KERNEL AND NEAREST NEIGHBOR ESTIMATION OF A CONDITIONAL QUANTILE

by

P.K. Bhattacharya
University of California, Davis

and

Ashis K. Gangopadhyay
University of North Carolina, Chapel Hill

ABSTRACT

Let (X_1,Z_1), $(X_2,Z_2),\dots,(X_n,Z_n)$ be i.i.d. as (X,Z), Z taking values in R^1, and for $0 < p < 1$, let $\xi_p(x)$ denote the conditional p–quantile of Z given X=x, i.e., $P(Z \le \xi_p(x)|X=x) = p$. In this paper, kernel and nearest neighbor estimators of $\xi_p(x)$ are proposed. As a first step in studying the asymptotics of these estimates, Bahadur type representations of the sample conditional quantile functions are obtained. These representations are used to examine the important issue of adaptive choice of the smoothing parameter by a local approach (for a fixed x) based on weak convergence of these estimators with varying k in k–nearest neighbor method and with varying h in kernel method with bandwidth h. These weak convergence results lead to asymptotic linear models which motivate certain estimators.

1. Introduction

Let (X_1,Z_1), (X_2,Z_2),... be two–dimensional random vectors which are iid as (X,Z), and for $0 < p < 1$, let $\xi_p(x)$ denote the conditional p–quantile of Z given X=x. We consider the problem of estimating $\xi_p(x)$ from the data $(X_1,Z_1),\ldots,(X_n,Z_n)$, and study the asymptotic properties of the kernel and nearest neighbor (NN) estimators, as $n \to \infty$.

Usefulness of conditional quantile functions as good descriptive statistics has been discussed by Hogg (1975) who calls them percentile regression lines. The problem of conditional quantile estimation has been investigated by Bhattacharya (1963) following the fractile approach, and is also included in the general scheme of nonparametric regression considered by Stone (1977). More recently asymptotic normality of estimators of conditional quantiles has been proved by Cheng (1983), who considered kernel estimators in the fixed design case, and by Stute (1986), who considered NN–type estimators in the random design case. However, both of these authors took the bandwidth h_n to be $O(n^{-\frac{1}{3}})$ which kept the bias smaller than the random error by an order of magnitude, and thereby made the rate of convergence slower than optimal.

In this paper, we obtain Bahadur–type representations [Bahadur (1966)] of NN and kernel estimators of conditional quantiles as a first step in studying their asymptotics (Theorems N1 and K1). Bias terms show up in these representations because of our choice of $k = k_n = O(n^{\frac{4}{5}})$ in the k–NN method and $h=h_n = O(n^{-\frac{1}{5}})$ in the kernel method (with uniform kernel), for the purpose of achieving optimum balance between bias and random error. These representations are then used to examine the important issue of adaptive choice of the smoothing parameters by a local approach (for fixed x). Weak convergence results (Theorems N2 and K2) are proved for these estimators with varying k in the NN method and with varying h in the kernel method. These weak convergence results are analogous to the one obtained by Bhattacharya and Mack (1987) for k–NN regression estimators with varying k and lead to asymptotic linear models similar to theirs, which motivate certain estimators.

2. The Main Results

For the random vector (X,Z), let f denote the pdf of X and $g(\cdot|x)$ the conditional pdf of Z given X=x, with corresponding conditional cdf $G(\cdot|x)$. We want to estimate $\xi_p(x_0)$, the conditional p–quantile of Z given $X=x_0$. Since $p \in (0,1)$ and x_0 will remain fixed throughout our discussion, we shall write $\xi_p(x_0) = \xi$.

The following regularity conditions are assumed.

1. (a) $f(x_0) > 0$.

   (b) $f''(x)$ exists in a neighborhood of x_0, and there exist $\epsilon > 0$ and $A < \infty$ such that $|x-x_0| \le \epsilon$ implies $|f''(x) - f''(x_0)| \le A|x-x_0|$.

2. (a) $g(\xi|x_0) > 0$, where $G(\xi|x_0) = p$.

   (b) The partial derivatives $g_x(z|x)$ and $g_{xx}(z|x)$ of $g(z|x)$ and $G_{xx}(z|x)$ of $G(z|x)$ exist in a neighborhood of (x_0,ξ), and there exist $\epsilon > 0$ and $A < \infty$ such that $|x-x_0| \le \epsilon$ and $|z-\xi| \le \epsilon$ together imply

$$|g_z(z|x)| \le A, \quad |g_x(z|x_0)| \le A, \quad |g_{xx}(z|x_0)| \le A,$$

$$|g_{xx}(z|x) - g_{xx}(z|x_0)| \le A|x-x_0|,$$

$$|G_{xx}(z|x) - G_{xx}(z|x_0)| \le A|x-x_0| .$$

By condition 2, ξ is uniquely defined by $G(\xi|x_0) = p$.

Now let $\{(X_i,Z_i),\ i=1,2,\ldots\}$ be iid as (X,Z), and let $Y_i = |X_i-x_0|$, so that $\{(Y_i,Z_i),\ i=1,2,\ldots\}$ are iid as $(Y = |X-x_0|, Z)$ with the pdf f_Y of Y, the conditional pdf $g^*(\cdot|y)$ of Z given Y=y and the corresponding conditional cdf $G^*(\cdot|y)$ given by

$$f_Y(y) = f(x_0 + y) + f(x_0 - y),$$

(1) $$g^*(z|y) = [f(x_0 + y)\, g(z|x_0 + y) + f(x_0 - y)\, g(z|x_0 - y)] \,/\, f_Y(y),$$

$$G^*(z|y) = [f(x_0 + y)\, G(z|x_0 + y) + f(x_0 - y)\, G(z|x_0 - y)] \,/\, f_Y(y).$$

Note that

$$g^*(z|0) = g(z|x_0) = g(z), \quad G^*(z|0) = G(z|x_0) = G(z).$$

Here and in what follows, we write $g(z|x_0) = g(z)$ and $G(z|x_0) = G(z)$ for simplicity.

Let $Y_{n1} < \dots < Y_{nn}$ denote the order statistics and $Z_{n1},\dots,Z_{nn}$ the induced order statistics of $(Y_1,Z_1),\dots,(Y_n,Z_n)$, i.e., $Z_{ni} = Z_j$ if $Y_{ni} = Y_j$. For any positive integer $k \le n$, the k–NN empirical cdf of Z (with respect to x_0) is now defined as:

(2) $$\hat{G}_{nk}(z) = k^{-1} \sum_{i=1}^{k} 1(Z_{ni} \le z),$$

where 1(S) denotes the indicator of the event S. The k–NN estimator of ξ can now be expressed as the p–quantile of $\hat{G}_{nk}$, i.e.,

(3) $$\begin{aligned}\hat{\xi}_{n,k} &= \text{the } [kp]^{th} \text{ order statistic of } Z_{n1},\dots,Z_{nk} \\ &= \inf\{z\colon \hat{G}_{nk}(z) \ge [kp]/k\}.\end{aligned}$$

The kernel estimator of ξ with uniform kernel and bandwidth h can also be expressed in the same manner, viz.,

(4) $$\begin{cases}\tilde{\xi}_{nh} = \inf\{z : \hat{G}_{n\,K_n(h)}(z) \ge [K_n(h)p] \,/\, K_n(h)\} \\ K_n(h) = \sum_{i=1}^{n} 1(Y_i \le h/2) = n\,\hat{F}_{Y,n}(h/2),\end{cases}$$

$\hat{F}_{Y,n}$ being the empirical cdf of $Y_1,\dots,Y_n$. The kernel estimators are thus related to the NN estimators by

(5) $$\tilde{\xi}_{nh} = \hat{\xi}_{n\,K_n(h)}$$

where $K_n(h)$ is the random integer given by (4).

We now state our main results in the following two theorems of which Theorem N1 gives a Bahadur–type representation for the k–NN estimator $\hat{\xi}_{nk}$ of ξ with k lying in

$$I_n(a,b) = \{k\colon k_0 = [n^{\frac{4}{5}} a] \le k \le [n^{\frac{4}{5}} b] = k_1\},\ 0 < a < b,$$

and Theorem K1 gives a corresponding representation for the kernel estimator $\tilde{\xi}_{nh}$ with h lying in

$$J_n(c,d) = [n^{-\frac{1}{5}} c,\ n^{-\frac{1}{5}} d],\ 0 < c < d.$$

<u>Theorem N1.</u> $\hat{\xi}_{nk} - \xi$

$$= \beta(\xi)(k/n)^2 + \{kg(\xi)\}^{-1} \sum_{i=1}^{k} [1(Z^*_{ni} > \xi) - (1-p)] + R_{nk},$$

where

$$\beta(\xi) = -[f(x_0)G_{xx}(\xi|x_0) + 2f'(x_0)\, G_x(\xi|x_0)] \,/\, \{24\, f^3(x_0)\, g(\xi)\},$$

$$Z^*_{ni} = G^{-1} \circ G^*(Z_{ni}|Y_{ni}),\ 1 \le i \le n,\ n \ge 1,$$

and

$$\max_{k \in I_n(a,b)} |R_{nk}| = O(n^{-\frac{3}{5}} \log n),\ \text{a.s.}$$

<u>Theorem K1.</u> $\tilde{\xi}_{nh} - \xi$

$$= \beta(\xi)\, f^2(x_0)\, h^2 + \{[n\, h\, f(x_0)]g(\xi)\}^{-1} \sum_{i=1}^{[nhf(x_0)]} [1(Z^*_{ni} > \xi) - (1-p)] + R^*_{nh}$$

where $\beta(\xi)$ and Z^*_{ni} are as in Theorem N1, and

$$\sup_{h \in J_n(c,d)} |R^*_{nh}| = O(n^{-\frac{3}{5}} \log n), \text{ a.s.}$$

Remarks.

1. Let $\mathcal{A} = \sigma\{Y_1, Y_2, \ldots\}$. Then $Z_{n1}, \ldots, Z_{nn}$ are conditionally independent given $\mathcal{A}$, with Z_{ni} having conditional cdf $G^*(\cdot\,|Y_{ni})$, as shown by Bhattacharya (1974). Hence $G^*(Z_{ni}|Y_{ni})$, $1 \le i \le n$, are conditionally independent and uniform (0,1) given $\mathcal{A}$, and therefore,

$$P[Z^*_{ni} \le z_i, 1 \le i \le n] = EP[G^*(Z_{ni}|Y_{ni}) \le G(z_i),\ 1 \le i \le n \,|\, \mathcal{A}] = \prod_{i=1}^{n} G(z_i).$$

Thus for each n, $Z^*_{n1}, \ldots, Z^*_{nn}$ are iid with cdf G. Since $G(\xi) = p$, it follows that for each n, the summands $1(Z^*_{ni} > \xi) - (1-p)$ in the above representations are independent random variables with mean 0 and variance $p(1-p)$.

2. The remainder terms in both theorems are $O(n^{-\frac{3}{5}} \log n)$, a.s. In Theorem N1, this corresponds to $O(k^{-\frac{3}{5}} \log k)$ with $k = O(n^{\frac{4}{5}})$, as one would expect. The same explanation applies to Theorem K1, because $[n\, h\, f(x_0)] = O(n^{\frac{4}{5}})$ for $h = O(n^{-\frac{1}{5}})$.

3. Weaker versions of the above theorems were proved by Gangopadhyay (1987). His remainder terms were $o(n^{-\frac{2}{5}})$, a.s. in Theorem N1 and $o_p(n^{-\frac{2}{5}})$ in Theorem K1.

3. Weak Convergence Properties of NN Estimators with Varying k and Kernel Estimators with Varying Bandwidth.

Consider the stochastic processes $\{\hat{\xi}_{nk}, k \in I_n(a,b)\}$ and $\{\tilde{\xi}_{nh}, h \in J_n(c,d)\}$. The two theorems in this section describe the weak convergence properties of suitably normalized versions of these processes, as $n \to \infty$. The symbol $\xrightarrow{\mathcal{D}}$ indicates convergence in distribution, i.e., weak convergence of the distributions of the stochastic processes (or random vectors) under consideration and $\{B(t),\ t \geq 0\}$ denotes a standard Brownian motion.

Theorem N2. Let $T_n(t) = \hat{\xi}_{n;[n^{\frac{4}{5}}t]}$. Then for any $0 < a < b$,

$$\{n^{\frac{2}{5}}[T_n(t) - \xi] - \beta t^2,\ a \leq t \leq b\} \overset{\mathscr{D}}{\to} \{\sigma\, t^{-1} B(t),\ a \leq t \leq b\}\ ,$$

where $\beta = \beta(\xi)$ given in Theorem N1 and $\sigma^2 = p(1-p)/g^2(\xi)$.

Proof. In the representation for $\hat{\xi}_{nk}$ given in Theorem N1, take $k = [n^{\frac{4}{5}}\, t] = n^{\frac{4}{5}}\, t + \epsilon_n(t)$ with $0 \leq \epsilon_n(t) < 1$. After a little rearrangement of terms, this leads to

$$n^{\frac{2}{5}}[T_n(t) - \xi] - \beta(\xi)t^2 = \{\sqrt{p(1-p)}\ /\ g(\xi)\}t^{-1} \cdot n^{-\frac{2}{5}} \sum_{1}^{[n^{\frac{4}{5}}t]} W_{ni} + \sum_{j=1}^{3} R_{nj}(t)\ ,$$

where

$$(6) \qquad W_{ni} = [1(Z^*_{ni} > \xi) - (1-p)]\ /\ \sqrt{p(1-p)}\ ,\quad 1 \leq i \leq n\ ,$$

are iid with mean 0 and variance 1 for each n in view of Remark 1. Of the three remainder terms, $R_{n1}(t) = n^{\frac{2}{5}} R_{n,[n^{\frac{4}{5}}t]}$ from Theorem N1, while $R_{n2}(t)$ and $R_{n3}(t)$ come from the discrepancy $0 \leq \epsilon_n(t) < 1$ due to replacing $k = [n^{\frac{4}{5}}t]$ by $n^{\frac{4}{5}}t$ in the first two terms of the representation. Hence

$$\sup_{a \leq t \leq b} |R_{n1}(t)| = n^{\frac{2}{5}} \max_{k \in I_n(a,b)} |R_{nk}| = O(n^{-\frac{1}{5}} \log n), \quad \text{a.s.},$$

while

$$\sup_{a\le t\le b} |R_{n2}(t) + R_{n3}(t)|$$

is easily seen to be $O_p(n^{-\frac{4}{5}})$ by virtue of $\sup_{a\le t\le b} |n^{-\frac{2}{5}} \sum_{1}^{[n^{\frac{4}{5}}t]} W_{ni}| = O_p(1)$. The three remainder terms are together $O_p(n^{-\frac{1}{5}} \log n)$ uniformly in $a \le t \le b$. We thus have, with $\sigma = \sqrt{p(1-p)} / g(\xi)$,

$$n^{\frac{2}{5}}[T_n(t) - \xi] - \beta(\xi)t^2 = \sigma t^{-1} \cdot n^{-\frac{2}{5}} \sum_{1}^{[n^{\frac{4}{5}}t]} W_{ni} + o_p(1)$$

uniformly in $a \le t \le b$. Now use Theorem 1, page 452 of Gikhman and Skorokhod (1969) to see that $\{n^{-\frac{2}{5}} \sum_{1}^{[n^{\frac{4}{5}}t]} W_{ni},\ a \le t \le b\} \overset{\mathcal{D}}{\to} \{B(t),\ a \le t \le b\}$. This proves the theorem. □

Theorem K2. Let $S_n(t) = \tilde{\xi}_{n,n^{-\frac{1}{5}}t}$. Then for any $0 < c < d$,

$$\{n^{\frac{2}{5}}[S_n(t)-\xi] - \gamma t^2,\ c \le t \le d\} \overset{\mathcal{D}}{\to} \{\tau t^{-1} B(t),\ c \le t \le d\},$$

where $\gamma = \beta f^2(x_0)$ and $\tau = \sigma/f(x_0)$, with β and σ as in Theorem N2.

Proof. In the representation for $\tilde{\xi}_{nh}$ given in Theorem K1, take $h = n^{-\frac{1}{5}} t$ and rearrange terms to obtain

$$n^{\frac{2}{5}}[S_n(t)-\xi] - \gamma t^2$$
$$= [1+O(n^{-\frac{4}{5}})]\tau t^{-1}\{n^{\frac{4}{5}} f(x_0)\}^{-\frac{1}{2}} \sum_{1}^{[n^{\frac{4}{5}}t\, f(x_0)]} W_{ni} + n^{\frac{2}{5}} R^*_{n,n^{-\frac{1}{5}}t},$$

where W_{ni}, $1 \le i \le n$, are as in the proof of Theorem N2 and $R^*_{n,n^{-\frac{1}{5}}t}$ is from Theorem K1. Thus $\sup_{c \le t \le d} n^{\frac{2}{5}} |R_{n,n^{-\frac{1}{5}}t}| = O(n^{-\frac{1}{5}} \log n)$, a.s., and the theorem follows from the triangular array version of Donsker's theorem referred to above. □

Remark. The uniformity of the order of magnitude of the remainder terms in Theorems N1 and K1 was crucial in proving Theorems N2 and K2.

From Theorems N2 and K2, it follows that $n^{\frac{2}{5}}[T_n(t)-\xi] \overset{\mathcal{D}}{\to} N(\beta\, t^2, \sigma^2\, t^{-1})$ and $n^{\frac{2}{5}}[S_n(t)-\xi] \overset{\mathcal{D}}{\to} N(\gamma\, t^2, \tau^2\, t^{-1})$ for each t, where $N(\mu,\sigma^2)$ denotes a Gaussian r.v. with mean μ and variance σ^2. Hence the asymptotic mean–squared errors (AMSE) of $T_n(t)$ is $n^{-\frac{4}{5}}(\beta^2\, t^4 + \sigma^2\, t^{-1})$, which is minimized at $t_1^* = \{\sigma^2/(4\,\beta^2)\}^{\frac{1}{5}}$ and the AMSE of $S_n(t)$ is $n^{-\frac{4}{5}}(\gamma^2\, t^4 + \tau^2\, t^{-1})$, which is minimized at $t_2^* = \{\tau^2/(4\,\gamma^2)\}^{\frac{1}{5}}$. However, these optimum t_1^* and t_2^* involve unknown quantities involving the marginal distribution of X and the conditional distribution of Z given $X = x_0$. Although one could attempt to use consistent (but possibly non–optimal) estimates of β, σ^2, γ and τ^2 to approximate t_1^* and t_2^* in the spirit of Woodroofe (1970) and Krieger and Pickands (1981), we shall take another approach in the next section by considering linear combinations of k–NN estimators with varying k and kernel estimators with varying bandwidth.

4. Asymptotic Linear Models and Linear Combinations of $\hat{\xi}_{nk}$ and $\tilde{\xi}_{nh}$.

Neglect the remainder form in Theorem N1 to obtain the following asymptotic linear model for $\{\hat{\xi}_{nk},\ k_0 \le k \le k_1\}$:

(7) $$\hat{\xi}_{nk} \simeq \xi + (k/n)^2\, \beta + \sigma\, \Delta_{nk},\ k_0 \le k \le k\,,$$

where β and σ are as in Theorem N2 and

$$\Delta_{nk} = k^{-1} \sum_{1}^{k} W_{ni} ,$$

in which W_{ni}, $1 \le i \le n$ are the iid r.v.'s with mean 0 and variance 1 given by (6). Hence

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

This is exactly like the asymptotic linear model for k–NN regression estimates obtained by Bhattacharya and Mack (1987). Following their approach, we can now obtain the BLUE of ξ in the model (7) and other suitably biased linear combinations of the $\hat{\xi}_{nk}$'s. The asymptotic distribution of the BLUE of ξ and its ARE with respect to $\hat{\xi}_{n,[n^{\frac{4}{5}}t_1^*]}$, where $[n^{\frac{4}{5}} t_1^*]$ is the optimum number of NN's, follow from Theorem N2 in exactly the same way as in the case of regression estimates discussed in [4].

The kernel estimators $\tilde{\xi}_{nh}$ with $n^{-\frac{1}{5}} c \le h \le n^{-\frac{1}{5}} d$ also satisfy an asymptotic linear model similar to (7). For this, first neglect the remainder term in Theorem K1 to obtain

$$\text{(8)} \qquad \tilde{\xi}_{nh} \simeq \xi + \beta f^2(x_0) h^2 + \sigma \Delta_{n,[nhf(x_0)]} , \;\; h \in J_n(c,d),$$

where β, σ and Δ_{nk} are as in (7). Although this linear model is indexed by a continuous parameter h, the problem of finding the BLUE and other suitably biased linear combinations of ξ in this model are essentially the same as the corresponding problems in (7). All questions about these linear combinations are thus covered by our discussion in the previous paragraph. To see this, let $m_0 = [n^{\frac{4}{5}} c\, f(x_0)]$, $m_1 = [n^{\frac{4}{5}} d\, f(x_0)]$, and define

$$h(m) = m/\{n\, f(x_0)\}, \;\; m = m_0, m_0 + 1, \ldots, m_1 .$$

Then for $h(m) \le h < h(m+1)$,

$$[n\,h\,f(x_0)] = m \quad \text{and} \quad f^2(x_0)h^2 = (m/n)^2 + O(n^{-\frac{6}{5}}) .$$

Using this in (8), we have

$$(9) \qquad \begin{cases} \tilde{\xi}_{nh} \simeq \tilde{\xi}_{nh(m)} , & h(m) \le h < h(m+1) \\ \tilde{\xi}_{nh(m)} \simeq \xi + (m/n)^2 \beta + \sigma \Delta_{nm} , & m_0 \le m \le m_1 \end{cases}$$

All linear combinations $\int_{n^{-\frac{1}{5}} c}^{n^{-\frac{1}{5}} d} \ell(h)\, \tilde{\xi}_{nh}\, dh$ of kernel estimators with varying bandwidth are, therefore, approximated by linear combinations of $\{\tilde{\xi}_{nh(m)}, m_0 \le m \le m_1\}$, for which the asymptotic linear model (9) is identical to (7). The asymptotic properties of these linear combinations follow from Theorem K2. The details are as in [7].

5. Proof of Theorem N1: Preliminary Lemmas.

The k–NN estimator $\hat{\xi}_{nk}$ of ξ is the p–quantile of the empirical cdf $\hat{G}_{nk}$ of $Z_{n1},\dots,Z_{nk}$, which are conditionally independent with Z_{ni} having conditional cdf $G^*(\cdot\,|\,Y_{ni})$. It is, therefore, natural to think of $\hat{\xi}_{nk}$ as an estimator of the p–quantile ξ_{nk} of the random cdf $k^{-1} \sum_1^k G^*(\cdot\,|\,Y_{ni})$. The discrepancy between this random cdf and the cdf $G^*(\cdot\,|\,0) = G(\cdot)$ of which ξ is the p–quantile, is going to give rise to a bias in addition to the random error in estimating ξ_{nk} by $\hat{\xi}_{nk}$. To facilitate the examination of this bias and the random error, we introduce some notations. Let

$$(10) \qquad \begin{aligned} & g^*(\cdot\,|\,Y_{ni}) = g_{ni}(\cdot), \; G^*(\cdot\,|\,Y_{ni}) = G_{ni}(\cdot) , \\ & \bar{g}_{nk}(\cdot) = k^{-1} \sum_1^k g_{ni}(\cdot), \; \bar{G}_{nk}(\cdot) = k^{-1} \sum_1^k G_{ni}(\cdot) . \end{aligned}$$

Then ξ_{nk}, which is the target of $\hat{\xi}_{nk}$ is given by

(11) $\bar{G}_{nk}(\xi_{nk}) = p = G(\xi)$.

To examine the asymptotic properties of $\hat{\xi}_{nk}-\xi_{nk}$ and $\xi_{nk}-\xi$ for $k \in I_n(a,b)$, we now analyze the corresponding properties of $\hat{G}_{nk}(\cdot) - \bar{G}_{nk}(\cdot)$ and $\bar{G}_{nk}(\cdot) - G(\cdot)$.

The order statistics $0 < Y_{n1} < \dots < Y_{n,[n^{\frac{4}{5}}b]}$ are of the order of $n^{-\frac{1}{5}}$. Consequently, for $k \in I_n(a,b)$, it should be possible to approximate the pdf's $\bar{g}_{nk}$ and the cdf's $\bar{G}_{nk}$ defined in (10), by the first few terms of their expansions in powers of Y_{ni}, $i \le [n^{\frac{4}{5}}b]$. To this end, we have the following lemmas.

<u>Lemma 1.</u> For $B > b/f(x_0)$ and for sufficiently large n,

$$P[Y_{n,[n^{\frac{4}{5}}b]} > n^{-\frac{1}{5}} B] \le \exp[-2\, n^{\frac{3}{5}}\{B\, f(x_0)-b\}^2].$$

<u>Proof</u>. This is proved in [4].

<u>Lemma 2.</u> $k^{-1} \sum_1^k Y_{ni}^2 = \{12\, f^2(x_0)\}^{-1}\ (k/n)^2 + R_{nk}$, where

$$\max_{k \le [n^{\frac{4}{5}}b]} |R_{nk}| = O(n^{-\frac{3}{5}}), \text{ a.s.}$$

<u>Proof</u>. Let $0 < U_{n1} < \dots < U_{nn} < 1$ denote the order statistics of a random sample of size n from uniform (0,1), and let $\varphi = F_Y^{-1}$. Then $\varphi(0) = \varphi''(0) = 0$, $\varphi'(0) = \{2\, f(x_0)\}^{-1}$ and φ''' is continuous at 0 by condition 1. Use this to expand $Y_{ni} = \varphi(U_{ni})$ to the cubic term and observe that $U_{n,[n^{\frac{4}{5}}b]} = O(n^{-\frac{1}{5}})$, a.s., by Lemma 1. This leads to

$$(12) \qquad k^{-1} \sum_1^k Y_{ni}^2 = \{2\, f(x_0)\}^{-2} \cdot k^{-1} \sum_1^k U_{ni}^2 + R_{nk}(1),$$

where $\max_{k \le [n^{\frac{4}{5}} b]} |R_{nk}(1)| = O(n^{-\frac{4}{5}})$, a.s. Now write

(13) $$k^{-1} \sum_1^k U_{ni}^2 = k^{-1} \sum_1^k (i/n)^2 + R_{nk}(2) = 3^{-1}(k/n)^2 + R_{nk}(2) + R_{nk}(3),$$

where

$$|R_{nk}(2)| \le 2\, k^{-1} \sum_1^k (i/n)|U_{ni} - i/n| + k^{-1} \sum_1^k |U_{ni} - i/n|^2$$

$$\le 2(k/n) \max_{1 \le i \le n} |U_{ni} - i/n| + \max_{1 \le i \le n} |U_{ni} - i/n|^2 ,$$

so that

$$\max_{k \le [n^{\frac{4}{5}} b]} |R_{nk}(2)| \le 2\, bn^{-\frac{1}{5}}\, O(n^{-\frac{1}{2}} \sqrt{\log\ \log\ n}) + O(n^{-1} \log \log n)$$

$$= O(n^{-\frac{7}{10}} \sqrt{\log\ \log\ n}), \text{ a.s.},$$

by the law of iterated logarithm (see [6], page 157), and

$$|R_{nk}(3)| = |k^{-1} \sum_1^k (i/k)^2 - \int_0^1 x^2 dx|(k/n)^2 \le (\sum_{i=1}^k 2ik^{-3})(k/n)^2 \le (k+1)/n^2 ,$$

so that $\max_{k \le [n^{\frac{4}{5}} b]} |R_{nk}| = O(n^{-\frac{6}{5}})$. Combining (12) and (13), the lemma is proved.

□

Lemma 3. The following expansions hold for the conditional pdf $g^*(z|y)$ and the conditional cdf $G^*(z|y)$:

$$g^*(z|y) = g(z) + \tfrac{1}{2} y^2\, q(z) + y^3\, r(y,z),$$

$$G^*(z|y) = G(z) + \tfrac{1}{2} y^2\, Q(z) + y^3\, R(y,z),$$

where

$$g(z) = g(z|x_0), \quad G(z) = G(z|x_0),$$

$$g(z) = g_{xx}(z|x_0) + 2\,f'(x_0)\,g_x(z|x_0)/f(x_0)\,,$$

$$Q(z) = G_{xx}(z|x_0) + 2\,f'(x_0)\,G_x(z|x_0)/f(x_0),$$

and there exist $\epsilon > 0$ and $M < \infty$ such that $|q(z)|$, $|Q(z)|$, $|r(y,z)|$ and $|R(y,z)|$ are all bounded by M for $0 \le y \le \epsilon$ and $|z-\xi| \le \epsilon$.

<u>Proof</u>. Expand $f(x_0 \pm y)$, $g(z|x_0 \pm y)$ and $G(z|x_0 \pm y)$ about 0 to obtain:

$$f(x_0 \pm y) = f(x_0) \pm y\,f'(x_0) + \tfrac{1}{2}\,y^2\{f''(x_0) + \Delta_1(y)\},$$

$$g(z|x_0 \pm y) = G(z|x_0) \pm y\,g_x(z|x_0) + \tfrac{1}{2}\,y^2\{g_{xx}(z|x_0) + \Delta_2(y,z)\},$$

$$G(z|x_0 \pm y) = G(z|x_0) \pm y\,G_x(z|x_0) + \tfrac{1}{2}\,y^2\{G_{xx}(z|x_0) + \Delta_3(y,z)\}\,,$$

with $\max\{|\Delta_1(y)|, |\Delta_2(y,z)|, |\Delta_3(y,z)|\} \le Ay$ for $0 \le y \le \epsilon$ and $|z-\xi| \le \epsilon$, where $\epsilon > 0$ and $A < \infty$ are as in conditions 1 and 2. To obtain the formulas for $g^*(z|y)$ and $G^*(z|y)$ stated in the lemma, use the above expansions in (1). It will be seen by appropriate arrangement of terms that in remainders $y^3r(y,z)$ for $g^*(y|z)$ and $y^3R(y,z)$ for $G^*(y,z)$, the quantities $|r(y,z)|$ and $|R(y,z)|$ for $0 \le y \le \epsilon$ and $|z-\xi| \le \epsilon$ will remain bounded by a constant M determined by $f(x_0)$, $f''(x_0)$, $g(\xi|x_0)$, $g_{xx}(\xi|x_0)$, $G_x(\xi|x_0)$, $G_{xx}(\xi|x_0)$ and A, where $\epsilon > 0$ and $A < \infty$ are as in conditions 1 and 2. The boundedness of $|q(z)|$ and $|Q(z)|$ for $|z-\xi| \le \epsilon$ also follows from Condition 2. □

6. <u>Proof of Theorem N1: Bias in $\hat{\xi}_{nk}$.</u>

Recall that the target of $\hat{\xi}_{nk}$ is ξ_{nk}, the p–quantile of the random cdf $\overline{G}_{nk}(\cdot) = k^{-1} \sum_1^k G^*(\cdot|Y_{ni})$, while ξ is the p–quantile of $G(\cdot)$. The leading term of $\xi_{nk} - \xi$ is non–stochastic with probability 1, which is determined in this section. This is the bias in $\hat{\xi}_{nk}$. We first use Lemma 3 to bound the discrepancy between $\overline{G}_{nk}(\cdot)$ and $G(\cdot)$ near ξ, in terms of Y_{nk}. This makes $\max_{k\in I_n(a,b)} |\xi_{nk}-\xi|$ small whenever $Y_{n,[n^{\frac{4}{5}}b]}$ is small, in the manner described in the following lemma. The almost sure order of magnitude of $\xi_{nk}-\xi$ is obtained as a corollary.

Lemma 4. For every B, there exist N and C such that in the sample space of infinite sequences $\{((y_1,z_1), (y_2,z_2),...) : y_i \geq 0, \ z_i \text{ real}\}$, $Y_{n,[n^{\frac{4}{5}}b]} \leq B\, n^{-\frac{1}{5}}$ implies $\max_{k\in I_n(a,b)} |\xi_{nk}-\xi| \leq C\, n^{-\frac{2}{5}}$ for all $n \geq N$.

Proof. Fix $B < \infty$ and $0 < a < b$. By (11), it is enough to show the existence of N and C such that for $n \geq N$ and $k \in I_n(a,b)$, $Y_{n,[n^{\frac{4}{5}}b]} \leq B\, n^{-\frac{1}{5}}$ implies

$$\overline{G}_{nk}(\xi - C\, n^{-\frac{2}{5}}) \leq G(\xi) \leq \overline{G}_{nk}(\xi + C\, n^{-\frac{2}{5}}) .$$

For this, assume $Y_{n,[n^{\frac{4}{5}}b]} \leq B$, choose N and C so that

$$\text{(14)} \qquad \max(B\, N^{-\frac{1}{5}},\ C\, N^{-\frac{2}{5}}) \leq \min(\epsilon,\tfrac{1}{2}) ,$$

and use Lemma 3 to obtain

$$|\overline{G}_{nk}(\xi \pm C\, n^{-\frac{2}{5}}) - G(\xi \pm C\, n^{-\frac{2}{5}})|$$

$$\text{(15)} \qquad \leq \tfrac{1}{2}\,(k^{-1} \sum_1^k Y_{ni}^2)\, |Q(\xi \pm C\, n^{-\frac{2}{5}})| + k^{-1} \sum_1^k Y_{ni}^3\, |R(Y_{ni}, \xi \pm C\, n^{-\frac{2}{5}})|$$

$$\leq M\, B^2\, n^{-\frac{2}{5}}$$

for $n \geq N$. Moreover, since by Condition 2, $G(\xi + C n^{-\frac{2}{5}}) \geq G(\xi) + \frac{1}{2} C n^{-\frac{2}{5}} g(\xi)$ and $G(\xi - C n^{-\frac{2}{5}}) \leq G(\xi) - \frac{1}{2} C n^{-\frac{2}{5}} g(\xi)$ for $C n^{-\frac{2}{5}} \leq \epsilon$, (14) implies

$$\overline{G}(\xi + C n^{-\frac{2}{5}}) \geq G(\xi) + \frac{1}{2} C n^{-\frac{2}{5}} g(\xi) - M B^2 n^{-\frac{2}{5}}$$

$$\overline{G}(\xi + C n^{-\frac{2}{5}}) \leq G(\xi) - \frac{1}{2} C n^{-\frac{2}{5}} g(\xi) + M B^2 n^{-\frac{2}{5}}$$

for $n \geq N$. The lemma is proved by choosing $C > 2M B^2/g(\xi)$ and then choosing N so as to satisfy (14). □

Corollary. For $0 < a < b$, $\max_{k \in I_n(a,b)} |\xi_{nk} - \xi| = O(n^{-\frac{2}{5}})$, a.s.

Proof. Take $B > b/f(x_0)$ and apply Lemma 4 using C and N appropriately determined by B. Then

$$\sum_{n=N}^{\infty} P[\max_{k \in I_n(a,b)} |\xi_{nk} - \xi| > C n^{-\frac{2}{5}}] \leq \sum_{n=N}^{\infty} P[Y_{n,[n^{\frac{4}{5}}b]} > B n^{-\frac{1}{5}}]$$

$$\leq \sum_{n=N}^{\infty} \exp[-2n^{\frac{3}{5}} \{B f(x_0) - b\}^2] < \infty. \qquad \square$$

We now determine the leading term of $\xi_{nk} - \xi$.

Lemma 5. For $0 < a < b$,

$$\max_{k \in I_n(a,b)} |\xi_{nk} - \xi - \beta(\xi)(k/n)^2| = O(n^{-\frac{3}{5}}), \text{ a.s.}$$

where $\beta(\xi) = -Q(\xi)\{24 f^2(x_0) g(\xi)\}^{-1}$.

Proof. By (11) and Lemma 3,

$$G(\xi) = \overline{G}_{nk}(\xi_{nk}) = k^{-1} \sum_1^k [G_{ni}(\xi) + (\xi_{nk}-\xi)\, g_{ni}(z_{ni})]$$

$$= G(z) + k^{-1} \sum_1^k \{\tfrac{1}{2} Y_{ni}^2\, Q(\xi) + Y_{ni}^3\, R(Y_{ni}, \xi)\}$$

$$+ (\xi_{nk}-\xi)[g(\xi) + k^{-1} \sum_1^k \{g(z_{ni}) - g(\xi)\} + \tfrac{1}{2} Y_{ni}^2\, q(z_{ni}) + Y_{ni}^3\, r(Y_{ni}, z_{ni})],$$

where

(16) $$\max_{k \in I_n(a,b)} \max_{1 \le i \le k} |z_{nk}-\xi| \le \max_{k \in I_n(a,b)} |\xi_{nk}-\xi| = O(n^{-\frac{2}{5}}), \text{ a.s.}$$

Hence

$$\xi_{nk}-\xi = -\frac{1}{2} \frac{Q(\xi) \cdot k^{-1} \sum_1^k Y_{ni}^2 + 2\, k^{-1} \sum_1^k Y_{ni}^3\, R(Y_{ni},\xi)}{g(\xi) + k^{-1} \sum_1^k \{g(z_{ni})-g(\xi)\} + \frac{1}{2} Y_{ni}^2\, q(z_{ni}) + Y_{ni}^3\, r(Y_{ni}, z_{ni})}$$

$$= -\frac{1}{2} \cdot \frac{Q(\xi)[\{12\, f^2(x_0)\}^{-1} (k/n)^2 + R_{nk}(1)] + R_{nk}(2)}{g(\xi) + R_{nk}(3)},$$

where $\max_{k \in I_n(a,b)} |R_{nk}(1)| = O(n^{-\frac{3}{5}})$, a.s., by Lemma 2, $\max_{k \in I_n(a,b)} |R_{nk}(2)| = O(n^{-\frac{3}{5}})$, a.s., by Lemmas 1 and 3, and $\max_{k \in I_n(a,b)} |R_{nk}(3)| = O(n^{-\frac{2}{5}})$, a.s., by (16), Lemmas 1 and 3, and condition 2. Since $(k/n)^2 \le n^{-\frac{2}{5}} b^2$ for $k \in I_n(a,b)$, the lemma is proved. □

7. Proof of Theorem N1: Conclusion

We shall first express $\hat{\xi}_{nk} - \xi_{nk}$ in a manner analogous to the familiar representation of sample quantiles [Bahadur (1966)]. The random variables involved in this representation are $1(Z_{ni} > \xi_{nk}) - (1-G_{ni}(\xi_{nk}))$, which are conditionally independent given $\mathscr{A} = \sigma\{Y_1, Y_2, \ldots\}$, and the remainder term is $O(n^{-\frac{3}{5}} \log n)$, a.s., which corresponds to $O(k^{-\frac{3}{5}} \log k)$ with $k = O(n^{\frac{4}{5}})$, as one would expect. This representation is then modified to another one in which the random variables $1(Z_{ni} > \xi_{nk}) - (1 - G_{ni}(\xi_{nk}))$ are replaced by $1(Z_{ni} > \xi) - (1 - G_{ni}(\xi))$, and to yet another one in which these random variables are replaced by $1(Z^*_{ni} > \xi) - (1 - G(\xi))$ where $Z^*_{ni} = G^{-1} \circ G_{ni}(Z_{ni})$, $1 \le i \le n$, are iid with cdf $G(\xi)$. Together with the bias term obtained in Lemma 5, this will establish the representation of $\hat{\xi}_{nk}$ given in Theorem N1. Having completed the proof of Theorem N1 in this section, we shall then prove Theorem K1 in the next section, via formula (5), establishing the corresponding representation for the kernel estimator $\tilde{\xi}_{nh}$ (with uniform kernel and bandwidth h). It should be emphasized that Theorem N1 is proved uniformly in $k \in I_n(a,b)$ with $0 < a < b$ and Theorem K1 is proved uniformly in $h \in J_n(c,d)$ with $0 < c < d$ where $I_n(a,b)$ and $J_n(c,d)$ are as given in Section 2.

The first representation of $\hat{\xi}_{nk} - \xi_{nk}$ rests on Lemmas 8 and 9, which run parallel to Bahadur's proof [1]. However, we start with a lemma which provides an exponential bound for the deviation of sums of independent Bernoulli variables from their mean and then prove another lemma dealing with fluctuations of $\hat{G}_{nk}(\cdot) - \overline{G}_{nk}(\cdot)$.

<u>Lemma 6</u>. Let $U_{n1}, \ldots, U_{mn}$ be independent Bernoulli variables with $P(U_{ni}=1) = \pi_{ni}$. Then

$$P[|n^{-1} \sum_1^n (U_{ni} - \pi_{ni})| > t_n] \le 2 \exp[-\tfrac{1}{2} n t_n^2 / \{\max_{1 \le i \le n} \pi_{ni} + t_n\}] .$$

In particular, if $t_n / \max_{1 \le i \le n} \pi_{ni} \to 0$, then for large n,

$$P[|n^{-1} \sum_1^n (U_{ni} - \pi_{ni})| > t_n] \le 2 \exp[-\tfrac{1}{4} n t_n^2 / \max_{1 \le i \le n} \pi_{ni}] .$$

and if $\max_{1 \le i \le n} \pi_{ni}/t_n \to 0$, then for large n,

$$P[|n^{-1} \sum_1^n (U_{ni} - \pi_{ni})| > t_n] \le 2 \exp[-\tfrac{1}{4} n\, t_n] .$$

Proof. The first inequality is a simplified version of Bernstein's inequality (see [14], page 205), from which the other two follow as special cases.

Lemma 7. Suppose ζ_{nk} are $\mathcal{A}$-measurable random variables with $|\zeta_{nk} - \xi_{nk}| \le C\, n^{-\frac{2}{5}} \log n = \epsilon_n(C)$. Then for any γ, there exists M such that

$$\sum_{n=1}^{\infty} n^{\gamma} \cdot \max_{k \in I_n(a,b)} P[|\{\hat{G}_{nk}(\xi_{nk}) - \hat{G}_{nk}(\xi_{nk})\} - \{\overline{G}_{nk}(\xi_{nk}) - \overline{G}_{nk}(\xi_{nk})\}| > M\, n^{-\frac{3}{5}} \log n] < \infty .$$

Proof. Write $U_{nki} = 1(Z_{ni} \le \zeta_{nk}) - 1(Z_{ni} \le \xi_{nk})$ and $\mu_{nki} = G_{ni}(\zeta_{nk}) - G_{ni}(\xi_{nk}) = E(U_{nki} | \mathcal{A})$. Then

$$\{\hat{G}_{nk}(\zeta_{nk}) - \hat{G}_{nk}(\xi_{nk})\} - \{\overline{G}_{nk}(\zeta_{nk}) - \overline{G}_{nk}(\xi_{nk})\} = k^{-1} \sum_{i=1}^{k} (U_{nki} - \mu_{nki}) .$$

Choose $B > b/f(x_0)$, and for each n, let $S_n = \{Y_{n,[n^{\frac{4}{5}} b]} \le B\, n^{-\frac{1}{5}}\}$. Since $|\zeta_{nk} - \xi_{nk}| \le \epsilon_n(C) = C\, n^{-\frac{2}{5}} \log n$, Lemma 4 implies that there exist C' and N such that for $n \ge N$ and for z lying between ζ_{nk} and ξ_{nk}, $|z - \xi| \le C\, n^{-\frac{2}{5}} \log n + C'\, n^{-\frac{2}{5}} \le 2\, \epsilon_n(C)$ holds on the set S_n. Using Lemma 3, we now conclude that when n is large, then on S_n,

$$\max_{1 \le i \le [n^{\frac{4}{5}} b]} |\mu_{nki}| \le \epsilon_n(C) \sup_{|z-\xi| \le 2\epsilon_n(C)} g(z) + B^2\, n^{-\frac{2}{5}} \sup_{|z-\xi| \le 2\epsilon_n(C)} |Q(z)|$$

$$+ 2 B^3 n^{-\frac{3}{5}} \sup_{0 \le y \le Bn^{-\frac{1}{5}},\ |z-\xi| \le 2\epsilon_n(C)} |R(y,z)| \ \le\ 2 g(\xi)\epsilon_n(C) .$$

It now follows from Lemma 6 that for sufficiently large n,

$$\max_{k \in I_n(a,b)} P[|k^{-1} \sum_1^k (U_{nki}-\mu_{nki})| > M n^{-\frac{3}{5}} \log n]$$

$$= \max_{k \in I_n(a,b)} E\, P[|k^{-1} \sum_1^k (U_{nki}-\mu_{nki}) > M n^{-\frac{3}{5}} \log n \mid \mathscr{A}]$$

$$\le 2 \exp[-M^2 a\{5 C g(\xi)\}^{-1} \log n] + P(S_n^c) .$$

To complete the proof, observe that $\sum_{n=1}^{\infty} n^{\gamma - M^2 a\{5Cg(\xi)\}^{-1}} < \infty$ for sufficiently large M, and $\sum_{n=1}^{\infty} P(S_n^c) < \infty$ by Lemma 1. □

We now define $a_n = n^{-\frac{2}{5}} \log n$, $b_n = n^{\frac{1}{5}}$ and divide the interval $[\xi_{nk}-a_n, \xi_{nk}+a_n]$ into $2b_n$ equal intervals:

$$J_{n,k,r} = [\xi_{nk} + r\, a_n/b_n, \xi_{nk} + (r+1)a_n/b_n] = [\eta_{nk,r}, \eta_{nk,r+1}] ,$$

$$r = -b_n,\ldots,-1,0,1,\ldots,b_n-1 ,$$

each of length $a_n/b_n = n^{-\frac{3}{5}} \log n$. Let

$$(17)\quad \begin{cases} H_{nk}(z) = \{\hat{G}_{nk}(z) - \hat{G}_{nk}(\xi_{nk})\} - \{\overline{G}_{nk}(z) - \overline{G}_{nk}(\xi_{nk})\} \\ H^*_{nk} = \sup_{|z-\xi_{nk}| \le a_n} |H_{nk}(z)| = \max_{-b_n \le r \le b_n-1} \sup_{z \in J_{nk,r}} |H_{nk}(z)| , \\ H^*_n = \max_{k \in I_n(a,b)} |H^*_{nk}| \end{cases}$$

Lemma 8. $P[\max_{k\in I_n(a,b)} |\hat{\xi}_{nk}-\xi_{nk}| \geq a_n \text{ i.o.}] = 0$

Proof. $\hat{\xi}_{nk} \le \xi_{nk}-a_n$ implies $k^{-1}\sum_1^k \{1(Z_{ni} \le \xi_{nk} - a_n) - G_{ni}(\xi_{nk}-a_n)\} \ge [kp]/k - \overline{G}_{nk}(\xi_{nk}-a_n)$. Fix $B > b/f(x_0)$ and let $S_n = \{Y_{n,[n^{\frac{4}{5}}b]} \le B\, n^{-\frac{1}{5}}\}$. Then by Lemma 3,

$$\min_{k\in I_n(a,b)} \{[kp]/k - \overline{G}_{nk}(\xi_{nk}-a_n)\} \ge \frac{1}{2}\, g(\xi)a_n$$

on the set S_n when n is large. Hence for large n,

$$P[\min_{k\in I_n(a,b)} (\hat{\xi}_{nk}-\xi_{nk}) \le -a_n]$$

$$\le \sum_{k\in I_n(a,b)} EP[|k^{-1}\sum_1^k \{1(Z_{ni} \le \xi_{nk}-a_n) - G_{ni}(\xi_{nk}-a_n)\}| \ge \frac{1}{2}\, g(\xi)a_n \,|\, \mathcal{A}]$$

$$+ P(S_n^c) \le 2\, n^{\frac{4}{5}}(b-a) \exp[-\frac{1}{2}\, a\, g^2(\xi)(\log n)^2] + P(S_n^c)$$

by Theorem 1 of Hoeffding (1963). Since $\sum_{n=1}^{\infty} P(S_n^c) < \infty$ by Lemma 1 and $\sum_{n=1}^{\infty} n^{\frac{4}{5}} \exp[-\frac{1}{2}\, a\, g^2(\xi)(\log n)^2] < \infty$,

$$P[\min_{k\in I_n(a,b)} (\hat{\xi}_{nk}-\xi_{nk}) \le -a_n \text{ i.o.}] = 0.$$

In the same way,

$$P[\max_{k\in I_n(a,b)} (\hat{\xi}_{nk}-\xi_{nk}) \ge a_n \text{ i.o.}] = 0,$$

and the lemma is proved.

Lemma 9. $P[H_n^* > C\, n^{-\frac{3}{5}} \log n \quad \text{i.o.}] = 0$ for large C.

Proof. It follows from the monotonicity of $\hat{G}_{nk}(\cdot)$ and $\overline{G}_{nk}(\cdot)$ that for $z \in J_{nk,r} = [\eta_{nk,r}, \eta_{nk,r+1}]$,

$$H_{nk}(\eta_{nk,r}) - \alpha_{nk,r} \le H_{nk}(z) \le H_{nk}(\eta_{nk,r+1}) + \alpha_{nk,r} ,$$

where $H_{nk}(\cdot)$ is given by (17), and

$$\alpha_{nk,r} = \overline{G}_{nk}(\eta_{nk,r+1}) - \overline{G}_{nk}(\eta_{nk,r}) .$$

Hence

$$H^*_{nk} = \sup_{|z-\xi_{nk}| \le a_n} |H_{nk}(z)|$$

$$\le \max_{-b_n \le r \le b_n} |H_{nk}(\eta_{nk,r})| + \max_{-b_n \le r \le b_n - 1} \alpha_{nk,r} .$$

Let $S_n = \{Y_{n,[n^{\frac{4}{5}}b]} \le B\, n^{-\frac{1}{5}}\}$ as in the previous proofs. Then by Lemmas 3 and 4,

$$\max_{-b_n \le r \le b_n - 1} \alpha_{nk,r} \le 2\, g(\xi) n^{-\frac{3}{5}} \log n$$

on the set S_n when n is large. Hence

$$P[H_n^* > \{M + 2g(\xi)\}\, n^{-\frac{3}{5}} \log n]$$

$$\le P[\max_{k \in I_n(a,b)} \max_{-b_n \le r \le b_n} |H_{nk}(\eta_{nk,r})| > M\, n^{-\frac{3}{5}} \log n] + P(S_n^c)$$

$$\le 2n(b-a) \max_{k \in I_n(a,b)} \max_{-b_n \le r \le b_n} P[|H_{nk}(\eta_{nk,r})| > M\, n^{-\frac{3}{5}} \log n] + P(S_n^c).$$

But $\max_{-b_n \le r \le b_n} |\eta_{nk,r} - \xi_{nk}| \le n^{-\frac{2}{5}} \log n$, and by Lemma 7,

$$\sum_{n=1}^{\infty} n \cdot \max_{k \in I_n(a,b)} \max_{-b_n \le r \le b_n} P[|H_{nk}(\eta_{nk,r})| > M\, b^{-\frac{3}{5}} \log n] < \infty,$$

while $\sum_{n=1}^{\infty} P(S_n^c) < \infty$ by Lemma 1. This completes the proof. □

By Lemmas 8 and 9, we now have:

$$\begin{aligned} p - \hat{G}_{nk}(\xi_{nk}) &= \overline{G}_{nk}(\hat{\xi}_{nk}) - \overline{G}_{nk}(\xi_{nk}) + R_{nk}(1) \\ &= (\hat{\xi}_{nk} - \xi_{nk})\bar{g}_{nk}(\xi^*_{nk}) + R_{nk}(1), \end{aligned} \tag{18}$$

where ξ^*_{nk} lies between $\hat{\xi}_{nk}$ and ξ_{nk}, and

$$\max_{k \in I_n(a,b)} |R_{nk}(1)| = O(n^{-\frac{2}{5}} \log n), \text{ a.s.} \tag{19}$$

By the corollary to Lemma 4 and Lemma 8, $\max_{k \in I_n(a,b)} |\xi^*_{nk} - \xi| = O(n^{-\frac{2}{5}} \log n)$, a.s.

Consequently, Lemma 1 and 3 now imply

$$\max_{k \in I_n(a,b)} |\bar{g}_{nk}(\xi^*_{nk}) - g(\xi)| = O(n^{-\frac{2}{5}} \log n), \text{ a.s.} \tag{20}$$

Again, letting $U_{nki} = 1(Z_{ni} \le \xi_{nk})$, we have

$$P[\max_{k \in I_n(a,b)} |p - \hat{G}_{nk}(\xi_{nk})| > n^{-\frac{2}{5}} \log n]$$

$$\le \sum_{k \in I_n(a,b)} EP[|k^{-1} \sum_{1}^{k} \{U_{nki} - E(U_{nki} | \mathcal{A})\}| > n^{-\frac{2}{5}} \log n | \mathcal{A}]$$

$$\leq 2(b-a)n^{\frac{4}{5}} \exp[-2(b-a)(\log n)^2],$$

by Theorem 1 of Hoeffding (1963), and $\sum_{n=1}^{\infty} n^{\frac{4}{5}} \exp[-2(b-a)(\log n)^2] < \infty.$

Hence

$$p - \hat{G}_{nk}(\xi_{nk}) = O(n^{-\frac{2}{5}} \log n), \text{ a.s.} \tag{21}$$

From (18), (19), (20), and (21), we have

$$\max_{k \in I_n(a,b)} |(\hat{\xi}_{nk} - \xi_{nk}) - \{g(\xi)\}^{-1} [p - \hat{G}_{nk}(\xi_{nk})]| = O(n^{-\frac{3}{5}} \log n), \text{ a.s.}$$

Since $p - \hat{G}_{nk}(\xi_{nk}) = k^{-1} \sum_1^k [1(Z_{ni} > \xi_{nk}) - \{1 - G_{ni}(\xi_{nk})\}]$,

we now have the following representation:

$$\begin{aligned} &\hat{\xi}_{nk} = \xi_{nk} + \{kg(\xi)\}^{-1} \sum_1^k [1(Z_{ni} > \xi_{nk}) - \{1 - G_{ni}(\xi_{nk})\}] + R_{nk}, \\ &\max_{k \in I_n(a,b)} |R_{nk}| = O(n^{-\frac{3}{5}} \log n), \text{ a.s.} \end{aligned} \tag{22a}$$

This representation can be easily modified to two other slightly different forms, viz.,

$$\hat{\xi}_{nk} = \xi_{nk} + \{k\, g(\xi)\}^{-1} \sum_1^k [1(Z_{ni} > \xi) - \{1 - G_{ni}(\xi)\}] + R_{nk} \tag{22b}$$

and

$$\hat{\xi}_{nk} = \xi_{nk} + \{k\, g(\xi)\}^{-1} \sum_1^k [1(Z_{ni}^* > \xi) - \{1 - G(\xi)\}] + R_{nk} \tag{22c}$$

where $\max_{k \in I_n(a,b)} |R_{nk}| = O(n^{-\frac{3}{5}} \log n)$, a.s., in both (22b) and (22c), and $Z^*_{ni} = G^{-1} \circ G_{ni}(Z_{ni})$ and $G(\cdot) = G(\cdot|x_0)$ is the conditional cdf of Z given $X = x_0$. Note that since $Z_{n1},\ldots,Z_{nn}$ are conditionally independent given $\mathscr{A}$, with Z_{ni} having conditional cdf G_{ni},

$$P[Z^*_{ni} \le z_i,\ i=1,\ldots,n] = E\, P[Z^*_{ni} \le z_i,\ i=1,\ldots,n \,|\, \mathscr{A}]$$

$$= E\, P[Z_{ni} \le G^{-1}_{ni} \circ G(z_i),\ i=1,\ldots,n \,|\, \mathscr{A}] = \prod_1^n G(z_i)\ ,$$

so that for each n, $Z^*_{n1},\ldots,Z^*_{nn}$ are iid with cdf $G(\cdot)$.

To obtain (22b) from (22a), use Lemma 7 with $\zeta_{nk} = \xi$ to conclude that

$$\max_{k \in I_n(a,b)} |k^{-1} \sum_1^k \{1(Z_{ni} \le \xi_{nk}) - 1(Z_{ni} \le \xi)\} - \{G_{ni}(\xi_{nk}) - G_{ni}(\xi)\}|$$

$$= O(n^{-\frac{3}{5}} \log n),\ \text{a.s.}$$

To obtain (22c) from (22b) it is enough to show that

$$\text{(23)} \qquad \max_{k \in I_n(a,b)} |k^{-1} \sum_1^k (U_{ni} - \mu_{ni})| = O(n^{-\frac{3}{5}} \log n),\ \text{a.s.},$$

where $U_{ni} = 1(Z^*_{ni} > G^{-1} \circ G_{ni}(\xi)) - 1(Z^*_{ni} > \xi)$, and $\mu_{ni} = G(\xi) - G_{ni}(\xi) = E(U_{ni} | \mathscr{A})$. For this, use Lemma 3 to observe that for large n,

$$\max_{i \le [n^{\frac{4}{5}} b]} |\mu_{ni}| \le B^2\, |Q(\xi)|\, n^{-\frac{2}{5}} \text{ on the set } S_n = \{Y_{n,[n^{\frac{4}{5}} b]} \le B n^{-\frac{1}{5}}\}\ ,$$

and then use Lemma 6 and Lemma 1 to obtain

$$\sum_{n=1}^{\infty} [\max_{k\in I_n(a,b)} |k^{-1}\sum_{1}^{k}(U_{ni}-\mu_{ni})| > n^{-\frac{3}{5}}\log n]$$

$$\le 2(b-a)\sum_{n=1}^{\infty} n^{\frac{4}{5}}\exp[-a\{4B^2|Q(\xi)|\}^{-1}(\log n)^2] + \sum_{n=1}^{\infty} P(S_n^c) < \infty .$$

This proves (23) and the representation (22c) is established.

Combine Lemma 5 with (22c) and note that $G(\xi) = p$ to complete the proof of Theorem N1.

8. Proof of Theorem K1.

The kernel estimator $\tilde{\xi}_{nh}$ can be regarded as the NN estimator $\hat{\xi}_{nK_n(h)}$ in which $K_n(h)$ is a random integer given by (4). A formal substitution for k by $K_n(h)$ in the representation given in Theorem N1 leads to

$$(24)\quad \tilde{\xi}_{nh}-\xi = \beta(\xi)\{K_n(h)/n\}^2 + \{K_n(h)g(\xi)\}^{-1}\sum_{1}^{K_n(h)}[1(Z^*_{ni} > \xi)-(1-p)] + R_{nK_n(h)} .$$

However, this is of no use unless we can show that

(a) $\sup_{h\in J_n(c,d)} |R_{nK_n(h)}|$ converges at a fast rate,

where $J_n(c,d) = [n^{-\frac{1}{5}}c,\ n^{-\frac{1}{5}}d]$, $0 < c < d$, and

(b) in the first two terms of the RHS of (24), $K_n(h)$ can be replaced by the leading term of its deterministic component without slowing down the rate of convergence of the remainder term.

We now proceed to establish (a) and (b).

First examine the magnitude of $K_n(h) - nhf(x_0)$ in the following lemma.

<u>Lemma 10.</u> Let $\Delta_n(h) = K_n(h) - nhf(x_0)$. Then

$$\sup_{h \in J_n(c,d)} |\Delta_n(h)| = O(n^{\frac{2}{5}} \log n), \text{ a.s.}$$

<u>Proof</u>. Let $\mu(h) = P(Y \le h/2)$ and write $\Delta_n(h) = \Delta_{n1}(h) + \Delta_{n2}(h)$, where $\Delta_{n1}(h) = \sum_1^n [1(Y_i \le h/2) - \mu(h)]$, $\Delta_{n2}(h) = n[\mu(h) - hf(x_0)]$. By condition 1, $|\mu(h) - h\, f(x_0)| \le (h^3/24) \sup_{|x-x_0| \le h/2} |f''(x_0)|$. Thus $\sup_{h \in J_n(c,d)} |\Delta_{n2}(h)| = O(n^{\frac{2}{5}})$.

Next consider $\Delta_{n1}(h)$, which is the difference of non–decreasing functions $\sum_1^n 1(Y_i \le h/2)$ and $n\mu(h)$. Since $\mu'(h) < 2\, f(x_0)$ on $J_n(c,d)$, we divide $J_n(c,d)$ into $\nu_n = 2(d-c)\, f(x_0) n^{\frac{2}{5}}/\log n$ equal intervals to ensure that $n\mu(h)$ increases by at most $n^{\frac{2}{5}} \log n$ in each of these intervals. Consequently, if $|\Delta_{n1}(h)| \le n^{\frac{2}{5}} \log n$ at the ν_n+1 endpoints of these intervals, then $\sup_{h \in J_n(c,d)} |\Delta_{n1}(h)| \le 2\, n^{\frac{2}{5}} \log n$. Thus

$$P[\sup_{h \in J_n(c,d)} |\Delta_{n1}(h)| > 2\, n^{\frac{2}{5}} \log n] \le (\nu_n+1) \sup_{h \in J_n(c,d)} P[|\Delta_{n1}(h)| > n^{\frac{2}{5}} \log n].$$

Finally, since $\mu(h) = hf(x_0) + n^{-1} \Delta_{n2}(h) \le 2\, d\, f(x_0) n^{-\frac{1}{5}}$ for all $h \in J_n(c,d)$,

$$\sup_{h \in J_n(c,d)} P[|\Delta_{n1}(h)| > n^{\frac{2}{5}} \log n] \le 2 \exp[-(\log n)^2/\{8\, df(x_0)\}]$$

by Bernstein's inequality, and $\sum_{n=1}^{\infty} n^{\frac{2}{5}} \exp[-a(\log n)^2] < \infty$ for all $a > 0$. Hence $\sup_{h \in J_n(c,d)} |\Delta_{n1}(h)| = O(n^{\frac{2}{5}} \log n)$, a.s., and the lemma is proved. □

The convergence rate of $\sup_{h \in J_n(c,d)} |R_{nK_n(h)}|$ is now determined in the following lemma.

<u>Lemma 11.</u> $\sup_{h \in J_n(c,d)} |R_{nK_n(h)}| = O(n^{-\frac{3}{5}} \log n)$, a.s.

<u>Proof</u>. For $0 < c < d$, let $0 < a = c\, f(x_0)/2 < 2\, d\, f(x_0) = b$, and let

$$A_n = \{ \sup_{h \in J_n(c,d)} |R_{nK_n(h)}| > M\, n^{-\frac{3}{5}} \log n \},$$

$$B_n = \{ \max_{k \in I_n(a,b)} |R_{nk}| > M\, n^{-\frac{3}{5}} \log n \},$$

$$C_n = \{ \max_{h \in J_n(c,d)} |\Delta_n(h)| > M\, n^{\frac{2}{5}} \log n \} .$$

There there exists $N_0 = N_0(M)$ such that for $n > N_0$, C_n^c implies $K_n(h) \in I_n(a,b)$ for all $h \in J_n(c,d)$. Thus

$$C_n^c \cap A_n \subset C_n^c \cap B_n \quad \text{for } n > N_0 .$$

It now follows that for sufficiently large M,

$$\begin{aligned} P[A_n \text{ i.o.}] &\le P[C_n \text{ i.o.}] + P[\bigcup_{N_0 \ge 1} \bigcap_{n \ge N_0} C_n^c \text{ and } A_n \text{ i.o.}] \\ &\le P[C_n \text{ i.o.}] + P[\bigcup_{N > N_0} \bigcap_{n \ge N_0} C_n^c \text{ and } B_n \text{ i.o.}] \\ &= 0, \end{aligned}$$

since for large M, $P[C_n \text{ i.o.}] = 0$ by Lemma 10 and $P[B_n \text{ i.o.}] = 0$ by Theorem N1.

□

We now consider the first two terms on the RHS of (24). Of these,

(25) $$\beta(\xi)\{K_n(h)/n\}^2 = \beta(\xi)f^2(x_0)h^2 + R'_{nh}$$

where $R'_{nh} = \beta(\xi)f^2(x_0)h^2\ [\Delta_n(h)/\{nhf(x_0)\}] \cdot [2 + \Delta_n(h)/\{nhf(x_0)\}]$, and by Lemma 10,

$$\sup_{h\in J_n(c,d)} |R'_{nh}| = O(n^{-\frac{4}{5}} \log n), \text{ a.s.}$$

To examine the other term, let

(26) $$U_{ni} = 1(Z^*_{ni} > \xi) - (1-p), \quad m_n(h) = [n\, h\, f(x_0)].$$

Then

$$\{K_n(h)g(\xi)\}^{-1} \sum_1^{K_n(h)} U_{ni} = \{m_n(h)g(\xi)\}^{-1} \sum_1^{m_n(h)} U_{ni} + R''_{nh} + R'''_{nh},$$

(27) $$R''_{nh} = -\{\Delta_n(h)/K_n(h)\}\ \{m_n(h)g(\xi)\}^{-1} \sum_1^{m_n(h)} U_{ni},$$

$$R'''_{nh} = \{1-\Delta_n(h)/K_n(h)\}\ \{m_n(h)\, g(\xi)\}^{-1} \left[\sum_1^{k_n(h)} U_{ni} - \sum_1^{m_n(h)} U_{ni}\right],$$

where $U_{n1},\ldots,U_{nn}$ are conditionally independent given $\mathcal{A}$, with $E(U_{ni}|\mathcal{A}) = 0$. By Lemma 10, $\sup_{h\in J_n(c,d)} |\Delta_n(h)/K_n(h)| = O(n^{-\frac{2}{5}} \log n)$, a.s., and

$$\sup_{h\in J_n(c,d)} |m_n(h) \sum_1^{m_n(h)} U_{ni}| = \max_{n^{\frac{4}{5}}cf(x_0) \le k \le n^{\frac{4}{5}}df(x_0)} |k^{-1} \sum_1^k U_{ni}| = O(n^{-\frac{1}{5}}), \text{ a.s.}$$

by an application of Theorem 1 of Hoeffding (1963). Hence

$$\sup_{h\in J_n(c,d)} |R''_{nh}| = O(n^{-\frac{3}{5}}\log n), \text{ a.s.}$$

Now consider the jump–points of $m_n(h) = [nhf(x_0)]$ in $J_n(c,d)$ together with the end–points $n^{-\frac{1}{5}}c$ and $n^{-\frac{1}{5}}d$, and call these points $n^{-\frac{1}{5}}c = h_{n0} < h_{n1} < \dots < h_{n\nu_n} = n^{-\frac{1}{5}}d$. Then

$$\nu_n \le n^{\frac{4}{5}}(d-c)\, f(x_0), \; h_{n,j+1} - h_{nj} \le \{n\, f(x_0)\}^{-1},$$

and $m_n(h)$ is constant on each of the ν_n intervals $[h_{nj}, h_{n,j+1})$. At the same time, $K_n(h)$ is also integer–valued and non–decreasing, and $|U_{ni}| \le 1$. Hence for each j and for all $h_{nj} \le h < h_{n,j+1}$,

$$\begin{aligned} &\Big|\sum_{1}^{K_n(h)} U_{ni} - \sum_{1}^{m_n(h)} U_{ni}\Big| = \Big|\sum_{1}^{K_n(h_{nj})} U_{ni} - \sum_{1}^{m_n(h_{nj})} U_{ni}\Big| \\ &\quad + \Big|\sum_{K_n(h_{nj})+1}^{K_n(h)} U_{ni}\Big| \le \Big|\sum_{1}^{K_n(h_{nj})} U_{ni} - \sum_{1}^{m_n(h_{nj})} U_{ni}\Big| + \{K_n(h_{n,j+1}) - K_n(h_{nj})\}. \end{aligned}$$

Therefore, if we can show that

$$\max_{0\le j\le \nu_n} \Big|\sum_{1}^{K_n(h_{nj})} U_{ni} - \sum_{1}^{m_n(h_{nj})} U_{ni}\Big| = O(n^{\frac{1}{5}}\log n), \text{ a.s.,} \tag{28}$$

$$\max_{0\le j\le \nu_n} \{K_n(h_{n,j+1}) - K_n(h_{nj})\} = O(n^{\frac{1}{5}}\log n), \text{ a.s.} \tag{29}$$

then in (27) $\sup_{h\in J_n(c,d)} |R'''_n(h)| = O(n^{-\frac{3}{5}}\log n)$, a.s., will follow because by Lemma 10,

$$\sup_{h \in J_n(c,d)} |1-\Delta_n(h)/K_n(h)| \{m_n(h)\}^{-1} = O(n^{-\frac{4}{5}}), \text{ a.s.}$$

To prove (29), note that $K_n(h_{n,j+1}) - K_n(h_{nj})$ is Binomial (n, π_{nj}), where

$$\pi_{nj} = P(h_{nj}/2 < Y_i \leq h_{n,j+1}/2) = n^{-1} [1+0(1)] ,$$

since $h_{n,j+1} - h_{nj} \leq \{nf(x_0)\}^{-1}$. Hence for large n and for all j,

$$P[K_n(h_{n,j+1}) - K_n(h_{nj}) \geq 2 M n^{\frac{1}{5}} \log n]$$

$$\leq P[n^{-1} \{K_n(h_{n,j+1}) - K_n(h_{nj})\} - \pi_{nj} \geq M n^{-\frac{4}{5}} \log n]$$

$$\leq \exp[-(M/4) n^{\frac{1}{5}} \log n]$$

by Lemma 6, and $\sum_{n=1}^{\infty} \nu_n \exp[-(\mu/4) n^{\frac{1}{5}} \log n] < \infty$, because $\nu_n = \text{const.}\ n^{\frac{4}{5}}$. This proves (29).

Finally, note that $\sum_1^{K_n(h_{nj})} U_{ni} - \sum_1^{m_n(h_{nj})} U_{ni}$ is a sum of $|K_n(h_{nj}) - m_n(h_{nj})|$ terms in which the summands U_{ni} given in (26), when $K_n(h_{nj}) > m_n(h_{nj})$, and $-U_{ni}$, when $K_n(h_{nj}) < m_n(h_{nj})$ are conditionally independent given $\mathcal{A}$, with $E(U_{ni} | \mathcal{A}) = 0$. Using Hoeffding's inequality, we therfore have

$$P[| \sum_1^{K_n(h_{nj})} U_{ni} - \sum_1^{m_n(h_{nj})} U_{ni} | > M n^{\frac{1}{5}} \log n]$$

$$\leq 2\, E \exp[-2 M^2 n^{\frac{2}{5}} (\log n)^2 / |K_n(h_{nj}) - m_n(h_{nj})|]$$

$$\le 2\,P[|K_n(h_{nj}) - m_n(h_{nj})| > M\, n^{\frac{2}{5}} \log n]$$

$$+\, 2 \exp[-2\, M \log n]$$

Since $\sum_{n=1}^{\infty} \nu_n \exp[-2\, M \log n] = \text{const.} \sum_{n=1}^{\infty} n^{\frac{4}{5}-2M} < \infty$ for sufficiently large M, we only need to show that

$$(30) \qquad \sum_{n=1}^{\infty} n^{\frac{4}{5}} P[|K_n(h_{nj}) - m_n(h_{nj})| > M\, n^{\frac{2}{5}} \log n] < \infty$$

for large M, in order to establish (28). But $K_n(h) - M_n(h) = K_n(h) - [nhf(x_0)]$ differs from $\Delta_n(h) = K_n(h) - nhf(x_0)$ by at most 1, and it was shown in the proof of Lemma 10, that $P[|\Delta_n(h)| > M\, n^{\frac{2}{5}} \log n] < \exp[-a(\log n)^2]$ for some $a > 0$, which implies (30) and thus (28) is established. We have now shown that in (27),

$$\sup_{h \in J_n(c,d)} |R'''_{nh}| = O(n^{-\frac{3}{5}} \log n), \text{ a.s.}$$

To complete the proof of Theorem K1, substitute the expressions in (25) and (27) for the first two terms on the RHS of (24). The remainder terms R'_{nh}, R''_{nh} and R'''_{nh} is these expressions have all been shown to be $O(n^{-\frac{3}{5}} \log n)$, a.s. and uniformly in $h \in J_n(c,d)$, and the other remainder term $R_{nK_n(h)}$ has also been shown to be of this order in Lemma 11. This establishes the order of magnitude of $R^*_{nh} = R'_{nh} + R''_{nh} + R'''_{nh} + R_{nK_n(h)}$ claimed in the theorem.

REFERENCES

(1) Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* 37, 577–580.

(2) Bhattacharya, P.K. (1963). On an analog of regression analysis. *Ann. Math. Statist.* 34, 1459–1473.

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

(4) Bhattacharya, P.K. and Mack Y.P. (1987). Weak convergence of k–NN density and regression estimators with varying k and applications. 15, 976–994.

(5) Cheng, K.F. (1983). Nonparametric estimators for percentile regression functions. *Commun.Statist.–Theor. Meth.* 12, 681–692.

(6) Csorgo, M. and Rëvesz, P. (1981). *Strong approximations in Probability and Statistics.* Academic Press, New York.

(7) Gangopadhyay, A.K. (1987). Nonparametric estimation of conditional quantile function. *Ph.D. dissertation at the University of California at Davis.*

(8) Gikhman I.I. and Skorokhod, A.V. (1969). *Introduction to the theory of random processes.* W.B. Saunders company, Philadelphia.

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

(10) Hogg, R.V. (1975). Estimates of percentile regression lines using salary data. *J. Amer. Statist. Assoc.* 70, 56–59.

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

(12) Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.* 5, 595–645.

(13) Stute, W. (1986). Conditional empirical processes. *Ann. Stistist.*, 14, 638–647.

(14) Uspensky, J.V. (1937). *Introduction to Mathematical Probability.* McGraw–Hill, New York.