Statistica Sinica 1(1991), 541-551

GLOBAL BEHAVIOR OF DECONVOLUTION KERNEL ESTIMATES

Jianqing Fan

University of North Carolina

Abstract: The desire to recover an unknown density when the data are contaminated with errors leads to nonparametric deconvolution problems. The difficulty of deconvolution depends on both the smoothness of the error distribution and the smoothness of the prior. Under certain smoothness constraints, we show that deconvolution kernel density estimates achieve the best global rates of convergence $n^{-\frac{k-l}{2(k+\beta)+1}}$ under an L_p $(1 \le p < \infty)$ norm, where l is the order of the derivative function of the unknown density to be estimated, k is the degree of smoothness constraints, and β is the degree of the smoothness of the error distribution. The results indicate that in the presence of errors, the bandwidth should be chosen larger than the ordinary density estimate. These results also constitute an extension of the ordinary kernel density estimates.

Key words and phrases: Deconvolution, Fourier transforms, kernel density estimates, L_p-norm, global rates of convergence, minimax risks.

1. Introduction

Deconvolution problems arise when direct observations are not possible. The basic model is as follows. We wish to estimate the unknown density of a random variable X, but the only data available are observations $Y_1, \ldots, Y_n$, which are contaminated with independent additive error ε, from the model

$$Y = X + \varepsilon. \tag{1.1}$$

In density function terms, we have realizations $Y_1, \ldots, Y_n$ from the density

$$f_Y(y) = \int_{-\infty}^{+\infty} f_X(y-x)dF_\varepsilon(x), \tag{1.2}$$

and wish to estimate the density f_X of the random variable X, where F_ε is the cumulative distribution function of the random variable ε.

Problems with contaminating error exist in many different fields (e.g., microfluorimetry, electrophoresis, biostatistics), and this model has been widely studied. Interesting applications can be found in Mendelson and Rice (1982), Wise et al. (1977), etc. In a Bayesian setting, the deconvolution problem is precisely the same as the empirical Bayesian estimation of a prior (Berger (1986)). Furthermore, deconvolution is the easiest model to understand the problem of estimating a mixture density (Zhang (1990)). Other fields of application include generalized linear measurement-error models (Anderson (1984), Bickel and Ritov (1987), Stefanski and Carroll (1987)) and nonparametric errors-in-variables regression (Fan, Truong and Wang (1990)), where the covariates are measured with errors. The deconvolution technique is used to recover the density of the covariates.

It is also of theoretical interest to discover and understand the "difficulty" of nonparametric estimation from indirect observations; the convolution model is perhaps the first simple model to try. Here, the "difficulty" of a nonparametric problem means roughly the best attainable rate of the problem. See Donoho and Liu (1987) for the exact meaning of the "difficulty" of a nonparametric problem.

The best local rates and strong consistency have been studied in Carroll and Hall (1988), Fan (1991), Liu and Taylor (1989), Stefanski and Carroll (1990), Zhang (1990), etc. Some simulations have been conducted to examine the effectiveness of deconvolution methods (Stefanski and Carroll (1990)). The results of Fan (1991) indicate clearly that the difficulty of deconvolution depends on the smoothness of the error distribution: the smoother the error distribution the harder the deconvolution will be.

In practice, it may be more interesting to understand how to estimate a whole density function, and how well an estimator behaves globally under certain global losses. The global loss functions we use here are those induced by a weighted L_p-norm:

$$\|f\|_{wp} = \left(\int_{-\infty}^{+\infty} |f(x)|^p w(x) dx \right)^{1/p}, \tag{1.3}$$

where $w(\cdot)$ is a weight function. When $w(x) \equiv 1$, denote $\|\cdot\|_{wp}$ by $\|\cdot\|_p$. A similar measure of global loss is used by Bickel and Rosenblatt (1973) and Stone (1982).

This paper focuses on studying how well a function $T \circ f = \sum_{j=0}^{l} a_j f^{(j)}(x)$ can be estimated, under a smoothness prior

$$\mathcal{F}_{m,\alpha,p,B,w} = \{f : \|f^{(m)}(x) - f^{(m)}(x+\delta)\|_{wp} \le B|\delta|^\alpha,\ |f| \le B\}, \tag{1.4}$$

where $1 > \alpha \ge 0$. When $w(x) \equiv 1$, denote $\mathcal{F}_{m,\alpha,p,B,w}$ by $\mathcal{F}_{m,\alpha,p,B}$. The class of densities here is larger than that formulated by Stone (1982). It includes interesting densities which are excluded by Stone (1982). Hence, when $\varepsilon \equiv 0$, our results are an extension of ordinary density estimates to a wider class of constraints.

We now give an estimating procedure for the deconvolution problem. Let ϕ_X, ϕ_Y, ϕ_ε denote respectively the characteristic functions of the random variables X, Y, ε. Then, $\phi_Y(t) = \phi_X(t)\phi_\varepsilon(t)$. If $\phi_Y(t)$ is known, then by Fourier inversion, f_X can be computed by

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-itx)\frac{\phi_Y(t)}{\phi_\varepsilon(t)}dt.$$

Thus, the problem is reduced to estimating $\phi_Y(t)$, which can be estimated by the empirical characteristic function:

$$\hat{\phi}_n(t) = \frac{1}{n}\sum_{j=1}^{n}\exp(-itY_j).$$

However, $\hat{\phi}_n(t)$ is not a good estimate of $\phi(t)$ at high frequencies. For this reason it is usual to incorporate a damping factor $\phi_K(h_n t)$, where ϕ_K, with $\phi_K(0) = 1$, is the Fourier transform of a kernel K, and h_n is a sequence tending to 0 (so that the damping factor $\phi_K(h_n t) \to 1$). Thus, a deconvoluted kernel density estimate is defined by

$$\hat{f}_n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-itx)\phi_K(th_n)\frac{\hat{\phi}_n(t)}{\phi_\varepsilon(t)}dt. \tag{1.5}$$

See Fan, Truong and Wang (1990) for a different derivation. The bandwidth h_n can be selected to minimize the risk of the estimator, and depends on the smoothness prior.

More generally, we define an estimator of $f^{(j)}(x)$ by

$$\hat{f}_n^{(j)}(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-itx)(-it)^j\phi_K(th_n)\frac{\hat{\phi}_n(t)}{\phi_\varepsilon(t)}dt. \tag{1.6}$$

Note that under some assumptions on integrability, the estimate (1.6) can be written as a kernel type of estimate:

$$\hat{f}_n^{(j)}(x) = \frac{1}{n}\sum_{a=1}^{n} K_{nj}\left(\frac{x - Y_a}{h_n}\right), \tag{1.7}$$

where

$$K_{nj}(x) = \frac{1}{2\pi h_n^{j+1}}\int_{-\infty}^{+\infty} \exp(-itx)\frac{(-it)^j\phi_K(t)}{\phi_\varepsilon(t/h_n)}dt. \tag{1.8}$$

Moreover, we shall use (1.6) to construct estimates of $T \circ f$.

2. Main Results

To discuss the asymptotic behavior of the deconvoluted kernel density estimate (1.6) we need the following assumptions on ϕ_ε and K.

A1) $\phi_\varepsilon(t) \neq 0$ for *any* t.

A2) $K(y)$ is bounded, continuous, and $\int_{-\infty}^{+\infty} |y|^{m+\alpha p}|K(y)|dy < \infty$.

A3) $\phi_K(t)$ is a symmetric function satisfying $\phi_K(t) = 1 + O(|t|^{m+\alpha})$, as $t \to 0$.

Note that condition A1) is sufficient to make the model (1.1) identifiable. When this condition fails the model might not be identifiable and the estimator (1.5) not well defined. Note that assumptions A2) and A3) say essentially that K is an $m+\alpha$ order kernel. Under these assumptions we have following results for biases.

Lemma 1. *Under assumptions* A1) – A3), *if*

$$\int_{-\infty}^{+\infty} \frac{|t^j \phi_K(th_n)|}{|\phi_\varepsilon(t)|} dt < \infty,$$

then

$$\sup_{f \in \mathcal{F}_{m,\alpha,p,B,w}} \|E\hat{f}_n^{(j)}(x) - f^{(j)}(x)\|_{wp} = O(h_n^{m+\alpha-j}) \quad (1 \le p < \infty), \tag{2.1}$$

where $\hat{f}_n^{(j)}$ *is defined by* (1.7) *and* $0 \le j < m+\alpha$.

When $w(x) \equiv 1$, we have the following results under L_2-loss.

Theorem 1. *Under the assumptions* A1) – A3) *with* $p = 2$, *and*

G1) $|\phi_\varepsilon(t)t^\beta| \ge d_0$ *(as* $t \to 0$*) for some positive constant* d_0,

G2) $\int_{-\infty}^{+\infty} |\phi_K(t)|^2 |t|^{2(\beta+l)} dt < \infty$ *and* $\int_{-\infty}^{+\infty} |\phi_K(t)|\,|t|^{\beta+l} dt < \infty$,

with bandwidth $h_n = dn^{-\frac{1}{2(m+\alpha+\beta)+1}}$ *where* $d > 0$, *we have*

$$\sup_{f \in \mathcal{F}_{m,\alpha,2,B}} E\Big\| \sum_{j=0}^{l} a_j \hat{f}_n^{(j)}(x) - \sum_{j=0}^{l} a_j f^{(j)}(x) \Big\|_2 = O\Big(n^{-\frac{m+\alpha-l}{2(m+\alpha+\beta)+1}}\Big),$$

where $l < m+\alpha$.

Under a weighted L_p-norm, we have the following results.

Theorem 2. *Under the assumptions* A1) – A3), *and*

G1)′ $|\phi_\varepsilon(t)t^\beta| \ge d_0$, $|\phi'_\varepsilon(t)t^{\beta+1}| \le d_1$, *as* $t \to \infty$ *with* $d_0 > 0$ *and* $d_1 \ge 0$,

G2)′ $\int_{-\infty}^{+\infty} (|\phi_K(t)| + |\phi'_K(t)|)|t|^{\beta+l} dt < \infty$, *with* $h_n = dn^{-\frac{1}{2(m+\alpha+\beta)+1}}$ *where* $d > 0$, *we have*

$$\sup_{f\in\mathcal{F}_{m,\alpha,p,B,w}} E\Big\|\sum_{j=0}^{l} a_j\hat{f}_n^{(j)}(x) - \sum_{j=0}^{l} a_j f^{(j)}(x)\Big\|_{wp} = O\Big(n^{-\frac{m+\alpha-l}{2(m+\alpha+\beta)+1}}\Big), l < m+\alpha$$

$(1 \le p < \infty)$, *provided that the weight function is integrable.*

Remark 1. The distributions satisfying G1)$'$ include the Gamma, symmetric Gamma, and double exponential distributions, which are called ordinary smooth distributions of order β (see Fan (1991) for a formal definition). Theorems 1 and 2 indicate that, in the presence of errors, we require a larger bandwidth than the ordinary density estimate (in the absence of errors). The smoother (larger β) the error distribution the larger the bandwidth required in order to balance the "bias" and "variance".

Remark 2. By using the idea of adaptively local 1-dimensional subproblems (Fan (1989)), the rates given above are optimal under some additional assumptions on the tail of ϕ_ε (see Fan (1989) for lower bounds). In particular, for estimating $f_X^{(l)}(x)$ under the constraint $\mathcal{F}_{m,\alpha,p,B,w}$, we have the following rates of convergence ($k = m+\alpha$):

error distributions	$\varepsilon \sim$ Gamma(β)	$\varepsilon \sim$ symmetric Gamma(β)	
		$\beta \neq 2j+1$ (j integer)	$\beta = 2j+1$ (j integer)
optimal global rates	$O(n^{-\frac{k-l}{2(k+\beta)+1}})$	$O(n^{-\frac{k-l}{2(k+\beta)+1}})$	$O(n^{-\frac{k-l}{2(k+\beta)+3}})$

Thus, the optimal global rate for estimating $f_X(x)$ is $O(n^{-\frac{k}{2k+5}})$, when the error is double exponential. The best rates above are a little worse than the ordinary density, but not impractical.

Remark 3. It is extremely difficult to do nonparametric deconvolution when the error distributions are normal and Cauchy (called supersmooth distributions, see Fan (1991) for a formal definition). Indeed, based on the traditional perturbation argument (e.g. Farrell (1972)), it has been shown by Fan (1988) and Zhang (1990) that the optimal global rates are extremely slow: $O((\log n)^{-\frac{m+\alpha-l}{\beta}})$ for supersmooth error of order β. In the terminologies of Donoho and Liu (1987), 1-dimensional subproblem is difficult enough to overcome the difficulty of estimating a *whole density* when the error distributions are *supersmooth*, whereas it is *not* the case when the error distributions are ordinary smooth. Indeed, the arguments of Fan (1989) indicate that it requires an $n^{\frac{1}{2(m+\alpha+\beta)+1}}$-dimensional subproblem in order to overcome the difficulty of estimating a whole density function.

Remark 4. The ordinary density estimation corresponds to the case $\beta = 0$ in our setting. Thus, our results are applicable to the ordinary density estimation with an extension to a wider class of constraints. It turns out that the kernel density estimator can also achieve the optimal rates of convergence under an L_p-norm $(1 \le p < \infty)$.

3. Proofs

Proof of Lemma 1. Note that

$$E\hat{f}_n^{(j)}(x) = \int_{-\infty}^{+\infty} f^{(j)}(x - h_n y)K(y)dy.$$

By using the integral form of the remainder term in the Taylor expansion of $f^{(j)}(x)$, we obtain that

$$\begin{aligned} & f^{(j)}(x - h_n y) \\ = & \sum_{i=0}^{m-j-1} \frac{(h_n y)^i}{i!} f^{(i+j)}(x) + \left[\int_0^1 \frac{(1-t)^{m-j-1}}{(m-j-1)!} f^{(m)}(x - th_n y)dt\right](h_n y)^{m-j}, \end{aligned}$$

and using the fact that

$$\int_{-\infty}^{+\infty} K(y)dy = 1, \quad \int_{-\infty}^{+\infty} y^i K(y)dy = 0, \quad i = 1, 2, \ldots, m,$$

we have that for each $f \in \mathcal{F}_{m,\alpha,p,B,w}$, the bias term

$$\begin{aligned} & |E\hat{f}_n^{(j)}(x) - f^{(j)}(x)| \\ \le & h_n^{m-j} \int_{-\infty}^{+\infty} \int_0^1 \frac{(1-t)^{m-j-1}}{(m-j-1)!} |f^{(m)}(x - th_n y) - f^{(m)}(x)|\, |y^{m-j} K(y)| dt\, dy. \end{aligned} \tag{3.1}$$

Let $C = \int_{-\infty}^{+\infty} \int_0^1 \frac{(1-t)^{m-j-1}}{(m-j-1)!} |y^{m-j} K(y)| dt\, dy < \infty$. Then, the function

$$\frac{(1-t)^{m-j-1}}{(m-j-1)!} |y^{m-j} K(y)|/C \quad (0 \le t \le 1, -\infty < y < \infty)$$

is a density function. Thus, by Jensen's inequality and Fubini's Theorem, the

L_p-norm of the second factor of (3.1) is

$$\left\| C \int_{-\infty}^{+\infty} \int_0^1 \frac{(1-t)^{m-j-1}}{(m-j-1)!} |y^{m-j} K(y)|/C |f^{(m)}(x - th_n y) - f^{(m)}(x)| dt\, dy \right\|_{wp}$$
$$\leq C \left\{ \int_{-\infty}^{+\infty} \int_0^1 \frac{(1-t)^{m-j-1}}{(m-j-1)!} |y^{m-j} K(y)| dt\, dy / C \right.$$
$$\left. \cdot \int_{-\infty}^{+\infty} |f^{(m)}(x - th_n y) - f^{(m)}(x)|^p w(x) dx \right\}^{1/p}$$
$$\leq BC^{1-1/p} h_n^{\alpha} \left\{ \int_{-\infty}^{+\infty} \int_0^1 \frac{(1-t)^{m-j-1} t^{\alpha p}}{(m-j-1)!} |y^{m+\alpha p - j} K(y)| dt\, dy \right\}^{1/p}, \tag{3.2}$$

where, in the last display, the inequality

$$\|f^{(m)}(x - th_n y) - f^{(m)}(x)\|_{wp} \leq B |th_n y|^{\alpha}$$

was used.

It follows from (3.1) and (3.2) that

$$\sup_{f \in \mathcal{F}_{m,\alpha,p,B,w}} \|E\hat{f}_n^{(j)}(x) - f^{(j)}(x)\|_{wp} = O(h_n^{m+\alpha-j}).$$

Proof of Theorem 1. With the bandwidth given in Theorem 1, by Lemma 1, we have the "bias"

$$\sup_{\mathcal{F}_{m,\alpha,2,B}} \|E\hat{f}_n^{(j)}(x) - f^{(j)}(x)\|_2 = O\left(n^{-\frac{m+\alpha-j}{2(m+\alpha+\beta)+1}}\right). \tag{3.3}$$

Thus, we need to compute the variance term. By Parseval's identity,

$$\int_{-\infty}^{+\infty} \operatorname{var}(\hat{f}_n^{(j)}(x)) dx = \frac{1}{2\pi n} \int_{-\infty}^{+\infty} \frac{|t|^{2j} |\phi_K(h_n t)|^2}{|\phi_\varepsilon(t)|^2} (1 - |\phi_Y(t)|^2) dt$$
$$\leq \frac{1}{\pi n h_n^{2j+1}} \int_0^{\infty} \frac{t^{2j} |\phi_K(t)|^2}{|\phi_\varepsilon(t/h_n)|^2} dt, \tag{3.4}$$

where $\phi_Y(t)$ is the characteristic function of $Y = X + \varepsilon$. By assumption G1), there exists an M such that, when $|t| > M$,

$$|\phi_\varepsilon(t) t^{\beta}| \geq d_0/2.$$

Hence, by (3.4),

$$\int_{-\infty}^{+\infty} \operatorname{var}(\hat{f}_n^{(j)}(x)) dx \leq \frac{1}{\pi n h_n^{2j+1}} \left[(2/d_0)^2 \int_{Mh_n}^{\infty} t^{2(j+\beta)} |\phi_K(t)|^2 h_n^{-2\beta} dt + O(Mh_n) \right]$$
$$= O\left(n^{-\frac{2(m+\alpha-j)}{2(m+\alpha+\beta)+1}}\right). \tag{3.5}$$

Thus, by the triangular inequality of the L_2-norm,

$$\sup_{f\in\mathcal{F}_{m,\alpha,2,B}} E\Big\|\sum_{j=0}^{l} a_j\hat{f}_n^{(j)}(x)-\sum_{j=0}^{l} a_j f^{(j)}(x)\Big\|_2 = O\Big(n^{-\frac{m+\alpha-l}{2(m+\alpha+\beta)+1}}\Big),$$

and the conclusion follows.

We need the following lemma to prove Theorem 2.

Lemma 2. *Under the assumptions of Theorem* 2

$$\sup_{f\in\mathcal{F}_{m,\alpha,p,B,w}}\sup_x E\Big|K_{nj}\Big(\frac{x-Y_1}{h_n}\Big)-EK_{nj}\Big(\frac{x-Y_1}{h_n}\Big)\Big|^r = O(h_n^{-r(j+\beta+1)+1}),$$
$$r=2,3,\ldots,\qquad j=1,\ldots,l, \tag{3.6}$$

where $K_{nj}(\cdot)$ *is defined by* (1.8).

Proof. Note that

$$E\Big|K_{nj}\Big(\frac{x-Y_1}{h_n}\Big)-EK_{nj}\Big(\frac{x-Y_1}{h_n}\Big)\Big|^r \le 2^r E\Big|K_{nj}\Big(\frac{x-Y_1}{h_n}\Big)\Big|^r.$$

Let f_Y be the density of Y. Then f_Y is bounded by the constant B for any $f\in\mathcal{F}_{m,\alpha,p,B,w}$. By the definition of K_{nj}

$$E\Big|K_{nj}\Big(\frac{x-Y_1}{h_n}\Big)\Big|^r = \int_{-\infty}^{+\infty}|K_{nj}(y)|^r h_n f_Y(x-h_ny)dy$$
$$\le Bh_n\int_{|y|\le 1}\Big[\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{|t|^j|\phi_K(t)|}{h_n^{j+1}|\phi_\varepsilon(t/h_n)|}\Big]^r dy + Bh_n\int_{|y|>1}|K_{nj}(y)|^r dy. \tag{3.7}$$

Arguing as in (3.4) and (3.5), it follows that the first term of (3.7) is of order $O(h_n^{-r(j+1+\beta)+1})$. Integration by parts yields

$$K_{nj}(y)=\frac{1}{2\pi h_n^{j+1}(iy)}\int_{-\infty}^{+\infty}\exp(-ity)\frac{d}{dt}\Big(\frac{(-it)^j\phi_K(t)}{\phi_\varepsilon(t/h_n)}\Big)dt.$$

Thus, arguing again as in (3.4) and (3.5), it follows that

$$|K_{nj}(y)|\le\frac{1}{2\pi h_n^{j+1}|y|}\int_{-\infty}^{+\infty}\Big|\frac{jt^{j-1}\phi_K(t)+t^j\phi_K'(t)}{\phi_\varepsilon(t/h_n)}-\frac{t^j\phi_K(t)\phi_\varepsilon'(t/h_n)}{h_n\phi_\varepsilon^2(t/h_n)}\Big|dt$$
$$\le\frac{D}{h_n^{j+1+\beta}|y|},$$

for some constant D. Consequently, the second term of (3.7) is of order $O(h_n^{-r(j+1+\beta)+1})$. The result follows from (3.7).

Proof of Theorem 2. We need only to prove that

$$E\|\hat{f}_n^{(j)}(x) - f^{(j)}(x)\|_{wp} = O\Big(n^{-\frac{m+\alpha-j}{2(m+\alpha+\beta)+1}}\Big).$$

By Jensen's inequality,

$$E\|\hat{f}_n^{(j)}(x) - f^{(j)}(x)\|_{wp} \le \Big(\int_{-\infty}^{+\infty} E|\hat{f}_n^{(j)}(x) - f^{(j)}(x)|^p w(x)dx\Big)^{1/p}$$

$$\le 2\Big[\int_{-\infty}^{+\infty} E|\hat{f}_n^{(j)}(x) - E\hat{f}_n^{(j)}(x)|^p w(x)dx + \int_{-\infty}^{+\infty} |E\hat{f}_n^{(j)}(x) - f^{(j)}(x)|^p w(x)dx\Big]^{1/p}. \tag{3.8}$$

By Lemma 1

$$\int_{-\infty}^{+\infty} |E\hat{f}_n^{(j)}(x) - f^{(j)}(x)|^p w(x)dx = O(h_n^{(m+\alpha-j)p}) = O\Big(n^{-\frac{(m+\alpha-j)p}{2(m+\alpha+\beta)+1}}\Big) \tag{3.9}$$

uniformly in $f \in \mathcal{F}_{m,\alpha,p,B,w}$. Thus, we need to calculate the first term of (3.8). By applying Jensen's inequality again,

$$E|\hat{f}_n^{(j)}(x) - f^{(j)}(x)|^p \le \Big(E|\hat{f}_n^{(j)}(x) - f^{(j)}(x)|^{2k}\Big)^{p/2k}, \tag{3.10}$$

where k is the smallest integer such that $2k \ge p$. Let

$$Z_{ni}(x) = K_{nj}\Big(\frac{x - Y_i}{h_n}\Big) - EK_{nj}\Big(\frac{x - Y_i}{h_n}\Big).$$

Then, $EZ_{ni}(x) = 0$. By (1.7) and (1.8),

$$\begin{aligned} E|\hat{f}_n^{(j)}(x) - f^{(j)}(x)|^{2k} &= \frac{1}{n^{2k}} \sum_{m_1=1}^{n} \cdots \sum_{m_{2k}=1}^{n} EZ_{nm_1}(x) \cdots Z_{nm_{2k}}(x) \\ &= \frac{1}{n^{2k}}\Big[n(n-1)\cdots(n-k+1)\big(EZ_{n1}^2(x)\big)^k + \cdots\Big] \\ &= O\Big(n^{-k} h_n^{k[-2(j+\beta+1)+1]}\Big) \end{aligned} \tag{3.11}$$

uniformly in $f \in \mathcal{F}_{m,\alpha,p,B,w}$ and x. The last expression follows from Lemma 2. For example, if

$$I = \{(m_1, \ldots, m_{2k}) : m_{i1} = m_{i2}, m_{i3} = m_{i4} = m_{i5}, \text{ the rest are equal}\}\ (k \ge 4),$$

then, by Lemma 2,

$$
\begin{aligned}
&\frac{1}{n^{2k}} \sum_{(m_1,\dots,m_{2k})\in I} EZ_{nm_1}(x)\cdots Z_{nm_{2k}}(x) \\
\le& \frac{n^3}{n^{2k}}\left[EZ_{n1}^2(x)E|Z_{n1}(x)|^3E|Z_{n1}(x)|^{2k-5}\right] \\
=& O\left(\frac{h_n^{-k[2(j+\beta)+1]}}{n^k}(nh_n)^{-(k-3)}\right) \\
=& o\left(n^k h_n^{-k[2(\beta+j)+1]}\right) \qquad (3.12)
\end{aligned}
$$

uniformly in x and $f \in \mathcal{F}_{m,\alpha,p,B,w}$, as $nh_n \to \infty$. Thus, by (3.10) and (3.11),

$$\sup_{f\in\mathcal{F}_{m,\alpha,p,B,w}} \sup_x E|\hat{f}_n^{(j)}(x) - f^{(j)}(x)|^p = O\left(\left[n^{-1}h_n^{-2(j+\beta)-1}\right]^{p/2}\right).$$

The result follows from (3.9) and the choice of the bandwidth.

Acknowledgement

The author is grateful to Professor J. S. Marron for his careful reading of the manuscripts. Thanks are also due to the referee for helpful comments that improved the readability of the paper.

References

Anderson, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* **12**, 1-45.

Berger, J. O. (1986). *Statistical Decision Theory.* Springer-Verlag, New York.

Bickel, P. J. and Ritov, Y. (1987). Efficient estimation in the errors in variables model. *Ann. Statist.* **15**, 513-540.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures on the deviations of density function estimates. *Ann. Statist.* **1**, 1071-1095.

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). Hardest 1-dimensional subproblems. Technical Report 105, Dept. of Statist., University of California, Berkeley.

Fan, J. (1988). Optimal global rates of convergence for nonparametric deconvolution problems. Technical Report 162, Dept. of Statist., University of California, Berkeley.

Fan, J. (1989). Adaptively local 1-dimensional subproblems. Institute of Statistics, Mimeo Series #2010, University of North Carolina, Chapel Hill.

Fan, J., Truong, Y. K. and Wang, Y. (1990). Nonparametric function estimation involving errors-in-variables. To appear in *Nonparametric Functional and Related Topics* (Edited by G. Roussas).

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. To appear in *Ann. Statist.*

Farrell, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43**, 170-180.

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

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

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. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040-1053.

Wise, G. L., 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**, 806-831.

Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, U.S.A.

(Received November 1989; accepted January 1991)