0\le X_i\le ph]} \\ &\le p^2h^2\mathrm{E}|d_n - d_{ni}| + Ap^3h^3\mathrm{E}|d_n^2 - d_{ni}^2|. \end{aligned}$$

Now (A.27) implies that

$$\begin{aligned} \mathrm{E}|d_n - d_{ni}| &\le (\mathrm{E}|d_n - d_{ni}|^2)^{\frac{1}{2}} \\ &= O\left(\frac{1}{nh^{3/2}}\right). \end{aligned} \tag{A.29}$$

Therefore, by (2.15), (A.27),(A.28) and (A.29), we have

$$\begin{aligned} \mathrm{E}|d_n^2 - d_{ni}^2| &= \mathrm{E}|d_n - d_{ni}||d_n + d_{ni}| \\ &= \mathrm{E}(|d_n - d_{ni}||d_n - d + d_{ni} - d + 2d|) \\ &\le \mathrm{E}(|d_n - d_{ni}||d_n - d|) + \mathrm{E}(|d_n - d_{ni}||d_{ni} - d|) + 2d\mathrm{E}|d_n - d_{ni}| \\ &\le (\mathrm{E}|d_n - d_{ni}|^2)^{\frac{1}{2}}(\mathrm{E}|d_n - d|^2)^{\frac{1}{2}} \\ &\quad + (\mathrm{E}|d_n - d_{ni}|^2)^{\frac{1}{2}}(\mathrm{E}|d_{ni} - d|^2)^{\frac{1}{2}} + 2d\mathrm{E}|d_n - d_{ni}| \\ &= O\left(\frac{1}{nh^{3/2}}\right). \end{aligned}$$

Then

$$\mathrm{E}|g_n(X_i) - g_{ni}(X_i))|I_{[0\le X_i\le ph]} = O\left(\frac{h^{1/2}}{n}\right). \tag{A.30}$$

Consequently,

$$\begin{aligned} |K_1| &= O\left(\frac{1}{nh^{3/2}}\right) \\ &= o(h^2). \end{aligned} \tag{A.31}$$

Similar to the proof of Lemma 2.1, we obtain

$$\mathrm{E}\bar{f}_{new}(x) = f(x)\int_{-1}^{c} K(t)dt - hf^{(1)}(x)\int_{-1}^{c} tK(t)dt$$
$$+\frac{f^{(2)}(x)}{2}h^2\int_{-1}^{c} t^2K(t)dt + \frac{1}{nh}\sum_{i=1}^{n}\mathrm{E}K\left(\frac{x+g_{ni}(X_i)}{h}\right) + o(h^2). \quad \text{(A.32)}$$

Note that

$$\mathrm{E}K\left(\frac{x+g_{n1}(X_1)}{h}\right) = \mathrm{E}\left\{\mathrm{E}\left[K\left(\frac{x+g_{n1}(X_1)}{h}\right)\bigg| d_{n1}\right]\right\} \quad \text{(A.33)}$$

and

$$\begin{aligned}
\mathrm{E}\left[K\left(\frac{x+g_{n1}(X_1)}{h}\right)\bigg| d_{n1}\right] &= \int_0^\infty K\left(\frac{x+g_{n1}(y)}{h}\right) f_{X_1|d_{n1}}(y)dy \\
&= \int_0^\infty K\left(\frac{x+g_{n1}(y)}{h}\right) f(y)dy \\
&= \frac{h\int_c^1 K(t)f(g_{n1}^{-1}(t-c)h)dt}{g_{n1}^{(1)}((g_{n1}^{-1}(t-c)h))}
\end{aligned}$$

$$\begin{aligned}
= \; & h\int_c^1 K(t)\left[\frac{f(g_{n1}^{-1}(0)}{g_{n1}^{(1)}(g_{n1}^{-1}(0))}\right. \\
& +(t-c)h\frac{g_{n1}^{(1)}(g_{n1}^{-1}(0))f^{(1)}(g_{n1}^{-1}(0)) - g_{n1}^{(2)}(g_{n1}^{-1}(0))f(g_{n1}^{-1}(0))}{[g_{n1}^{(1)}(g_{n1}^{-1}(0))]^3} \\
& +\frac{h^2(t-c)^2}{2}\left\{\frac{g_{n1}^{(1)}(g_{n1}^{-1}(\theta))f^{(2)}(g_{n1}^{-1}(\theta)) - g_{n1}^{(3)}(g_{n1}^{-1}(\theta))f(g_{n1}^{-1}(\theta))}{[g_{n1}^{(1)}(g_{n1}^{-1}(\theta))]^4}\right. \\
& \left.\left. - \frac{3g_{n1}^{(2)}(g_{n1}^{-1}(\theta))[g_{n1}^{(1)}(g_{n1}^{-1}(\theta))f^{(1)}(g_{n1}^{-1}(\theta)) - g_{n1}^{(2)}(g_{n1}^{-1}(\theta))f(g_{n1}^{-1}(\theta))]}{[g_{n1}^{(1)}((g_{n1}^{-1}(\theta)]^5}\right\}\right] dt \\
= \; & hf(0)\int_c^1 K(t)dt + h^2\int_c^1 (t-c)K(t)dt[f^{(1)}(0) - 2d_{n1}f(0)] \\
& +\frac{h^3}{2}\int_c^1 (t-c)^2K(t)\Delta dt,
\end{aligned}$$

where $0 \le \theta \le (t-c)h \le h$.

By the monotonicity of g_{n1} and the fact $g_{n1}(x) \ge qx$, $q = (4A-1)/(4A)$, we know that for any $0 \le \theta^* \le h$, $g_{n1}^{-1}(\theta^*) \le \frac{h}{q} \to 0$, as $n \to \infty$. Since $g_{n1}^{(1)}(x) =$

$1 + 2d_{n1}x + 3Ad_{n1}^2x^2 \geq \frac{3A-1}{3A} > 0$, for $A > 1/3$, $g_{n1}^{(2)}(x) = 2d_{n1} + 6Ad_{n1}^2x$ and $g_{n1}^{(3)}(x) = 6Ad_{n1}^2$, simple algebra leads to $|\Delta| \leq \Delta_1 = O(|d_{n1}|^k)$ for some $k \geq 2$. By (A.28), $\mathrm{E}|d_{n1}^k| = \mathrm{E}|d_{n1} - d + d|^k \leq C_k\mathrm{E}|d_{n1} - d|^k + C_k|d|^k = O(h^k) + C_k|d|^k$. Therefore, for sufficiently large n, $\mathrm{E}\Delta_1$ is bounded. By (A.33), $\mathrm{E}K\left(\frac{x+g_{n1}(X_1)}{h}\right) = hf(0)\int_c^1 K(t)dt + h^2\int_c^1(t-c)K(t)dt[f^{(1)}(0) - 2\mathrm{E}d_{n1}f(0)] + \frac{h^3}{2}\mathrm{E}\int_c^1(t-c)^2K(t)\Delta dt$. Since $|\mathrm{E}\int_c^1(t-c)^2K(t)\Delta dt| \leq \mathrm{E}\int_c^1(t-c)^2K(t)|\Delta|dt \leq \int_c^1(t-c)^2K(t)dt\mathrm{E}\Delta_1$, $\mathrm{E}\int_c^1(t-c)^2K(t)\Delta dt$ is bounded. Then

$$\mathrm{E}K\left(\frac{x+g_{n1}(X_1)}{h}\right) = hf(0)\int_c^1 K(t)dt$$
$$+h^2\int_c^1(t-c)K(t)dt[f^{(1)}(0) - 2\mathrm{E}d_{n1}f(0)] + O(h^3). \qquad \text{(A.34)}$$

Substituting (A.34) into (A.32), we have

$$\mathrm{E}\bar{f}_{new}(x) = f(x)\int_{-1}^c K(t)dt - hf^{(1)}(x)\int_{-1}^c tK(t)dt + h^2\frac{f^{(2)}(x)}{2}\int_{-1}^c t^2K(t)dt$$
$$+f(0)\int_c^1 K(t)dt + h\int_c^1(t-c)K(t)dt\left[f^{(1)}(0) - 2f(0)\frac{\sum_{i=1}^n d_{ni}}{n}\right] + O(h^2).$$

(A.5) shows that

$$\begin{aligned}\mathrm{E}\hat{f}_n(x) &= f(x)\int_{-1}^c K(t)dt - hf^{(1)}(x)\int_{-1}^c tK(t)dt + h^2\frac{f^{(2)}(x)}{2}\int_{-1}^c t^2K(t)dt \\ &\quad +f(0)\int_c^1 K(t)dt + h\int_c^1(t-c)K(t)dt[f^{(1)}(0) - 2f(0)d] + O(h^2).\end{aligned}$$

Hence

$$\begin{aligned}|K_2| &= |\mathrm{E}\bar{f}_{new}(x) - \mathrm{E}\hat{f}_n(x)| \\ &= \left| -2h\int_c^1(t-c)K(t)dtf(0)\frac{\sum_{i=1}^n(\mathrm{E}d_{ni} - d)}{n}\right| \\ &= O(h^2),\end{aligned}$$

since $|\mathrm{E}d_{ni} - d| \leq \mathrm{E}|d_{ni} - d| \leq (\mathrm{E}|d_{ni} - d|^2)^{1/2} = O(h)$ by (A.28). But (2.8) proved that $|K_3| = O(h^2)$. Now combining above results complete the proof of (2.18).

Now we prove (2.19). Using straightforward calculations,

$$\begin{aligned}
\mathrm{Var}\hat{f}_{new}(x) &= \frac{1}{n^2h^2}\mathrm{Var}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h}\right)+K\left(\frac{x+g_n(X_i)}{h}\right)\right\} \\
&= \frac{1}{n^2h^2}\mathrm{Var}\left\{\sum_{i=1}^{n}\left[K\left(\frac{x-X_i}{h}\right)+K\left(\frac{x+g(X_i)}{h}\right)\right]\right. \\
&\quad +\sum_{i=1}^{n}\left[K\left(\frac{x+g_n(X_i)}{h}\right)-K\left(\frac{x+g_{ni}(X_i)}{h}\right)\right] \\
&\quad \left.+\sum_{i=1}^{n}\left[K\left(\frac{x+g_{ni}(X_i)}{h}\right)-K\left(\frac{x+g(X_i)}{h}\right)\right]\right\} \\
&\leq \frac{4}{n^2h^2}\mathrm{Var}\sum_{i=1}^{n}\left[K\left(\frac{x-X_i}{h}\right)+K\left(\frac{x+g(X_i)}{h}\right)\right] \\
&\quad +\frac{4}{n^2h^2}\mathrm{Var}\sum_{i=1}^{n}\left[K\left(\frac{x+g_n(X_i)}{h}\right)-K\left(\frac{x+g_{ni}(X_i)}{h}\right)\right] \\
&\quad +\frac{4}{n^2h^2}\mathrm{Var}\sum_{i=1}^{n}\left[K\left(\frac{x+g_{ni}(X_i)}{h}\right)-K\left(\frac{x+g(X_i)}{h}\right)\right] \qquad (A.35) \\
&= M_1+M_2+M_3,
\end{aligned}$$

where M_i $(i = 1, 2, 3)$ denotes the i^{th} term on the right hand side of the inequality (A.35). Note that (2.3) shows that $M_1 = O(\frac{1}{nh})$. Then it is enough to prove that $M_2 = O(\frac{1}{nh})$ and $M_3 = O(\frac{1}{nh})$. Consider

$$\begin{aligned}
M_2 &\leq \frac{4}{n^2h^2}\mathrm{E}\left[\sum_{i=1}^{n}\left\{K\left(\frac{x+g_n(X_i)}{h}\right)-K\left(\frac{x+g_{ni}(X_i)}{h}\right)\right\}\right]^2 \\
&= \frac{4}{n^2h^2}\mathrm{E}\left\{\sum_{i=1}^{n}\frac{g_n(X_i)-g_{ni}(X_i)}{h}K^{(1)}\left(\frac{x+g_{ni}(X_i)-\alpha(g_n(X_i)-g_{ni}(X_i))}{h}\right)\right\}^2 \\
&\leq \frac{4}{n^2h^2}\sum_{i=1}^{n}\mathrm{E}[g_n(X_i)-g_{ni}(X_i)]^2\left[K^{(1)}\left(\frac{x+\alpha g_n(X_i)+(1-\alpha)g_{ni}(X_i)}{h}\right)\right]^2 \\
&\leq \frac{C}{n^2h^2}\sum_{i=1}^{n}\mathrm{E}[g_n(X_i)-g_{ni}(X_i)]^2 I_{[0\leq X_i\leq ph]},
\end{aligned}$$

where $0 < \alpha < 1$ and C is a positive constant. Similar to (A.30), we can prove

$$\mathrm{E}[g_n(X_i)-g_{ni}(X_i)]^2 I_{[0\leq X_i\leq ph]} = O\left(\frac{h}{n^2}\right).$$

Then

$$M_2 = O\left(\frac{1}{n^2h^3}\right) = o\left(\frac{1}{nh}\right).$$

Again by straightforward algebra,

$$
\begin{aligned}
M_3 &= \frac{4}{n^2h^2}\mathrm{Var}\sum_{i=1}^{n}\frac{g_{ni}(X_i)-g(X_i)}{h}K^{(1)}\left(\frac{x+\delta g_{ni}(X_i)+(1-\delta)g(X_i)}{h}\right)\\
&\le \frac{4}{n^2h^4}\mathrm{E}\left\{\sum_{i=1}^{n}(g_{ni}(X_i)-g(X_i))I_{[0\le X_i\le ph]}K^{(1)}\left(\frac{x+\delta g_{ni}(X_i)+(1-\delta)g(X_i)}{h}\right)\right\}^2\\
&= \frac{4}{n^2h^4}\mathrm{E}\sum_{i=1}^{n}[g_{ni}(X_i)-g(X_i]^2I_{[0\le X_i\le ph]}\left[K^{(1)}\left(\frac{x+\delta g_{ni}(X_i)+(1-\delta)g(X_i)}{h}\right)\right]^2\\
&+\frac{8}{n^2h^4}\sum_{1\le i<j\le n}\mathrm{E}(g_{ni}(X_i)-g(X_i))(g_{nj}(X_j)-g(X_j))I_{[0\le X_i\le ph]}I_{[0\le X_j\le ph]}\\
&K^{(1)}\left(\frac{x+\delta g_{ni}(X_i)+(1-\delta)g(X_i)}{h}\right)K^{(1)}\left(\frac{x+\delta g_{nj}(X_j)+(1-\delta)g(X_j)}{h}\right)\\
&\le \frac{4}{n^2h^4}\sum_{i=1}^{n}\mathrm{E}[g_{ni}(X_i)-g(X_i)]^2I_{[0\le X_i\le ph]}\\
&+\frac{8}{n^2h^4}\sum_{1\le i<j\le n}\{\mathrm{E}\left[(g_{ni}(X_i)-g(X_i))\right.\\
&\left.(g_{nj}(X_j)-g(X_j))I_{[0\le X_i\le ph]}I_{[0\le X_j\le ph]}\right]^{\frac{1+\alpha}{\alpha}}\Big\}^{\frac{\alpha}{1+\alpha}}\\
&\Big\{\mathrm{E}\left[K^{(1)}\left(\frac{x+\delta g_{ni}(X_i)+(1-\delta)g(X_i)}{h}\right)\right.\\
&\left.K^{(1)}\left(\frac{x+\delta g_{nj}(X_j)+(1-\delta)g(X_j)}{h}\right)\right]^{1+\alpha}\Big\}^{\frac{1}{1+\alpha}}\\
&= \frac{4}{n^2h^4}\sum_{i=1}^{n}M_{3i}+\frac{8}{n^2h^4}\sum_{1\le i<j\le n}(M_{3ij})^{\frac{\alpha}{1+\alpha}}(M'_{3ij})^{\frac{1}{1+\alpha}}, \qquad \text{(A.36)}
\end{aligned}
$$

where α is any constant which is greater than zero.

For any i, j, consider

$$
\begin{aligned}
M_{3ij} &= \mathrm{E}\Big\{[((d_{ni}-d)X_i^2+AX_i^3(d_{ni}^2-d^2)]\\
&\quad [(d_{nj}-d)X_j^2+AX_j^3(d_{nj}^2-d^2)]I_{[0\le X_i\le ph]}I_{[0\le X_j\le ph]}\Big\}^{\frac{1+\alpha}{\alpha}}\\
&= \mathrm{E}\Big\{[((d_{ni}-d)(d_{nj}-d)X_i^2X_j^2+A(d_{ni}^2-d^2)(d_{nj}-d)X_i^3X_j^2\\
&\quad +A(d_{ni}-d)(d_{nj}^2-d^2)X_i^2X_j^3\\
&\quad +A^2X_i^3X_j^3(d_{ni}^2-d^2)(d_{nj}^2-d^2)]I_{[0\le X_i\le ph]}I_{[0\le X_j\le ph]}\Big\}^{\frac{1+\alpha}{\alpha}}
\end{aligned}
$$

$$\leq C_1 h^{\frac{4(1+\alpha)}{\alpha}} \mathrm{E}\left[|d_{ni}-d||d_{nj}-d|\right]^{\frac{1+\alpha}{\alpha}} + C_2 h^{\frac{5(1+\alpha)}{\alpha}} \mathrm{E}\left[|d_{ni}-d||d_{nj}^2-d^2|\right]^{\frac{1+\alpha}{\alpha}}$$
$$+C_3 h^{\frac{5(1+\alpha)}{\alpha}} \mathrm{E}\left[|d_{ni}^2-d^2||d_{nj}-d|\right]^{\frac{1+\alpha}{\alpha}} + C_4 \mathrm{E}\left[|d_{ni}^2-d^2||d_{nj}^2-d^2|\right]^{\frac{1+\alpha}{\alpha}},$$

where C_1, C_2, C_3 and C_4 are positive constants. It is easy to see that

$$\mathrm{E}\left[|d_{ni}-d||d_{nj}^2-d^2|\right]^{\frac{1+\alpha}{\alpha}} = O\left(\mathrm{E}\left[|d_{ni}-d||d_{nj}-d|\right]^{\frac{1+\alpha}{\alpha}}\right)$$

$$\mathrm{E}\left[|d_{ni}^2-d^2||d_{nj}^2-d^2|\right]^{\frac{1+\alpha}{\alpha}} = O\left(\mathrm{E}\left[|d_{ni}-d||d_{nj}-d|\right]^{\frac{1+\alpha}{\alpha}}\right)$$

$$\mathrm{E}\left[|d_{ni}-d||d_{nj}-d|\right]^{\frac{1+\alpha}{\alpha}} = O\left(h^{\frac{2(1+\alpha)}{\alpha}}\right).$$

Therefore

$$M_{3ij} = O\left(h^{\frac{6(1+\alpha)}{\alpha}}\right). \tag{A.37}$$

Since for $i = 1, ..., n$, there exists a positive constant α_i such that

$$K^{(1)}\left(\frac{x+\delta g_{ni}(X_i)+(1-\delta)g(X_i)}{h}\right) = K^{(1)}\left(\frac{x+g(X_i)}{h}\right)$$
$$+\delta\frac{(g_{ni}(X_i)-g(X_i))}{h}K^{(2)}\left(\frac{x+g(X_i)}{h}+\frac{\alpha_i\delta(g_{ni}(X_i)-g(X_i))}{h}\right),$$

we obtain

$$\begin{aligned} M'_{3ij} &= M'_{3ij} I_{[0\leq X_i\leq ph]} I_{[0\leq X_j\leq ph]} \\ &= \mathrm{E}\left[K^{(1)}\left(\frac{x+g(X_i)}{h}\right)+\delta\frac{(g_{ni}(X_i)-g(X_i))}{h}\right. \\ &\quad \left. K^{(2)}\left(\frac{x+g(X_i)}{h}+\frac{\alpha_i\delta(g_{ni}(X_i)-g(X_i))}{h}\right)\right]^{1+\alpha} \\ &\quad \left[K^{(1)}\left(\frac{x+g(X_j)}{h}\right)+\delta\frac{(g_{nj}(X_j)-g(X_j))}{h}\right. \\ &\quad \left. K^{(2)}\left(\frac{x+g(X_j)}{h}+\frac{\alpha_j\delta(g_{nj}(X_i)-g(X_j))}{h}\right)\right]^{1+\alpha} I_{[0\leq X_i\leq ph]} I_{[0\leq X_j\leq ph]} \end{aligned}$$

$$
\begin{aligned}
&\le C_1 \mathrm{E}\left|K^{(1)}\left(\frac{x+g(X_i)}{h}\right)K^{(1)}\left(\frac{x+g(X_j)}{h}\right)\right|^{1+\alpha} \\
&\quad +C_2\mathrm{E}\left|K^{(1)}\left(\frac{x+g(X_i)}{h}\right)\delta\frac{(g_{nj}(X_j)-g(X_j))}{h}\right. \\
&\quad \left. K^{(2)}\left(\frac{x+g(X_j)}{h}+\frac{\alpha_j\delta(g_{nj}(X_i)-g(X_j))}{h}\right)\right|^{1+\alpha} I_{[0\le X_j\le ph]} \\
&\quad +C_3\mathrm{E}\left|K^{(1)}\left(\frac{x+g(X_j)}{h}\right)\delta\frac{(g_{ni}(X_i)-g(X_i))}{h}\right. \\
&\quad \left. K^{(2)}\left(\frac{x+g(X_i)}{h}+\frac{\alpha_i\delta(g_{ni}(X_i)-g(X_i))}{h}\right)\right|^{1+\alpha} I_{[0\le X_i\le ph]} \\
&\quad +C_4\mathrm{E}\left|\delta^2\frac{(g_{ni}(X_i)-g(X_i))(g_{nj}(X_i)-g(X_j))}{h^2}\right. \\
&\quad K^{(2)}\left(\frac{x+g(X_i)}{h}+\frac{\alpha_i\delta(g_{ni}(X_i)-g(X_i))}{h}\right) \\
&\quad \left. K^{(2)}\left(\frac{x+g(X_j)}{h}+\frac{\alpha_j\delta(g_{nj}(X_i)-g(X_j))}{h}\right)\right|^{1+\alpha} I_{[0\le X_i\le ph]}I_{[0\le X_j\le ph]} \\
&\le C_1 \mathrm{E}\left|K^{(1)}\left(\frac{x+g(X_i)}{h}\right)K^{(1)}\left(\frac{x+g(X_j)}{h}\right)\right|^{1+\alpha} \\
&\quad +C_2h^{-(1+\alpha)}\mathrm{E}\left|(g_{nj}(X_j)-g(X_j))I_{[0\le X_j\le ph]}\right|^{1+\alpha} \\
&\quad +C_3h^{-(1+\alpha)}\mathrm{E}\left|(g_{ni}(X_i)-g(X_i))I_{[0\le X_i\le ph]}\right|^{1+\alpha} \\
&\quad +C_4h^{-2(1+\alpha)}\mathrm{E}\left|(g_{ni}(X_i)-g(X_i))(g_{nj}(X_j)-g(X_j))\right. \\
&\quad \left. I_{[0\le X_i\le ph]}I_{[0\le X_j\le ph]}\right|^{1+\alpha} \\
&= L_1+L_2+L_3+L_4, \qquad \text{(A.38)}
\end{aligned}
$$

where L_i ($i = 1, ..., 4$) denotes the i^{th}-term on the right hand side of (A.38). Consider

$$
\begin{aligned}
L_1 &= c_1\int\left|K^{(1)}\left(\frac{x+g(y_i)}{h}\right)K^{(1)}\left(\frac{x+g(y_j)}{h}\right)\right|^{1+\alpha} f(y_1)\cdots f(y_n)dy_1\cdots dy_n \\
&= c_1h^2\int\left|K^{(1)}(t_i)K^{(1)}(t_j)\right|^{1+\alpha}\frac{f(g^{-1}(ht_i-x))}{g^{(1)}(g^{-1}(ht_i-x))}\frac{f(g^{-1}(ht_j-x))}{g^{(1)}(g^{-1}(ht_j-x))} \\
&\quad f(y_1)\cdots f(y_{i-1})f(y_{i+1})\cdots f(y_{j-1})f(y_{j+1})\cdots f(y_n)dy_1\cdots dy_n \\
&\le c_1\left(\frac{3A}{3A-1}\right)^2\int\left|K^{(1)}(t_i)K^{(1)}(t_j)\right|^{1+\alpha} f(g^{-1}(ht_i-x))f(g^{-1}(ht_j-x))dy_idy_j
\end{aligned}
$$

$$= O(h^2).$$

By (A.28), it is easy to see that

$$L_2 = O(h^{2(1+\alpha)}).$$

Similarly we can prove

$$L_3 = O(h^{2(1+\alpha)}) \text{ and } L_4 = O(h^{4(1+\alpha)}).$$

Therefore by (A.38) and the above results,

$$M'_{3ij} = O(h^2). \tag{A.39}$$

(A.36) combining with (A.37) and (A.39) leads to

$$M_3 = O(h^{2+\frac{2}{1+\alpha}}) \sim O(h^4) = O\left(\frac{1}{nh}\right), \tag{A.40}$$

since α can be arbitrarily small. Now combine the results on M_1, M_2 and M_3 to complete the proof of (2.19).

References

Cheng, M. Y., Fan, J., and Marron, J. S. (1995), "On Automatic Boundary Corrections," Technical report.

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

Cowling, A., and Hall, P. (1996), "On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation," *Journal of the Royal Statiscal Society*, Series B, 58, 551-563.

Gasser, T., and Müller, H. G. (1979), "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation*, Lecture Notes in Matehmatics 757, eds. T. Gasser and M. Rosenblatt, Heidelberg: Springer-Verla g, pp. 23-68.

Gasser, T., Müller, H. G., and Mammitzsch, V. (1985), " Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society*, Series B, 47, 238-256.

Jones, J. C. (1993), "Simple boundary correction for kernel density estimation" *Statist. Comput.*, 3, 135-146.

Jones, J. C., and Foster, P. J. (1995), "A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation," Technical report.

Loève, M. (1963). *Probalility Theory* I, 4^{th} ed., Springer-Verlag, New York.

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

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

Schuster, E. F. (1985), "Incorporating Support Constraints into Nonparametric Estimators of Densities," *Communications in Statistics- Theory and Methods*, 14, 1123-1136.

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

Zhang, S., and Karunamuni, R. J. (1995), " On Kernel Density Estimation Near Endpoints with an Application to Line Transect Sampling," Technical Report, Department of Mathematical Sciences, Statistics Center, Univ. of Alberta.

Zhang, S., and Karunamuni, R. J. (1996), "On Nonparametric Density Estimation at the Boundary," Technical Report, Department of Mathematical Sciences, Statistics Center, Univ. of Alberta.

TABLE 1

With 500 repetitions, the global optimal bandwidth was used.

1. For the model f(x)=exp(-x), f'(0)<0

	Bias	Var	MSE
New method	-0.04454214	0.007457099	0.009426187
Boundary kernel method	-0.07177556	0.01111894	0.01624844
C & H method	-0.05825679	0.008871693	0.0122478
J & F method	-0.006896141	0.03411436	0.03409369

2. For the model f(x)=5exp(-5x), f'(0)<0

	Bias	Var	MSE
New method	-0.2227107	0.1864274	0.2356547
Boundary kernel method	-0.3588778	0.2779735	0.4062109
C & H method	-0.2912841	0.2217923	0.3061951
J & F method	-0.03448071	0.8528591	0.8523423

3. For the model f(x)=15exp(-15x), f'(0)<0

	Bias	Var	MSE
New method	-0.6681323	1.677847	2.120892
Boundary kernel method	-1.076634	2.501762	3.655898
C & H method	-0.8737638	1.995983	2.755454
J & F method	-0.1034421	7.675731	7.67108

4. For the model f(x)=6(1-x)^5, f'(0)<0

	Bias	Var	MSE
New method	-0.1861442	0.1360133	0.1703909
Boundary kernel method	-0.3479006	0.2468442	0.3673853
C & H method	-0.6440013	0.1460156	0.5604613
J & F method	0.02200559	0.7895212	0.7884264

5. For the model f(x)=sqrt(2/pi)exp(-x^2/2), f'(0)=0

	Bias	Var	MSE
New method	0.09082599	0.01349354	0.02171592
Boundary kernel method	0.05161071	0.01002975	0.01267336
C & H method	0.01826397	0.01525244	0.01555551
J & F method	0.02728512	0.02267466	0.02337379

6. For the model f(x)=exp(-x)/2+xexp(-x)/2, f'(0)=0

	Bias	Var	MSE
New method	0.0477912	0.00516555	0.007439219
Boundary kernel method	0.02323203	0.004166113	0.004697508
C & H method	0.03753807	0.007836216	0.00922965
J & F method	0.02570031	0.01019376	0.01083388

7. For the model f(x)=exp(-x)/4+3xexp(-x)/4, f'(0)>0

	Bias	Var	MSE
New method	0.06465811	0.0045507	0.008722269
Boundary kernel method	0.0580326	0.004872798	0.008230835
C & H method	0.06956179	0.008274144	0.01309644
J & F method	0.04134119	0.00756584	0.009259803

8. For the model f(x)=5*exp(-5x)/4+75/4*x*exp(-5x), f'(0)>0

	Bias	Var	MSE
New method	0.332224	0.1194492	0.2295831
Boundary kernel method	0.3012219	0.1230829	0.1230829
C & H method	0.3738701	0.2396643	0.3789639
J & F method	0.196568	0.2030669	0.2412998

9. For the model f(x)=(5*exp(-5x)+36*x*exp(-6x))/2, f'(0)>0

	Bias	Var	MSE
New method	0.2939535	0.2308418	0.3167888
Boundary kernel method	0.2209048	0.1775493	0.2259932
C & H method	0.4152819	0.4969886	0.6684536
J & F method	0.1599774	0.3615393	0.386409

10. For the model f(x)=x*exp(-x), f(0)=0, 5% of the estimates from boundary kernel method are negative.

	Bias	Var	MSE
New method	0.1095876	0.001382451	0.01338913
Boundary kernel method	0.07477185	0.002321634	0.007907819
C & H method	0.1100184	0.001066763	0.01316868
J & F method	0.07747921	0.001576301	0.007576176

11. For the model f(x)=x^2*exp(-x)/2, f(0)=0, 85% of the estimates from boundary kernel method are negative.

	Bias	Var	MSE
New method	0.021378	7.386959e-05	0.0005307406
Boundary kernel method	-0.01810241	0.000332969	0.0006600002
C & H method	0.02672185	8.986076e-05	0.0008037386
J & F method	0.01108219	5.065929e-05	0.0001733728

12. For the model f(x)=x^4*exp(-x)/24, f(0)=0, 100% of the estimates from boundary kernel method are negative.

	Bias	Var	MSE
New method	0.002667142	3.297202e-06	1.040425e-05
Boundary kernel method	-0.02237713	3.486008e-05	0.0005355261
C & H method	0.004770325	4.361464e-06	2.710874e-05
J & F method	0.000905797	1.460361e-06	2.277909e-06

13. For the model f(x)=1/2/sqrt(x) with subjective bandwidth h = 0.18

	Mean	Var
New method	3.250777	0.04907193
Boundary kernel method	4.001732	0.1853219
C & H method	2.874924	0.08416587
J & F method	6.832344	1.67953

Figure 1. Estimates of 5exp(-5x) (density (2)) with the optimal global bandwidth = 0.1368511.

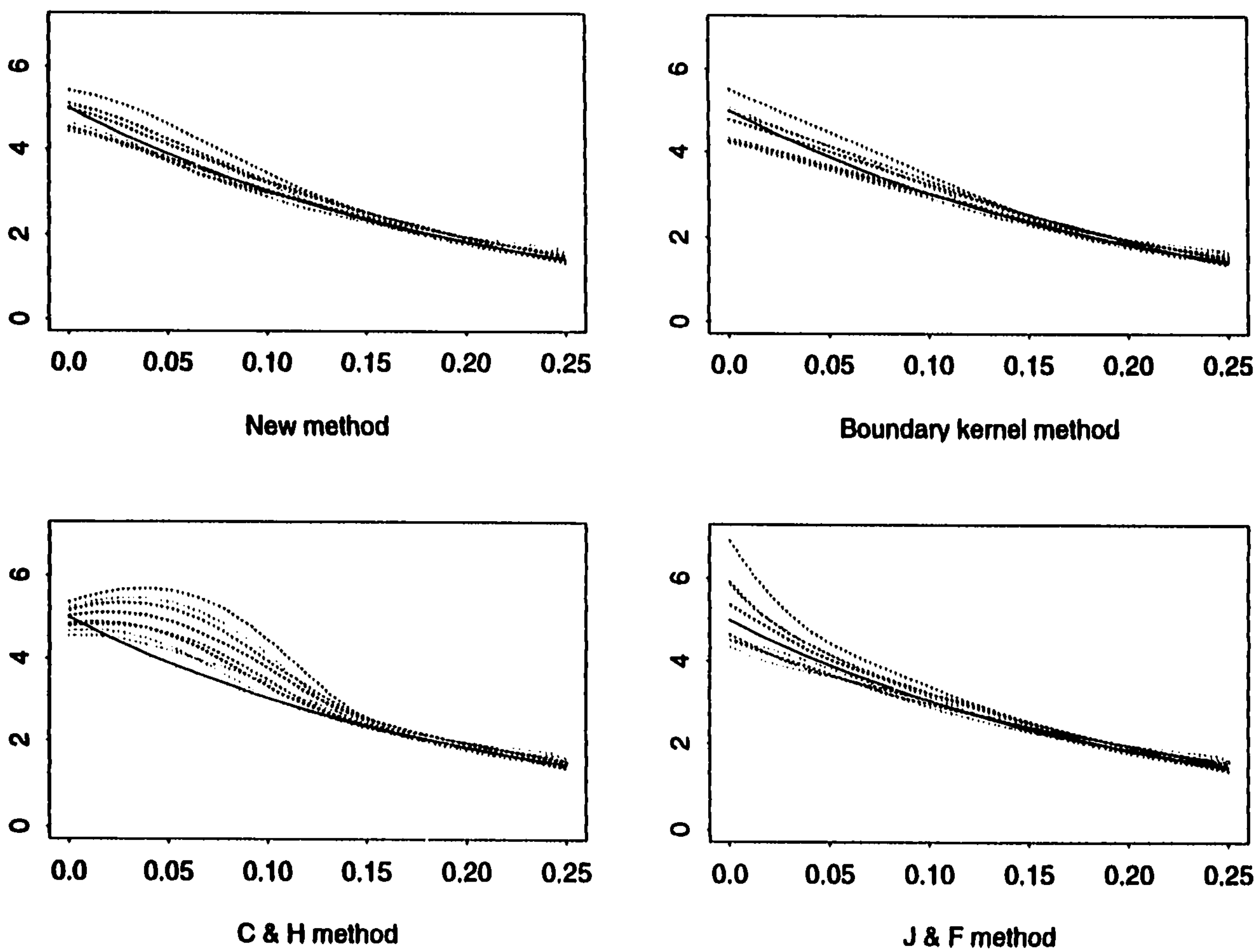


Figure 2. Estimates of exp(-x)/2+x*exp(-x)/2 (density (6)) with the optimal global bandwidth = 1.037137.

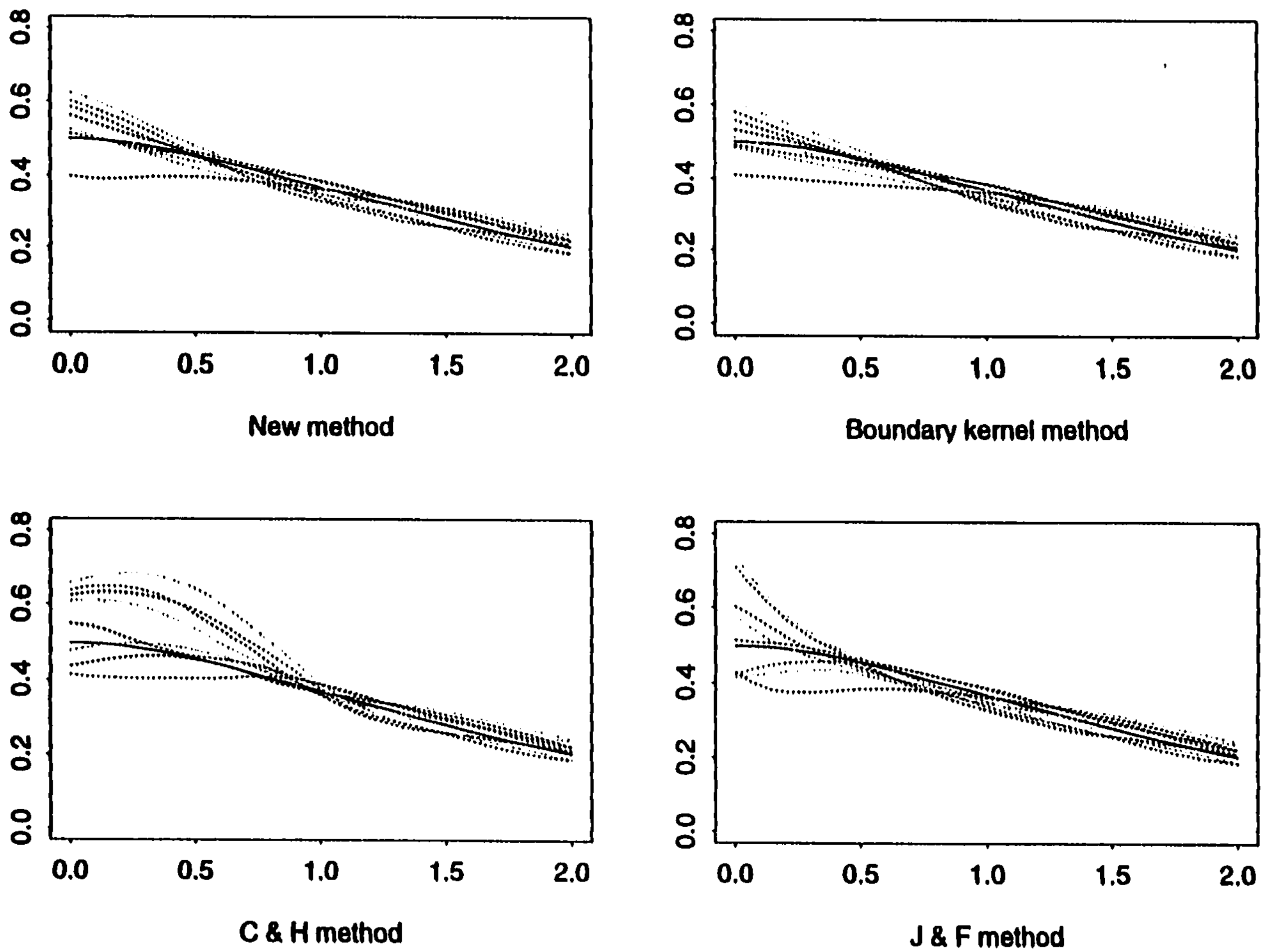


Figure 3. Estimates of (5exp(-5x)+36*x*exp(-6x))/2 (density (9)) with the optimal global bandwidth = 0.1453146.

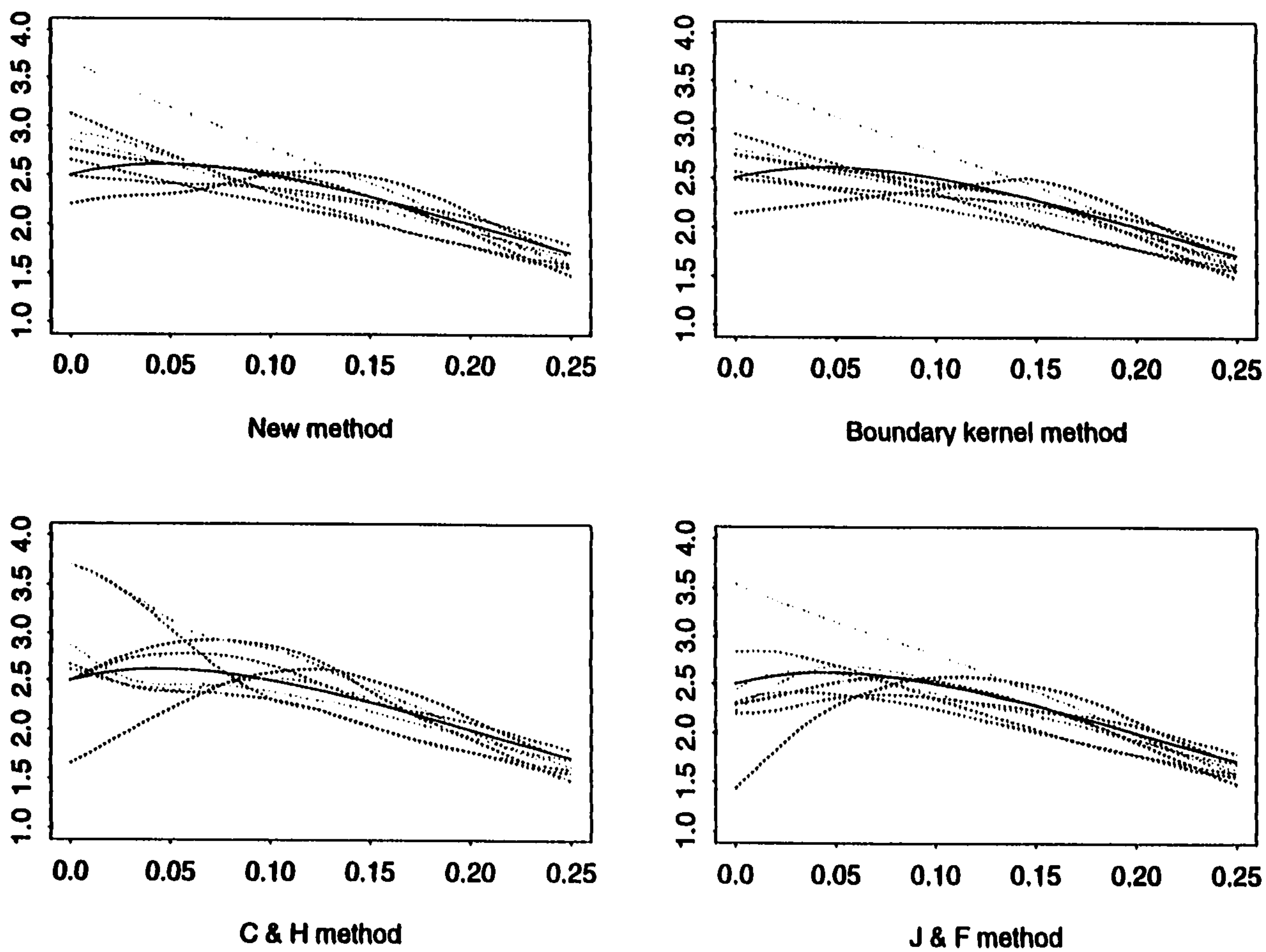


Figure 4. Estimates of x^4exp(-x)/24 (density (12) with the optimal global bandwidth = 1.449559.

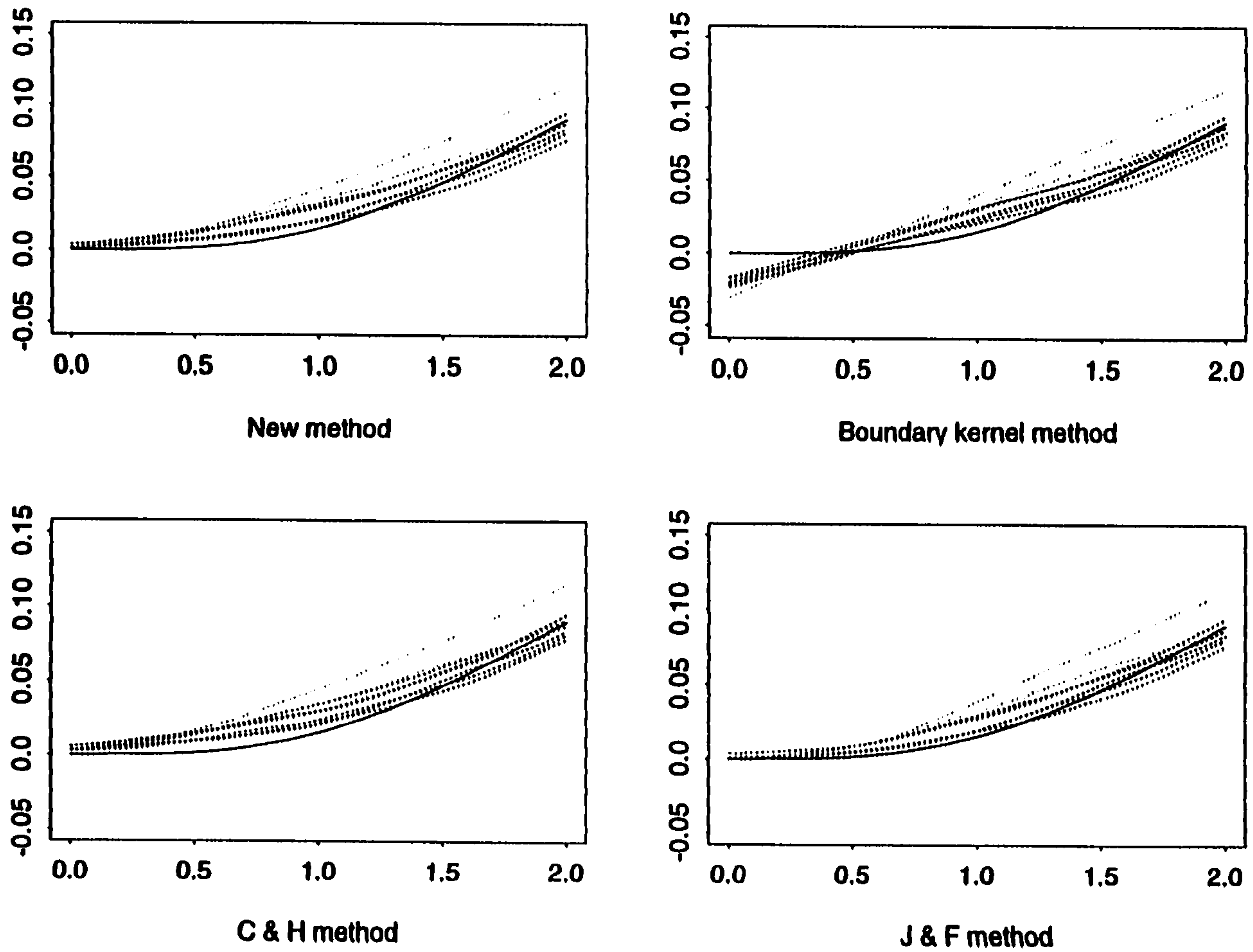


Figure 5. Estimates of 1/2/sqrt(x) (density (13)) with bandwidth = 0.18.