# Global Behavior of Deconvolution Kernel Estimates

Jianqing Fan

Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

November 17, 1989

**Abstract**

The desire to recover the unknown density when data are contaminated with errors leads to nonparametric deconvolution problems. The difficulty of deconvolution depends on both the smoothness of error distribution and the smoothness of the priori. Under a general class of smoothness constraints, we show that deconvolution kernel density estimates achieve the best attainable global rates of convergence $n^{-\frac{k-l}{2(k+\beta)+1}}$ under $L_p$ $(1 \le p < \infty)$ norm, where $l$ is the order of the derivative function of the unknown density to be estimated, $k$ is the degrees of smoothness constraints, and $\beta$ is the degree of the smoothness of the error distribution. Our results indicate that in present 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.

---

[0] *Abbreviated title*: global rates of deconvolution.

*AMS 1980 subject classification.* Primary 62G20. Secondary 62G05.

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

# 1 Introduction

The deconvolution problems arise when direct observations are not possible. The basic model is as follows. We want 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 $\varepsilon$, from the model

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

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

$$f_Y(y) = \int f_X(y - x) dF_\varepsilon(x), \tag{1.2}$$

and we want to estimate the density $f_X$ of the random variable $X$, where $F_\varepsilon$ is the cumulative distribution function of the random variable $\varepsilon$.

Problems with contaminating error exist in many different fields (e.g., microfluorimetry, electrophoresis, biostatistics), and the model has been widely studied. Interesting applications to real data problems are included in Mendelson and Rice (1982), Wise *et al.* (1977), etc. In 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 (1989)). Another field of application is the generalized linear measurement-error models (Anderson (1984), Bickel and Ritov (1987), Stefanski and Carroll (1987a)), where the covariates are measured with error. 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, and 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 (Carroll and Hall (1988), Fan (1988a), Liu and Taylor (1987a), Stefanski and Carroll (1987b), Zhang (1989), etc.).

Some simulations have been conducted to see how an estimate behaves (Stefanski and Carroll (1987b), Liu and Taylor (1987b)). The results of Fan (1988a) indicate clearly that the difficulty of deconvolution depends on the smoothness of the error distribution: the smoother the error distribution is, 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:

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

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), Stone (1982).

The 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 *priori*

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

where and hereafter $1 > \alpha \geq 0$. 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 those 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 will demonstrate that the kernel density estimates achieve the best global rates of convergence. For a nice kernel function $K$, let $\phi_K(t)$ be its Fourier transform with $\phi_K(0) = 1$. Then, a kernel density estimate is defined by

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

where $\hat{\phi}_n(t)$ is an empirical characteristic function defined by

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

More generally, we define

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

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

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

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 will use (1.6) to construct the estimates of $T \circ f$.

## 2 Main Results

To discuss the asymptotic behavior of kernel density estimate (1.6), we need the following assumptions on $\phi_\varepsilon$ 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 the assumptions A2), and A3) say essentially that $K$ is an $m+\alpha$ order kernel. Under these assumptions, we have following results of biases.

**Lemma 1.** Under the assumptions A1) $\sim$ A3), if

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

then

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

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 the $L_2$-loss.

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

G1) $|\phi_\varepsilon(t)t^\beta| \geq d_0$ (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 $\int_{-\infty}^{+\infty} |\phi_K(t)||t|^{\beta+l}dt < \infty$,

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

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

where $l < m + \alpha$.

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

**Theorem 2.** Under the assumptions A1) $\sim$ A3), and

G1)' $|\phi_\varepsilon(t)t^\beta| \geq d_0$, and $|\phi_\varepsilon'(t)t^{\beta+1}| \leq d_1$, as $t \to \infty$ with $d_0 > 0$, and $d_1 \geq 0$,

G2)' $\int_{-\infty}^{+\infty} (|\phi_K(t)| + |\phi_K'(t)|)\,|t|^{\beta+l}dt < \infty$,

then by choosing $h_n = dn^{-\frac{1}{2(m+\alpha+\beta)+1}}$ with some $d > 0$, we have

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

provided that the weight function is integrable, where $l < m + \alpha$.

**Remark 1.** The distributions satisfying G1)' include Gamma distributions, symmetric gamma distributions, and double exponential distributions, which are called ordinary smooth distributions of order $\beta$ (Fan (1988a)). Theorem 1 & 2 indicates that, in present of errors, we have to use a larger bandwidth than the ordinary density estimate (in absent of the errors). The smoother (the larger $\beta$) is, the larger the bandwidth we have to choose in order to balance the "bias" and "variance".

**Remark 2.** By using the adaptively local 1-dimensional idea (Fan (1989)), the rates given above are optimal under some additional assumptions on the tail of $\phi_\varepsilon$ (see Fan (1989) for lower bounds). Specially, 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($\beta$) | $\varepsilon \sim$ symmetric Gamma($\beta$) | |
|---|---|---|---|
| | | $\beta \neq 2j+1$, ($j$ 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 rates for estimating $f_X(x)$ is $O(n^{-\frac{k}{2k+5}})$, when error is double exponential. The best rates above are a little bit worse than the ordinary density, but not too worse to be impractical.

**Remark 3.** It is extremely difficult (hence, impractical) to do nonparametric deconvolution, when the error distributions are normal, and Cauchy, called supersmooth distributions (Fan (1988a)). Indeed, based on the traditional perturbation argument (e.g. Farrell (1972)), it has been shown by Fan (1988b), Zhang (1989) that the optimal global rates are extremely slow $O((\log n)^{\frac{m+\alpha-l}{\beta}})$ for supersmooth error of order $\beta$ (Fan (1988a)). In the terminologies of Donoho (1987), 1-dimensional subproblem is difficult enough to capture the difficulty of estimating a *whole density* when error distributions are *supersmooth*, while it is *not* the case when error distribution is 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 capture the difficulty of estimating a whole density function.

**Remark 4.** The ordinary density estimation corresponds to the case that $\beta = 0$ in our setting. Thus, our results are applicable to the ordinary density estimation with an extension to a wider class of constraints.

## 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 Taylor expansion of $f^{(j)}(x)$,

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

and using the fact that

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

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

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

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

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

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

$$\begin{aligned} & ||C \int_{-\infty}^{+\infty} \int_0^1 \frac{(1-t)^{m-j-1}}{(m-j-1)!} |y^{m-j} K(y)|/C |f^{(m)}(x - t h_n y) - f^{(m)}(x)| dt dy||_{wp} \\ \leq \ & C \left\{ \int_{-\infty}^{+\infty} \int_0^1 \frac{(1-t)^{m-j-1}}{(m-j-1)!} |y^{m-j} K(y)| dt \, dy / C \int_{-\infty}^{+\infty} |f^{(m)}(x - t h_n y) - f^{(m)}(x)|^p w(x) dx \right\}^{1/p} \\ \leq \ & B C^{1-1/p} h_n^{\alpha} \left\{ \int_{-\infty}^{+\infty} \int_0^1 \frac{(1-t)^{m-j-1} t^{\alpha p}}{(m-j-1)!} |y^{m+\alpha p - j} K(y)| dt \, dy \right\}^{1/p}, \quad (3.2) \end{aligned}$$

where in the last display, the inequality

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

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 by Theorem 1, by Lemma 1, we have the "bias"

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

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

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

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

$$|\phi_\varepsilon(t)t^\beta| \ge d_0/2.$$

Hence, by (3.4),

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

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

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

The conclusion follows.

We need the following Lemma to prove Theorem 2.

**Lemma 2**. Under the assumptions of Theorem 2, we have

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

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

**Proof.** Note that

$$E\left[K_{nj}(\frac{x-Y_1}{h_n})-EK_{nj}(\frac{x-Y_1}{h_n})\right]^{2r}\le 2^{2r}E\left[K_{nj}(\frac{x-Y_1}{h_n})\right]^{2r}.$$

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 $K_{nj}$, we have

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

By the arguments from (3.4) to (3.5), we have that the first term of (3.7) is of order $O(h_n^{-2r(j+1+\beta)+1})$. By integration by parts, we have

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

Thus, using the arguments from (3.4) to (3.5) again, we have

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

for some constant D. Consequently, the second term of (3.7) is bounded by $O(h_n^{-2r(j+1+\beta)+1})$. The result follows from (3.7)

**Proof of Theorem 2.** We need only to prove that

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

By Jensen's inequality,

$$\begin{aligned}
E||\hat f_n^{(j)}(x)-f^{(j)}(x)||_{wp} &\le \left(\int_{-\infty}^{+\infty}E|\hat f_n^{(j)}(x)-f^{(j)}(x)|^p w(x)dx\right)^{1/p}\\
&\le 2(\int_{-\infty}^{+\infty}E|\hat f_n^{(j)}(x)-E\hat f_n^{(j)}(x)|^p w(x)dx\\
&\quad+\int_{-\infty}^{+\infty}|E\hat f_n^{(j)}(x)-f^{(j)}(x)|^p w(x)dx)^{1/p}. \qquad (3.8)
\end{aligned}$$

By Lemma 1, we have that

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

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

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

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

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

By (1.7) and (1.8),

$$\begin{aligned}
&\sup_x E|\hat{f}_n^{(j)}(x) - f^{(j)}(x)|^{2k} \\
= \;& \frac{1}{n^{2k}} \sum_{m_1=1}^{n} \cdots \sum_{m_k=1}^{n} EZ_{nm_1}^2(x) \cdots Z_{nm_k}^2(x) \\
= \;& \frac{1}{n^{2k}} \left[n(n-1)\cdots(n-k+1)\left(EZ_{n1}^2(x)\right)^k + \cdots\right] \\
= \;& O\left(n^{-k} h_n^{k[-2(j+\beta+1)+1]}\right), \qquad (3.11)
\end{aligned}$$

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

$$I = \{(m_1, \ldots, m_k):\ m_{i1} = m_{i2}, m_{i3} = m_{i4} = m_{i5}, \text{ the rest are not equal}\},$$

then

$$\begin{aligned}
&\sum_{(m_1,\ldots,m_k)\in I} EZ_{nm_1}^2(x) \cdots Z_{nm_k}^2(x) \\
\le \;& n^k \left[(EZ_{n1}^2(x))^{k-5}(n^{-1}EZ_{n1}^4(x))(n^{-2}EZ_{n1}^6)\right] \\
= \;& n^k O\left(h_n^{-k[2(\beta+j)+1]}(nh_n)^{-1}(nh_n)^{-2}\right) \\
= \;& o\left(n^k h_n^{-k[2(\beta+j)+1]}\right), \qquad (3.12)
\end{aligned}$$

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), we have

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

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

**Acknowledgement**. The author gratefully acknowledges Professor J. S. Marron for his careful reading of the manuscripts and fruitful discussions.

# References

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

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

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

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

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

[6] Donoho, D. L. and Liu, R. C. (1987). Hardest 1-dimensional subproblems. *Tech. Report 105*, Dept. of Statist., University of California, Berkeley.

[7] Fan, J. (1988a). On the optimal rates of convergence for nonparametric deconvolution problem. *Tech. Report 157*, Dept. of Statistics, University of California, Berkeley.

[8] Fan, J. (1988b). Optimal global rates of convergence for nonparametric deconvolution problems. *Tech. Report 162*, Dept. of Statist., Univ. of California, Berkeley.

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

[10] 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.

[11] Liu, M. C. and Taylor, R. L. (1987a). A consistent nonparametric density estimator for the deconvolution problem. *Tech. Report 73*, Dept. of Statist., Univ. of Georgia.

[12] Liu, M. C. and Taylor, R. L. (1987b). Simulations and computations of nonparametric density estimates for the deconvolution problem. *Tech. Report 74*, Dept. of Statist., Univ. of Georgia.

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

[14] Stefanski, L. A. and Carroll, R. J. (1987a). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, **74**, 703–716.

[15] Stefanski, L. A. and Carroll, R. J. (1987b). Deconvoluting kernel density estimators. *Institute of Statistics Mimeo Series #1623*, Univ. of North Carolina, Chapel Hill.

[16] Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040-1053.

[17] Wise, G. L., Traganitis, A. D., and Thomas, J. B. (1977). The estimation of a probability density from measurements corrupted by Poisson noise. *IEEE Trans. Inform. Theory*, **23**, 764–766.

[18] Zhang, C. H. (1989). Fourier methods for estimating mixing densities and distributions. *Preprint.*