Ann. Inst. Statist. Math.
Vol. 52, No. 4, 612–629 (2000)

BOUNDARY BIAS CORRECTION FOR NONPARAMETRIC DECONVOLUTION*

SHUNPU ZHANG** AND ROHANA J. KARUNAMUNI

Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

(Received September 10, 1997; revised February 22, 1999)

Abstract. In this paper we consider the deconvolution problem in nonparametric density estimation. That is, one wishes to estimate the unknown density of a random variable X, say f_X, based on the observed variables Y's, where $Y = X + \epsilon$ with ϵ being the error. Previous results on this problem have considered the estimation of f_X at interior points. Here we study the deconvolution problem for boundary points. A kernel-type estimator is proposed, and its mean squared error properties, including the rates of convergence, are investigated for supersmooth and ordinary smooth error distributions. Results of a simulation study are also presented.

Key words and phrases: Deconvolution, density estimation, boundary effects, bandwidth variation.

1. Introduction

The deconvolution problem in nonparametric density estimation has received considerable attention in recent years. Deconvolution arises when direct observation is not possible due to the measurement error or the nature of the environment. The basic model is as follows. One wishes to estimate the unknown density of a random variable X, but the only data available are observations $Y_1, \dots, Y_n$, which are contaminated with independent additive error ϵ, from the model $Y = X + \epsilon$. In density function forms, the problem is to estimate $f_X(x)$ using data $Y_1, \dots, Y_n$ from the density

$$f_Y(y) = \int f_X(y - x) dF_\epsilon(x), \tag{1.1}$$

where F_ϵ is a known cumulative distribution function of ϵ. The conventional kernel estimator of $f_X^{(l)}$, the l-th derivative of f_X, is defined as

$$\hat{f}_n^{(l)}(x) = \frac{1}{nh^{l+1}} \sum_{j=1}^{n} g_n^{(l)}\left(\frac{x - Y_j}{h}\right) \tag{1.2}$$

if the function $\phi_K(t)/\phi_\epsilon(t/h)$ is integrable, where

$$g_n^{(l)}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (-it)^l \exp(-itx) \frac{\phi_K(t)}{\phi_\epsilon(t/h)} dt \tag{1.3}$$

* The research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

** Now at Department of Mathematical Sciences, University of Alaska Fairbanks, Fairbanks, AK 99775-6660, U.S.A.

with $h = h(n)$ is the bandwidth ($h \to 0$ as $n \to \infty$), ϕ_K is the Fourier transform of a kernel function K (which needs to be chosen) and ϕ_ϵ is the characteristic function of ϵ.

There have been a number of papers in literature setting out various properties of (1.2) with different rates of convergence relative to important discrepancy "measures" and under variety of conditions on K, h and the smoothness of f_X and F_ϵ. See, for example, the work of Carroll and Hall (1988), Devroye (1989), Diggle and Hall (1993), Fan (1991a, 1991b, 1991c, 1992), Liu and Taylor (1989), Mendelsohn and Rice (1982), Penskaya (1988), Piterbarg and Penskaya (1993), Stenfanski (1990), Stenfanski and Carroll (1990) and Zhang (1990), among others. Barry and Diggle (1995) considered the problem of choosing the bandwidth for nonparametric deconvolution.

The above papers discuss the estimation of f_X at its interior points. However, it is of theoretical and practical interest to study the deconvolution problem for "truncated" density functions at boundary points as well. Our motivation to study the problem stems from our experience with the following application. Biometricians and ecologists often needs to estimate the density or abundance of biological populations such as animals in an area; see, e.g., the monograph by Buckland *et al.* (1993). Line and point transect sampling are the primary "distance" methods that are employed in such situations. In a typical application, an observer walks a straight path of length L, noting all the animals seen (n) and their right-angle distances from the transect line (X). Given various assumptions which can be found in Buckland *et al.* (1993), it can be shown that an estimator of density of the animal population (D) is given by $\hat{D} = n\hat{f}_X(0)/2L$, where $\hat{f}_X(0)$ is an estimate of $f_X(0)$ with f_X being the probability density function of X, the right-angle distance from the line. Note that the support of f_X is $[0, \infty)$. The preceding formula of $\hat{D}$ is based on the assumption that the perpendicular distances are recorded without error. From a practical point of view, however, it is reasonable to assume the presence of measurement errors. So, the basic problem becomes an estimation problem of $f_X(0)$ based on the Y's such that $Y = X + \epsilon$, where Y's and ϵ's are the recorded right-angle distances and the measurement errors, respectively. There are more examples that would fit in with this paper such as lifetimes for survival data.

The above discussion shows that a careful study of the deconvolution problem for truncated densities should be of considerable interest in applications. To achieve the above is precisely the objective of this paper. We shall assume that the support of f_X is $[0, \infty)$, although our estimator of f_X can be easily modified for compact support cases as well. Finite endpoints of the support add a degree of complexity to the development of consistent estimators. In Section 2, we propose the estimator of f_X. The asymptotic properties of our proposed estimator are given in Section 3. We obtain expressions for the mean squared error (MSE) of our estimator, including the rates of convergence, for two types of error distributions: ordinary smooth and supersmooth distributions. We argue that our rates are in the best possible form. Section 4 discusses the bandwidth choice problem in the present context. The special case of $f_X^{(1)}(0) = 0$ is examined in Section 5. Finally, Section 6 contains results of a simulation study.

2. The estimator

In this section, we study the asymptotic properties of (1.2). The performance of (1.2) will depend on the smoothness of the error distribution and of the density function f_X.

ASSUMPTION 1. (i) $|\phi_\epsilon(t)| > 0$, for all t. (ii) $f_X^{(k)}(x)$ is continuous on $[0, \infty)$.

Condition (i) ensures that the estimator (1.2) is well defined. Condition (ii) is analogous to that required in the ordinary density estimation. The assumptions on the kernel K are stated on its Fourier domain:

ASSUMPTION 2.
(1) $\phi_K(t)$ is a symmetric function, having $m+2$ bounded integrable derivatives on $(-\infty, \infty)$;
(2) $\phi_K(t) = 1 + O(|t|^m)$ as $t \to 0$,
where m is a positive integer.

Note that Assumption 2 implies that K is a k-th-order kernel satisfying

$$\int_{-\infty}^{\infty} K(t)dt = 1,$$
$$\int_{-\infty}^{\infty} t^j K(t)dt = 0, \qquad \text{for} \quad j = 1, \dots, m-1,$$
$$\int_{-\infty}^{\infty} t^m K(t)dt \neq 0.$$

Assume that a bandwidth h is used in (1.2), then for $0 \le x = ch$, $c \ge 0$,

$$\begin{aligned} E\hat{f}_n^{(l)}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} (-it)^l \exp(-itx)\phi_K(th)\phi_X(t)dt \\ &= \int_{-\infty}^{c} f_X^{(l)}(x-hy)K(y)dy \\ &= f_X^{(l)}(x)\int_{-\infty}^{c} K(t)dt - hf_X^{(l+1)}(x)\int_{-\infty}^{c} tK(t)dt + \cdots \\ &\quad + (-h)^{k-l}\frac{f_X^{(k)}(x)}{(k-l)!}\int_{-\infty}^{c} t^{k-l}K(t)dt + o(h^{k-l}). \end{aligned} \tag{2.1}$$

If the support of K is $[-1,1]$, then for $0 \le c < 1$, $\int_{-\infty}^{c} K(t)dt < 1$, $\int_{-\infty}^{c} t^j K(t)dt \neq 0$, for $j = 1, \dots, k-l-1$. Therefore, $\hat{f}_n^{(l)}(x)$ is not asymptotically consistent for $f_X^{(l)}(x)$ for any $0 \le x < h$. Such points are called boundary points. If the support of K is $(-\infty, \infty)$, we can see that $\hat{f}_n^{(l)}(x)$ is not a consistent estimator of $f_X^{(l)}(x)$ for all $x = ch$, $c \ge 0$.

Remark 1. From (2.1), it seems that an order $(0, k-l)$ boundary kernel K which satisfies

$$\int_{-\infty}^{c} K(t)dt = 1,$$
$$\int_{-\infty}^{c} t^j K(t)dt = 0, \qquad \text{for} \quad j = 1, \dots, k-l-1,$$
$$\int_{-\infty}^{c} t^{k-l} K(t)dt \neq 0,$$

would be sufficent to guarantee the consistency of the estimator (1.2). Unfortunately, such a kernel with asymmetric support $(-\infty, c]$ would yield an unsymmetric ϕ_K (the Fourier transform of K). This will make (1.2) take complex values.

In view of the fact that

$$E\hat{f}_n^{(l+1)}(x) = f_X^{(l+1)}(x)\int_{-\infty}^{c} K(t)dt - hf_X^{(l+2)}(x)\int_{-\infty}^{c} tK(t)dt + \cdots + (-h)^{k-l-1}\frac{f_X^{(k)}(x)}{(k-l-1)!}\int_{-\infty}^{c} t^{k-l-1}K(t)dt + o(h^{k-l-1}).$$

$$\cdots$$

$$E\hat{f}_n^{(k-1)}(x) = f_X^{(k-1)}(x)\int_{-\infty}^{c} K(t)dt - hf_X^{(k)}(x)\int_{-\infty}^{c} tK(t)dt + o(h),$$

we define our new estimator as a linear combination of $\hat{f}_n^{(l)}, \ldots, \hat{f}_n^{(k-1)}$. That is,

$$\tilde{f}^{(l)}(x) = \sum_{j=l}^{k-1}(-h)^{j-l}a_{j-l}\hat{f}_n^{(j)}(x), \tag{2.2}$$

where $a_0, \ldots, a_{k-l-1}$ are coefficients to be determined in order to remove boundary effects.

It is easy to see that

$$\begin{aligned} E\tilde{f}_n^{(l)}(x) = {} & a_0 f_X^{(l)}(x)\int_{-\infty}^{c} K(t)dt - \cdots + a_0(-h)^{k-l}\frac{f_X^{(k)}(x)}{(k-l)!}\int_{-\infty}^{c} t^{k-l}K(t)dt \\ & - a_1 h f_X^{(l+1)}(x)\int_{-\infty}^{c} K(t)dt + \cdots \\ & + a_1(-h)^{k-l}\frac{f_X^{(k)}(x)}{(k-l-1)!}\int_{-\infty}^{c} t^{k-l-1}K(t)dt + \cdots \\ & + a_{k-l-1}(-h)^{k-l-1}f_X^{(k-1)}(x)\int_{-\infty}^{c} K(t)dt \\ & + a_{k-l-1}(-h)^{k-l}f_X^{(k)}(x)\int_{-\infty}^{c} tK(t)dt + o(h^{k-l}). \end{aligned} \tag{2.3}$$

So, if $a_0, \ldots, a_{k-l-1}$ satisfy

$$\begin{cases} a_0\int_{-\infty}^{c} K(t)dt = 1 \\ a_0\int_{-\infty}^{c} tK(t)dt + a_1\int_{-\infty}^{c} K(t)dt = 0, \\ \cdots \\ a_0\frac{\int_{-\infty}^{c} t^{k-l-1}K(t)dt}{(k-l-1)!} + a_1\frac{\int_{-\infty}^{c} t^{k-l-2}K(t)dt}{(k-l-2)!} + \cdots + a_{k-l-1}\int_{-\infty}^{c} K(t)dt = 0, \end{cases} \tag{2.4}$$

then $\tilde{f}_n^{(l)}(x)$ has the usual bias expansion. Write

$$b_i = \frac{\int_{-\infty}^{c} t^i K(t)dt}{i!}, \qquad i = 0, 1, \ldots, k-l.$$

Then (2.4) can be written as

$$\begin{pmatrix} b_0 & 0 & \cdots & 0 \\ b_1 & b_0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ b_{k-l-1} & b_{k-l-2} & \cdots & b_0 \end{pmatrix}\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-l-1} \end{pmatrix} = e_1, \tag{2.5}$$

where $e_1 = (1, 0, \ldots, 0)^T$, T indicates the transpose of a vector or a matrix. Denote the matrix of the left-hand side of (2.5) by S_c^T. Then

$$(a_0, a_1, \ldots, a_{k-l-1}) = e_1^T S_c^{-1}. \tag{2.6}$$

Note that S_c is invertible since $\int_{-\infty}^{c} K(t)dt \neq 0$. Therefore, for $x \geq 0$, the boundary corrected estimator of $f_X^{(l)}(x)$ is

$$\begin{aligned}
\tilde{f}_n^{(l)}(x) &= (a_0, a_1, \ldots, a_{k-l-1}) \operatorname{Diag}(1, -h, \ldots, (-h)^{k-l-1}) \begin{pmatrix} \hat{f}_n^{(l)}(x) \\ \hat{f}_n^{(l+1)}(x) \\ \vdots \\ \hat{f}_n^{(k-1)}(x) \end{pmatrix} \\
&= e_1^T S_c^{-1} \operatorname{Diag}(1, -h, \ldots, (-h)^{k-l-1}) \begin{pmatrix} \hat{f}_n^{(l)}(x) \\ \hat{f}_n^{(l+1)}(x) \\ \vdots \\ \hat{f}_n^{(k-1)}(x) \end{pmatrix} \\
&= e_1^T S_c^{-1} \operatorname{Diag}(1, -h, \ldots, (-h)^{k-l-1}) \begin{pmatrix} \frac{1}{nh^{l+1}} \sum_{j=1}^{n} g_n^{(l)}\left(\frac{x-Y_j}{h}\right) \\ \frac{1}{nh^{l+2}} \sum_{j=1}^{n} g_n^{(l+1)}\left(\frac{x-Y_j}{h}\right) \\ \vdots \\ \frac{1}{nh^{k}} \sum_{j=1}^{n} g_n^{(k-1)}\left(\frac{x-Y_j}{h}\right) \end{pmatrix} \\
&= \frac{1}{nh^{l+1}} \sum_{j=1}^{n} e_1^T S_c^{-1} \begin{pmatrix} g_n^{(l)}\left(\frac{x-Y_j}{h}\right) \\ -g_n^{(l+1)}\left(\frac{x-Y_j}{h}\right) \\ \vdots \\ (-1)^{k-l-1} g_n^{(k-1)}\left(\frac{x-Y_j}{h}\right) \end{pmatrix}.
\end{aligned} \tag{2.7}$$

Remark 2. For the case that the support of f_X has a right endpoint b, boundary points are $b - ch$, $0 \leq c \leq 1$, if the support of the kernel K is $[-1, 1]$. The estimator (2.7) can be modified to this case simply by replacing the integral interval $(-\infty, c]$ in the entries of S_c by $[-c, \infty)$, see circa (2.5).

3. Asymptotic results

In this section, we study the asymptotic properties of the estimator (2.7) proposed in Section 2. For this purpose, we need following lemmas.

LEMMA 1. *Assume that Assumptions* 1 *and* 2 *hold with* $m = k - l$. *Further assume that* ϕ_K *has a bounded support* $[-M_0, M_0]$ *for some positive constant* M_0. *Then, for* $x = ch$, $c \geq 0$, *we have*

$$E\hat{f}_n^{(l)}(x) - f_X^{(l)}(x) = (-h)^{k-l} e_1^T S_c^{-1} S_c^* f_X^{(k)}(x)(1 + o(1)), \tag{3.1}$$

where $S_c^* = (b_{k-l}, b_{k-l-1}, \dots, b_0)^T$.

PROOF. Note that

$$E\tilde{f}_n^{(l)}(x) = \frac{1}{h^{l+1}} e_1^T S_c^{-1} \begin{pmatrix} Eg_n^{(l)}\left(\frac{x-Y_j}{h}\right) \\ -Eg_n^{(l+1)}\left(\frac{x-Y_j}{h}\right) \\ \vdots \\ (-1)^{k-l-1} Eg_n^{(k-l)}\left(\frac{x-Y_j}{h}\right) \end{pmatrix}$$

$$= \frac{1}{h^{l+1}} e_1^T$$

$$\cdot S_c^{-1} \begin{pmatrix} h^{l+1}[b_0 f_X^{(l)}(x) - hb_1 f_X^{(l+1)}(x) + \cdots + (-h)^{k-l} b_{k-l} f_X^{(k)}(x) + o(h^{k-l})] \\ -h^{l+2}[b_0 f_X^{(l+1)}(x) - hb_1 f_X^{(l+2)}(x) + \cdots + (-h)^{k-l-1} b_{k-l-1} f_X^{(k)}(x) + o(h^{k-l-1})] \\ \vdots \\ (-1)^{k-l-1} h^k [b_0 f_X^{(k-1)}(x) - hb_1 f_X^{(k)}(x) + o(h)] \end{pmatrix}$$

$$= \frac{1}{h^{l+1}} e_1^T S_c^{-1} \begin{pmatrix} b_0 & b_1 & \cdots & b_{k-l} \\ 0 & b_0 & \cdots & b_{k-l-1} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & b_1 \end{pmatrix} \begin{pmatrix} h^{l+1} f_X^{(l)}(x) \\ h^{l+2} f_X^{(l+1)}(x) \\ \vdots \\ (-1)^{k-l} h^{k+1} f_X^{(k)}(x) \end{pmatrix} (1 + o(1))$$

$$= e_1^T (I, S_c^{-1} S_c^*) \begin{pmatrix} f_X^{(l)}(x) \\ -h f_X^{(l+1)}(x) \\ \vdots \\ (-1)^{k-l} h^{k-l} f_X^{(k)}(x) \end{pmatrix} (1 + o(1))$$

$$= [f_X^{(l)}(x) + (-h)^{k-l} e_1^T S_c^{-1} S_c^* f_X^{(k)}(x)](1 + o(1)).$$

Therefore,

$$E\tilde{f}_n^{(l)}(x) - f_X^{(l)}(x) = (-h)^{k-l} e_1^T S_c^{-1} S_c^* f_X^{(k)}(x)(1 + o(1)).$$

This completes the proof of Lemma 1.

Remark 3. For estimating $f_X^{(l)}$, the kernel needs to be only of the order $(0, k-l)$, not of the order $(0, k)$ as claimed by Fan (1991*a*).

We now discuss the variance problem. Fan (1991*a*) defined two classes of errors and showed that the variance heavily depends on the tail of the distribution of error variable ϵ. The two classes of errors are defined as follows:

I. ϵ is supersmooth of order β, if

$$d_0 |t|^{\beta_0} \exp(-|t|^\beta/\gamma) \le |\phi_\epsilon(t)| \le d_1 |t|^{\beta_1} \exp(-|t|^\beta/\gamma), \quad \text{as} \quad t \to \infty,$$

for some positive constants d_0, d_1, β, γ and constants β_0, β_1, and

II. ϵ is ordinary smooth of order β, if

$$d_0 |t|^{-\beta} \le |\phi_\epsilon(t)| \le d_1 |t|^{-\beta}, \quad \text{as} \quad t \to \infty,$$

for some positive constants d_0, d_1 and β.

First, let's discuss the supersmooth error case. The following lemma is a direct result of Fan (1991*a*, 1991*b*).

LEMMA 2. *Assume that Assumptions* 1 *and* 2 *hold with* $m = k - l$. *Further assume that* ϵ *is supersmooth of order* β *and* ϕ_K *has a bounded support* $[-M_0, M_0]$ *for some positive constant* M_0. *Then, by choosing the bandwidth* $h = w(\log n)^{-1/\beta}$ *with* $w > M_0(2/\gamma)^{1/\beta}$, *for* $x = ch$, $c \geq 0$, *we have*

$$\operatorname{Var} \tilde{f}_n^{(l)}(x) = o(h^{2(k-l)}). \tag{3.2}$$

When the error is ordinary smooth, we have the following lemma.

LEMMA 3. *Assume that Assumptions* 1 *and* 2 *hold with* $m = k-l$. *Further assume that* ϵ *is ordinary smooth of order* β, *and*

(i) $|\phi_\epsilon^{(1)}(t)t^{\beta+1}| = O(1)$, *as* $t \to \infty$,

(ii) $\phi_\epsilon(t)t^\beta \to \alpha$ *as* $t \to \infty$ *for some constant* $\alpha \neq 0$,

(iii) $|t|^{k+\beta-1}|\phi_K(t)| \to 0$ *and* $\int_{-\infty}^{\infty}[\phi_K(t) + |\phi_K^{(1)}(t)|]|t|^{\beta+k-1}dt < \infty$,

(iv) $\int_{-\infty}^{\infty}[\int_{-\infty}^{\infty} \exp(-itu)(-it)^{j+l}t^\beta\phi_K(t)dt][\int_{-\infty}^{\infty} \exp(-itu)(-it)^{m+l}t^\beta\phi_K(t)dt]du < \infty$, *for* $0 \leq j$, $m \leq k-l-1$. *Then, for* $x = ch$, $c \geq 0$, *we have*

$$\operatorname{Var} \tilde{f}_n^{(l)}(x) = \frac{f_Y(x)}{nh^{2(l+\beta)+1}} e_1^T S_c^{-1} M (S_c^{-1})^T e_1 (1 + o(1)), \tag{3.3}$$

where $M = (m_{jm})_{0 \leq j,\, m \leq k-l-1}$ *and*

$$m_{jm} = \frac{(-1)^{j+m}}{(2\pi\alpha)^2} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \exp(-itu)(-it)^{j+l}t^\beta\phi_K(t)dt\right] \cdot \left[\int_{-\infty}^{\infty} \exp(-itu)(-it)^{m+l}t^\beta\phi_K(t)dt\right] du.$$

PROOF. Note that

$$\operatorname{Var} \tilde{f}_n^{(l)}(x) = \frac{1}{nh^{2l+2}} e_1^T S_c^{-1} \cdot \left(\operatorname{cov}\left((-1)^j g_n^{(l+j)}\left(\frac{x - Y_1}{h}\right), (-1)^m g_n^{(l+m)}\left(\frac{x - Y_1}{h}\right)\right)\right)_{0 \leq j,\, m \leq k-l-1} \cdot (S_c^{-1})^T e_1.$$

Also observe that

$$\begin{aligned} &\operatorname{cov}\left((-1)^j g_n^{(l+j)}\left(\frac{x - Y_1}{h}\right), (-1)^m g_n^{(l+m)}\left(\frac{x - Y_1}{h}\right)\right) \\ &= (-1)^{m+j}\left[E g_n^{(l+j)}\left(\frac{x - Y_1}{h}\right) g_n^{(l+m)}\left(\frac{x - Y_1}{h}\right) - E g_n^{(l+j)}\left(\frac{x - Y_1}{h}\right) E g_n^{(l+m)}\left(\frac{x - Y_1}{h}\right)\right]. \end{aligned}$$

Simple algebra leads to

$$E g_n^{(l+w)}\left(\frac{x - Y_1}{h}\right) = h^{1+l+w} f_X^{(l+w)}(x) \int_{-\infty}^{c} K(t)dt(1 + o(1)), \quad \text{for} \quad 0 \leq w \leq k-l-1. \tag{3.4}$$

Since ϵ is ordinary smooth, there exist positive constants c_1, M and β such that $|\phi_\epsilon(t)| \geq c_1|t|^{-\beta}$ for $|t| > M$. By (iii), for any $0 \leq r \leq k-l-1$, we have

$$\begin{aligned}
(3.5) \quad |h^\beta g_n^{(l+r)}(u)| &\leq \frac{h^\beta}{2\pi}\int_{-\infty}^{\infty}\frac{|t|^{l+r}|\phi_K(t)|}{|\phi_\epsilon(t/h)|}dt \\
&\leq \frac{h^\beta}{2\pi}\left[\int_{|t|\leq Mh}\frac{|t|^{l+r}|\phi_K(t)|}{|\phi_\epsilon(t/h)|}dt + h^{-\beta}\int_{|t|>Mh}\frac{|t|^{l+r+\beta}|\phi_K(t)|}{c_1}dt\right] \\
&\leq \frac{h^\beta}{2\pi\min|\phi_\epsilon(t)|}\int_{-\infty}^{\infty}|t|^{l+r}|\phi_K(t)|dt + \frac{1}{2c_1\pi}\int_{-\infty}^{\infty}|t|^{l+r+\beta}|\phi_K(t)|dt \\
&= O(1).
\end{aligned}$$

Also, by integration by parts and assumptions (i) and (iii), we have

$$\begin{aligned}
(3.6) \quad |h^\beta g_n^{(l+r)}(u)| &= \frac{h^\beta}{2\pi}\left|\int_{-\infty}^{\infty}(-it)^{l+r}\exp(-itu)\frac{\phi_K(t)}{\phi_\epsilon(t/h)}dt\right| \\
&\leq \frac{h^\beta}{2\pi|u|}\int_{-\infty}^{\infty}\left|\left[\frac{t^{l+r}\phi_K(t)}{\phi_\epsilon(t/h)}\right]^{(1)}\right|dt \\
&\leq \frac{ch^\beta}{|u|} \\
&\quad\cdot\int_{-\infty}^{\infty}\left|\frac{[(r+l)t^{l+r-1}\phi_K(t)+t^{l+r}\phi_K^{(1)}(t)]\phi_\epsilon(t/h)+t^{l+r}\phi_K(t)\phi_\epsilon^{(1)}(t/h)/h}{[\phi_\epsilon(t/h)]^2}\right|dt \\
&\leq \frac{c}{|u|}\left\{\int_{|t|\geq Mh}[d_1|t|^{l+r+\beta-1}|\phi_K(t)|+d_2|t|^{l+r+\beta}|\phi_K^{(1)}(t)|]dt\right. \\
&\quad + d_3\int_{|t|\geq Mh}\left|\frac{t^{\beta+1}\phi_\epsilon^{(1)}(t/h)}{h^{\beta+1}}t^{l+r+\beta-1}\phi_K(t)\right|dt \\
&\quad + \frac{h^\beta}{\min|\phi_\epsilon(t)|^2}\int_{|t|<Mh}|[(r+l)t^{l+r-1}\phi_K(t) \\
&\quad \left. + t^{l+r}\phi_K^{(1)}(t)]\phi_\epsilon(t/h)+t^{l+r}\phi_K(t)\phi_\epsilon^{(1)}(t/h)/h|dt\right\} \\
&\leq \frac{c}{|u|},
\end{aligned}$$

where c, d_1, d_2 and d_3 are some positive constants, which may take different values at different places.

But, (3.5) and (3.6) imply that $|h^\beta g_n^{(l+r)}(u)|^2 \leq \frac{c}{1+u^2}$. Therefore,

$$(3.7) \qquad \int_{-\infty}^{\infty}|h^\beta g_n^{(l+r)}(u)|^2du < \infty, \qquad \text{for} \quad 0 \leq r \leq k-l-1.$$

By applying the Lebesgue's Dominated Convergence Theorem together with (3.5) and assumption (ii), we obtain

$$(3.8) \qquad h^\beta g_n^{(l+r)}(u) \to \frac{1}{2\pi\alpha}\int_{-\infty}^{\infty}\exp(-itu)(-it)^{l+r}t^\beta\phi_K(t)dt.$$

By Lemma 2.1 of Fan (1991b), (3.6), (3.7) and (3.8), for $0 \le j,\ m \le k-l-1$, we have

$$
\begin{aligned}
(3.9)\quad & Eg_n^{(l+j)}\left(\frac{x-Y_1}{h}\right) g_n^{(l+m)}\left(\frac{x-Y_1}{h}\right)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_n^{(l+j)}\left(\frac{x-u-v}{h}\right) g_n^{(l+m)}\left(\frac{x-u-v}{h}\right) f_X(u)dF_\epsilon(v)du\\
&= h\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_n^{(l+j)}(u) g_n^{(l+j)}(v) f_X(x-v-uh)dF_\epsilon(v)du\\
&= \frac{1}{(2\pi\alpha)^2 h^{2\beta-1}}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\exp(-itu)(-it)^{j+l}t^\beta\phi_K(t)dt\right]\\
&\quad\cdot\left[\int_{-\infty}^{\infty}\exp(-itu)(-it)^{m+l}t^\beta\phi_K(t)dt\right] du(1+o(1))\\
&= \frac{f_Y(X)}{h^{2\beta-1}} m_{jm}(1+o(1)).
\end{aligned}
$$

Combining (3.4) with (3.9) leads to

$$
\operatorname{cov}\left((-1)^j g_n^{(l+j)}\left(\frac{x-Y_1}{h}\right), (-1)^m g_n^{(l+m)}\left(\frac{x-Y_1}{h}\right)\right) = \frac{f_Y(X)}{h^{2\beta-1}} m_{jm}(1+o(1)).
$$

Hence

$$
\operatorname{Var} \tilde{f}_n^{(l)}(x) = \frac{f_Y(x)}{nh^{2(l+\beta)+1}} e_1^T S_c^{-1} M (S_c^{-1})^T e_1
$$

This completes the proof of Lemma 3.

THEOREM 1. *Assume that the assumptions of Lemmas* 1 *and* 2 *hold. By choosing the bandwidth* $h = w(\log n)^{1/\beta}$ *with* $w > M_0(2/\gamma)^{1/\beta}$, *for* $x = ch$, $c \ge 0$, *we have*

$$
(3.10)\qquad E|\tilde{f}_n^{(l)}(x) - f_X^{(l)}(x)|^2 = h^{2(k-l)}\{e_1^T S_c^{-1} S_c^*\}^2 [f_X^{(k)}(x)]^2 (1+o(1)).
$$

THEOREM 2. *Assume that the assumptions of Lemmas* 1 *and* 3 *hold. By choosing the bandwidth* $h = dn^{-1/(2(k+\beta)+1)}$, $d > 0$, *for* $x = ch$, $0 \le c \le 1$, *we have*

$$
\begin{aligned}
(3.11)\quad E|\tilde{f}_n^{(l)}(x) - f_X^{(l)}(x)|^2 &= \{d^{2(k-l)}[e_1^T S_c^{-1} S_c^*]^2 [f_X^{(k)}(x)]^2\\
&\quad + d^{-2(l+\beta)+1} f_Y(x)(1,0,\ldots,0) S_c^{-1} M (S_c^{-1})^T e_1\} n^{-(2(k-l))/(2(k+\beta)+1)}(1+o(1)).
\end{aligned}
$$

The above two theorems are direct consequences of Lemmas 1, 2 and 3. The proofs are omitted.

Remark 4. Fan (1991a) showed that the rates in our Theorems 1 and 2 are optimal in the case that the estimated density $f_X(x)$ has support $(-\infty,\infty)$. As it is obvious that the deconvolution problem can't be easier for a truncated density than for a density with support $(-\infty,\infty)$, the rates in Theorems 1 and 2 are also optimal.

Remark 5. When the error is supersmooth, the rates of convergence are too slow to be practical. However, there are some situations in which the rates of convergence can be greatly improved. One situation is when the noise level of the supersmooth error ϵ is small, the rates of convergence will be comparable to that from ordinary smooth error, see Fan (1992). Another case is when the data are only partially contaminated, as it is very common in practice. More specifically, assume that the data are $100p\%$ $(0 \le p \le 1)$ contaminated. Namely,

$$Y = X + \epsilon, \tag{3.12}$$

with $P(\epsilon = 0) = 1 - p$ and $P(\epsilon = \tilde{\epsilon}) = p$, $\tilde{\epsilon}$ is a random variable having a supersmooth distribution and with the characteristic function $\phi_{\tilde{\epsilon}}(t)$. Assume that $Re(\phi_{\tilde{\epsilon}}(t)) \ge 0$, for all t, where $Re(\cdot)$ is the real part of the expression. Then the characteristic function of ϵ is

$$\phi_\epsilon(t) = (1-p) + p\phi_{\tilde{\epsilon}}(t). \tag{3.13}$$

It can be proved that (3.13) satisfies the conditions of Lemma 3 with $\beta = 0$. So, the model (3.12) is a special case of Theorem 2.

4. The local optimal bandwidth

When the error is supersmooth, the convergence rates of the density estimator is extremely slow. Since the local bandwidth choice does not change the rates of convergence, we shall not discuss the bandwidth choice problem for supersmooth errors in this paper.

In the following, we shall only consider the bandwidth choice problem for ordinary smooth errors. By (3.1) and (3.3), for $x = ch$,

$$\begin{aligned}\mathrm{MSE}(\tilde{f}_n^{(l)}, x) = {} & h^{2(k-l)}[e_1^T S_c^{-1} S_c^*]^2 [f_X^{(k)}(x)]^2 \\ & + \frac{1}{nh^{2(l+\beta)+1}} f_Y(x) e_1^T S_c^{-1} M [S_c^{-1}]^T e_1.\end{aligned} \tag{4.1}$$

For $x = h$, the asymptotic MSE of $\tilde{f}_n^{(l)}$ is

$$\begin{aligned}\mathrm{MSE}(\tilde{f}_n^{(l)}, h) \sim {} & h^{2(k-l)}[e_1^T S_1^{-1} S_1^*]^2 [f_X^{(k)}(0)]^2 \\ & + \frac{1}{nh^{2(l+\beta)+1}} f_Y(0) e_1^T S_1^{-1} M [S_1^{-1}]^T e_1.\end{aligned} \tag{4.2}$$

Assume that the local optimal bandwidth is adapted at $x = h$. Then

$$h = \left[\frac{[2(l+\beta)+1] e_1^T S_1^{-1} M [S_1^{-1}]^T e_1 f_Y(0)}{2(k-l)[e_1^T S_1^{-1} S_1^*]^2 [f_X^{(k)}(0)]^2}\right]^{1/(2(k+\beta)+1)} n^{-1/(2(k+\beta)+1)}. \tag{4.3}$$

Since h defined by (4.3) is only optimal at $x = h$, this h will not be a good choice of bandwidth for points other than $x = h$, especially for the points $x = ch$, $0 \le c < 1$. For the points $x = ch$, $c > 1$, we assume that this h is not far from the optimal global bandwidth. At the points $x = ch$ $(0 \le c < 1)$, we suggest the use of a bandwidth variation function. Denote $b(c)$ the bandwidth variation function which satisfies $b(c) > 0$ for $0 \le c < 1$ and $b(1) = 1$. At $x = ch$, assume that the bandwidth $h(c) = b(c)h$ is used. From (4.1), for $x = ch$ $(0 \le c \le 1)$

$$\begin{aligned}\mathrm{MSE}(\tilde{f}_n^{(l)}, x) = {} & b(c)^{2(k-l)} h^{2(k-l)} [e_1^T S_{c/b(c)}^{-1} S_{c/b(c)}^*]^2 [f_X^{(k)}(0)]^2 \\ & + \frac{1}{nb(c)^{2(l+\beta)+1} h^{2(l+\beta)+1}} f_Y(0) e_1^T S_{c/b(c)}^{-1} M [S_{c/b(c)}^{-1}]^T e_1.\end{aligned} \tag{4.4}$$

In general, the uniformly optimal $b(c)$ which minimizes (4.4) is not available since $(1,0,\ldots,0)S_c^{-1}S_c^*$ has a zero for some $0 < c < 1$. One way to get around this problem is to obtain the optimal bandwidth variation function at $x = 0$, i.e., $b(0)$ which minimizes (4.4) for $c = 0$, and define

$$b(c) = b(1) - (c-1)(b(0) - b(1)) \tag{4.5}$$

as the (suboptimal) bandwidth variation function. For more details on the choice of bandwidth variation function, see Zhang and Karunamuni (1998). This issue will be further discussed in Section 6.

5. The case of $f_X^{(1)}(0) = 0$

In this section, we shall consider estimation of $f_X(x)$ by an order (0, 2) kernel. For the estimator defined by (1.2) with $l = 0$, we shall denote $\hat{f}_n^{(0)}$ by $\hat{f}_n$. Since for $x = ch$, $c \geq 0$,

$$E\hat{f}_n(x) = f_X(x)\int_{-\infty}^{c} K(t)dt + \frac{f_X^{(2)}(x)}{2!}h^2\int_{-\infty}^{c} t^2K(t)dt + o(h^2). \tag{5.1}$$

Intuitively, we can define

$$f_n^+(x) = \begin{cases} \dfrac{1}{nh\int_{-\infty}^{c} K(t)dt}\displaystyle\sum_{j=1}^{n} g_n\left(\frac{x-Y_j}{h}\right) & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{5.2}$$

as the estimator of $f_X(x)$ at $x = ch$, $c \geq 0$, where

$$g_n(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-ity)\frac{\phi_K(t)}{\phi_\epsilon(t/h)}dt. \tag{5.3}$$

It is easy to see that the asymptotic bias of f_n^+ at $x = ch$ is

$$Ef_n^+(x) - f_X(x) = \frac{f_X^{(2)}(x)}{2\int_{-\infty}^{c} K(t)dt}h^2\int_{-\infty}^{c} t^2K(t)dt + o(h^2). \tag{5.4}$$

When the error ϵ is supersmooth of order β, we can show that (see proof of Lemma 2) for $x = ch$,

$$\mathrm{Var}\, f_n^+(x) \leq O\left(\frac{\exp(|M_0/h|^\beta/\gamma)}{nh^{2+2|\beta_0|}[\int_{-\infty}^{c} K(t)dt]^2}\right) = o(h^4). \tag{5.5}$$

THEOREM 3. *Assume that the assumptions of Lemmas 1 and 2 hold with $k = 2$ and $l = 0$. Then by choosing the bandwidth $h = w(\log n)^{1/\beta}$ with $w > M_0(2/\gamma)^{1/\beta}$, for $x = ch$, $c \geq 0$, we have*

$$E|f_n^+(x) - f_X(x)|^2 = h^4\frac{[f_X^{(2)}(x)]^2}{4[\int_{-\infty}^{c} K(t)dt]^2}\left[\int_{-\infty}^{c} t^2K(t)dt\right]^2(1+o(1)). \tag{5.6}$$

When the error ϵ is ordinary smooth of order β, we can show that (see proof of Lemma 3) for $x = ch$,

$$\operatorname{Var} f_n^+(x) = \frac{1}{nh^2[\int_{-\infty}^{c} K(t)dt]^2} \operatorname{Var} g_n\left(\frac{x - Y_1}{h}\right)$$

with

$$\operatorname{Var} g_n\left(\frac{x-Y_1}{h}\right) = \frac{f_Y(x)}{2\pi\alpha^2 h^{2\beta-1}} \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} \exp(-itu)t^{\beta}\phi_K(t)dt\right]^2 du(1+o(1)). \tag{5.7}$$

Therefore, by Parseval's identity we obtain from (5.7) that

$$\operatorname{Var} f_n^+(x) = \frac{f_Y(x)}{nh^{2\beta+1}2\pi\alpha^2[\int_{-\infty}^{c} K(t)dt]^2} \int_{-\infty}^{\infty} t^{2\beta}|\phi_K(t)|^2 dt(1+o(1)). \tag{5.8}$$

THEOREM 4. *Assume that the assumptions of Lemmas* 1 *and* 3 *hold with* $k = 2$ *and* $l = 0$. *Then by choosing the bandwidth* $h = dn^{-1/(2\beta+5)}$, $d > 0$, *for* $x = ch$, $c \geq 0$, *we have*

$$E|\hat{f}_n^{(l)}(x) - f_X^{(l)}(x)|^2 = \left\{d^4 \frac{[f_X^{(2)}(x)]^2}{4[\int_{-\infty}^{c/b(c)} K(t)dt]^2}\left[\int_{-\infty}^{c/b(c)} t^2 K(t)dt\right]^2 + \frac{f_Y(x)}{2\pi\alpha^2[\int_{-\infty}^{c/b(c)} K(t)dt]^2}\int_{-\infty}^{\infty} t^{2\beta}|\phi_K(t)|^2 dt\right\} n^{-4/(2\beta+5)}. \tag{5.9}$$

For the ordinary smooth error case, we still need to choose a bandwidth variation function for the points $x = ch$, $0 \leq c < 1$. Denote the bandwidth variation function by $b(c)$ which satisfies $b(c) > 0$ and $b(1) = 1$. Under the assumptions of Theorem 4, at $x = ch$, the explicit form of the asymptotic MSE is

$$\operatorname{MSE}(f_n^+, x) = b(c)^4 h^4 \frac{[f_X^{(2)}(x)]^2}{4[\int_{-\infty}^{c/b(c)} K(t)dt]^2}\left[\int_{-\infty}^{c/b(c)} t^2 K(t)dt\right]^2 + \frac{f_Y(x)}{nb(c)^{2\beta+1}h^{2\beta+1}2\pi\alpha^2[\int_{-\infty}^{c/b(c)} K(t)dt]^2}\int_{-\infty}^{\infty} t^{2\beta}|\phi_K(t)|^2 dt. \tag{5.10}$$

Assume that the local optimal bandwidth is chosen at $x = h$, then

$$h = \left\{\frac{(2\beta+1)\int_{-\infty}^{\infty} t^{2\beta}|\phi_K(t)|^2 dt f_Y(0)}{2\pi\alpha^2[\int_{-\infty}^{1} t^2 K(t)dt]^2[f_X^{(2)}(0)]^2}\right\}^{1/(2\beta+5)} n^{-1/(2\beta+5)}. \tag{5.11}$$

Similar to Section 4, a suboptimal bandwidth variation function $b(c)$ will be employed for any point $x = ch$, $0 \leq c \leq 1$, where $b(c) = b(1) - (c-1)(b(0) - b(1))$ with

$$b(0) = \left\{\frac{[\int_{-\infty}^{1} t^2 K(t)dt]^2}{[\int_{-\infty}^{0} t^2 K(t)dt]^2}\right\}^{1/(2\beta+5)}. \tag{5.12}$$

6. Simulations

Example 1. Let X have a standard exponential distribution with density

$$f_X(x) = \begin{cases} \exp(-x) & x \geq 0 \\ 0 & x < 0. \end{cases} \tag{6.1}$$

Let the error distribution be double exponential with density

$$f_\epsilon(x) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|x|), \qquad -\infty < x < \infty, \tag{6.2}$$

and the characteristic function

$$\phi_\epsilon(t) = \left(1 + \frac{t^2}{2}\right)^{-1}. \tag{6.3}$$

Therefore $\beta = 2$, $\alpha = 2$. For $x = ch$, the boundary corrected estimator $\tilde{f}_n$ of f_X defined by (2.7) is

$$\tilde{f}_n(x) = \frac{1}{nh[\int_{-\infty}^{c} K(t)dt]^2} \sum_{j=1}^{n} \left\{ \int_{-\infty}^{c} K(t)dt g_n^{(0)}\left(\frac{x - Y_j}{h}\right) - \int_{-\infty}^{c} tK(t)dt g_n^{(1)}\left(\frac{x - Y_j}{h}\right) \right\}, \tag{6.4}$$

where $g_n^{(0)}$ and $g_n^{(1)}$ are defined by (1.3). For the Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

we have

$$g_n^{(l)}(x) = \frac{1}{\sqrt{2\pi}} (-x)^l \exp\left(-\frac{x^2}{2}\right) \left(1 - \frac{x^2 - 3^l}{2h^2}\right). \tag{6.5}$$

By simple algebra, the local optimal bandwidth at $x = h$ defined by (4.3) is

$$h = \left[\frac{15\left(2N(1)^2 + \frac{5}{2\pi \exp(1)}\right)}{256\sqrt{2\pi}(\sqrt{2}+1)\left(\frac{N(1)^2}{2} - \frac{N(1)}{2\sqrt{2\pi \exp(1)}} - \frac{1}{2\pi \exp(1)}\right)^2} \right]^{1/9} n^{-1/9}, \tag{6.6}$$

where $N(1) = P(Z \leq 1)$, $Z \sim N(0,1)$. Similarly, we can obtain $b(0)$ which minimizes (4.4) at $c = 0$:

$$b(0) = \left\{ \frac{\left[2N(0)^2 + \frac{5}{2\pi}\right] \left[\frac{N(1)^2}{2} - \frac{N(1)}{2\sqrt{2\pi \exp(1)}} - \frac{1}{2\pi \exp(1)}\right]^2}{\left[2N(1)^2 + \frac{5}{2\pi \exp(1)}\right] \left[\frac{N(0)^2}{2} - \frac{1}{2\pi}\right]^2} \right\}^{1/9}. \tag{6.7}$$

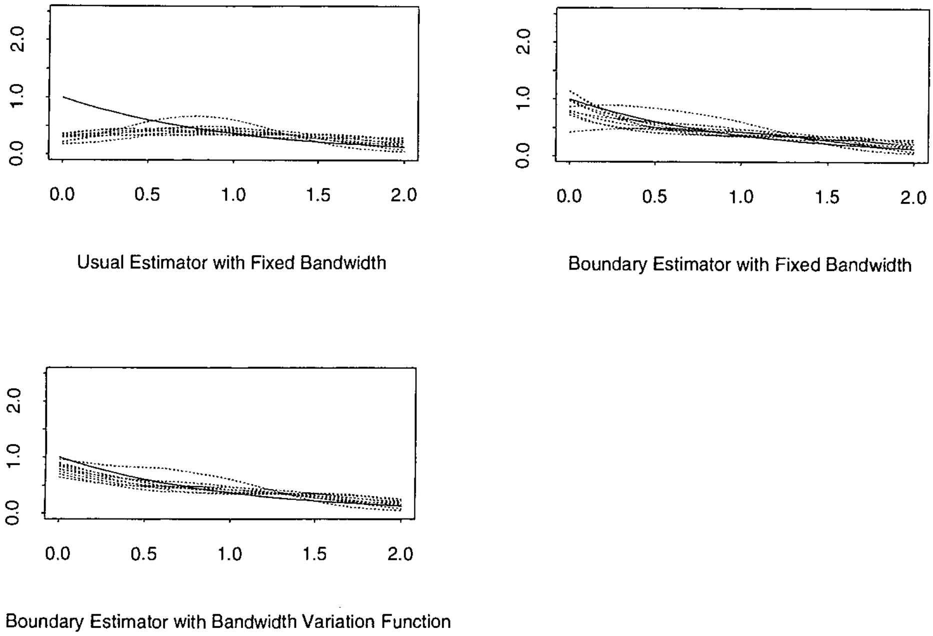

Fig. 1. Estimates of $\exp(-x)$.

Table 1. MSE values at some points from different methods for Example 1.

	$x = 0.0$	$x = 0.5$	$x = 1.0$	$x = 1.5$	$x = 2.0$
Usual estimator with fixed bandwidth	0.4350	0.0291	0.0055	0.0048	0.0046
Boundary method with fixed bandwidth	0.0657	0.0139	0.0059	0.0048	0.0046
Boundary method with bandwidth variation function	0.0328	0.0130	0.0059	0.0048	0.0046

In Fig. 1, we compared the performance of the conventional estimator (without boundary correction), the boundary corrected estimator without use of a bandwidth variation function and the boundary corrected estimator with the use of a bandwidth variation function for estimating density (6.1). The sample size was 100. The bandwidth variation function we employed in our simulation is given by (4.5). Ten typical estimates of density (6.1) were plotted in each case. In each graph the solid curve represents the true density curve. It is obvious that our boundary corrected estimator corrects the boundary effect of the conventional estimator. It can also be seen that the use of the the bandwidth variation function greatly improves the performance of the boundary corrected estimator.

Table 1 represents the MSE values for the three methods at $x = 0.0$, 0.5, 1.0, 1.5 and 2.0 for this example. The bandwidth used was 0.5474, and $b(0) = 1.4259$. We also calculated the mean integrated square error (MISE) from the above methods. The results were the averages of 100 repetitions for the sample size $n = 100$. The MISE values from the conventional estimator, the boundary corrected estimator without use

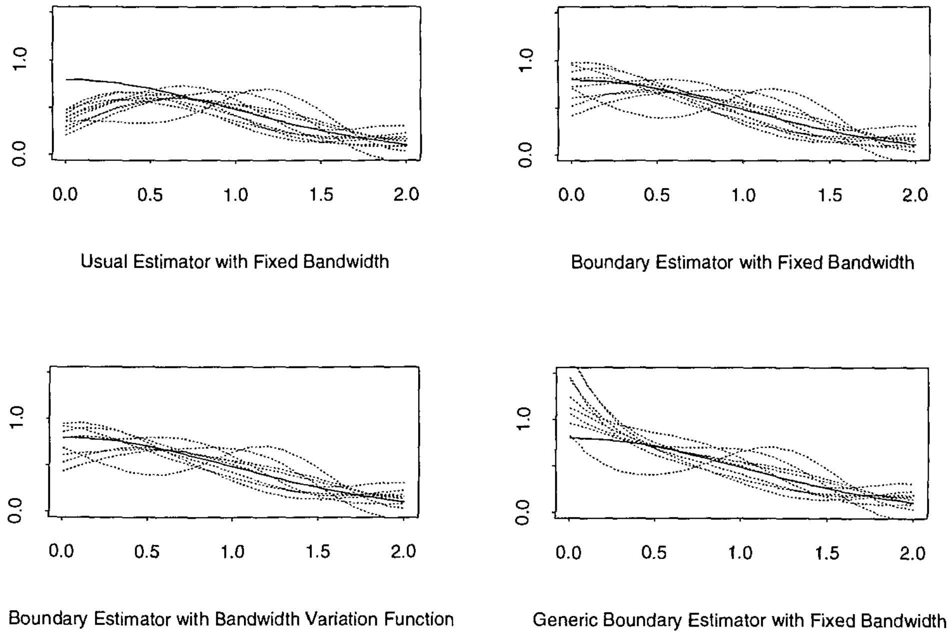

Fig. 2. Estimates of $(2\pi)^{-1/2}\exp(-x^2/2)$.

Table 2. MSE values at some points from different methods for Example 2.

	$x = 0.0$	$x = 0.5$	$x = 1.0$	$x = 1.5$	$x = 2.0$
Usual estimator with fixed bandwidth	0.2099	0.0453	0.0088	0.0072	0.0058
Boundary method with fixed bandwidth	0.0385	0.0183	0.0087	0.0073	0.0058
Boundary method with bandwidth variation function	0.0352	0.0183	0.0087	0.0073	0.0058
Generic boundary method with fixed bandwidth	0.1533	0.0180	0.0092	0.0072	0.0058

of a bandwidth variation function and the boundary corrected estimator with the use of a bandwidth variation function were 0.1151676, 0.0309845 and 0.0263469, respectively. The preceding values show that the boundary corrected estimator with the use of a bandwidth variation function performs the best, followed by the boundary corrected estimator without use of a bandwidth variation function. The conventional estimator was the worst among the three. The MSE values in Table 1 also shows the same behavior of the three estimators near the endpoint.

Example 2. Let X have a half normal distribution with density

$$f_X(x) = \begin{cases} \sqrt{\frac{2}{\pi}}\exp(-x^2/2) & x \geq 0 \\ 0 & x < 0. \end{cases} \tag{6.8}$$

We shall still use the double exponential error and the Gaussian kernel as in Example 1 above. For $x = ch$, the boundary corrected estimator f_n^+ of f_X defined by (5.2) is

$$f_n^+(x) = \frac{1}{nh\int_{-\infty}^{c} K(t)dt} \sum_{j=1}^{n} g_n^{(0)}\left(\frac{x - Y_j}{h}\right), \tag{6.9}$$

where $g_n^{(0)}$ is defined by (6.5).

By simple algebra, the local optimal bandwidth at $x = h$ defined by (5.11) is

$$h = \left\{ \frac{15\sqrt{2\pi}\exp(1)N(-\sqrt{2})}{64\left[N(1) - \frac{1}{\sqrt{2\pi\exp(1)}}\right]^2} \right\}^{1/9} n^{-1/9}. \tag{6.10}$$

Similarly, we can obtain $b(0)$ which minimizes (5.9) at $c = 0$:

$$b(0) = \left[\frac{N(1) - \frac{1}{\sqrt{2\pi\exp(1)}}}{N(0)} \right]^{-1/9}.$$

In Fig. 2, we compared the performance of the conventional estimator (without boundary correction), the generic boundary method (which does not take the assumption $f_X^{(1)}(0) = 0$ into consideration), the boundary corrected estimator without use of a bandwidth variation function and the boundary corrected estimator with the use of a bandwidth variation function for estimating density (6.8). The bandwidth variation function we employed in our simulation is given by (4.5). Ten typical estimates of density (6.8) were plotted in each case, again the true curve is given by the solid curve.

Table 2 gives the MSE values for the four methods at $x = 0.0$, 0.5, 1.0, 1.5 and 2.0 for this example. The bandwidth used was 0.5334, and $b(0) = 1.0411$. For this example, we again calculated the MISE values for the four methods. The sample size and number of repetitions were the same as in Example 1. The MISE from the conventional estimator, the generic boundary method, the boundary corrected estimator without use of a bandwidth variation function and the boundary corrected estimator with the use of a bandwidth variation function were 0.08803471, 0.07092552, 0.04961943, and 0.04835677, respectively. Again, we see that all the boundary corrected estimators perform better than the conventional estimator. The generic boundary method is better than the conventional method, but is much worse than the other two boundary corrected methods, especially near the endpoint.

The performance of the boundary corrected estimator with the use of a bandwidth variation function is very similar to that of the boundary corrected estimator without use of a bandwidth variation function. This is because $b(0) = 1.0411$ in this case. Therefore, for a density function satisfying $f_X^{(1)}(0) = 0$, the use of the bandwidth variation function is not necessary.

Example 3. The Wooden Stakes data set described in Burnham *et al.* ((1980), pp. 61–63), was obtained from a line transect sampling experiment in a sagebrush desert in which wooden stakes had been placed at a known density, D (see Section 1 above). Although a fixed width transect was used, it was wide enough to ignore the estimation complications of truncation (Buckland *et al.* (1993)). The true form of f_X, the pdf of the true perpendicular distances X_i from the transect line, was unknown and varied from observer to observer. The sample data Y_i on perpendicular distances for the sample size $n = 68$ stakes found are shown in Table 6 of Burnham *et al.* ((1980), p. 62). The actual value of D was known to be 0.00375 stakes/ha, with the line length $L = 1000$m and width of 20m. So, the actual value of $f_X(0)$ is given by the formula $f_X(0) = 2LD/n = 0.1102941$.

For this data set, we computed the value of $f_X(0)$ using the four methods discussed in Example 2 above. We assumed that the error distribution is double-exponential. The half normal model fits the data quite well. Therefore, we also made the assumption that $f_X(x) = (2\pi\sigma^2)^{-1/2}\exp(-x^2/2\sigma^2)$ for $x \geq 0$ and 0 otherwise. As a result, the value of the bandwidth defined by (5.11) was 2.635. The computed values of $f_X(0)$ from the usual method, generic method, the boundary corrected method without bandwidth variation and the boundary corrected method with bandwidth variation were 0.05299, 0.17630, 0.105978 and 0.105815, respectively. Thus, both the boundary corrected estimators estimate the actual value of $f_X(0)$ somewhat accurately. Incidentally, the maximum likelihood and the Fourier series estimates (Burnham *et al.* (1980)) of $f_X(0)$ are 0.104299 and 0.114803, respectively.

We also computed $f_X(0)$ assuming that f_X has an exponential form as in Example 1 above, and again with the error distribution is double-exponential. The computed values of the conventional estimator, the boundary corrected estimator without bandwidth variation and the boundary corrected estimator with bandwidth variation were 0.05409039, 0.178819 and 0.1749348, respectively. It appears that these values are quite off from the actual value as we expected.

Acknowledgements

We thank two referees for their helpful suggestions and comments which led to a much improved presentation, among other things.

References

Barry, J. and Diggle, P. (1995). Choosing the smoothing parameter in a Fourier approach to nonparametric deconvolution of a density estimate, *J. Nonparametr. Statist.*, **4**, 223–232.

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

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

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

Devroye, P. L. (1989). Consistent deconvolution in density estimation, *Canad. J. Statist.*, **17**, 235–239.

Diggle, P. J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate, *J. Roy. Statist. Soc. Ser. B*, **55**, 523–531.

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

Fan, J. (1991*b*). Asymptotic normality for deconvolution kernel density estimators, *Sankhyā Ser. A*, **53**, 97–100.

Fan, J. (1991*c*). Global behavior of deconvolution kernel estimators, *Statist. Sinica*, **1**, 541–551.

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

Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem, *Canad. J. Statist.*, **17**, 427–438.

Mendelsohn J. and Rice, J. (1982). Deconvolution of microfluorometric histograms with B splines, *J. Amer. Stat. Assoc.*, 77, 748–753.

Penskaya, M. Ya. (1988). Estimation of a prior density function and functionals thereof, Ph. D Thesis, Moscow University Press, Moscow (in Russian).

Piterbarg, V. F. and Penskaya, M. Ya. (1993). On asymptotic distribution of integrated squared error of an estimate of a component of a convolution, *Math. Methods Statist.*, **2**, 30–41.

Stefanski, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems, *Statist. Probab. Lett.*, **9**, 229–235.

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

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

Zhang, S. and Karunamuni, R. J. (1998). On kernel density estimation near endpoints, *J. Statist. Plann. Inference*, **70**, 301–316.