

## ON THE OPTIMAL RATES OF CONVERGENCE FOR NONPARAMETRIC DECONVOLUTION PROBLEMS

BY JIANQING FAN

*University of North Carolina*

Deconvolution problems arise in a variety of situations in statistics. An interesting problem is to estimate the density  $f$  of a random variable  $X$  based on  $n$  i.i.d. observations from  $Y = X + \varepsilon$ , where  $\varepsilon$  is a measurement error with a known distribution. In this paper, the effect of errors in variables of nonparametric deconvolution is examined. Insights are gained by showing that the difficulty of deconvolution depends on the smoothness of error distributions: the smoother, the harder. In fact, there are two types of optimal rates of convergence according to whether the error distribution is ordinary smooth or supersmooth. It is shown that optimal rates of convergence can be achieved by deconvolution kernel density estimators.

**1. Introduction.** Suppose we have  $n$  i.i.d. observations  $Y_1, \dots, Y_n$  having the same distribution as that of  $Y$  available to estimate the unknown density  $f(x)$  of a random variable  $X$ , where

$$(1.1) \quad Y = X + \varepsilon$$

with a measurement error  $\varepsilon$  of a known distribution. Assume furthermore that the random variables  $X$  and  $\varepsilon$  are independent. We will discuss herein how well the unknown density and its cumulative distribution function (cdf) can be estimated nonparametrically under certain smoothness conditions.

The usual smoothness condition imposed on the unknown density  $f$  is that  $f$  is in the set

$$(1.2) \quad \mathcal{C}_{m, \alpha, B} = \{f(x) : |f^{(m)}(x) - f^{(m)}(x + \delta)| \leq B\delta^\alpha\},$$

where  $m$ ,  $B$  and  $0 \leq \alpha < 1$  are known constants. The functionals we want to estimate are  $T(f) = f^{(l)}(x)$  ( $l = 0$ , density function).

Such a model of measurements being contaminated with error exists in many different fields and has been widely studied. Recent related works include Carroll and Hall (1988), Devroye (1989), Fan (1989), Mendelsohn and Rice (1982), Liu and Taylor (1989), Stefanski and Carroll (1990), Stefanski (1990) and Zhang (1990). The applications of the model in a theoretical setting and an applied setting are discussed by Carroll and Hall (1988) and the other papers cited above. Most of the papers cited above address how to estimate the unknown density and compute the rates of convergence for some specific error

---

Received July 1988; revised April 1990.

AMS 1980 subject classifications. Primary 62G05; secondary 62C25.

*Key words and phrases.* Deconvolution, nonparametric density estimation, estimation of distribution, optimal rates of convergence, kernel estimate, Fourier transformation, smoothness of error distributions.

distributions. Yet, few results discuss the issue of how difficult the deconvolution is. It is of theoretical and practical interest to ask the following questions: What are the best estimators? [in terms of the rates of convergence, according to Stone's (1983) definition]; What are the optimal rates of convergence? What is the difficulty of the problem? Where does the difficulty come from? Attempting to answer these questions forms the core of the paper.

The insights of the nonparametric deconvolution are gained by our study. The optimal rates of convergence can be characterized by two types of error distributions: ordinary smooth and supersmooth distributions. We show that the difficulty of deconvolution depends heavily on the smoothness of the distribution of the error variable  $\varepsilon$ , and on the smoothness of the object being estimated: the smoother the error distribution is, the harder the deconvolution will be. By the smoothness of the error distribution, we mean the order of the characteristic function  $\phi_\varepsilon(t)$  of the random variable  $\varepsilon$  as  $t \rightarrow \infty$ . We will call the distribution of a random variable  $\varepsilon$  supersmooth of order  $\beta$  if its characteristic function  $\phi_\varepsilon(t)$  satisfies

$$(1.3) \quad d_0 |t|^{\beta_0} \exp(-|t|^\beta/\gamma) \leq |\phi_\varepsilon(t)| \leq d_1 |t|^{\beta_1} \exp(-|t|^\beta/\gamma) \quad \text{as } t \rightarrow \infty,$$

for some positive constants  $d_0, d_1, \beta, \gamma$  and constants  $\beta_0$  and  $\beta_1$  (note that the density of  $\varepsilon$  has all finite derivatives). We will call the distribution of a random variable  $\varepsilon$  ordinary smooth of order  $\beta$  if its characteristic function  $\phi_\varepsilon(t)$  satisfies

$$(1.4) \quad d_0 |t|^{-\beta} \leq |\phi_\varepsilon(t)| \leq d_1 |t|^{-\beta} \quad \text{as } t \rightarrow \infty,$$

for some positive constants  $d_0, d_1, \beta$ . The examples of supersmooth distributions are normal, mixture normal, Cauchy, etc. The examples of ordinary smooth distributions include gamma, double exponential and symmetric gamma distributions.

Carroll and Hall (1988) give the optimal rates of estimating density at a point when the error is normal. Zhang (1990) discusses the optimal rates of convergence under the  $L_2$ -norm and computes both upper bounds and lower bounds on rates under his formulation. The results of both papers are very interesting. However, the two papers do not show exactly why the lower bounds on rates depend on the tail of  $\phi_\varepsilon$ , the characteristic function of the error distribution and in particular do not find attainable lower bounds for many interesting error distributions (e.g., gamma, double exponential, etc.). The reason for this dependence is clearly stated in Section 3, and the optimal rates of convergence are obtained for both types of error distributions. The applications of the achievements can be found in Fan (1989), where global rates of convergence under the weighted  $L_p$ -norm are devoted via introducing a new technique, and Fan and Truong (1990) which addresses both optimal local and global rates for error-in-variable nonparametric regression.

We will use kernel density estimators to estimate the unknown density  $f$ , as well as its derivatives. A similar construction is used by Stefanski and Carroll (1990) and Zhang (1990). For a nice kernel function  $K(x)$ , let  $\phi_K(t)$  be its Fourier transform with  $\phi_K(0) = 1$ . Then the kernel density estimator is

defined by

$$(1.5) \quad \hat{f}_n^{(l)}(x) \equiv \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx) (-it)^l \phi_K(th_n) \frac{\hat{\phi}_n(t)}{\phi_\varepsilon(t)} dt,$$

for suitable choice of a bandwidth  $h_n$  and a kernel function  $K$ , where  $\hat{\phi}_n(t)$  is the empirical characteristic function defined by

$$(1.6) \quad \hat{\phi}_n(t) = \frac{1}{n} \sum_1^n \exp(itY_n).$$

We will rewrite  $\hat{f}_n^{(0)}(x)$  by  $\hat{f}_n(x)$ . Note that  $\hat{f}_n^{(l)}(x)$  is real and can be rewritten into kernel form [see (2.2) below].

The paper is organized as follows. In Section 2, we will exhibit the rates of deconvolution kernel density estimators, which are optimal in terms of rates of convergence. In Section 3, we will state the results on lower bounds and give their heuristic arguments. Relevant issues are discussed in Section 4. Results are proved in Section 5.

**2. Kernel density estimators.** We will start with the kernel density estimator (1.5). Let  $K(t)$  be the Fourier inversion of  $\phi_K(t)$  defined by

$$(2.1) \quad K(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx) \phi_K(t) dt.$$

Then (1.5) can be rewritten as a kernel type of estimator:

$$(2.2) \quad \hat{f}_n(x_0) = \frac{1}{n} \sum_{j=1}^n \frac{1}{h_n} g_n \left( \frac{x_0 - Y_j}{h_n} \right)$$

if the function  $\phi_K/\phi_\varepsilon(t/h_n)$  is integrable, where

$$(2.3) \quad g_n(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx) \frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)} dt.$$

To compute the mean square error of a kernel density estimator, we first compute the bias of the estimator and see what kind of kernel we should use. For  $x_0 \in (-\infty, \infty)$ , under the assumptions below, we have

$$(2.4) \quad \begin{aligned} E\hat{f}_n(x_0) - f(x_0) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx_0) \phi_K(th_n) \phi_X(t) dt - f(x_0) \\ &= f(x) * \frac{1}{h_n} K \left( \frac{x}{h_n} \right) \Big|_{x_0} - f(x_0). \end{aligned}$$

The last expression does not depend on the error distribution. Thus the kernel function  $K$  should be imposed to satisfy the conditions of a kernel in the ordinary density estimation [see Prakasa Rao (1983)]. We state them on its

Fourier domain:

- (A1)  $\phi_K(t)$  is a symmetric function, having  $m + 2$  bounded integrable derivatives on  $(-\infty, +\infty)$ ;  
 (A2)  $\phi_K(t) = 1 + O(|t|^m)$  as  $t \rightarrow 0$ .

In addition, to develop upper bounds, we assume that

- (A3)  $\phi_\varepsilon(t) \neq 0$  for any  $t$ .

For the case that the distribution of error  $\varepsilon$  is no smoother than super-smooth (exponential decay), we have the following rates of convergence.

**THEOREM 1.** *Under assumptions (A1) to (A3) and*

- (E1)  $\phi_K(t) = 0$  for  $|t| \geq 1$ ,  
 (E2)  $|\phi_\varepsilon(t)| |t|^{-\beta_0} \exp(|t|^\beta/\gamma) \geq d_0$  (as  $t \rightarrow \infty$ ) for some positive constants  $\beta, \gamma, d_0$ , and a constant  $\beta_0$ ,

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

$$(2.5) \quad \sup_{f \in \mathcal{C}_{m,\alpha,B}} E(\hat{f}_n^{(l)}(x_0) - f^{(l)}(x_0))^2 = O((\log n)^{-2(m+\alpha-l)/\beta}),$$

$$l = 0, \dots, m - 1,$$

where  $\hat{f}_n(x_0)$  is defined by (1.5).

For the case of geometric decay of  $\phi_\varepsilon$  (ordinary smooth), we have the following result.

**THEOREM 2.** *Under assumptions (A1) to (A3) and*

- (G1)  $|\phi_\varepsilon(t)t^\beta| \geq d_0$  as  $t \rightarrow \infty$ , for some positive constant  $d_0$ ,  
 (G2)  $\int_{-\infty}^{+\infty} |\phi_K(t)t^{\beta+l}| dt < \infty$  and  $\int_{-\infty}^{+\infty} |\phi_K(t)t^{\beta+l}|^2 dt < \infty$ ,

by choosing the bandwidth  $h_n = dn^{-1/(2(m+\alpha+\beta)+1)}$  for some  $d > 0$ , we have

$$(2.6) \quad \sup_{f \in \mathcal{C}_{m,\alpha,B}} E(\hat{f}_n^{(l)}(x_0) - f^{(l)}(x_0))^2 = O(n^{-2(m+\alpha-l)/[2(m+\alpha+\beta)+1]}),$$

$$l = 0, \dots, m - 1.$$

Define an estimator of the cdf  $F(x_0)$  of the random variable  $X$  by

$$(2.7) \quad \hat{F}_n(x_0) = \int_{-M_n}^{x_0} \hat{f}_n(t) dt,$$

where  $\hat{f}_n(t)$  is the kernel density estimator given by (1.5), and  $M_n (\rightarrow \infty)$  is a sequence of constants.

**THEOREM 3.** *Under assumptions (E1), (E2) and (A3) of Theorem 1, suppose that  $\phi_K(t)$  is a symmetric function, having  $m + 3$  bounded integrable*

derivatives on  $(-\infty, +\infty)$ , and  $\phi_K(t) = 1 + O(|t|^{m+1})$ , as  $t \rightarrow 0$ . Then by choosing the same bandwidth as for Theorem 1 and  $M_n = n^{1/3}$ , we have

$$\sup_{f \in C'_{m, \alpha, B}} E_f(\hat{F}_n(x_0) - F(x_0))^2 = O((\log n)^{-2(m+\alpha+1)/\beta}),$$

where

$$C'_{m, \alpha, B} = \left\{ f \in \mathcal{C}_{m, \alpha, B} : F(-n) \leq D(\log n)^{-(m+2)/\beta} \right\}.$$

EXAMPLE. Suppose that we have  $n$  i.i.d. observations from the convolution model (1.1). We want to estimate the  $l$ th derivative  $T(f) = f^{(l)}(x_0)$  under the constraint that  $f \in \mathcal{C}_{m, \alpha, B}$ . Write  $k = m + \alpha$ . By applying the results of the upper bounds in this section and the lower bounds developed in the next section, without any extra calculation, we have the following results.

Optimal rates	Error distribution	Optimal rates	Error distribution
$O((\log n)^{-(k-l)/2})$	$\varepsilon \sim N(0, 1)$	$O(n^{-(k-l)/[2(k+\beta)+1]})$	$\varepsilon \sim \text{Gamma}(\beta)$
Same as above	$0.7N(1, 1) + 0.3N(-1, 1)$	Same as above	$\varepsilon \sim \text{symmetric Gamma}(\beta)$ $\beta \neq 2j + 1$ ( $j$ integer)
$O((\log n)^{-(k-l)})$	$\varepsilon \sim \text{Cauchy}(0, 1)$	$O(n^{-(k-l)/(2k+4j+5)})$	$\varepsilon \sim \text{symmetric Gamma}(\beta)$ $\beta = 2j + 1$ ( $j$ integer)

Specifically, if  $\varepsilon$  is double exponential (corresponding to a symmetric gamma distribution with  $\beta = 1$ ), the optimal rate is  $O(n^{-(k-l)/(2k+5)})$ . Note that the error distributions in the left column are supersmooth, while those in the right column are ordinary smooth. For estimating a cdf in the supersmooth cases, the optimal rates of convergence are listed above by plugging  $l = -1$ . For estimating a cdf in the ordinary smooth case, the lower rates are listed above by applying  $l = -1$ , but the lower rates are too small to be attainable (see Remark 3 for further discussion).

**3. Lower bounds.** In this section, we will find lower bounds for estimating densities and their cdf's. To begin with, suppose the functional of interest is  $T(f) = f(x_0)$ , density at a point. We will give a heuristic argument to show why the results of lower bounds should depend on the tail of  $\phi_\varepsilon$ , the smoothness condition of the error distribution. Rigorous proof will be given in Section 5, which involves more mathematical details and more careful constructions.

We will assume without loss of generality that  $x_0 = 0$  by relocating  $x_0$  to the origin. To derive a lower bound for estimating  $T(f) = f(0)$ , we take a pair  $f_0 \in \mathcal{C}_{m, \alpha, B}$ ,  $f_n \in \mathcal{C}_{m, \alpha, B}$ , for which

$$(3.1) \quad f_n(x) = f_0(x) + \delta_n^k H(x/\delta_n),$$

where  $k = m + \alpha$ ,  $H(0) \neq 0$ ,  $\int_{-\infty}^{+\infty} H(x) dx = 0$  and the  $m$ th derivative of  $H(x)$  satisfies Lipschitz's condition of order  $\alpha$ . Then by suitable choice of the tail of  $H$  and  $f_0$ , the function  $f_n$  will be a density in  $\mathcal{C}_{m, \alpha, B}$  for small  $\delta_n$ .  $\delta_n$  is

chosen such that the  $\chi^2$ -distance

$$(3.2) \quad \int_{-\infty}^{+\infty} (f_{Y_0} - f_{Y_n})^2 f_{Y_0}^{-1} dx \leq \frac{c}{n},$$

for some constant  $c > 0$ , where  $f_{Y_0}$  and  $f_{Y_n}$  are the density functions of the  $Y$ -variable under (1.1) with  $X$  distributed as  $f_0$  and  $f_n$ , respectively. Then it is proved by Ibragimov, Nemirovskii and Khas'minskii (1986) and Donoho and Liu (1987, 1991a, b) that a lower bound of estimating  $T(f)$  is for any estimator  $\hat{T}_n$ ,

$$(3.3) \quad \sup_{f \in \{f_0, f_n\}} P_f \{ |\hat{T}_n - T(f)| > |T(f_0) - T(f_n)|/2 \} > d_1,$$

and, consequently,

$$(3.4) \quad \sup_{f \in \mathcal{E}_{m, \alpha, B}} E_f (\hat{T}_n - T(f))^2 > \frac{d_1}{4} |T(f_0) - T(f_n)|^2,$$

for some positive constant  $d_1 > 0$ . In other words, the order of

$$(3.5) \quad |T(f_0) - T(f_n)| = \delta_n^k |H(0)|$$

provides a lower bound for estimating  $T(f)$ . Thus we have to find  $\delta_n$  as large as possible such that (3.2) holds, or equivalently by (3.1) with changes of variables  $x' = x/\delta_n$  and  $y' = y/\delta_n$ , such that

$$(3.6) \quad \delta_n^{2k+1} \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} H(x' - y') dF_\varepsilon(\delta_n y') \right)^2 g_0^{-1}(\delta_n x') dx' \leq \frac{c}{n},$$

where  $F_\varepsilon$  is the distribution function of the random variable  $\varepsilon$ , and  $g_0 = f_{Y_0} = f_0 * F_\varepsilon$ .

Suppose we can prove that as  $\delta_n \rightarrow 0$ ,

$$(3.7) \quad \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{+\infty} H(x - y) dF_\varepsilon(\delta_n y) \right]^2 g_0^{-1}(\delta_n x) dx \leq C \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{+\infty} H(x - y) dF_\varepsilon(\delta_n y) \right]^2 dx,$$

where  $C$  is a constant independent of  $n$ . Then by Parseval's identity, to make (3.6) hold, we have to choose  $\delta_n$  such that

$$(3.8) \quad \delta_n^{2k+1} \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} H(x - y) dF_\varepsilon(\delta_n y) \right)^2 dx < \frac{c}{nC},$$

or equivalently such that

$$(3.9) \quad \frac{1}{2\pi} \delta_n^{2k+1} \int_{-\infty}^{+\infty} |\phi_H(t) \phi_\varepsilon(t/\delta_n)|^2 dt \leq \frac{c}{nC},$$

where  $\phi_H$  is the Fourier transformation of  $H$ . Thus the resulting  $\delta_n$  will depend on the tail of  $\phi_\varepsilon$  only, and consequently the lower bound (3.5) will depend on the tail of  $\phi_\varepsilon$ .

Although  $\chi^2$  distance is used for this paper, other distances are also possible [see Donoho and Liu (1987, 1991a, b)].

For the case that the error distribution is smoother than supersmooth, we have the following lower bound.

**THEOREM 4.** *Suppose that the tail of  $\phi_\varepsilon$  satisfies*

$$|\phi_\varepsilon(t)| |t|^{-\beta_1} \exp(|t|^\beta/\gamma) \leq d_1 \text{ (as } t \rightarrow \infty)$$

*with  $\beta, \gamma > 0, d_1 \geq 0$  and  $\beta_1$  a constant,*

*and  $P\{|\varepsilon - x| \leq |x|^{\alpha_0}\} = O(|x|^{-(a-\alpha_0)})$  (as  $x \rightarrow \pm\infty$ ) for some  $0 < \alpha_0 < 1, a > 1 + \alpha_0$ . Then no estimator can estimate  $T(f) = f_n^{(l)}(x_0)$  with the constraint  $f \in \mathcal{C}_{m,\alpha,B}$  faster than  $O((\log n)^{-(m+\alpha-l)/\beta})$  in the sense that for any estimator  $\hat{T}_n$ ,*

$$(3.10) \quad \sup_{f \in \mathcal{C}_{m,\alpha,B}} E_f(\hat{T}_n - T(f))^2 > d(\log n)^{-2(m+\alpha-l)/\beta},$$

*for some  $d > 0$  ( $d$  is also independent of  $\hat{T}_n$ ).*

The technical condition “ $P\{|\varepsilon - x| \leq |x|^{\alpha_0}\} = O(|x|^{-(a-\alpha_0)})$ ” is used to ensure that the tail of convolution function decays at the speed of (see Lemma 5.2)

$$\int_{-\infty}^{+\infty} H(x - y) dF_\varepsilon(y) = O(|x|^{-(a-\alpha_0)}), \quad |x| \rightarrow \infty.$$

Note that the condition holds, if the density  $f_\varepsilon$  (exists for all supersmooth distributions) of the random variable  $\varepsilon$  satisfies  $f_\varepsilon(y) = O(|y|^{-a})$  (as  $|y| \rightarrow \infty$ ) for some  $a > 1$ . This condition can be replaced by the function  $|\phi_\varepsilon''(t)|$  bounded [which is similar to assumption (G3) below]. However, the condition on boundedness of  $|\phi_\varepsilon''|$  excludes Cauchy distributions. Hence the condition “ $P\{|\varepsilon - x| \leq |x|^{\alpha_0}\} = O(|x|^{-(a-\alpha_0)})$ ” is stated in Theorem 4.

For the case that the error distribution is smoother than ordinary smooth, we have the lower bounds as follows.

**THEOREM 5.** *Suppose that the tail of  $\phi_\varepsilon$  satisfies*

(G3)  $|t^{-\beta-j} \phi_\varepsilon^{(j)}(t)| \leq d_j$  as  $t \rightarrow \infty$ , for  $j = 0, 1, 2$ , where  $d_j$  is a positive constant.

*Then no estimator can estimate  $T(f) = f^{(l)}(x_0)$ , under the constraint that  $f \in \mathcal{C}_{m,\alpha,B}$ , faster than  $O(n^{-(m+\alpha-l)/(2m+2\alpha+2\beta+1)})$  in the sense that for any estimator  $\hat{T}_n$ ,*

$$(3.11) \quad \sup_{f \in \mathcal{C}_{m,\alpha,B}} E_f(\hat{T}_n - T(f))^2 > dn^{-2(m+\alpha-l)/(2m+2\alpha+2\beta+1)},$$

*for some  $d > 0$ , where  $\phi_\varepsilon^{(j)}(t)$  is the  $j$ th derivative of  $\phi_\varepsilon$ .*

Thus we have found both lower and upper bounds for the ordinary smooth cases and the supersmooth cases. In practice, those conditions are easy to check. The cases of error distributions satisfying Theorems 2 and 5 include gamma distribution, double exponential distribution, etc., and the cases of error distributions satisfying Theorems 1 and 4 are normal, Cauchy, mixture normal, and many other distributions. Now, we state some lower bounds for estimating the cumulative distribution functions.

**THEOREM 6.** *Under the conditions of Theorem 4, no estimator can estimate the cdf of the random variable  $X$  at a point under the constraint (1.2) faster than  $O((\log n)^{-(m+\alpha+1)/\beta})$  in the sense of (3.10), and under the assumptions of Theorem 5, no estimator can estimate the cdf of the random variable  $X$  under the constraint (1.2) faster than  $O(n^{-(m+\alpha+1)/(2m+2\alpha+2\beta+1)})$  in the sense of (3.11).*

The optimal rates of convergence are investigated by Zhang (1990) for normal and Cauchy errors with  $l = 0$  or  $-1$ ,  $m = 2$ ,  $\alpha = 0$ , and by Carroll and Hall (1988) for normal error with  $l = \alpha = 0$ . Our results for the supersmooth case are more general and compatible with theirs. While for the ordinary smooth cases, the lower bound is better, we can obtain the optimal rates, while they cannot.

#### 4. Discussion.

**REMARK 1.** We have shown that the supersmooth error is much harder to deconvolve than the ordinary smooth error, and the higher the order of the smoothness is, the harder the deconvolution will be for both ordinary smooth and supersmooth cases. If we want to estimate  $T(f) = \sum_0^l a_j f^{(j)}(x_0)$  in  $\mathcal{E}_{m,\alpha,B}$ , then the kernel density estimator  $T(\hat{f}_n) = \sum_1^l a_j \hat{f}_n^{(j)}(x_0)$ ,  $a_l \neq 0$ , achieves the optimal rate  $O((\log n)^{-(m+\alpha-l)/\beta})$  or  $O(n^{-(m+\alpha-l)/(2(m+\alpha+\beta)+1)})$ , depending on the rate of the tail of  $\phi_\varepsilon$ . As the optimal rates of convergence are extremely slow, practically it should be very cautious to deconvolve with supersmooth errors; while it is possible to use the deconvolution techniques for the ordinary smooth cases (e.g., double exponential error).

**REMARK 2.** For estimating the cdf of the random variable  $X$ , we imposed an extra condition that the unknown cdf satisfies

$$(4.1) \quad F(-n) \leq D(\log n)^{-(m+2)/\beta} \quad \text{for some } D > 0,$$

which seems to be uncheckable, as a referee pointed out. However, the condition (4.1) does not really restrict our class of unknown density to a much smaller class. If we want to estimate the probability functional  $T(f) = \int_a^b f(x) dx$ , the conclusions of Theorems 3 and 6 still hold without the extra condition (4.1). We only use such a condition to establish the upper bound, not the lower bound. Under this extra condition, we can show that estimating a cdf is easier than estimating a density in the supersmooth cases, and the result is



heuristically the same as estimating “-1 derivative” of the unknown density. Without this condition, we are not able to identify exactly the optimal rate for estimating a cdf by the proposed procedure.

REMARK 3. For the ordinary smooth case, we give a lower bound of estimating a cdf, which is of order  $O(n^{-(m+\alpha+1)/(2m+2\alpha+2\beta+1)})$ . Exhibiting the corresponding kernel-type of estimator, we can show that an attainable rate of estimating a cdf is  $O(n^{-(m+\alpha+1)/2(m+\alpha+\beta+1)})$  under some assumptions on the derivatives of  $\phi_\varepsilon(t)$ . We conjecture that the latter rate is the best attainable one [ $\beta = 0$  corresponding to  $O(n^{-1/2})$ ]. The reason is that in the current situation, we conjecture that a pair of densities cannot capture the difficulty of the full problem for estimating a cdf [i.e., modulus bound (4.2) below is too small]. To construct an attainable lower bound, one might need to test two highly composite hypotheses as Stone (1982).

REMARK 4. The optimal rates of convergence above are addressed in terms of mean squared errors, which are slightly different from the definition of Stone (1980), who defines them in terms of convergence in probability. However, the results continue to hold under Stone's definition. To see this, note that the upper bound is obtained in the mean square error, which implies the convergence in probability. On the other hand, from (3.3), one can obtain a similar lower bound in terms of convergence in probability.

## 5. Proofs.

PROOF OF THEOREM 1. According to our remark in Section 2, the function  $K(t)$  satisfies the conditions of a kernel function in the ordinary density estimation. Thus we can apply the result of the ordinary kernel density estimation to (2.4) [see Prakasa Rao (1983), pages 46 and 47], and it follows that

$$\begin{aligned} & \sup_{f \in \mathcal{C}_{m,\alpha,B}} \left| E\hat{f}_n^{(l)}(x_0) - f^{(l)}(x_0) \right| \\ &= \sup_{f \in \mathcal{C}_{m,\alpha,B}} \left| \int_{-\infty}^{+\infty} f^{(l)}(x_0 - y) \frac{1}{h_n} K\left(\frac{y}{h_n}\right) dy - f^{(l)}(x_0) \right| \leq Ch_n^{k-l}, \end{aligned}$$

for some constant  $C$ , where  $k = m + \alpha$ . Now the variance of  $\hat{f}_n^{(l)}(x_0)$  is

$$\begin{aligned} \text{var}(\hat{f}_n^{(l)}(x_0)) &\leq \frac{1}{(2\pi)^2 n} E \left| \int_{-\infty}^{+\infty} (-it)^l \exp(-it(x_0 - Y_1)) \frac{\phi_K(th_n)}{\phi_\varepsilon(t)} dt \right|^2 \\ (5.1) \quad &\leq \frac{1}{(2\pi)^2 nh_n^{2+2l}} \left[ \int_{-1}^{+1} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|} dt \right]^2. \end{aligned}$$

By assumption (E2), when  $Mh_n \leq |t| \leq 1$  (for large but fixed  $M$ ),

$$|\phi_\varepsilon(t/h_n)| \geq \frac{d_0}{2} (t/h_n)^{\beta_0} \exp(-h_n^{-\beta}/\gamma).$$

Moreover, by (A3),

$$|\phi_\varepsilon(t/h_n)| \geq \min_{|t| \leq M} |\phi_\varepsilon(t)| > 0 \quad \text{when } |t| \leq Mh_n.$$

Thus, by (5.1),

$$\text{var}(\hat{f}_n^{(l)}(x_0)) \leq \frac{1}{(2\pi)^2 nh_n^{a+2l}} O(\exp(2h_n^{-\beta}/\gamma)) = o(n^{-1/3})$$

by choosing the bandwidth  $h_n = (4/\gamma)^{1/\beta}(\log n)^{-1/\beta}$ , where  $a = 2$  if  $\beta_0 \geq 0$ , and  $a = 2 - \beta_0$  if  $\beta_0 < 0$ . The conclusion follows.  $\square$

PROOF OF THEOREM 2. By choosing the bandwidth as given by Theorem 2 and by the calculation of Theorem 1, we have

$$\sup_{f \in \mathcal{E}_{m,\alpha,B}} |E\hat{f}_n^{(l)}(x_0) - f^{(l)}(x_0)| = O(h_n^{k-l}) = O(n^{-(k-l)/(2k+2\beta+1)}),$$

where  $k = m + \alpha$ . Now, we need only to compute the variance of the estimator. Let

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

Then

$$\hat{f}_n^{(l)}(x_0) = \frac{1}{n} \sum_1^n \frac{1}{h_n^{l+1}} g_{nl}\left(\frac{x_0 - Y_j}{h_n}\right)$$

and

$$(5.2) \quad \text{var}(\hat{f}_n^{(l)}(x_0)) \leq \frac{1}{nh_n^{2+2l}} E g_{nl}^2\left(\frac{x_0 - Y_1}{h_n}\right).$$

Let  $f_Y(y)$  be the density of  $Y = X + \varepsilon$ . Then it is proved by Bickel and Ritov (1988) that  $\sup_{f \in \mathcal{E}_{m,\alpha,B}} f(x) \leq C$ , for some positive constant  $C$ . By Parseval's identity, we have

$$\begin{aligned} E g_{nl}^2\left(\frac{x_0 - Y_1}{h_n}\right) &= h_n \int_{-\infty}^{+\infty} g_{nl}^2(y) f_Y(x_0 - h_n y) dy \\ &\leq \frac{Ch_n}{2\pi} \int_{-\infty}^{+\infty} \frac{|\phi_K(t)t^l|^2}{|\phi_\varepsilon(t/h_n)|^2} dt. \end{aligned}$$

By the similar argument to Theorem 1, the last term is of order  $O(h_n^{-2\beta+1})$ . Hence, by (5.2), we get the desired conclusion.  $\square$

PROOF OF THEOREM 3. Note that

$$\begin{aligned} E\hat{F}_n(x_0) &= \int_{-n^{1/3}}^{x_0} \int_{-\infty}^{+\infty} f(u-y) \frac{1}{h_n} K\left(\frac{y}{h_n}\right) dy du \\ &= \int_{-\infty}^{+\infty} \frac{1}{h_n} (F(x_0-y) - F(-n^{1/3}-y)) K(y/h_n) dy. \end{aligned}$$

Now by a standard argument, we can show that

$$\begin{aligned} &\sup_{f \in C'_{m,\alpha,B}} |E\hat{F}_n(x_0) - F(x_0)| \\ &\leq \sup_{f \in C'_{m,\alpha,B}} \left| \int_{-\infty}^{+\infty} F(x_0-y) \frac{1}{h_n} K(y/h_n) dy - F(x_0) \right| \\ &\quad + \sup_{f \in C'_{m,\alpha,B}} \int_{-\infty}^{+\infty} F(-n^{1/3}-h_n y) |K(y)| dy \\ &\leq O(h_n^{m+\alpha+1}) + O(F(-n^{1/3}[1-h_n])) + \int_{-\infty}^{-n^{1/3}} |K(y)| dy \\ &= O((\log n)^{-(m+\alpha+1)/\beta}), \end{aligned}$$

by using the fact that  $|K(y)| \leq D|y|^{-m-2}$  for some  $D$ . On the other hand, the variance of  $\hat{F}_n(x_0)$  is

$$\begin{aligned} \text{var}(\hat{F}_n(x_0)) &\leq (n^{1/3} + |x_0|) \frac{1}{(2\pi)^2 n h_n^2} \left[ \int_{-\infty}^{+\infty} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|} dt \right]^2 \\ &\leq O\left( n^{1/3} \frac{1}{n h_n^\alpha} \exp(2h_n^{-\beta}/\gamma) \right), \end{aligned}$$

where  $\alpha$  is the same as that in Theorem 1. The proof is complete.  $\square$

We need the following lemmas in order to prove Theorems 4 to 6.

LEMMA 5.1. *Suppose that  $F$  is a cumulative distribution function. Then the convolution density*

$$g_0(x) = \int_{-\infty}^{+\infty} \frac{C_r}{(1+(x-y)^2)^r} dF(y)$$

satisfies

$$g_0(x) \geq D|x|^{-2r} \quad \text{as } x \rightarrow \infty,$$

for some  $D > 0$ , where  $C_r$  is a constant such that  $C_r(1+x^2)^{-r}$  is a density function ( $r > 0.5$ ).

LEMMA 5.2. Suppose  $P\{|\varepsilon - x| \leq |x|^{\alpha_0}\} = O(|x|^{-(a-\alpha_0)})$  (as  $|x| \rightarrow \infty$ ) for some  $0 < \alpha_0 < 1$  and  $a > 1 + \alpha_0$ , and  $H(x)$  is bounded with  $H(x) = O(|x|^{-m_0})$  (as  $|x| \rightarrow \infty$ ). Then there exists a large  $M$  and a constant  $C$  such that when  $|\delta x| \geq M$ ,

$$\int_{-\infty}^{+\infty} H(x-y) dF_\varepsilon(\delta y) \leq C(\delta|x|)^{-a+\alpha_0} \quad \text{for all } \delta \leq 1,$$

provided  $(m_0 + 1)\alpha_0 > a$ , where  $F_\varepsilon$  is the cdf of the random variable  $\varepsilon$ .

PROOF. Divide the real line into two parts:

$$I_1 = \{y: |x - y/\delta| \leq |x|^{\alpha_0}\} \quad \text{and} \quad I_2 = \{y: |x - y/\delta| > |x|^{\alpha_0}\}.$$

Then, by simple algebra, when  $\delta x$  is large enough,

$$\begin{aligned} \int_{-\infty}^{+\infty} H(x-y) dF_\varepsilon(\delta y) &\leq \int_{I_1} + \int_{I_2} H(x-y/\delta) dF_\varepsilon(y) \\ &\leq O((\delta|x|)^{-(a-\alpha_0)}) + O(|x|^{-m_0\alpha_0}), \end{aligned}$$

as having to be shown.  $\square$

PROOF OF THEOREM 4. By relocating  $x_0$  to the origin, without loss of generality assume that  $x_0 = 0$ . Denote  $k = m + \alpha$ . Take a real function  $H(\cdot)$  satisfying the following conditions:

1.  $H^{(l)}(0) \neq 0$ ;
2.  $H^{(j)}(x)$  is bounded continuous for each  $j$ ;
3.  $H(x) = O(x^{-m_0})$  as  $|x| \rightarrow \infty$ , for some given  $m_0$  such that  $(m_0 + 1)\alpha_0 > a$ ;
4.  $\int_{-\infty}^{+\infty} H(x) dx = 0$ ;
5.  $\int_{-\infty}^0 H(x) dx \neq 0$ ;
6.  $\phi_H(t) = 0$  when  $|t|$  is outside  $[1, 2]$ , where  $\phi_H$  is the Fourier transformation of  $H$ .

To see why such a function  $H(\cdot)$  exists, we will take a nonnegative symmetric function  $\phi(t)$  which vanishes outside  $[1, 2]$  when  $t \geq 0$  and has continuous first  $m_0$  bounded derivatives. Moreover,  $\phi(t)$  satisfies

$$(5.3) \quad h^{(l)}(0) \neq h^{(l)}(1)$$

and

$$\int_1^2 \frac{\sin t}{t} \phi(t) dt \neq 0,$$

where  $h(x)$  is the Fourier inversion of  $\phi(t)$  defined by

$$h(x) = \frac{1}{\pi} \int_1^2 \cos(tx) \phi(t) dt.$$

Such a  $\phi(\cdot)$  exists because all functions satisfying the above conditions are infinite dimensional. Let  $H(x) = h(x) - h(x+1)$ . Then its Fourier transformation  $\phi_H(t) = (1 - e^{-it})\phi(t)$ , and  $H(x)$  satisfies conditions 1 to 6.

Now take a pair of densities

$$f_0 = \frac{C_r}{(1+x^2)^r} \quad \text{and} \quad f_n = f_0 + c\delta_n^k H(\cdot/\delta_n),$$

where  $r$  satisfies  $0.5 < r < \min[1, a - \alpha_0 - 0.5]$ . By choosing  $r$  close to 0.5 and  $c$  close to 0, the densities  $f_0$  and  $f_n \in \mathcal{C}_{m, \alpha, B}$ , for all small  $\delta_n$ .

Denote  $g_0 = f_0 * F_\varepsilon$ . Now the  $\chi^2$ -distance between the pair of densities in the convolution space is given by (3.6). Note that by Parseval's identity, we have

$$\begin{aligned} \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} H(x-y) dF_\varepsilon(\delta_n y) \right)^2 dx &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\phi_H(t)\phi_\varepsilon(t/\delta_n)|^2 dt \\ (5.4) \qquad \qquad \qquad &= \frac{1}{\pi} \int_1^2 |\phi_H(t)\phi_\varepsilon(t/\delta_n)|^2 dt \\ &= O(\delta_n^{-2\beta_1} \exp(-2\delta_n^{-\beta}/\gamma)), \end{aligned}$$

uniformly in small  $\delta_n$ . Consequently, by Lemmas 5.1 and 5.2 and (5.4),

$$\begin{aligned} &\int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} H(x-y) dF_\varepsilon(\delta_n y) \right)^2 g_0^{-1}(\delta_n x) dx \\ &\leq \int_{|\delta_n x| \leq M_n} + \int_{|\delta_n x| > M_n} \left( \int_{-\infty}^{+\infty} H(x-y) dF_\varepsilon(\delta_n y) \right)^2 g_0^{-1}(\delta_n x) dx \\ &\leq \frac{M_n^{2r}}{D} \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} H(x-y) dF_\varepsilon(\delta_n y) \right)^2 dx + \int_{|\delta_n x| > M_n} \frac{C^2 |\delta_n x|^{-2a+2\alpha_0}}{D |\delta_n x|^{-2r}} dx \\ &= O(\delta_n^{-2\beta_1} M_n^{2r} \exp(-2\delta_n^{-\beta}/\gamma) + \delta_n^{-1} M_n^{-\varepsilon_0}) \\ &= o(\exp(-\varepsilon_1 \delta_n^{-\beta})), \end{aligned}$$

where  $M_n = \exp(\delta_n^{-\beta}/\gamma)$ ,  $\varepsilon_0 = 2(a - \alpha_0 - r) - 1$  and  $\varepsilon_1 = \min((1-r)/\gamma, \varepsilon_0/2\gamma) > 0$  (by the choice of  $r$ ). Consequently, by (3.6),

$$(5.5) \quad \int_{-\infty}^{+\infty} (f_{Y1} - f_{Y2})^2 (f_{Y1})^{-1} dx = o(\delta_n^{2(m+\alpha)+1} \exp(-\varepsilon_1 \delta_n^{-\beta})).$$

Taking

$$(5.6) \quad \delta_n = \varepsilon_1^{1/\beta} (\log n)^{-1/\beta},$$

we conclude that the right-hand side of (5.5) is of order  $o(1/n)$ , and the order of the lower bound is [see (3.4)]

$$|f_n^{(l)}(0) - f_0^{(l)}(0)| = O(\delta_n^{(k-l)}) = O((\log n)^{-(k-l)/\beta}).$$

The conclusion follows.  $\square$

PROOF OF THEOREM 5. Use the same notation as in the proof of Theorem 5. Take the same  $\phi(t)$  except only the first two continuous derivatives are

required in this case. Now take a pair of densities

$$f_0 = C_r(1 + x^2)^{-r} \quad \text{and} \quad f_n = f_0 + c\delta_n^k H(\cdot/\delta_n).$$

Let  $\phi_H(t) = (1 - e^{-it})\phi(t)$  be the Fourier transformation of  $H(x)$ , and define

$$\phi_{\delta_n}(t) = \frac{d^2(\phi_H(t)\phi_\varepsilon(t/\delta_n))}{dt^2}.$$

By assumption (G3) for  $1 \leq t \leq 2$ , we have (as  $\delta_n \rightarrow 0$ ),

$$(5.7) \quad \delta_n^{-\beta} |\phi_{\delta_n}(t)| \leq C_1/2,$$

for some  $C_1 > 0$ . Now, by the Fourier inversion formula,

$$(5.8) \quad \begin{aligned} \int_{-\infty}^{+\infty} H(x - y) dF_\varepsilon(\delta_n y) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx) \phi_H(t) \phi_\varepsilon(t/\delta_n) dt \\ &= -\frac{1}{2\pi x^2} \int_{1 \leq |t| \leq 2} e^{-itx} \phi_{\delta_n}(t) dt. \end{aligned}$$

Now, we are ready to compute the left-hand side of (3.6). By Parseval's identity, when  $\delta_n$  is small,

$$\begin{aligned} I_1 &\equiv \int_{|x| \leq 1} \left( \int_{-\infty}^{+\infty} H(x - y) dF_\varepsilon(\delta_n y) \right)^2 g_0^{-1}(\delta_n x) dx \\ &\leq C_2 \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} H(x - y) dF_\varepsilon(\delta_n y) \right)^2 dx \\ &= 2C_2 \int_1^2 |\phi_H(t)\phi_\varepsilon(t/\delta_n)|^2 dt \\ &= O(\delta_n^{2\beta}), \end{aligned}$$

where  $g_0 = f_0 * F_\varepsilon$  does not vanish, and hence  $C_2 = \max_{x \in [-1, 1]} g_0^{-1}(x)$  is a finite constant. By Lemma 5.1, (5.7) and (5.8), we have

$$\begin{aligned} I_2 &\equiv \int_{|x| \geq 1} \left( \int_{-\infty}^{+\infty} H(x - y) dF_\varepsilon(\delta_n y) \right)^2 g_0^{-1}(\delta_n x) dx \\ &\leq \delta_n^{2\beta} \int_{|x| \geq 1} (2\pi x^2)^{-2} C_1^2 g_0^{-1}(\delta_n x) dx \\ &= O(\delta_n^{2\beta}). \end{aligned}$$

Consequently, the  $\chi^2$ -distance of the pair of densities [see (3.6)] is of order

$$\delta_n^{2(m+\alpha)+1} (I_1 + I_2) = O(n^{-1}),$$

by taking  $\delta_n = n^{-1/[2(m+\alpha)+2\beta+1]}$ , and the lower bound is of order [see (3.4)]

$$\begin{aligned} |T(f_n) - T(f_0)| &= \delta_n^{m+\alpha-l} |h^{(l)}(1) - h^{(l)}(0)| \\ &= |h^{(l)}(1) - h^{(l)}(0)| n^{-(m+\alpha-l)/[2(m+\alpha+\beta)+1]}. \end{aligned}$$

This completes the proof.  $\square$

**PROOF OF THEOREM 6.** By translation, without loss of generality assume that  $x_0 = 0$ . Take the same least favorable pairs as used in Theorems 4 and 5. Then the lower bound is of order

$$|F_n(0) - F_0(0)| = \delta_n^{m+\alpha} \left| \int_{-\infty}^0 H(x/\delta_n) dx \right| = O(\delta_n^{m+\alpha+1}),$$

where  $\delta_n$  is given by (5.6) for the first conclusion, and  $\delta_n = n^{-1/[2(m+\alpha+\beta)+1]}$  for the second conclusion.  $\square$

**Acknowledgments.** This work is part of the author's Ph.D. dissertation at the University of California, Berkeley, written under the supervision of Professors P. J. Bickel and D. L. Donoho, whose generous guidance and suggestions are gratefully acknowledged and appreciated. The author would also like to express his sincere thanks to an Associate Editor and the referees for bringing to his attention the work of Zhang (1990) and for helpful suggestions and comments that improved the results and presentation. The author is in debt to Thomas Billings for a careful reading of the manuscript and useful suggestions.

## REFERENCES

- BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order convergence estimates. *Sankhyā Ser. A* **50** 381–393.
- CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.
- DONOHO, D. L. and LIU, R. C. (1987). Geometrizing rates of convergence, I. Technical Report 137a, Dept. Statistics, Univ. California, Berkeley.
- DONOHO, D. L. and LIU, R. C. (1991a). Geometrizing rates of convergence, II. *Ann. Statist.* **19** 633–667.
- DONOHO, D. L. and LIU, R. C. (1991b). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668–701.
- DEVROYE, P. L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.* **17** 235–239.
- FAN, J. (1989). Adaptively local one-dimensional subproblems. Mimeo Series 2010, Inst. Statistics, Univ. North Carolina, Chapel Hill.
- FAN, J. and TRUONG, Y. K. (1990). Nonparametric regression with errors in variables. Mimeo Series 2023, Inst. Statistics, Univ. North Carolina, Chapel Hill.
- IBRAGIMOV, I. A., NEMIROVSKII, A. S. and KHAS'MINSKII, R. Z. (1986). Some problems on nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.
- LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.
- LE CAM, L. (1985). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

- LIU, M. C. and TAYLOR, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* **17** 427–438.
- MENDELSON, J. and RICE, J. (1982). Deconvolution of microfluorometric histograms with  $B$  splines. *J. Amer. Statist. Assoc.* **77** 748–753.
- PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic, New York.
- 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. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74** 703–716.
- STEFANSKI, L. A. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184.
- STONE, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.
- STONE, C. (1983). Optimal uniform rates of convergence for nonparametric estimators of a density function or its derivatives. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday* (M. H. Rezvi, J. S. Rustagi and D. Siegmund, eds.) 393–406. Academic, New York.
- WISE, G., TRAGANITIS, A. P. and THOMAS, J. B. (1977). The estimation of a probability density function from measurements corrupted by Poisson noise. *IEEE Trans. Inform. Theory* **23** 764–766.
- ZHANG, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–830.

DEPARTMENT OF STATISTICS  
UNIVERSITY OF NORTH CAROLINA  
CHAPEL HILL, NORTH CAROLINA 27514