ASYMPTOTIC NORMALITY FOR DECONVOLVING KERNEL DENSITY ESTIMATORS

Jianqing Fan

Department of Statistics
University of North Carolina, Chapel Hill
Chapel Hill, NC 27514

ABSTRACT

Suppose that we have n observations from the convolution model $Y = X + \varepsilon$, where X and ε are the independent unobservable random variables, and ε is measurement error with a known distribution. We will discuss the asymptotic normality for deconvolving kernel density estimators of the unknown density $f_X(\cdot)$ of X by assuming either the tail of the characteristic function of ε behaves as $|t|^{\beta_0}\exp(-|t|^{\beta}/\gamma)$ *as* $t \to \infty$ (which is called supersmooth error), or the tail of the characteristic function is of order $O(t^{-\beta})$ (called ordinary smooth error). Asymptotic normality of estimating the functional $T(f) = \sum_1^l a_j f^{(j)}(x_0)$ is also addressed.

KEY WORDS: Asymptotic normality, deconvolution, nonparametric kernel density estimate, Fourier transformation, smoothness of error.

Abbreviated title: Asymptotic normality for deconvolution

AMS(1980) subject classification: Primary 62G05; secondary 60F05.

1. INTRODUCTION

Suppose that we have n observations $y_1, \cdots, y_n$ i.i.d. from the convolution model

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

we want to estimate the unknown density $f_X(x)$ of X nonparametrically, where ε is the measurement error with a known distribution.

Such a model with contaminated error exists in many different fields and has been widely studied. The model arises from microfluorimetry, electrophoresis, biostatistics, and some other fields, where the measurements can not be observed directly. For example, in AIDS study, Y may be the time from some starting point to the time that symptoms appear, ε might be the time from the starting point to the time that infection occurs, and X is the incubation period (the time from the occurrence of infection to the time of symptoms). Obviously, there are many examples of this kind in the biological studies. Additional practical problems of deconvolution can be found in Medgyessy (1977).

Applying the model to Bayesian setting, the deconvolution problem is precisely the same as the empirical Bayesian estimation of prior (Berger(1986)). Another possible field of application is the generalized linear measurement-error model (Anderson (1984), Bickel and Ritov (1987), Schafer (1987), Stefanski and Carroll (1987)), where in the structural model, deconvolution provides a non-parametric way of estimating the unknown density of predictor and hence the conditional first two moments in Schafer's EM algorithm and the unknown density in the efficiency score defined by Stefanski and Carroll(1987) can be estimated non-parametrically. Another interesting application is using deconvolution techniques to estimate a monotone density (Donoho and Liu (1987,1988), Grenander (1956), Greoneboom (1985), etc.) where nonparametric maximum likelihood method and kernel density estimators are typically used. Note that a random variable Y having a monotone density iff $Y \stackrel{d}{=} X\varepsilon$, where X and ε are independent, and ε is distributed uniformly on [0, 1] (hence log ε is distributed as an exponential distribution so that the density of X can be explicitly deconvolved). The approach appears to be new in the literature.

Related works on the deconvolution problem include Carroll and Hall (1988), Fan (1988a, b), Kuelbs (1976), Liu and Taylor (1987a, b), Mendelsohn and Rice (1982), Nadaraja (1965), Schwartz (1969), Stefanski and Carroll (1987), Stefanski (1988), Wise *et al* (1977), and Zhang (1989). Some optimal rate (local and global) results and the uniform consistency results can be found in the related works cited above. The procedure for estimating the unknown density is the kernel density estimator, which achieves the optimal rate of convergence for an appropriate kernel and bandwidth. As in the ordinary kernel density estimation, it appears necessary to find the asymptotic distributions of kernel density estimators so that we can construct the confident intervals and do some other statistical inferences on the unknown f_X.

The kernel density estimator of estimating the unknown density of X is defined by using Fourier inversion, i.e.

$$\hat{f}_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) \frac{\phi_K(th_n)\, \hat{\phi}_n(t)}{\phi_\varepsilon(t)}\, dt, \tag{1.2}$$

where ϕ_K is a known function with $\phi_K(0) = 1$, which can be thought as a characteristic function of a kernel $K(\cdot)$, and $\hat{\phi}_n(t)$ is the empirical characteristic function of the observations defined by

$$\hat{\phi}_n(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(ity_j). \tag{1.3}$$

To see why the estimator is a kernel density estimator, note that $\hat{f}_n(x)$ can be rewritten as

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n}\, g_n\left(\frac{x - y_i}{h_n}\right), \tag{1.4}$$

where

$$g_n(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ity)\, \frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)} dt. \tag{1.5}$$

More generally, we want to find the limit distribution of estimating the lth derivative, $f_X^{(l)}(x)$, of the unknown density $f_X(x)$ under certain conditions. A natural estimator is the lth derivative, defined by

$$\hat{f}_n^{(l)}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h_n^{1+l}} g_n^{(l)}(\frac{x - y_i}{h_n}), \tag{1.6}$$

where

$$g_n^{(l)}(y) \triangleq \frac{1}{2\pi} \int_{-\infty}^{\infty} (-it)^l \exp(-ity)\phi_K(t)/\phi_\varepsilon(t/h_n)\, dt. \tag{1.7}$$

Similar to the results given by Carroll and Hall (1987), Fan (1988a, b), the limiting behavior of the estimator (1.6) will heavily depend on the tail of ϕ_ε. We will separate the problem into two cases: ordinary smooth and supersmooth. Roughly speaking, we will assume that the tail of ϕ_ε behaves as either

i) $\phi_\varepsilon(t) = O(t^{-\beta})$, $t \rightarrow +\infty$ for some $\beta \geq 0$ (ordinary smooth case) or

ii) $|\phi_\varepsilon(t)| = O(t^{\beta_0}\exp(-t^\beta/\gamma))$, $t \rightarrow +\infty$ for some $\beta > 0$, $\gamma > 0$ and real number β_0 (supersmooth case).

The typical examples of super-smooth error distributions are normal, Cauchy, and their mixtures. Examples of ordinary smooth error distributions include gamma, symmetric gamma, double exponential, and their mixtures.

2. ASYMPTOTIC NORMALITY

To discuss asymptotic normality of the estimator (1.6), note that (1.6) is the sum of an i.i.d. sequence. Thus, we need only check the usual triangular CLT conditions hold. A sufficient condition for asymptotical normality

$$\frac{\hat{f}_n^{(l)}(x) - E\hat{f}_n^{(l)}(x)}{\sqrt{\mathrm{var}(\hat{f}_n^{(l)}(x))}} \xrightarrow{L} N(0, 1) \tag{2.1}$$

is that Lyapounov's condition holds, i.e. for some $\delta > 0$,

$$\frac{E\,|Z_{n1} - EZ_{n1}|^{2+\delta}}{n^{\delta/2}[\mathrm{var}(Z_{n1})]^{1+\delta/2}} \longrightarrow 0, \tag{2.2}$$

where

$$Z_{nj} = \frac{1}{h_n^{1+l}} g_n^{(l)}\left(\frac{x - y_j}{h_n}\right). \tag{2.3}$$

To check (2.2), we have to give a lower bound for $\mathrm{var}(Z_{n1})$. For the ordinary smooth case, we need the following lemma, which generalizes the Theorem A1 of Parzen (1962).

Lemma 2.1: Suppose that $K_n(\cdot)$ is a sequence of Borel functions satisfying

$$K_n(y) \longrightarrow K(y), \quad \text{and} \quad \sup_n |K_n(y)| \le K^*(y),$$

where $K^*(y)$ satisfies

$$\int_{-\infty}^{\infty} K^*(y)\, dy < \infty \quad \text{and} \lim_{y \to \infty} |yK^*(y)| = 0.$$

If x is a continuity point of a density $f(\cdot)$, then for any sequence $h_n \longrightarrow 0$, we have

$$\lim_{n \to \infty} \frac{1}{h_n} \int_{-\infty}^{\infty} K_n\left(\frac{x-y}{h_n}\right) f(y)\, dy = f(x) \int_{-\infty}^{\infty} K(y) dy.$$

Case I (ordinary smooth case): For this case, we will need the following assumptions on the tail of characteristic function ϕ_ε and kernel function ϕ_K, call Assumption $A_{m,l}$.

Assumption $\mathbf{A_{m,l}}$:

i) $\phi_\varepsilon(t)\, t^\beta \longrightarrow c$, $\phi'_\varepsilon(t) t^{\beta+1} \longrightarrow -\beta c$, (as $t \longrightarrow +\infty$), with some constant $c \ne 0$ and $\beta \ge 0$. Moreover $\phi_\varepsilon(t) \ne 0$, for all t,

ii) $\phi_K(t)$ is a symmetric function, having $m+2$ bounded integrable derivatives, $\phi_K(0) = 1$ and $\phi_K(t) = 1 + O(|t|^m)$, as $t \to 0$,

iii) $\int_{-\infty}^{\infty} [|\phi_K(t)| + |\phi_K'(t)|]\, |t|^{\beta+l}\, dt < \infty$, $\int_{-\infty}^{\infty} |t|^{2(\beta+l)} |\phi_K(t)|^2\, dt < \infty$,

iv) The mth derivative of the unknown $f_X(\cdot)$ exists and is continuous.

Note that the assumption $A_{m,l}$ is stronger than the assumption $A_{k,l}$ if $m > k$. Note also that the assumption $A_{m,l}$ ii) implies the kernel function

$$K(y) = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{ity} \phi_K(t) dt \tag{2.4}$$

is a real-valued function integrating to unity, and satisfying

$$|K(y)| \le D(1 + |y|^{m+2})^{-1} \tag{2.5}$$

for some positive constant D and

$$\int_{-\infty}^{\infty} y^j K(y)\, dy = 0, \text{ for } 1 \le j \le m-1, \tag{2.6}$$

i.e., $K(\cdot)$ satisfies the ordinary kernel conditions (Prakasa Rao (1983), page 46). Under these assumptions, it can be proved (Lemma 3.1) that if $h_n \to 0$, then

$$h_n^{2\beta} [g_n^{(l)}(x)]^2 \le \frac{C_1^2}{x^2}, \tag{2.7}$$

for some constant C_1, which is independent of n and x.

Theorem 2.1: Under the assumption $A_{l,l}$, if $h_n \to 0$ and $nh_n \to \infty$, then

$$\frac{\hat{f}_n^{(l)}(x) - E\hat{f}_n^{(l)}(x)}{\sqrt{\text{var}(\hat{f}_n^{(l)}(x))}} \xrightarrow{L} N(0,\, 1).$$

Remark 1: The assumption that $\phi_K(t) = 1 + O(t^m)$ as $t \longrightarrow 0$ is not used in the proof of Theorem 2.1 and 2.2. But, it will be used in Corollary 2.1 & 2.2 to ensure that the bias of estimator goes to 0 sufficiently fast so that we can replace $E\hat{f}_n^{(l)}(x)$ by $f_X^{(l)}(x)$.

Even though we obtain the asymptotic normality for a kernel density estimator, its rescaling still involves the unknown quantity $\text{var}(\hat{f}_n^{(l)}(x))$. We have to estimate the scale from the data. Note that the unknown quantity from the proof of Theorem 2.1 is

$$\text{var}(\hat{f}_n^{(l)}(x)) = \frac{1}{n} h_n^{-2(1+l+\beta)+1} \frac{f_Y(x)}{2\pi |c|^2} \int_{-\infty}^{\infty} |t|^{2(l+\beta)} |\phi_K(t)|^2 \, dt (1 + o(1)).$$

Here, the unknown f_Y can be replaced by any of its consistent estimators. For example, it can be replaced by $\hat{f}_Y = \hat{f}_n * F_\varepsilon$, where $\hat{f}_n$ is the kernel density estimator given by (1.2) and F_ε denotes the cdf of ε. But, it involves some extra computation for the convolution and the integration. Naturally, we would rather use sample variance to estimate it. In what follows, we need to check that the sample variance converges in probability to var(Z_{n1}), where Z_{nj} is defined by (2.3). As $E(Z_{n1}^2)$ goes to ∞, and EZ_{n1} is bounded, it is not important to know that whether the sample mean converges to EZ_{n1} or not. In fact, it is not generally true unless h_n tends to 0 sufficiently slow.

Lemma 2.2: Under the assumptions of Theorem 2.1,

$$\frac{\sum_1^n Z_{nj}^2}{nEZ_{n1}^2} \xrightarrow{P} 1, \tag{2.8}$$

and if moreover $nh_n^{2(l+\beta)+1} \longrightarrow \infty$, then

$$\frac{1}{n}\sum_1^n Z_{nj} - EZ_{n1} \xrightarrow{P} 0, \tag{2.9}$$

where Z_{nj} is defined by (2.3).

Corollary 2.1: Under Assumption $A_{m,l}$ $(m > l)$, if $h_n = o(n^{-\frac{1}{2m+2\beta+1}})$ and $f_X^{(m)}(\cdot)$ is bounded, then

$$\sqrt{n}\ \frac{\hat{f}_n^{(l)}(x) - f_X^{(l)}(x)}{s_n} \xrightarrow{L} N(0, 1),$$

where s_n is given by either $s_n^2 = \frac{1}{n}\sum_1^n Z_{nj}^2$ or the sample variance defined by $s_n^2 = \frac{1}{n}\sum_1^n (Z_{nj} - \frac{1}{n}\sum_1^n Z_{nj})^2$ provided that $nh_n^{2(l+\beta)+1} \longrightarrow \infty$, where Z_{nj} is given by (2.3).

Remark 2: By the results of Fan (1988a), to achieve the optimal rates (singular) of convergence under quadratic loss, the optimal rate of bandwidth h_n must be of order $n^{-\frac{1}{2m+2\beta+1}}$. To assure the the bias is a negligible quantity in comparison with the variance of the estimator, we have to let bandwidth go to 0 faster than the optimal one.

Case II (Supersmooth Case): In this case, we will assume that $\phi_\varepsilon(t)$ and $\phi_K(t)$ satisfies the following assumptions, called Assumption $B_{m,l}$.

Assumption $\mathbf{B_{m,l}}$:

i) $c_2 \geq |\phi_\varepsilon(t)||t|^{-\beta_0}\exp(|t|^\beta/\gamma) \geq c_1$, as $t \longrightarrow +\infty$ with $\beta, \gamma, c_1, c_2 > 0$ and some real number β_0. $\phi_\varepsilon(t) \neq 0$ for all t. Write $\phi_\varepsilon(t) = R_\varepsilon(t) + i\, I_\varepsilon(t)$, where $R_\varepsilon(t)$ and $I_\varepsilon(t)$ are the real part and the imaginary part of the characteristic function of ε. Assume furthermore that either $I_\varepsilon(t) = o(R_\varepsilon(t))$ or $R_\varepsilon(t) = o(I_\varepsilon(t))$ as $t \longrightarrow \infty$.

ii) $\phi_K(t)$ is a symmetric and supported with $[-1, 1]$, having the first $m+2$ continuous derivatives. Moreover, $\phi_K(t) > c_3(1-t)^{m+3}$, for $t \in [1-\delta, 1)$ for some $\delta > 0$, and $c_3 > 0$.

iii) $\phi_K(0) = 1$ and $\phi_K(t) = 1 + O(|t|^m)$, as $t \longrightarrow \infty$.

iv) The mth derivative of unknown $f_X(\cdot)$ exists and is continuous.

Note that for this case, it is hard (maybe impossible) to find a simple expression and the exact order for the function $g_n^{(l)}(y)$ defined by (1.7). Consequently, the essential difficulty is in finding a lower bound of EZ_{n1}^2. This is why we impose so many assumptions. Fortunely, these assumptions do not exclude the commonly-used error distributions, such as normal, Cauchy and mixtures of normal, etc.

The assumption i) essentially says that the error distribution is supersmooth, and at the tail, its characteristic function is essentially pure real or pure imaginary. The assumption ii) is imposed to make (1.7) converge. The smoothness condition on $\phi_K(\cdot)$ is imposed to make (2.5) valid. Because of the smoothness conditions of $\phi_K(\cdot)$, the tail condition that $\phi_K(t) \geq c_3(1-t)^{m+3}$ as $t \to 1$ is the simplest one we can impose. The assumption iii) is made so that kernel function satisfies (2.6). It will not be used in proving Theorem 2.2.

Lemma 2.3: Under the assumption $B_{l,l}$, as $n \longrightarrow \infty$,

$$|g_n^{(l)}(y)| \geq c_4 q_l(y) \exp\left(\frac{(1-b_n)^\beta}{\gamma h_n^\beta}\right) h_n^{\beta_0} b_n^{m+4}, \tag{2.10}$$

uniformly over $y \in [0, \pi/2]$, where $b_n = h_n^{\beta/(2(m+5))}$ and c_4 is a positive constant, and

$$q_l(y) = \begin{cases} \left| \dfrac{d^l(\cos y)}{dy^l} \right|, & \text{if } I_\varepsilon(t) = o(R_\varepsilon(t)) \\ \left| \dfrac{d^l(\sin y)}{dy^l} \right|, & \text{if } R_\varepsilon(t) = o(I_\varepsilon(t)) \end{cases}.$$

Now, we are ready to state and prove the asymptotic normality for a kernel density estimator.

Theorem 2.2: Under the assumption $B_{l,l}$, then we have

$$\frac{\hat{f}_n^{(l)}(x) - E\hat{f}_n^{(l)}(x)}{\sqrt{\mathrm{var}(\hat{f}_n^{(l)}(x))}} \xrightarrow{L} N(0, 1),$$

provided that $h_n \sim c(\log n)^{-1/\beta}$, for some $c > 0$.

To show the legitimacy of replacing $\mathrm{var}(Z_{n1})$ by the sample variance, we need the following Lemma.

Lemma 2.4: Under the assumption $B_{l,l}$, if $h_n \sim c\ (\log n)^{-\frac{1}{\beta}}$, then

$$\frac{\sum_1^n Z_{ni}^2}{nEZ_{n1}^2} \xrightarrow{P} 1 \tag{2.11}$$

in probability and if moreover $h_n \sim (a\frac{\gamma}{2}\log n)^{-1/\beta}$ for some $0 < a < 1$, then

$$\frac{1}{n}\sum_1^n Z_{nj} - EZ_{n1} \xrightarrow{P} 0, \tag{2.12}$$

where Z_{nj} is defined by (2.3).

Corollary 2.2: Under the assumption $B_{m,l}$ $(m > l)$, if $h_n \sim (a\frac{\gamma}{2}\log n)^{-1/\beta}$ for some $a > 1$ and $f^{(m)}(\cdot)$ is bounded, then

$$\sqrt{n}\ \frac{\hat{f}_n^{(l)}(x) - f^{(l)}(x)}{s_n} \xrightarrow{L} N(0, 1), \tag{2.13}$$

where $s_n^2 = \frac{1}{n} \sum_1^n Z_{ni}^2$ is defined similarly to Corollary 2.1.

Remark 3: The optimal bandwidth in the supersmooth case is $h_n \sim d\,(\log n)^{-1/\beta}$ for some large d $(> (\gamma/2)^{-1/\beta})$. The constant term d is important for the order of magnitude of the variance term, and the variance can even go to infinite if $d < (\gamma/2)^{-\beta/2}$ (see Lemma 2.3 and (3.7)). By choosing a sufficiently small constant, the bias is a negligible term in comparison to its variance. However, the situation changes for a sufficiently large constant. For a large constant d, the variance will be a small order term. Thus, the mean square error of a kernel estimator and the result of Corollary 2.2 is quite sensitive to choosing a suitable constant. As the optimal rate of the problem is only $O((\log n)^{-\frac{m-l}{\beta}})$ (Fan (1988a, b), we cannot expect that (2.13) converges reasonably fast for moderate sample sizes. The final remark is that the sample mean $\frac{1}{n} \sum_1^n Z_{nj}$ may not converge to EZ_{n1} for the bandwidth given by Corollary 2.2. This is why we don't use the sample variance to estimate var(Z_{n1}) in Corollary 2.2.

3. PROOFS

Proof of Lemma 2.1:

The proof is quite standard (Parzen (1962)). Let $\delta > 0$, and split the region of integration into two parts: $|y| \le \delta$ and $|y| > \delta$. Then

$$\left| \frac{1}{h_n} \int_{-\infty}^{\infty} K_n\left(\frac{x-y}{h_n}\right) f(y)dy - f(x) \int_{-\infty}^{\infty} K(y)dy \right|$$

$$\le \left| \int_{-\infty}^{\infty} [f(x-y) - f(x)] \frac{1}{h_n} K_n(y/h_n)\, dy \right| + f(x) \left| \int_{-\infty}^{\infty} (K_n(y) - K(y))dy \right|$$

$$\le \max_{|y| \le \delta} |f(x-y) - f(x)| \int_{-\infty}^{\infty} K^*(y)dy + \frac{1}{\delta} \sup_{|y| \ge \delta/h_n} |yK^*(y)|$$

$$+ f(x) \int_{|y| \ge \delta/h_n} |K^*(z)|dz + f(x) \left| \int_{-\infty}^{\infty} (K_n(y) - K(y))dy \right|.$$

By Lesbegue's dominated convergence Theorem, and the assumptions, the last three terms tend to 0, as $n \to \infty$. Letting $\delta \to 0$, we conclude the result.

We need the following Lemma to prove Theorem 2.

Lemma 3.1: Under assumption $A_{l,l}$,

$$h_n^{2\beta}[g_n^{(l)}(x)]^2 \le \frac{C_1}{x^2},$$

for some constant C_1.

Proof of Lemma 3.1:

By integration by parts,

$$g_n^{(l)}(x) = \frac{1}{ix} \int_{-\infty}^{\infty} \exp(-itx)\Big[(-it)^l \frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}\Big]' dt.$$

Thus, by the usual argument (see (3.1)),

$$h_n^{2\beta}[g_n^{(l)}(x)]^2 \le \frac{1}{x^2} \left\{ \int_{-\infty}^{\infty} h_n^{\beta} \left| \Big[\frac{t^l \phi_K(t)}{\phi_\varepsilon(t/h_n)}\Big]' \right| dt \right\}^2$$

$$\le \frac{C}{x^2}\{ \int_{-\infty}^{\infty} (|\phi_K(t)| + |\phi'_K(t)|)\, |t|^{\beta+l} dt \}^2,$$

for some constant C. Hence, the assertion follows.

Proof of Theorem 2.1:

Note that the assumption $A_{l,l}$ i) implies that

$$\left| \frac{h_n^{\beta} \phi_K(t)}{\phi_\varepsilon(t/h_n)} \right| \le \max\left\{2 \left| \frac{t^{\beta}\phi_K(t)}{c} \right| 1_{[|t| \ge Mh_n]},\ \frac{\max|\phi_K(t)|}{\min_{|t| \le M} |\phi_\varepsilon(t)|} 1_{[|t| \le Mh_n]}\right\}$$

$$\le g_0(t), \tag{3.1}$$

where g_0 is a positive integrable function such that $\int_{-\infty}^{\infty} |t|^l g_0(t)dt < \infty$, 1_A is the indicator function for a set A, and M is a large enough (but fixed) such that

$$|t^{-\beta} \phi_\varepsilon(t)| \geq \frac{|c|}{2}, \text{ when } |t| \geq M.$$

By Lesbegue's dominated convergence theorem(see (1.7) & (3.1)), we have

$$h_n^\beta g_n^{(l)}(y) \longrightarrow \frac{1}{2\pi c} \int_{-\infty}^{\infty} \exp(-ity)(-it)^l t^\beta \phi_K(t)\, dt,$$

where $g_n^{(l)}(\cdot)$ is the "deconvolution kernel" function defined by (1.7). A consequence of (3.1) is that

$$|h_n^\beta g_n^{(l)}(y)| \leq \int_{-\infty}^{\infty} |t|^l g_0(t)dt \triangleq C_2 < \infty.$$

Combining with (2.7) (see Lemma 3.1), we have

$$h_n^\beta |g_n^{(l)}(y)| \leq \min(C_2, C_1/y). \tag{3.2}$$

Let $f_Y(\cdot)$ be the density of $Y = X + \varepsilon$. Then by the assumption $A_{l,l}$ iv), $f_Y(\cdot)$ is continuous. By Lemma 2.1 and Parseval's identity, we have

$$\begin{aligned} EZ_{n1}^2 &= \frac{1}{h_n^{2(1+l)}} \int_{-\infty}^{\infty} [g_n^{(l)}(\frac{x-y}{h_n})]^2 f_Y(y)dy \\ &= \frac{f_Y(x)}{h_n^{2(1+l+\beta)-1}} \int_{-\infty}^{\infty} \Big[\frac{1}{2\pi c} \int_{-\infty}^{\infty} \exp(-ity)(-it)^l t^\beta \phi_K(t)dt\Big]^2 dy\,(1+o(1)) \\ &= \frac{f_Y(x)}{2\pi |c|^2} \int_{-\infty}^{\infty} |t|^{2(l+\beta)} |\phi_K(t)|^2\, dt\; h_n^{-2(1+l+\beta)+1}(1+o(1)), \end{aligned} \tag{3.3}$$

and by Fubini's Theorem and Lemma 2.1,

$$EZ_{n1} = \int_{-\infty}^{\infty} f_X^{(l)}(x-y)\frac{1}{h_n}K(\frac{y}{h_n})dy \longrightarrow f_X^{(l)}(x),$$

where K is given by (2.4). Similarly, by (3.2) and Lemma 2.1, we conclude that

$$E\,|Z_{n1}|^{2+\delta} = O(h_n^{-(2+\delta)(1+\beta+l)+1}).$$

Thus, Lyapounov's condition (2.2) holds and the conclusion follows from Lyapounov's Theorem.

Proof of Lemma 2.2:

According to the Weak Law of Large Numbers (Chow & Teicher (1978), P328), (2.8) holds if for each $\varepsilon > 0$

$$\frac{1}{EZ_{n1}^2} E\left(Z_{n1}^2 1_{[|Z_{n1}|^2 \geq \varepsilon n\, EZ_{n1}^2]}\right) \longrightarrow 0. \tag{3.4}$$

By (2.3), (3.2) and (3.3), we have

$$|Z_{n1}|^2 \leq \left(\frac{C_2}{h_n^{1+l+\beta}}\right)^2 < \varepsilon\, n\, EZ_{n1}^2 \quad (a.s.)$$

when n is large enough, by using the fact that $nh_n \longrightarrow \infty$. Consequently, (3.4) holds. Now by (3.3), if $nh_n^{2(l+\beta)+1} \longrightarrow \infty$, then

$$n^{-1}\,\mathrm{var}(Z_{n1}) = O(n^{-1}h_n^{-2(1+l+\beta)+1}) \longrightarrow 0,$$

which implies (2.9).

Proof of Corollary 2.1:

By (2.4) and (2.6), the kernel function $K(\cdot)$ satisfies the classical kernel conditions (Prakasa Rao (1983), P46-47). Thus

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

Consequently by Theorem 2.1, (3.3) and the assumption of h_n, $E\hat{f}_n^{(l)}(x)$ can be replaced by $f^{(l)}(x)$ in (2.1). Note that

$$\mathrm{var}(Z_{n1})/EZ_{n1}^2 \longrightarrow 1.$$

By Lemma 2.2,

$$\mathrm{var}(Z_{n1})/s_n^2 \xrightarrow{P} 1.$$

Hence, the result follows.

Proof of Lemma 2.3:

Split the integral region of $g_n^{(l)}(y)$ in (1.7) into three parts: $I_1 = \{t: |t| \le 1 - a_n\}$, $I_2 = \{t: 1 - a_n < |t| \le 1 - b_n\}$, and $I_3 = \{t: 1 \ge |t| > 1 - b_n\}$, where $a_n = b_n + b_n^{m+5}$. Let the results of the integral over I_1 , I_2, and I_3 be J_1, J_2, and J_3, respectively. We will show that J_3 is the dominant term, i.e. $J_1 = o(J_3)$, $J_2 = o(J_3)$, as we expect intuitively.

According to our assumption on the tail of $|\phi_\varepsilon(t)|$, there exists an $M > 0$ such that when $|t| \ge M$,

$$2c_2 \ge |\phi_\varepsilon(t)|\ |t|^{-\beta_0} \exp(|t|^\beta/\gamma) \ge \frac{c_1}{2}. \tag{3.5}$$

Hence,

$$|J_1| \triangleq \left| \int_{I_1} \exp(-ity)(-ity)^l \frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}\, dt \right|$$

$$\le \frac{\max|\phi_K(t)|}{\min\limits_{|t| \le M} |\phi_\varepsilon(t)|} 2Mh_n + \max|\phi_K(t)|$$

$$\times \int_{1 - a_n \ge |t| \ge Mh_n} \frac{2}{c_1} \exp(\frac{|t|^\beta}{\gamma h_n^\beta}) \frac{h_n^{\beta_0}}{|t|^{\beta_0}}\, dt$$

$$= O(\exp(\frac{(1 - a_n)^\beta}{\gamma h_n^\beta})c_n),$$

where

$$c_n = \begin{cases} 1, & \text{if } \beta_0 \ge 0 \\ h_n^{\beta_0}, & \text{if } \beta_0 < 0 \end{cases}.$$

As $b_n \longrightarrow 0$, when n is large enough, we have

$$(1 - a_n)^\beta \doteq (1 - b_n - b_n^{m+5})^\beta \le (1 - b_n)^\beta - \frac{\beta}{2} b_n^{m+5}.$$

Thus,

$$|J_1| \le O(\exp(\frac{(1-b_n)^\beta}{\gamma h_n^\beta})\exp(-\frac{\beta}{2\gamma h_n^{\beta/2}}))$$

$$= o(\exp(\frac{(1-b_n)^\beta}{\gamma h_n^\beta})h_n^{\beta_0}b_n^{m+5}).$$

Similarly, by the assumption on the tail of ϕ_ε, we have

$$|J_2| \triangleq |\int_{1-b_n \ge |t| > 1-a_n} \exp(-ity)(-it)^l \frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}dt|$$

$$\le O(\exp(\frac{(1-b_n)^\beta}{\gamma h_n^\beta})h_n^{\beta_0}b_n^{m+5}),$$

as we integrate over the intervals of length $2(a_n - b_n) = 2b_n^{m+5}$. Finally, we have to evaluate J_3. Note that by the fact that $\phi_\varepsilon(t)$ is a characteristic function, we can rewrite

$$J_3 = \Big[\frac{1}{\pi}\int_{1-b_n<t<1} \phi_K(t)\,(\cos ty\frac{R_\varepsilon(t/h_n)}{|\phi_\varepsilon(t/h_n)|^2} + \sin ty\frac{I_\varepsilon(t/h_n)}{|\phi_\varepsilon(t/h_n)|^2}\,dt\Big]^{(l)}, \tag{3.6}$$

by symmetry. For simplicity, assume $l = 0$ and $I_\varepsilon(t/h_n) = o(R_\varepsilon(t/h_n))$ (for other cases, the proof is essentially the same).

Note that when $1 - b_n < t < 1$ and n is large,

$$\cos ty = \cos y + o(1)$$

uniformly in $y \in [0, \pi/2]$. By assumption $B_{l,J}$ i), $R_\varepsilon(t/h_n)$ can not change its sign over $[\frac{1}{2}, 1]$ when n is large (otherwise, $\phi_\varepsilon(t)$ has zero point). Thus by (3.5) and (3.6), when n is sufficiently large

$$|J_3| \ge \frac{1}{2\pi}\int_{1-b_n}^{1} \phi_K(t)\cos y\frac{1}{2c_2}\exp(\frac{|t|^\beta}{\gamma h_n^\beta})\,h_n^{\beta_0}dt$$

$$> \frac{c_3\cos y}{4\pi c_2}\exp(\frac{(1-b_n)^\beta}{\gamma h_n^\beta})\,h_n^{\beta_0}\int_{1-b_n}^{1}(1-t)^{m+3}\,dt$$

$$\ge c_4\exp(\frac{(1-b_n)^\beta}{\gamma h_n^\beta})\,b_n^{m+4}\,h_n^{\beta_0}\cos y$$

for some constant c_4. The conclusion follows.

Proof of Theorem 2.2:

By Lemma 2.3, when n is sufficiently large

$$EZ_{n1}^2 = \frac{1}{h_n^{1+2l}} \int_{-\infty}^{\infty} [g_n(y)]^2 f_Y(x - h_n y)\, dy$$

$$\geq \frac{1}{h_n^{1+2l}} \int_0^{\pi/2} \left[c_4 q_l(y) \exp\left(\frac{(1-b_n)^\beta}{\gamma h_n^\beta}\right) h_n^{\beta_0} b_n^{m+4}\right]^2 f_Y(x - h_n y) dy$$

$$\geq c_5 \frac{f_Y(x)}{h_n^{1+2l}} \left[\exp\left(\frac{(1-b_n)^\beta}{\gamma h_n^\beta}\right) h_n^{\beta_0} b_n^{m+4}\right]^2$$

$$\geq c_5 f_Y(x) h_n^{c_6} \exp\left(\frac{2}{\gamma h_n^\beta} - \frac{4\beta b_n}{\gamma h_n^\beta}\right), \tag{3.7}$$

for some constants $c_5 > 0$ and c_6. Note that EZ_{n1} is bounded as shown in Theorem 2.1. Consequently, (3.7) is also a lower bound for $\mathrm{var}(Z_{n1})$. By the definition of $g_n^{(l)}$ and (3.5), it is easy to check that

$$|g_n^{(l)}(y)| \leq O\left(\exp\left(\frac{1}{\gamma h_n^\beta}\right) c_n\right), \tag{3.8}$$

where c_n is defined in the proof of Lemma 2.3. By (1.7), (2.3), and (3.8),

$$E\,|Z_{n1}|^{2+\delta} \leq O\left(h_n^{-(2+\delta)(1+l)} \exp\left(\frac{2+\delta}{\gamma h_n^\beta}\right) c_n^{2+\delta}\right). \tag{3.9}$$

Consequently, Lyapounov's condition holds for any $\delta > 0$ by choosing $h_n \sim c\,(\log n)^{-\frac{1}{\beta}}$. The assertion follows.

Proof of Lemma 2.4:

Similar to the proof of Lemma 2.2, we need only check (3.4) and $\frac{1}{n}\mathrm{var}(Z_{n1}) \longrightarrow 0$. By (3.7) and (3.8), it is easy to see the event $\{Z_{n1}^2 \geq n\,\varepsilon\, EZ_{n1}^2\}$ is an empty set when n is sufficiently large, and hence (3.4) holds. By (3.9) with $\delta = 0$, we conclude the second result.

Acknowledgements. The author gratefully acknowledges Prof. P.J. Bickel for his fruitful discussions and encouragement. Thanks are also due to the referees for their constructive comments on a previous draft that improved the readability of the paper.

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. Statisti. Assoc*, **83**, 1184-1186.
6. Chow, Y.S. & Teicher, H. (1978): *Probability Theory.* Springer-Verlag, New York.
7. Crump, J.G. and Seinfeld, J.H. (1982): A new algorithm for inversion of aerosol size distribution data. *Aerosol Science and Technology*, **1**, 15-34.
8. Donoho, D.L. and Liu, R.C. (1987): Geometrizing rate of convergence I. *Tech. Report 137a*, Dept. of Statist., University of California, Berkeley.
9. Donoho, D.L. and Liu, R.C. (1988): Geometrizing Rate of Convergence III. *Tech. Report 105*, 138 Dept. of Statist., University of California, Berkeley.
10. Fan, J. (1988a): On the optimal rates of convergence for non-parametric deconvolution problem. *Tech. Report. 157*, Dept. of Statist., University of California, Berkeley.
11. Fan, J. (1988b): Optimal global rates of convergence for non-parametric deconvolution problem, *Tech. Report. 162*, Dept. of Statist., University of California, Berkeley.
12. Farrell, R.H. (1972): On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Stat.*, **43**, 170-180.

13. Grenander, U. (1956): On the theory of mortality measurement, Part II. *Skand. Akt.*, 39, 125-153.

14. Greoneboom, P. (1985): Estimating a monotone density. *Preceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol II, (L.M. Le Cam and R.A. Olshen eds.), 539-555.

15. Greoneboom, P. and Pyke, R. (1983): Asymptotic normality of statistics based on the convex minorants of empirical distribution functions. *Ann. Probab.* **11**, 328-345.

16. Kuelbs, J.D. (1976): Estimation of the multidimensional probability density function. *MRC Tech. Report*. No. 1646, University of Wisconsin, Madison.

17. Le Cam, L. (1985): *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York-Berlin-Heidelberg.

18. 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.

19. 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.

20. Medgyessy, P. (1977): *Decomposition of Superpositions of Density Functions and Discrete Distributions*. John Wiley, New York.

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

22. Nadaraja, E. (1965): On nonparametric estimation of density function and regression. *Theory Probab. Appl.*, **10**, 186-190.

23. Prakasa Rao, B.L.S. (1983): *Nonparametric Functional Estimation. Academic Press, New York.*

24. Rosenblatt, M. (1976): On the maximum deviation of k-dimensional density estimates. *Ann. Statist.*, **4**, 1009-1015.

25. Schafer, D.W. (1987): Covariate measurememt error in generalized linear models. *Biometrika*, **74**, 385-391.

26. Schwartz, S.C. (1969): On the estimation of Gaussian convolution probability density. *SIAM J. Appl. Math.*, **17**, 447-453.

27. Stefanski, L.A. and Carroll, R. J. (1987): Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, **74** 703-716.

28. Stefanski, L.A. and Carroll, R.J. (1987): Deconvoluting kernel density estimators. *University of North Carolina Mimeo Series* #1623.

29. Stefanski, L.A. (1988): Rates of convergence of some estimators in a class of deconvolution problems. *Manuscript.*

30. Stone, C. (1980): Optimal rates of convergence for nonparametric estimators. *Ann. Stat.* **8**, 1348-1360.

31. Wise, G.L., Tragentis, 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.