The Annals of Statistics
1987, Vol. 15, No. 3, 976–994

WEAK CONVERGENCE OF k-NN DENSITY AND REGRESSION ESTIMATORS WITH VARYING k AND APPLICATIONS

By P. K. Bhattacharya[1] and Y. P. Mack[2]

University of California, Davis

In both density and regression estimation problems, the k-nearest neighbor estimators with k varying in an appropriate range, when transformed to continuous time stochastic processes, are shown to have a common limiting structure under the usual second-order smoothness conditions as the sample size tends to ∞. These results lead to asymptotic linear models in which BLUE's and suitably biased linear combinations are considered.

1. Introduction. In the area of nonparametric density and regression estimation, appropriate choice of the smoothing parameter has always remained a key issue. Specifically, we are thinking of the window-width h_n used in kernel estimators and the integer k_n in k-nearest neighbor (k-NN) estimators, where n denotes the sample size.

Depending on the smoothness class of functions among which estimation is attempted, the appropriate rate at which h_n (or k_n) should tend to 0 (or ∞) as $n \to \infty$ as well as the rate of convergence of the mean squared error (MSE) of the resulting estimators, is well known from the works of Rosenblatt (1956), Parzen (1962), Bartlett (1963), Mack and Rosenblatt (1979) and Mack (1981), while the optimality of these rates of convergence was established by Farrell (1972), Wahba (1975) and Stone (1980).

At a rate appropriate for a certain order of smoothness of the functions to be estimated, one still needs to know the actual value of h_n or k_n to be used for a given set of data. Several adaptive methods have been developed for this purpose based on two main approaches. One of these [considered by Woodroofe (1970) and Krieger and Pickands (1981) in the context of kernel estimators of a density f at a given x] is to use consistent but possibly nonoptimal initial estimators of $f(x)$ and $f''(x)$ in the formula for the optimum bandwidth and to show that the resulting estimator of $f(x)$ is asymptotically efficient. The second approach is a global one, in which minimization of a performance criterion such as the mean integrated squared error or Kullback–Leibler information number, etc., is attempted through cross-validation. Asymptotic efficiency of this approach has been established by Hall (1982), Stone (1984), Marron (1985) and others in varying degrees of generality for kernel estimation of density and by Härdle and Marron (1985) for kernel estimation of regression.

Received December 1985; revised September 1986.

[1] Research supported by NSF Grants MCS 80-02179 and MCS 81-01976.

[2] Research supported by NSF Grant MCS 82-13801.

AMS 1980 *subject classifications*. Primary 62G05, 62J02; secondary 62G20, 62G30, 60F17.

Key words and phrases. Density estimation, regression estimation, nearest neighbor, asymptotic linear model, order statistics, induced order statistics, weak convergence.

Although the subject of bandwidth selection for kernel methods has generated much research in recent years, the corresponding problem for nearest neighbor methods remains virtually unexplored except for a consistency result due to Li (1984). In this paper, we examine the behavior of k-NN density and regression estimators at a given point, as k varies. To keep to the main issue, we consider the problem under a second-order smoothness condition in the one-dimensional case. (Generalization of these results involves further technicalities which will be treated in a future paper).

Our main results are two weak convergence theorems in Section 2, revealing the same formal limiting structure of the k-NN estimators in the density as well as the regression problem, as k varies from $[n^{4/5}a]$ to $[n^{4/5}b]$ for arbitrary $0 < a < b$. This limiting structure also turns out to be the same as the one derived by Krieger and Pickands (1981) for kernel density estimators with varying bandwidth. In Section 3 we consider linear combinations of k-NN estimators with varying k in asymptotic linear models motivated by the weak convergence theorems, obtain formulas for the best linear unbiased estimator (BLUE) in these models and derive their asymptotic distributions. In Section 4 we show that

(i) the method of substituting initial estimators of relevant quantities in formulas for optimum k_n works.

We also show that

(ii) in some situations, the BLUE's in the asymptotic linear models of Section 3 can attain smaller asymptotic MSE's (AMSE) (i.e., MSE's in their asymptotic distributions) than the estimators with the theoretically optimal number k_n^* of nearest neighbors, and

(iii) it is possible to construct appropriately biased linear estimators in these models which are guaranteed to attain smaller AMSE's than the k_n^*-NN estimators.

The proofs of the weak convergence theorems of Section 2 are given in Sections 5 and 6. The asymptotics of the k-NN density estimators with k varying between $n^{4/5}a$ and $n^{4/5}b$ are simplified by the usual device of relating order statistics to the partial sum process for exponentials. Donsker's theorem is applied to this process with time scaled by $n^{4/5}$, providing information about the local behavior of the empirical process. In the regression problem, the asymptotics of partial sums of induced order statistics are handled by a conditional Skorokhod embedding.

2. The main results. Let (X_i, Z_i), $i = 1, 2, \dots,$ be the independent two-dimensional random vectors distributed as (X, Z), where X has marginal cdf F with pdf f and the regression of Z on X is $\mu(x) = E(Z|X = x)$ with residual variance $\sigma^2(x) = \mathrm{Var}(Z|X = x)$ and conditional fourth central moment $\xi(x) = E[\{Z - \mu(x)\}^4|X = x]$. For a fixed x, let $Y_i = |X_i - x|$ and let $0 < Y_{n1} < \cdots < Y_{nn}$ denote the order statistics and $Z_{n1}, \dots, Z_{nn}$ the induced order statistics in $(Y_1, Z_1), \dots, (Y_n, Z_n)$, i.e., $Z_{ni} = Z_j$ if $Y_{ni} = Y_j$. We denote the cdf and the pdf of Y by F_Y and f_Y, respectively, and the regression of Z on Y by $m(y) = E(Z| |X - x| = y)$, i.e.,

$$\begin{aligned} F_Y(y) &= F(x+y) - F(x-y), \\ f_Y(y) &= f(x+y) + f(x-y), \\ m(y) &= [f(x+y)\mu(x+y) + f(x-y)\mu(x-y)]/f_Y(y). \end{aligned}$$

The residual variance and the conditional fourth central moment of Z, given Y, are denoted by

$$\begin{aligned} s^2(y) &= \mathrm{Var}(Z|Y=y), \\ \tau(y) &= E\left[\{Z - m(y)\}^4 | Y = y\right]. \end{aligned}$$

The k-NN estimator of $f(x)$ corresponding to the uniform kernel is

$$f_{nk}(x) = (k-1)/(2nY_{nk}) \tag{1}$$

and the k-NN estimator of $\mu(x)$ with uniform weights is

$$\mu_{nk}(x) = k^{-1} \sum_{j=1}^{k} Z_{nj}. \tag{2}$$

In this section we shall discuss the limiting behaviors of the stochastic processes $\{f_{nk}(x)\}$ and $\{\mu_{nk}(x)\}$ indexed by k, as $n \to \infty$. More precisely, we represent the discrete parameter processes indexed by

$$k_0 = [n^{4/5}a] \le k \le k_1 = [n^{4/5}b], \qquad 0 < a < b, \tag{3a}$$

by letting

$$k = [n^{4/5}t], \qquad a \le t \le b \tag{3b}$$

and defining

$$T_n(t) = f_{n,[n^{4/5}t]}(x), \qquad S_n(t) = \mu_{n,[n^{4/5}t]}(x). \tag{4}$$

We then derive the weak convergence properties of the stochastic processes $\{n^{2/5}[T_n(t) - f(x)],\ a \le t \le b\}$ and $\{n^{2/5}[S_n(t) - \mu(x)],\ a \le t \le b\}$ for $0 < a < b < \infty$.

We shall make the following assumptions:

1. $f(x) > 0$ and f'' is continuous at x.
2. μ'' is continuous at x.
3. The residual variance σ^2 is continuous at x.
4. The conditional fourth central moment ξ is either bounded or Lipschitz.

We now state our main results in the following two theorems of which Theorem D.1 deals with density estimators and Theorem R.1 deals with regression estimators. The symbol $\to_{\mathscr{D}}$ indicates convergence in distribution, i.e., weak convergence of the distributions of the stochastic processes (or random vectors) under consideration and $\{B(t),\ t \ge 0\}$ denotes a standard Brownian motion.

THEOREM D.1. *Under Assumption* 1, *for any* $0 < a < b < \infty$,

$$\{n^{2/5}[T_n(t) - \alpha_D] - \beta_D t^2,\ a \le t \le b\} \to_{\mathscr{D}} \{\alpha_D t^{-1}B(t),\ a \le t \le b\},$$

where

$$\alpha_D = f(x) \quad \text{and} \quad \beta_D = f''(x)/\{24 f^2(x)\}. \tag{5}$$

THEOREM R.1. *Under Assumptions* 1, 2, 3 *and* 4, *for any* $0 < a < b < \infty$,

$$\{n^{2/5}[S_n(t) - \alpha_R] - \beta_R t^2,\ a \le t \le b\} \to_{\mathscr{D}} \{\sigma(x) t^{-1}B(t),\ a \le t \le b\},$$

where

$$\alpha_R = \mu(x) \quad \text{and} \quad \beta_R = \{f(x)\mu''(x) + 2f'(x)\mu'(x)\}/\{24 f^3(x)\}. \tag{6}$$

REMARK 2.1. From the above theorems we see that in the limit, the stochastic processes $T_n(t)$ and $S_n(t)$ have the same formal structure, the only differences being in the formulas for the constants α and β in the deterministic part and the scale factor in the random part. This formal structure was also obtained by Krieger and Pickands (1981) for kernel estimators of density with varying window-width.

3. Asymptotic linear models and linear combinations of k-NN density and regression estimators. Use (3b) to rewrite t in terms of k and (4) to rewrite $T_n(t)$ and $S_n(t)$ in terms of $f_{nk}(x)$ and $\mu_{nk}(x)$. Theorems D.1 and R.1 then suggest the following asymptotic linear models for $f_{nk}(x)$ and $\mu_{nk}(x)$ as n gets large:

$$\begin{aligned} f_{nk}(x) &\simeq \alpha_D + n^{-2/5}\beta_D (kn^{-4/5})^2 + n^{-2/5}\alpha_D (kn^{-4/5})^{-1} B(kn^{-4/5}) \\ &= \alpha_D + (k/n)^2 \beta_D + \alpha_D \Delta_{nk}, \qquad k_0 \le k \le k_1, \end{aligned} \tag{7a}$$

where k_0 and k_1 are given by (3a) and the errors $\Delta_{nk} = n^{2/5}k^{-1}B(kn^{-4/5})$ have

$$E(\Delta_{nk}) = 0, \qquad \operatorname{Cov}(\Delta_{nj}, \Delta_{nk}) = \min(j^{-1}, k^{-1}).$$

Similarly,

$$\mu_{nk}(x) \simeq \alpha_R + (k/n)^2 \beta_R + \sigma(x)\Delta_{nk}, \qquad k_0 \le k \le k_1. \tag{7b}$$

Because of the similarity between the two models, we shall examine the BLUE's of α_D and β_D in (7a) and their asymptotic distribution, and the corresponding results in the other model will be immediate.

First note that due to the covariance structure of $\{\Delta_{nk}\}$,

$$\begin{aligned} \varepsilon_{nk} &= \sqrt{k(k+1)}\,(\Delta_{n,k+1} - \Delta_{nk}), \qquad k_0 \le k \le k_1 - 1, \\ \varepsilon_{nk_1} &= \sqrt{k_1}\,\Delta_{nk_1} \end{aligned} \tag{8}$$

are mutually uncorrelated with mean 0 and variance 1. Taking normalized differences in (7a), we thus have

$$
\begin{aligned}
V_{nk} &= \sqrt{k(k+1)}\,\{f_{n,k+1}(x) - f_{nk}(x)\}\\
&\simeq u_{nk}\beta_D + \alpha_D\varepsilon_{nk}, \qquad k_0 \le k \le k_1 - 1,\\
V_{nk_1} &= \sqrt{k_1}\, f_{nk_1}(x)\\
&\simeq \sqrt{k_1}\,\alpha_D + u_{nk_1}\beta_D + \alpha_D\varepsilon_{nk_1},
\end{aligned} \tag{9a}
$$

where

$$
\begin{aligned}
u_{nk} &= \sqrt{k(k+1)}\,(2k+1)n^{-2}, \qquad k_0 \le k \le k_1 - 1,\\
u_{nk_1} &= k_1^{5/2} n^{-2}.
\end{aligned} \tag{10}
$$

REMARK 3.1. In the linear model given by (9a) and (10), the form of the design matrix and the order of magnitude of its elements are analogous to a simple linear regression in which all but one of approximately $n^{4/5}(b-a)$ observations correspond to a regression through the origin and the regressors u_{nk} are $O(n^{-2/5})$, so that $\Sigma u_{nk}^2 = O(1)$ for these observations. Consequently, the slope β_D cannot be consistently estimated (which, not surprisingly, will show up in the covariance structure of Theorem D.2). The important thing is that these observations still provide enough information about β_D to improve upon the crude estimator $f_{nk_1}(x) = V_{nk_1}/\sqrt{k_1}$ of α_D by bias correction.

The BLUE's of α_D and β_D in the asymptotic linear model given by (8), (9a) and (10), i.e., the BLUE's of these parameters if this linear model were exact, are, respectively,

$$
\hat\alpha_D = f_{nk_1}(x) - \hat\beta_D(k_1/n)^2, \qquad \hat\beta_D = \sum_{k=k_0}^{k_1-1} u_{nk}V_{nk} \Big/ \sum_{k=k_0}^{k_1-1} u_{nk}^2. \tag{11a}
$$

To derive the asymptotic joint distribution of α_D and β_D, note that

$$
\begin{aligned}
\sum_{k=k_0}^{k_1-1} u_{nk}^2 &= n^{-4}\sum_{k=k_0}^{k_1-1} k(k+1)(2k+1)^2 = \int_a^b 4t^4\,dt + O(n^{-4/5})\\
&= \tfrac{4}{5}(b^5 - a^5) + O(n^{-4/5}),
\end{aligned} \tag{12}
$$

and

$$
\begin{aligned}
\sum_{k=k_0}^{k_1-1} u_{nk}V_{nk} &= n^{-2}\sum_{k=k_0}^{k_1-1} k(k+1)(2k+1)\{f_{n,k+1}(x) - f_{nk}(x)\}\\
&= n^{-2}\Bigg[-k_0(k_0+1)(2k_0+1)f_{nk_0}(x) + (k_1-1)k_1(2k_1-1)f_{nk_1}(x)\\
&\qquad\qquad - 6\sum_{k=k_0}^{k_1-1} k^2 f_{nk}(x)\Bigg]\\
&= n^{2/5}\left[-2a^3T_n(a) + 2b^3T_n(b) - 6\int_a^b W_n(t)T_n(t)\,dt\right] + O_p(n^{-2/5}),
\end{aligned}
$$

where $W_n(t) = ([n^{4/5}t]/n^{4/5})^2 \to t^2$ uniformly in $a \le t \le b$. By virtue of the weak convergence of $T_n(t)$ given in Theorem D.1 the above expression further simplifies to

$$\sum_{k=k_0}^{k_1-1} u_{nk}V_{nk} = \tfrac{4}{5}(b^5 - a^5)\beta_D + \alpha_D\left[-6\int_a^b tB(t)\,dt + 2b^2B(b) - 2a^2B(a)\right] + o_p(1). \tag{13a}$$

Substituting (12) and (13a) in (11a), using Theorem D.1 again on $f_{nk_1}(x) = T_n(b)$, and carrying out some algebraic simplification, we arrive at

$$n^{2/5}(\hat{\alpha}_D - \alpha_D) = \alpha_D\xi + o_p(1),$$
$$\hat{\beta}_D - \beta_D = \alpha_D\eta + o_p(1),$$

where

$$\begin{aligned}\eta &= -2Ab^{-5}\left[3\int_a^b tB(t)\,dt - b^2B(b) + a^2B(a)\right],\\ \xi &= -b^2\eta + b^{-1}B(b),\\ A &= \tfrac{5}{4}\left[1 - (a/b)^5\right]^{-1}.\end{aligned} \tag{14}$$

Clearly, ξ and η follow a bivariate normal distribution with mean vector (0, 0) and it is easy to verify that

$$\operatorname{Var}(\eta) = Ab^{-5}, \qquad \operatorname{Var}(b^{-1}B(b)) = b^{-1}, \qquad \operatorname{Cov}(\eta, b^{-1}B(b)) = 0,$$

so that

$$\operatorname{Var}(\xi) = (A+1)b^{-1}, \qquad \operatorname{Cov}(\xi, \eta) = -Ab^{-3}.$$

In the regression problem, we proceed analogously by taking normalized differences in (7b) to arrive at the asymptotic linear model

$$\begin{aligned}V_{nk}^* &= \sqrt{k(k+1)}\,\{\mu_{n,k+1}(x) - \mu_{nk}(x)\}\\ &\simeq u_{nk}\beta_R + \sigma(x)\varepsilon_{nk}, \qquad k_0 \le k \le k_1 - 1,\\ V_{nk_1}^* &= \sqrt{k_1}\,\mu_{nk_1}(x)\\ &\simeq \sqrt{k_1}\,\alpha_R + u_{nk_1}\beta_R + \sigma(x)\varepsilon_{nk_1},\end{aligned} \tag{9b}$$

where the u_{nk}'s and ε_{nk}'s are as in (8) and (10). The BLUE's of α_R and β_R in this asymptotic linear model are, respectively,

$$\hat{\alpha}_R = \mu_{nk_1}(x) - \hat{\beta}_R(k_1/n)^2, \qquad \hat{\beta}_R = \sum_{k=k_0}^{k_1-1} u_{nk}V_{nk}^* \Big/ \sum_{k=k_0}^{k_1-1} u_{nk}^2, \tag{11b}$$

and analogous to (13a) we have

$$\sum_{k=k_0}^{k_1-1} u_{nk}V_{nk}^* \tag{13b}$$
$$= \tfrac{4}{5}(b^5 - a^5)\beta_R + \sigma(x)\left[-6\int_a^b tB(t)\,dt + 2b^2B(b) - 2a^2B(a)\right] + o_p(1).$$

Using (12) and (13b) in (11b), and applying Theorem R.2, we now have

$$n^{2/5}(\hat{\alpha}_R - \alpha_R) = \sigma(x)\xi + o_p(1),$$
$$\hat{\beta}_R - \beta_R = \sigma(x)\eta + o_p(1),$$

where ξ and η are as in (14).

These results are summarized in the following theorems.

THEOREM D.2. *The linear combinations $\hat{\alpha}_D$ and $\hat{\beta}_D$ of $\{f_{nk}(x),\ k_0 \le k \le k_1\}$ given by* (11a) *are the BLUE's of α_D and β_D, respectively, in the asymptotic linear model given by* (8), (9a) *and* (10). *Under Assumption* 1,

$$\left(n^{2/5}(\hat{\alpha}_D - \alpha_D), (\hat{\beta}_D - \beta_D)\right) \to_{\mathscr{D}} \alpha_D(\xi, \eta),$$

where (ξ, n) is bivariate normal with mean vector $(0,0)$ *and covariance matrix*

$$\begin{pmatrix} \sigma_{\xi\xi} & \sigma_{\xi\eta} \\ \sigma_{\xi\eta} & \sigma_{\eta\eta} \end{pmatrix} = \begin{pmatrix} (A+1)b^{-1} & -Ab^{-3} \\ -Ab^{-3} & Ab^{-5} \end{pmatrix}, \qquad A = \tfrac{5}{4}\left[1 - (a/b)^5\right]^{-1}.$$

THEOREM R.2. *The linear combinations $\hat{\alpha}_R$ and $\hat{\beta}_R$ of $\{\mu_{nk}(x),\ k_0 \le k \le k_1\}$ given by* (11*b*) *are the BLUE's of α_R and β_R, respectively, in the asymptotic linear model given by* (8), (9b) *and* (10). *Under Assumptions* 1, 2, 3 *and* 4,

$$\left(n^{2/5}(\hat{\alpha}_R - \alpha_R), (\hat{\beta}_R - \beta_R)\right) \to_{\mathscr{D}} \sigma(x)(\xi, \eta),$$

where (ξ, η) follows the same bivariate normal distribution as in Theorem D.2.

REMARK 3.2. Theorem D.1 and the asymptotic distribution of $\hat{\alpha}_D$ were announced by Bhattacharya and Mack (1985). Corresponding results for kernel estimators of density with bandwidth varying over a finite set were also obtained by Yang and Cox (1984).

REMARK 3.3. For the density estimation problem, the BLUE of α_D in the linear model (9a) does not make use of the fact that the residuals have variance α_D^2. One obvious way to incorporate this information is to consider the usual estimator of the residual variance and take its square root, viz.,

$$\hat{\hat{\alpha}}_D = \left[(k_1 - k_0 - 1)^{-1} \sum_{k=k_0}^{k_1-1} \left(V_{nk} - \hat{\beta}_D u_{nk}\right)^2\right]^{1/2},$$

which is easily shown to have the convergence property

$$n^{2/5}\left(\hat{\hat{\alpha}}_D - \alpha_D\right) \to_{\mathscr{D}} \alpha_D \zeta,$$

where ζ is normally distributed with mean 0 and variance $\sigma_{\zeta\zeta} = \{2(b-a)\}^{-1}$, and is independent of the normal random variable ξ in Theorem D.2. The appropriate combination of $\hat{\alpha}_D$ and $\hat{\hat{\alpha}}_D$ is

$$\alpha_D^* = \frac{\sigma_{\xi\xi}^{-1}\hat{\alpha}_D + \sigma_{\zeta\zeta}^{-1}\hat{\hat{\alpha}}_D}{\sigma_{\xi\xi}^{-1} + \sigma_{\zeta\zeta}^{-1}},$$

having convergence property

$$n^{2/5}(\alpha_D^* - \alpha_D) \to_{\mathscr{D}} \alpha_D \omega,$$

where ω is normally distributed with mean 0 and variance $\sigma_{\omega\omega} = (\sigma_{\xi\xi}^{-1} + \sigma_{\zeta\zeta}^{-1})^{-1}$. The estimator α_D^* thus improves upon the BLUE $\hat{\alpha}_D$. One could also consider the maximum likelihood estimator of (α_D, β_D) for the model (9a) with Gaussian errors. The likelihood equations are a bit messy, and due to inconsistency of the estimator of β_D, the asymptotics for the estimator of α_D become complicated. However, the Fisher-information matrix for (9a) with Gaussian errors is

$$\sum_{k=k_0}^{k_1} I_{nk} = \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\beta} \\ I_{\alpha\beta} & I_{\beta\beta} \end{pmatrix} = \alpha_D^{-2} \begin{pmatrix} 3k_1 - 2k_0 + 2 & \sqrt{k_1}\, u_{nk_1} \\ \sqrt{k_1}\, u_{nk_1} & k_1^3 n^{-2} \end{pmatrix},$$

from which the element $I^{\alpha\alpha}$ in $(\sum_{k=k_0}^{k_1} I_{nk})^{-1}$ is seen to be

$$I^{\alpha\alpha} = n^{-4/5}\sigma_D^2[2(b-a) + b/(A+1)]^{-1}\{1 + o(1)\} = n^{-4/5}\alpha_D^2\sigma_{\omega\omega}\{1 + o(1)\}.$$

Hence the estimator α_D^* has the asymptotic efficiency one would expect the maximum likelihood estimator to have.

REMARK 3.4. The technique described in this section would lead to estimators of α_D and α_R whose MSE's tend to 0 at a rate faster than $n^{-4/5}$ if Assumptions 1 and 2 are strengthened by requiring f'' and μ'' to be Lipschitz of order r, i.e.,

$$|f''(x+h) - f''(x)| \le M|h|^r, \qquad |\mu''(x+h) - \mu''(x)| \le M|h|^r,$$

for all sufficiently small h and for some $M < \infty$ and $0 < r < 1$. For this, consider $\{f_{nk}(x)\}$ and $\{\mu_{nk}(x)\}$ as k varies from $k_0' = [n^{4/5+2\delta}a]$ to $k_1' = [n^{4/5+2\delta}b]$ with $0 < \delta < r/\{5(5+2r)\}$, and define

$$T_n'(t) = f_{n,[n^{4/5+2\delta}t]}(x), \qquad S_n'(t) = \mu_{n,[n^{4/5+2\delta}t]}(x).$$

Then on $a \le t \le b$, the stochastic processes

$$\left\{n^{2/5+\delta}[T_n'(t) - \alpha_D] - n^{5\delta}\beta_D t^2\right\} \quad \text{and} \quad \left\{n^{2/5+\delta}[S_n'(t) - \alpha_R] - n^{5\delta}\beta_R t^2\right\}$$

converge in distribution to $\{\alpha_D t^{-1}B(t)\}$ and $\{\sigma(x)t^{-1}B(t)\}$, respectively. Consequently, the asymptotic linear models (9a) and (9b) hold for $k_0' \le k \le k_1'$, and the BLUE's $(\hat{\alpha}_D, \hat{\beta}_D)$, $(\hat{\alpha}_R, \hat{\beta}_R)$ of the parameters in these models have the same asymptotic distributions as in Theorems D.2 and R.2 with normalizing constants

$n^{2/5+\delta}$ for $\hat{\alpha}_D - \alpha_D$ and $\hat{\alpha}_R - \alpha_R$, and $n^{5\delta}$ for $\hat{\beta}_D - \beta_D$ and $\hat{\beta}_R - \beta_R$. For different values of the exponent $0 < r < 1$ in the Lipschitz conditions previously mentioned, a continuum of smoothness classes is generated as in Farrell's (1972) Case I for density estimation, and by taking δ arbitrarily close to (but less than) $r/\{5(5 + 2r)\}$, the rate of convergence of the MSE of $\hat{\alpha}_D$ or $\hat{\alpha}_R$ can be made to approach the optimal threshold rate of $n^{-(4+2r)/(5+2r)}$. A similar improvement in the rate of convergence of kernel density estimation by introducing Lipschitz condition has been discussed by Ibragimov and Has'minskii (1981), page 235.

4. Applications. From Theorems D.1 and R.1, it follows that $n^{2/5}[T_n(t) - \alpha_D] \to_{\mathscr{D}} N(\beta_D t^2, \alpha_D^2 t^{-1})$ and $n^{2/5}[S_n(t) - \alpha_R] \to_{\mathscr{D}} N(\beta_R t^2, \sigma^2(x) t^{-1})$ for each t, where $N(\mu, \sigma^2)$ denotes a Gaussian r.v. with mean μ and variance σ^2. Hence the asymptotic MSE's (AMSE) of $T_n(t)$ and $S_n(t)$, i.e., MSE's in their asymptotic distributions, are $n^{-4/5}(\beta_D^2 t^4 + \alpha_D^2 t^{-1})$ and $n^{-4/5}(\beta_R^2 t^4 + \sigma^2(x) t^{-1})$, respectively. These AMSE's are minimized at $t_D = \{\alpha_D^2/(4\beta_D^2)\}^{1/5}$ in the density problem, $n^{4/5}\text{AMSE}(T_n(t_D))$ being $\frac{5}{4}\alpha_D^2 t_D^{-1}$ and at $t_R = \{\sigma^2(x)/4\beta_R^2)\}^{1/5}$ in the regression problem, $n^{4/5}\text{AMSE}(S_n(t_R))$ being $\frac{5}{4}\sigma^2(x) t_R^{-1}$. However, we cannot put the estimators $T_n(t_D)$ and $S_n(t_R)$ into practice, because t_D and t_R involve unknown quantities.

Let us define the asymptotic relative efficiency (ARE) of a given density (or regression) estimator with respect to $T_n(t_D)$ [or $S_n(t_R)$] as the ratio of the AMSE of $T_n(t_D)$ [or $S_n(t_R)$] to that of the given estimator. In this section, we construct estimators in the density and regression problems whose ARE's are equal to 1, may exceed 1 in some situations or are guaranteed to exceed 1.

4.1. *Substituting initial estimators in the formulae for optimal t.* By standard weak convergence arguments [in particular, using Theorem 4.4 of Billingsley (1968)] it follows from Theorem D.1 that if $\hat{t}_D$ is a consistent estimator of the optimal t_D in the density problem, then

$$n^{2/5}\left[T_n(\hat{t}_D) - \alpha_D\right] \to_{\mathscr{D}} \beta_D t_D^2 + \alpha_D t_D^{-1} B(t_D).$$

Thus, $T_n(\hat{t}_D)$ has the same asymptotic distribution as $T_n(t_D)$, and, therefore, its ARE equals 1. For the same reason, in the regression problem it follows from Theorem R.1, that if $\hat{t}_R$ is a consistent estimator of t_R, then the ARE of $S_n(\hat{t}_R)$ equals 1. These results parallel the results of Woodroofe (1970) and Krieger and Pickands (1981).

The optimal t_D and t_R are continuous functions of $f(x)$, $\mu(x)$, their first two derivatives and $\sigma^2(x)$, and from consistent estimators of these quantities, $\hat{t}_D$ and $\hat{t}_R$ can be obtained for the preceding purpose. This can easily be accomplished under Assumptions 1, 2 and 3. For example, take the kernel estimators

$$\hat{f}_n(x) = (nh_n)^{-1} \sum_1^n K((x - X_i)/h_n),$$

$$\hat{\mu}_n(x) = (nh_n)^{-1} \sum_1^n Z_i K((x - X_i)/h_n)/\hat{f}_n(x),$$

with $K(u) = C1(|u| < 1)\exp[-u^2/(1-u^2)]$, where C is such that K is a pdf and let $h_n \downarrow 0$ and $nh_n^3 \to \infty$ as $n \to \infty$. Then $\hat{f}_n^{(r)}(x)$ and $\hat{\mu}_n^{(r)}(x)$, $r = 0, 1, 2$, serve our purpose. Finally, the estimated residual variance

$$\hat{\sigma}^2(x) = (k_1 - k_0 - 1)^{-1} \sum_{k=k_0}^{k_1-1} \left(V_{nk}^* - \hat{\beta}_R u_{nk}\right)^2,$$

in the regression problem is a consistent estimator of $\sigma^2(x)$.

4.2. *ARE's of $\hat{\alpha}_D$ and $\hat{\alpha}_R$.* From Theorem D.2, the AMSE of $\hat{\alpha}_D$ is $n^{-4/5}\alpha_D^2(A+1)b^{-1}$. Thus, the ARE of $\hat{\alpha}_D$ with respect to $T_n(t_D)$ is

$$\tfrac{5}{4}t_D^{-1}/\{(A+1)b^{-1}\} = bt_D^{-1}\left[\tfrac{4}{5} + \left\{1 - (a/b)^5\right\}^{-1}\right].$$

The ARE of $\hat{\alpha}_R$ with respect to $S_n(t_R)$ also has the same expression with t_D replaced by t_R. Using consistent estimators, $\hat{t}_D$ and $\hat{t}_R$, we can choose b sufficiently large for any $a/b < 1$ so as to make these ARE's arbitrarily large. However, due to the practical limitation imposed by $k_1 = [n^{4/5}b] \le n$, the choice of b is restricted by a finite quantity for any given sample size. Moreover, values of b near this upper bound fall outside the scope of our theorems. Extensions of Theorems D.1 and R.1 for $b \to \infty$ may give us a better understanding of this point. The situation here has some similarity with the one considered by Abramson (1982).

4.3. *Biased linear combinations of k-NN estimators.* We now consider estimators of the form $\hat{\alpha}_D + cn^{-2/5}\hat{\beta}_D$ for α_D with suitably chosen c. These are the BLUE's of $\alpha_D + cn^{-2/5}\beta_D$ and, therefore, have smaller AMSE than any other linear combinations of $\{f_{nk}(x)\}$ with the same amounts of bias. [The terms "bias" and "MSE" apply here to $\hat{\alpha}_D + cn^{-2/5}\hat{\beta}_D$, or to arbitrary linear combinations of $\{f_{nk}(x)\}$, as estimators of α_D.] In particular, with $c = t_D^2$,

$$\text{AMSE}\left(\hat{\alpha}_D + n^{-2/5}\hat{\beta}_D t_D^2\right) \le \text{AMSE}(T_n(t_D)),$$

provided that $a < t_D < b$, because then $T_n(t_D)$ is a linear combination in the class of estimators under consideration, having a bias of $n^{-2/5}\beta_D t_D^2$. However, $\hat{\alpha}_D + n^{-2/5}\hat{\beta}_D t_D^2$ involves the unknown t_D and a and b have to be chosen so that $a < t_D < b$. To this end, we choose two numbers $0 < \gamma_1 < \gamma_2 < 1$, determine $\hat{a}_D, \hat{b}_D$ by $\hat{b}_D = \hat{t}_D'/\gamma_2$, $\hat{a}_D = \gamma_1\hat{b}_D$, where $\hat{t}_D'$ is a consistent estimator of t_D obtained, as in Section 4.1, with arbitrary $a < b$ and then obtain the BLUE's $\hat{\alpha}_D, \hat{\beta}_D$ in the asymptotic linear model with $a = \hat{a}_D$ and $b = \hat{b}_D$. Finally, let $\hat{t}_D = \{\hat{\alpha}_D^2/(4\hat{\beta}_D^2)\}^{1/5}$ and consider the estimator

$$\tilde{\alpha}_D = \hat{\alpha}_D + n^{-2/5}\hat{\beta}_D\hat{t}_D^2. \tag{15}$$

Then $(\hat{\alpha}_D, \hat{\beta}_D)$ has the same asymptotic joint distribution as given in Theorem D.2 with $a/b = \gamma_1$ and $b = t_D\gamma_2^{-1}$, and

$$\begin{aligned} n^{2/5}(\tilde{\alpha}_D - \alpha_D) &= n^{2/5}(\hat{\alpha}_D - \alpha_D) + \hat{\beta}_D\left\{t_D^2 + o_p(1)\right\} \\ &\to_{\mathscr{D}} N\left(\beta_D t_D^2, \alpha_D^2\left(\alpha_{\xi\xi} + t_D^4\sigma_{\eta\eta} + 2t_D^2\sigma_{\xi\eta}\right)\right). \end{aligned}$$

Hence, using $t_D/b = \gamma_2$ and $A = \frac{5}{4}(1-\gamma_1^5)^{-1}$,

$$n^{4/5}\text{AMSE}(\tilde{\alpha}_D) = \beta_D^2 t_D^4 + \alpha_D^2\left(\sigma_{\xi\xi} + t_D^4\sigma_{\eta\eta} + 2t_D^2\sigma_{\xi\eta}\right)$$
$$= \tfrac{5}{4}\left(\alpha_D^2/t_D\right)\left[1 - \tfrac{4}{5}(1-\gamma_2) + \gamma_2\left(1-\gamma_2^2\right)^2/\left(1-\gamma_1^5\right)\right].$$

Thus, the ARE of $\tilde{\alpha}_D$ with respect to $T_n(t_D)$ is

$$\left[1 - (1-\gamma_2)\left\{\tfrac{4}{5} - \gamma_2(1-\gamma_2)(1+\gamma_2)^2/\left(1-\gamma_1^5\right)\right\}\right]^{-1}. \tag{16}$$

If, in the regression problem, we determine $\hat{a}_R$, $\hat{b}_R$ in the same manner and then construct $\tilde{\alpha}_R = \hat{\alpha}_R + n^{-2/5}\hat{\beta}_R\hat{t}_R^2$ analogous to $\tilde{\alpha}_D$ given in (15), then the ARE of $\tilde{\alpha}_R$ with respect to $S_n(t_R)$ is also given by (16). This, ARE > 1 because $0 < \gamma_1 < \gamma_2 < 1$ implies

$$(1-\gamma_2)\left\{\tfrac{4}{5} - \gamma_2(1-\gamma_2)(1+\gamma_2)^2/\left(1-\gamma_1^5\right)\right\}$$
$$> (1-\gamma_2)\left\{\tfrac{4}{5} - \gamma_2(1-\gamma_2)(1+\gamma_2)^2/(1-\gamma_2)^5\right\}$$
$$= (1-\gamma_2)^4\left(4 + 7\gamma_2 + 4\gamma_2^2\right)/\left\{5\left(1-\gamma_2^5\right)\right\} > 0.$$

Getting back to the more general type of biased estimator of α_D, we first choose $\hat{a}_D$ and $\hat{b}_D$, as previously explained, and then for the BLUE's $\hat{\alpha}_D$, $\hat{\beta}_D$ obtained with this $\hat{a}_D < \hat{b}_D$, we minimize

$$n^{4/5}\text{AMSE}\left(\hat{\alpha}_D + cn^{-2/5}\hat{\beta}_D\right)$$
$$= \alpha_D^2\left[\tfrac{1}{4}c^2t_D^{-5} + (A+1)b^{-1} + c^2Ab^{-5} - 2cAb^{-3}\right],$$

with respect to c. This requires $c_D = At_D^2\gamma_2^3/(\frac{1}{4} + A\gamma_2^5)$, of which a consistent estimator $\hat{c}_D$ is obtained by substituting $\hat{t}_D$ for t_D. The resulting estimator $\hat{\alpha}_D + n^{-2/5}\hat{c}_D\hat{\beta}_D$ will then be an improvement upon $\tilde{\alpha}_D$ defined earlier. Analogous improvement can also be achieved in the regression problem.

5. Proof of Theorem D.1. Some properties of the quantile function of $Y = |X - x|$ and the regression of Z on Y follow from our basic assumptions by elementary calculations. These properties are stated without proof in the following lemma.

LEMMA 1. *By Assumption* 1,

(i) $g(u) = F_Y^{-1}(u)$ *is defined for* $0 \le u \le \varepsilon$ *and for some* $\varepsilon > 0$ *as the unique solution of* $F_Y[g(u)] = u$,

(ii) g''' *is continuous at* 0,

(iii) $g(0) = g''(0) = 0$, $g'(0) = \{2f(x)\}^{-1}$, $g'''(0) = -f''(x)/\{8f^4(x)\}$.

Moreover, *by Assumptions* 1 *and* 2,

(iv) m'' *is continuous at* 0,

(v) $m(0) = \mu(x)$, $m'(0) = 0$, $m''(0) = \mu''(x) + 2f'(x)\mu'(x)/f(x)$.

Finally, by Assumptions 2, 3 *and* 4,

(vi) s^2 *is continuous at* 0, *and*
(vii) $\tau(y) \le M_1 + M_2 y$ *for all* $y > 0$ *and some* $M_1, M_2 < \infty$.

Let $U_{n1} < \cdots < U_{nn}$ denote the order statistics in a random sample $U_1, \dots, U_n$ of size n from the uniform distribution on $(0, 1)$. Recall that $Y_{n1} < \cdots < Y_{nn}$ are the order statistics in $Y_i = |X_i - x|$, $i = 1, \dots, n$, and $g = F_Y^{-1}$ is the quantile function of Y. Thus $Y_{ni} = g(U_{ni})$. Since $Y_{n,[\phi(n)t]}$ with $\phi(n) \to \infty$ and $n^{-1}\phi(n) \to 0$ as $n \to \infty$ are the key elements in k-NN density and regression estimation, we need the following properties of these order statistics. By $1(S)$ we denote the indicator function of a set S and for simplicity of notation, we write $Y_{\phi t}$ and $U_{\phi t}$ for $Y_{n,[\phi(n)t]}$ and $U_{n,[\phi(n)t]}$, respectively.

LEMMA 2. *Let* $\phi(n) \to \infty$ *and* $n^{-1}\phi(n) \to 0$ *as* $n \to \infty$. *Then*

(i) *for* $B > b$ *and sufficiently large* n,

$$P\left[U_{\phi b} > n^{-1}\phi(n)B\right] \le \exp\left[-2n^{-1}\phi^2(n)(B - b)^2\right];$$

(ii) *for* $B > b/f(x)$ *and sufficiently large* n,

$$P\left[Y_{\phi b} > n^{-1}\phi(n)B\right] \le \exp\left[-2n^{-1}\phi^2(n)\{Bf(x) - b\}^2\right].$$

PROOF. We prove (ii) and indicate during the proof how it should be modified for (i). Since $Y_{\phi b} > n^{-1}\phi(n)B$ if and only if $\sum_1^n 1(Y_j \le n^{-1}\phi(n)B) < [\phi(n)b]$, we have

$$P\left[Y_{\phi b} > n^{-1}\phi(n)B\right] = P\Bigg[\sum_1^n \left\{1\left(Y_j \le n^{-1}\phi(n)B\right) - E1\left(Y \le n^{-1}\phi(n)B\right)\right\}$$
$$< -n\left\{E1\left(Y \le n^{-1}\phi(n)B\right) - n^{-1}[\phi(n)b]\right\}\Bigg].$$

Now for large n,

$$\begin{aligned} E1\left(Y \le n^{-1}\phi(n)B\right) - n^{-1}[\phi(n)b] &= \int_{x - n^{-1}\phi(n)B}^{x + n^{-1}\phi(n)B} f(t)\,dt - n^{-1}[\phi(n)b] \\ &> n^{-1}\phi(n)Bf(x) - n^{-1}\phi(n)b \\ &= n^{-1}\phi(n)\{Bf(x) - b\} > 0. \end{aligned}$$

[For part (i), $E1(U \le n^{-1}\phi(n)B) - n^{-1}[\phi(n)b] > n^{-1}\phi(n)(B - b) > 0$.] It now follows by an application of Theorem 1 of Hoeffding (1963) that

$$\begin{aligned} P\left[Y_{\phi b} > n^{-1}\phi(n)B\right] &\le \exp\left[-2n\left\{E1\left(Y \le n^{-1}\phi(n)B\right) - n^{-1}[\phi(n)b]\right\}^2\right] \\ &< \exp\left[-2n^{-1}\phi^2(n)\{Bf(x) - b\}^2\right]. \end{aligned}$$ □

COROLLARY 1. *If* $\phi(n) = n^{\varepsilon}$, $\frac{1}{2} < \varepsilon < 1$, *then* $U_{\phi b}$ *and* $Y_{\phi b}$ *are* $O_p(n^{-1}\phi(n))$.

PROOF. For each B, the previous bounds can be made smaller than arbitrary $\varepsilon > 0$ by making n sufficiently large. □

COROLLARY 2. *If* $\phi(n) = n^{\varepsilon}$, $\frac{1}{2} < \varepsilon < 1$, *then for sufficiently large* n,

$$E\left[Y_{\phi b}\right] \leq 2n^{-1}\phi(n)b/f(x).$$

PROOF.

$$\begin{aligned}
E\left[Y_{\phi b}\right] &= n^{-1}\phi(n)\int_0^{\infty} P\left[n\phi(n)^{-1}Y_{\phi b} > y\right] dy \\
&\leq n^{-1}\phi(n)\left[\int_0^{b/f(x)} 1\, dy + \int_{b/f(x)}^{\infty} \exp\left[-2n^{-1}\phi^2(n)f^2(x)\right.\right. \\
&\qquad\qquad \left.\left.\times\{y - b/f(x)\}^2\right] dy\right] \\
&= n^{-1}\phi(n)f(x)^{-1}\left[b + n^{1/2}\phi(n)^{-1}(\pi/8)^{1/2}\right] \\
&< 2n^{-1}\phi(n)b/f(x),
\end{aligned}$$

by a change of variable, when n is large. □

Before proceeding to the proof of Theorem D.1, we state a well-known representation of the uniform order statistics in the following lemma [see, e.g., Bickel and Doksum (1977), page 44]. We use the symbol $\{X_\lambda,\ \lambda \in \Lambda\} =_{\mathscr{D}} \{X'_\lambda,\ \lambda \in \Lambda\}$ to indicate that two collections of random variables have the same joint distribution.

LEMMA 3.

$$\begin{aligned}
\{U_{nk}, 1 \leq k \leq n\} &=_{\mathscr{D}} \left\{\sum_1^k W_i \Big/ \sum_1^{n+1} W_i,\ 1 \leq k \leq n\right\} \\
&= \left\{k(n+1)^{-1}\left(1 + k^{-1}\sum_1^k W_i^*\right)\right. \\
&\qquad \left.\times\left(1 + (n+1)^{-1}\sum_1^{n+1} W_i^*\right)^{-1}, 1 \leq k \leq n\right\},
\end{aligned}$$

where $W_1, \ldots, W_{n+1}$ *are iid negative exponential rv's with mean* 1 *and* $W_i^* = W_i - 1$ *are iid with mean* 0 *and variance* 1.

From now on we treat $n^{4/5}t = \phi(n)t$ as an integer to avoid unnecessary complications. Thus, $T_n(t)$ given by (1) and (4) becomes

$$T_n(t) = \{\phi(n)t - 1\}\left\{2ng(U_{\phi t})\right\}^{-1}, \tag{17}$$

where $g = F_Y^{-1}$ is the quantile function of $Y = |X - x|$.

In the Taylor expansion of $g(U_{\phi t})$ around 0, use Lemma 1(iii) and (5) to obtain

$$
\begin{aligned}
\{2g(U_{\phi t})\}^{-1} &= \alpha_D U_{\phi t}^{-1}\left[1 - \alpha_D^{-1}\beta_D U_{\phi t}^2 + R_n(t)U_{\phi t}^2\right]^{-1} \\
&= \alpha_D U_{\phi t}^{-1}\left[1 + \alpha_D^{-1}\beta_D U_{\phi t}^2 + R_n'(t)U_{\phi t}^2\right] \qquad (18)\\
&= \alpha_D U_{\phi t}^{-1} + \beta_D U_{\phi t} + \alpha_D R_n'(t) U_{\phi t},
\end{aligned}
$$

where $R_n(t) = 3^{-1}\alpha_D\{g'''(\lambda U_{\phi t}) - g'''(0)\}$, $0 \le \lambda \le 1$, and $R_n'(t)$ is obtained in terms of $R_n(t)$ and $U_{\phi t}^2$ by comparing the second and third expressions in (18). Substituting this in (17), we have

$$
\begin{aligned}
T_n(t) = \Big[\alpha_D\{n^{-1}\phi(n)tU_{\phi t}^{-1}\} + \beta_D\{n^{-1}\phi(n)tU_{\phi t}\} \qquad (19)\\
+ R_n''(t)\Big]\left[1 + O(n^{-4/5})\right],
\end{aligned}
$$

with $R_n''(t) = \alpha_D\{n^{-1}\phi(n)tU_{\phi t}\}R_n'(t)$. We now examine the terms in (19). By Lemma 3, using the standard calculus of o_p and O_p [see Pratt (1959)],

$$
\begin{aligned}
n^{-1}\phi(n)tU_{\phi t}^{-1} &=_{\mathscr{D}} (1 + n^{-1})\left[1 + n^{-2/5}t^{-1}\phi(n)^{-1/2}\sum_1^{\phi(n)t} W_i^*\right]^{-1} \\
&\quad \times \left[1 + (n+1)^{-1}\sum_1^{n+1} W_i^*\right] \qquad (20)\\
&= \left[1 - n^{-2/5}t^{-1}\phi(n)^{-1/2}\sum_1^{\Phi(n)t} W_i^* + o_p(n^{-2/5})\right]\left[1 + O_p(n^{-1/2})\right] \\
&= 1 - n^{-2/5}t^{-1}\phi(n)^{-1/2}\sum_1^{\phi(n)t} W_i^* + o_p(n^{-2/5}),
\end{aligned}
$$

and, similarly,

$$
\begin{aligned}
n^{-1}\phi(n)tU_{\phi t} &=_{\mathscr{D}} \{n^{-1}\phi(n)t\}^2\left[1 + \{\phi(n)t\}^{-1}\sum_1^{\phi(n)t} W_i^*\right]\left[1 + O_p(n^{-1/2})\right] \qquad (21)\\
&= n^{-2/5}t^2 + O_p(n^{-4/5}) = n^{-2/5}t^2 + o_p(n^{-2/5}),
\end{aligned}
$$

where the o_p-terms in (20) and (21) are uniform in $0 < a \le t \le b < \infty$ by virtue of

$$
\sup_{a \le t \le b}\left|\Phi(n)^{-1/2}\sum_1^{\phi(n)t} W_i^*\right| = O_p(1).
$$

For the remainder term, we first look at $R_n(t)$ in (18). Under Assumption 1, g''' is continuous at 0 by Lemma 1(ii) and $U_{\phi t} = O_p(n^{-1}\phi(n)) = O_p(n^{-1/5})$, by Corollary 1 to Lemma 2. Hence, $R_n(t) = o_p(1)$, so that $R_n'(t)$ is also $o_p(1)$, which leads to $R_n''(t) = o_p(n^{-2/5})$. Moreover, as in (20) and (21), this o_p-term is also uniform in $a \le t \le b$, because $U_{\phi t} \le U_{\phi b}$ for all $t \le b$. Using this in conjunction with (20) and (21), we rewrite (19) as

$$n^{2/5}[T_n(t) - \alpha_D] =_{\mathscr{D}} \beta_D t^2 - \alpha_D t^{-1}\phi(n)^{-1/2} \sum_1^{\phi(n)t} W_i^* + o_p(1), \tag{22}$$

uniformly in $a \le t \le b$. Theorem D.1 now follows from (22), because

$$\left\{-\phi(n)^{-1/2} \sum_1^{\phi(n)t} W_i^*,\ a \le t \le b\right\} \to_{\mathscr{D}} \{B(t),\ a \le t \le b\},$$

by Donsker's theorem.

6. Proof of Theorem R.1. We write

$$S_n(t) = \{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} m(Y_{nj}) + n^{-2/5}\sigma(x)t^{-1}\Psi_n(t), \tag{23}$$

of which the first term is examined in Lemma 4 and

$$\Psi_n(t) = \phi(n)^{-1/2}\sigma(x)^{-1} \sum_1^{\phi(n)t} \{Z_{nj} - m(Y_{nj})\}$$

is analyzed in Lemmas 5, 6 and 7. As before, $\phi(n) = n^{4/5}$.

LEMMA 4.

$$\{\phi(n)t\}^{-1} \sum_1^{\Phi(n)t} m(Y_{nj}) = \alpha_R + n^{-2/5}\beta_R t^2 + R_n(t),$$

where $\sup_{a \le t \le b}|R_n(t)| = o_p(n^{-2/5})$.

PROOF. By Lemma 1(iv, v) and (6), for $\phi(n)a \le j \le \phi(n)b$,

$$m(Y_{nj}) = \alpha_R + 12f^2(x)\beta_R Y_{nj}^2 + R_{nj}, \tag{24}$$

where $|R_{nj}| = 2^{-1}|m''(\lambda Y_{nj}) - m''(0)|Y_{nj}^2$, $0 \le \lambda \le 1$. By continuity of m'' and Corollary 1 to Lemma 2, $\sup_{\phi(n)a \le j \le \phi(n)b}|R_{nj}| = o_p(n^{-2/5})$. Thus,

$$\{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} m(Y_{nj}) = \alpha_R + 12f^2(x)\beta_R\{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} Y_{nj}^2 + \overline{R}_n(t), \tag{25}$$

where $\sup_{a \le t \le b}|\overline{R}_n(t)| = o_p(n^{-2/5})$. Now by Lemma 1(ii) and Corollary 1 to Lemma 2,

$$\begin{aligned} \{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} Y_{nj}^2 &= \{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} g^2(U_{nj}) \\ &= \{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} \left[g'(0)U_{nj} + (3!)^{-1}g'''(\lambda U_{nj})U_{nj}^3\right]^2 \qquad (26) \\ &= \{g'(0)\}^2\{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} U_{nj}^2 + o_p(n^{-2/5}). \end{aligned}$$

Finally, letting F_n denote the empirical cdf of $U_1, \dots, U_n$,

$$\max_{1 \le j \le n} |U_{nj} - jn^{-1}| = \max_{1 \le j \le n} |U_{nj} - F_n(U_{nj})|$$
$$\le \sup_{0 \le u \le 1} |F_n(u) - u| = O_p(n^{-1/2}),$$

so that $U_{nj}^2 = (j/n)^2 + o_p(n^{-1/2})$ uniformly in $\phi(n)a \le j \le \phi(n)b$. Hence,

$$\{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} U_{nj}^2 = \{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} (j/n)^2 + o_p(n^{-1/2})$$
$$= \{n^{-1}\phi(n)\}^2 t^{-1}\phi(n)^{-1} \sum_1^{\phi(n)t} \{j/\phi(n)\}^2 + o_p(n^{-1/2})$$
$$= n^{-2/5} t^{-1} \left[\int_0^t s^2 \, ds + O(n^{-4/5}) \right] + o_p(n^{-1/2})$$
$$= n^{-2/5} 3^{-1} t^2 + o_p(n^{-2/5}).$$

Thus,

$$\{\phi(n)t\}^{-1} \sum_1^{\phi(n)t} m(Y_{nj}) = \alpha_R + 12 f^2(x)\beta_R\{g'(0)\}^2 n^{-2/5} 3^{-1} t^2 + o_p(n^{-2/5})$$
$$= \alpha_R + n^{-2/5}\beta_R t^2 + o_p(n^{-2/5}),$$

as was to be proved. □

LEMMA 5. *Under Assumption* 3,

$$\sup_{\phi(n)a \le k \le \phi(n)b} \left| k^{-1} \sum_1^k s^2(Y_{nj}) - s^2(0) \right| = o_p(1).$$

PROOF. By Lemma 1(vi), $s^2(y)$ is continuous at 0, i.e., $s^2(y) = s^2(0) + o(1)$ as $y \to 0$. Since Y_{nj}, $j \le \phi(n)b$ are uniformly $O_p(n^{-1/5}) = o_p(1)$ by Corollary 1 to Lemma 2, we have $s^2(Y_{nj}) = s^2(0) + o_p(1)$ uniformly in $j \le \phi(n)b$, which implies the results. □

Without any loss of generality, assume that there is a B.M. $\{B(t),\ t \ge 0\}$ on the same space on which $(X_1, Z_1), (X_2, Z_2), \dots$ are defined and let $\mathscr{A}$ denote the σ-field of $Y_i = |X_i - x|$, $i = 1, 2, \dots,$ in this space.

LEMMA 6. *For every n, there exist stopping times* $T_{n1}, \dots, T_{nn}$ *of the Brownian motion* $\{B(t),\ t \ge 0\}$ *such that*

(i) $\{\sum_1^k [Z_{nj} - m(Y_{nj})],\ 1 \le k \le n\} =_{\mathscr{D}} \{B(T_{n1} + \cdots + T_{nk}),\ 1 \le k \le n\}$,
(ii) $T_{n1}, \dots, T_{nk}$ *are conditionally independent given* $\mathscr{A}$ *a.s.*,
(iii) $E(T_{nk} | \mathscr{A}) = s^2(Y_{nk})$ *a.s.*,
(iv) $E(T_{nk}^2 | \mathscr{A}) \le C\tau(Y_{nk})$ *a.s., where C is a constant.*

PROOF. By Lemma 1 of Bhattacharya (1974),

$$P\left[Z_{nj} \le z_j,\ j = 1, \dots, n | \mathscr{A}\right] = \prod_{j=1}^{n} H_{Y_{nj}}(z_j) \quad \text{a.s.},$$

where $H_y(z) = P(Z \le z | Y = y)$. Thus, $Z_{nj} - m(Y_{nj})$, $1 \le j \le n$, are conditionally independent given $\mathscr{A}$ with mean 0, variance $s^2(Y_{nj})$ and fourth moment $\tau(Y_{nj})$. In the conditional argument given $\mathscr{A}$, the lemma is thus a special case of the well-known theorem of Skorokhod (1965), page 163. □

LEMMA 7. $\{\Psi_n(t),\ a \le t \le b\} \to_{\mathscr{D}} \{B(t),\ a \le t \le b\}$.

PROOF. By Lemma 6(i),

$$\begin{aligned}\{\Psi_n(t),\ a \le t \le b\} &= \left\{\phi(n)^{-1/2}\sigma(x)^{-1}\sum_1^{\phi(n)t}\left[Z_{nj} - m(Y_{nj})\right],\ a \le t \le b\right\}\\ &=_{\mathscr{D}} \left\{\phi(n)^{-1/2}\sigma(x)^{-1}B\left(T_{n1} + \cdots + T_{n,\Phi(n)t}\right),\ a \le t \le b\right\}\\ &=_{\mathscr{D}} \left\{B\left(\left(\phi(n)\sigma^2(x)\right)^{-1}\sum_1^{\phi(n)t} T_{nj}\right),\ a \le t \le b\right\}.\end{aligned}$$

To complete the proof, we shall show that

$$\sup_{a \le t \le b}\left|\left\{\phi(n)\sigma^2(x)\right\}^{-1}\sum_1^{\phi(n)t} T_{nj} - t\right| \to_p 0, \tag{27}$$

because, by arguing as in Theorem 13.8 of Breiman (1968), (27) would imply that along all sufficiently rapidly increasing subsequences $\{n_i\}$, the expression in (27) converges to 0 a.s. and the desired weak convergence will follow by Theorem 13.12 of Breiman (1968). To prove (27), note that

$$\begin{aligned}&\sigma^2(x)\sup_{a \le t \le b}\left|\left\{\phi(n)\sigma^2(x)\right\}^{-1}\sum_1^{\phi(n)t} T_{nj} - t\right|\\ &\quad\le \sup_{\phi(n)a \le k \le \phi(n)b}\left|\Phi(n)^{-1}\sum_1^{k}\left\{T_{nj} - \sigma^2(x)\right\}\right| + o(1)\\ &\quad\le b \sup_{\Phi(n)a \le k \le \phi(n)b}\left|k^{-1}\sum_1^{k}\left\{T_{nj} - \sigma^2(x)\right\}\right| + o(1)\\ &\quad\le b \sup_{\phi(n)a \le k \le \Phi(n)b}\left|k^{-1}\sum_1^{k}\left\{T_{nj} - s^2(Y_{nj})\right\}\right| + o_p(1),\end{aligned}$$

by Lemma 5, since $\sigma^2(x) = s^2(0)$. By Lemma 6 and the Hájek–Rényi inequality, we now have

$$P\left[\sup_{\phi(n)a\le k\le\phi(n)b}\left|k^{-1}\sum_1^k\left\{T_{nj}-s^2(Y_{nj})\right\}\right|>\varepsilon\right]$$

$$\le C\varepsilon^{-2}\left[\{\phi(n)a\}^{-2}\sum_1^{\phi(n)a}\tau(Y_{nk})+\sum_1^{\Phi(n)b}k^{-2}\tau(Y_{nk})\right],\quad \text{a.s.}$$

$$\le 2C\varepsilon^{-2}\{\phi(n)a\}^{-1}\max_{1\le k\le\phi(n)b}\tau(Y_{nk})$$

$$\le 2C\varepsilon^{-2}\{\phi(n)a\}^{-1}\left[M_1+M_2Y_{\phi b}\right],$$

using Lemma 1(vii) in the last step. Hence,

$$P\left[\sup_{\phi(n)a\le k\le\phi(n)b}\left|k^{-1}\sum_1^{\phi(n)a}\left\{T_{nj}-s^2(Y_{nj})\right\}\right|>\varepsilon\right]$$

$$\le 2C\varepsilon^{-2}\{\phi(n)a\}^{-1}\left[M_1+M_2E(Y_{\phi b})\right],$$

and the proof is completed by an application of Corollary 2 to Lemma 2.

The proof of Theorem R.1 is now accomplished by using the results of Lemmas 4 and 7 in (23). □

Acknowledgments. We gratefully acknowledge a suggestion of Peter Hall which initiated Remark 3.4 and the helpful comments of the referees and the Associate Editor.

REFERENCES

ABRAMSON, I. S. (1982). On bandwidth variation in kernel estimates—a square-root law. *Ann. Statist.* **10** 1217–1223.

BARTLETT, M. S. (1963). Statistical estimation of density functions. *Sankhyā Ser. A* **25** 245–254.

BHATTACHARYA, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *Ann. Statist.* **2** 1034–1039.

BHATTACHARYA, P. K. and MACK, Y. P. (1985). A two-stage procedure for nonparametric estimation. *Statist. Decisions* (*Suppl.*) **2** 143–153.

BICKEL, P. J. and DOKSUM, K. A. (1977). *Mathematical Statistics*: *Basic Ideas and Selected Topics*. Holden-Day, San Francisco.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

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.

HALL, P. (1982). Cross-validation in density estimation. *Biometrika* **69** 383–390.

HÄRDLE, W. and MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13** 1465–1481.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, Berlin.

KRIEGER, A. M. and PICKANDS, J. (1981). Weak convergence and efficient density estimation at a point. *Ann. Statist.* **9** 1066–1078.

LI, K.-C. (1984). Cross-validated nearest neighbor estimates. *Ann. Statist.* **12** 230–240.

MACK, Y. P. (1981). Local properties of k-NN regression estimates. *SIAM J. Algebraic Discrete Methods* **2** 311–323.

MACK, Y. P. and ROSENBLATT, M. (1979). Multivariate k-nearest neighbor density estimates. *J. Multivariate Anal.* **9** 1–15.

MARRON, J. S. (1985). An asymptotically efficient solution to the bandwidth problem of kernel density estimation. *Ann. Statist.* **13** 1011–1023.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

PRATT, J. W. (1959). On a general concept of "in probability." *Ann. Math. Statist.* **30** 549–558.

ROSENBLATT, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

SKOROKHOD, A. V. (1965). *Studies in the Theory of Random Processes*. Addison-Wesley, Reading, Mass.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.* **12** 1285–1297.

WAHBA, G. (1975). Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation. *Ann. Statist.* **3** 15–29.

WOODROOFE, M. (1970). On choosing a delta-sequence. *Ann. Math. Statist.* **41** 1665–1671.

YANG, Z. H. and COX, D. D. (1984). The least squares estimation of a probability density function and its derivatives. Technical Report No. 753, Dept. of Statistics, Univ. of Wisconsin, Madison.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616