Nonparametric Estimation of a Regression Function: Limiting Distribution

by

Kuang-Fu Cheng and Pi-Erh Lin

FSU Statistics Report No. 2518
ONR Technical Report No. 148

September, 1979
Department of Statistics
The Florida State University
Tallahassee, Florida 32306

research supported by the Army, Navy and Air Force under Office of Naval Research Contract No. N00014-76-C-0608. Reproduction in whole or in part is permitted for any purpose of the United States Government.
Nonparametric Estimation of a Regression Function: Limiting Distribution

by

Kuang-Fu Cheng\textsuperscript{1,2} and Pi-Erh Lin\textsuperscript{1}
The Florida State University

Summary

Consider the regression model $Y_i = g(x_i) + e_i$, $i = 1, \ldots, n$, where $g$ is an unknown function defined on $[0, 1]$, $0 = x_0 < x_1 < \ldots < x_n \leq 1$ are chosen so that $\max_{1 \leq i \leq n} (x_i - x_{i-1}) = o(n^{-1})$, and where $\{e_i\}$ are i.i.d. with $Ee_i = 0$ and $\text{Var} e_i = \sigma^2$. In a previous paper, Cheng and Lin (1979) study three estimators of $g$, namely, $g_{1n}$ of Cheng and Lin (1979), $g_{2n}$ of Clark (1977), and $g_{3n}$ of Priestley and Chao (1972). Consistent results are established and rates of strong uniform convergence are obtained. In the current investigation the limiting distribution of $g_{1n}$, $i = 1, 2, 3$, and that of the isotonic estimator $g^{**}$ are considered.

\textsuperscript{1} Research supported by the Army, Navy and Air Force under Office of Naval Research Contract No. N00014-76-C-0608. Reproduction in whole or in part is permitted for any purpose of the United States Government.

\textsuperscript{2} Now at the SUNY at Buffalo, Amherst, New York 14226, USA

AMS 1970 subject classifications: Primary 62E20; Secondary 62G05

Key Words and Phrases: Asymptotic normality; Berry-Esseen bound; isotonic; kernel function; Liapunov's theorem; and Lipschitz.
1. **Introduction.** Let \((x_1, Y_1), \ldots, (x_n, Y_n)\) be \(n\) independent pairs of observations where \(0 = x_0 < x_1 < \ldots < x_n \leq 1\) are fixed predictor variables such that \(\max_{1 \leq i \leq n} (x_i - x_{i-1}) = o(n^{-1})\) and \(Y_1, \ldots, Y_n\) are observations on the criterion variable \(Y\) according to the nonlinear regression model

\[
Y_i = g(x_i) + e_i, \ i = 1, \ldots, n,
\]

where \(g\) is an unknown function defined on \([0, 1]\) and \(\{e_i\}\) are i.i.d. random variables with \(\mathbb{E}e_i = 0\) and \(\text{Var} e_i = \sigma^2 < \infty\). In a previous paper, Cheng and Lin (1979) study stochastic properties for three estimators of \(g\): namely, \(g_{1n}\) of Cheng and Lin (1979), \(g_{2n}\) of Clark (1977), and \(g_{3n}\) of Priestley and Chao (1972). Also considered is an isotonic estimator \(g^{**}_n\) corresponding to \(g_{1n}\) when \(g\) is known to be nondecreasing. These estimators are constructed utilizing a kernel function \(k(z)\) and a sequence \(\{a_n\}\) of positive constants converging to 0 as \(n \to \infty\). For simplicity, the kernel function \(k(z)\) is specialized to satisfy (i) \(k(z) \geq 0\) for all \(z \in (-\infty, \infty)\), (ii) \(k(z) = 0\) for all \(z \notin [-L, L]\) for some positive constant \(-L\), and \(\int k(z)dz = 1\). (Throughout the study no limits of integration will be given whenever the integration extends over \((-\infty, \infty)\).) Under mild regularity conditions, these estimators have been shown to enjoy various weak and strong uniform convergence properties.

In this paper we will study the limiting distributions of \(g_{1n}, \ g_{2n}, \ g_{3n}, \) and of \(g^{**}_n\); the former will be presented in Section 2 and the latter in Section 3.
For the benefit of the reader, the estimators under consideration are given below:

\begin{equation}
(1.2) \quad g_{1n}(x) = \sum_{i=1}^{n} Y_i \int_{x_{i-1}}^{x_i} a_n^{-1} k((x - z)/a_n) \, dz;
\end{equation}

\begin{equation}
(1.3) \quad g_{2n}(x) = \sum_{i=1}^{n} c_{ni} Y_i
\end{equation}

where, for all \( x \in (0, 1) \) and for \( n \) sufficiently large,

\begin{equation}
(1.4) \quad c_{ni} = \begin{cases} 
\int_{x_{i-1}}^{x_i} a_n^{-1} k \left( \frac{x-z}{a_n} \right) \, dz + \int_{x_{i-1}}^{x_{i+1}} a_n^{-1} k \left( \frac{x-z}{a_n} \right) \, dz, & i = 1 \\
\int_{x_{i-1}}^{x_{i+1}} a_n^{-1} k \left( \frac{x-z}{a_n} \right) \, dz + \int_{x_{i+1}}^{x_{i+1}} a_n^{-1} k \left( \frac{x-z}{a_n} \right) \, dz, & i = 2, \ldots, n-1 \\
\int_{x_{n-1}}^{x_{n}} a_n^{-1} k \left( \frac{x-z}{a_n} \right) \, dz, & i = n;
\end{cases}
\end{equation}

\begin{equation}
(1.5) \quad g_{3n} = \sum_{i=1}^{n} Y_i (x_i - x_{i-1}) a_n^{-1} k \left( \frac{x-x_i}{a_n} \right);
\end{equation}

and

\begin{equation}
(1.6) \quad g^{**}(x) = \max \min \left[ \sum_{s \leq t}^{t} \sum_{j=s}^{t} \frac{g_{1n}(t_{nj}) w(t_{nj})}{\sum_{j=s}^{t} w(t_{nj})} \right]
\end{equation}

for all \( x \in [t_{ni}, t_{n(i+1)}], \) \( i = 1, \ldots, n-1, \) where \( w \) is a positive weight function defined on \([0, 1]\) and \( 0 < a \leq t_{n1} < t_{n2} < \ldots < t_{nn} \leq b < 1.\)

2. Limiting distributions of \( g_{1n}(x). \) Benedetti (1977) obtained the asymptotic normality of \( g_{3n} \) which, for completeness, is stated below:
Lemma 2.1. Assume the following conditions:

(i) $k(z)$ is continuous and nondecreasing for $z < 0$, nonincreasing for $z > 0$ and $\int k^3(z)dz < \infty$;

(ii) $\Delta_1/n \leq x_{i} - x_{i-1} \leq \Delta_2/n$ for some $0 < \Delta_1 \leq \Delta_2$, $i = 1, \ldots, n$, and $na_n \to \infty$ as $n \to \infty$; and

(iii) $\gamma = E|e|^3 < \infty$.

Then

$$
\frac{G_{3n}(x) - E G_{3n}(x)}{[\text{Var} \ G_{3n}(x)]^{1/2}} \xrightarrow{L} N(0, 1), \text{ as } n \to \infty,
$$

for all $x \in (0, 1)$. Furthermore, if $x_{i} - x_{i-1} = n^{-1}$, $i = 1, \ldots, n$, then

$$
\text{Var}[G_{3n}(x)] = (na_n)^{-1} \sigma^2 \int k^2(z)dz.
$$

In view of Lemma 1 of Cheng and Lin (1979), if

$$
||g||_{[0,1]} \equiv \sup_{0 \leq x \leq 1} |g(x)| < \infty, \ k(z) \in \text{Lip}(\beta), \text{ and } na_n^{(2\beta+1)/(2\beta-1)} \to 0 \text{ as } n \to \infty,
$$

then conditions (i), (ii), and (iii) of Lemma 2.1 imply that

$$
\frac{G_{1n}(x) - E G_{1n}(x)}{[\text{Var} \ G_{1n}(x)]^{1/2}} \xrightarrow{L} N(0, 1), \text{ as } n \to \infty,
$$

for all $x \in (0, 1)$. This immediate result seems attractive but, in fact, it is obtained under rather restrictive conditions on the kernel function $k(z)$ and the sequence $\{a_n\}$. In the following theorem, we will present theory on the asymptotic distributions of $G_{1n}(x)$ under weaker conditions on $k$ and $a_n$. The limiting distribution of $g_{2n}(x)$ will also be characterized. Later, a consistent estimator of $\sigma^2$ will be suggested. This, together with an appropriate convergence rate of the bias, $E G_{1n}(x) - g(x)$, established by
Cheng and Lin (1979), an approximate (1 - \(\alpha\))100% confidence interval for \(g(x)\) can easily be constructed using the asymptotic distribution of \(g_{1n}(x)\).

**Theorem 2.2.** Assume conditions (ii) and (iii) of Lemma 2.1. If \(k(z) \in \text{Lip}(\beta)\) for some \(\beta > 0\), then

\[
\frac{g_{1n}(x) - Eg_{1n}(x)}{[\text{Var } g_{1n}(x)]^{\frac{1}{2}}} \xrightarrow{L} N(0, 1), \text{ as } n \to \infty,
\]

for all \(x \in (0, 1)\). If, in addition, \(x_i - x_{i-1} = \frac{1}{n}\), for all \(i\), then

\[
\frac{g_{2n}(x) - Eg_{2n}(x)}{[\text{Var } g_{2n}(x)]^{\frac{1}{2}}} \xrightarrow{L} N(0, 1), \text{ as } n \to \infty,
\]

for all \(x \in (0, 1)\).

**Proof.** Since \(g_{1n}(x)\) is a weighted sum of \(n\) independent random variables, the asymptotic normality of \(g_{1n}(x) - Eg_{1n}(x)\) may be established by verifying the Berry-Esseen bound. In so doing, it is necessary to evaluate the second and third moments of the independent random variables

\[
(2.6) \quad g_{1n}(x, i) = \sum_{i=1}^{x_i} \frac{x_i - x_{i-1}}{n} k\left(\frac{x_i - z}{a_n}\right) dz, \quad i = 1, ..., n.
\]

Consider the following approximation to \(\text{Var } g_{1n}(x)\):

\[
(2.7) \quad |\text{Var } g_{1n}(x) - \sigma^2| \leq \sigma^2 \left| \sum_{i=1}^{x_i} \left( \frac{x_i - x_{i-1}}{a_n} \right) a_n^{-1} \int_{x_{i-1}}^{x_i} k^2\left(\frac{x-z}{a_n}\right) dz \right|
\]

\[
\leq \sigma^2 \left| \sum_{i=1}^{x_i} \left( \frac{x_i - x_{i-1}}{a_n} \right) a_n^{-1} \int_{x_{i-1}}^{x_i} k\left(\frac{x-z}{a_n}\right) dz \right|^2 - \sigma^2 \left| \sum_{i=1}^{x_i} \left( \frac{x_i - x_{i-1}}{a_n} \right) a_n^{-1} \int_{x_{i-1}}^{x_i} k\left(\frac{x-z}{a_n}\right) dz \right| \int_{x_{i-1}}^{x_i} \frac{x-z}{a_n}^2 dz
\]

\[
+ \sigma^2 \left| \sum_{i=1}^{x_i} \left( \frac{x_i - x_{i-1}}{a_n} \right) a_n^{-1} \int_{x_{i-1}}^{x_i} k\left(\frac{x-z}{a_n}\right) dz \right| - \sigma^2 \left| \sum_{i=1}^{x_i} \left( \frac{x_i - x_{i-1}}{a_n} \right) a_n^{-1} \int_{x_{i-1}}^{x_i} k\left(\frac{x-z}{a_n}\right) dz \right| \int_{x_{i-1}}^{x_i} \frac{x-z}{a_n}^2 dz
\]
\begin{align*}
\leq \sigma^2 \sum_{i=1}^{n} \left[ \int_{x_{i-1}}^{x_i} a^{-1}_n k \left( \frac{x-z}{a_n} \right) dz \right] \int_{x_{i-1}}^{x_i} a^{-1}_n \left| k \left( \frac{x-z}{a_n} \right) - k \left( \frac{x-x_i}{a_n} \right) \right| dz \\
+ \sigma^2 \sum_{i=1}^{n} \left( \frac{x_i-x_{i-1}}{a_n} \right) \int_{x_{i-1}}^{x_i} a^{-1}_n \left| k \left( \frac{x-z}{a_n} \right) - k \left( \frac{x-x_i}{a_n} \right) \right| dz \\
\leq C_1 (\delta_n a^{-1}_n)^{\beta+1}, \quad 0 < C_1 < \infty, \text{ for all } x \in (0, 1).
\end{align*}

To evaluate the third absolute moment, we recall (2.6) and condition (iii) of Lemma 2.1. Then, for all $x \in (0, 1)$,

\begin{align*}
(2.8) \quad & \left| \sum_{i=1}^{n} E \left[ g_{1n}(1,x) - E g_{1n}(1,x) \right] \right|^3 - \gamma \left| \sum_{i=1}^{n} \left( \frac{x_i-x_{i-1}}{a_n} \right)^2 a^{-1}_n \int_{x_{i-1}}^{x_i} k \left( \frac{x-z}{a_n} \right) dz \right| \\
= \gamma \left| \sum_{i=1}^{n} \left[ \int_{x_{i-1}}^{x_i} a^{-1}_n k \left( \frac{x-z}{a_n} \right) dz \right] - \sum_{i=1}^{n} \left( \frac{x_i-x_{i-1}}{a_n} \right)^2 a^{-1}_n \int_{x_{i-1}}^{x_i} k \left( \frac{x-z}{a_n} \right) dz \right| \\
\leq \gamma \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} a^{-1}_n k \left( \frac{x-z}{a_n} \right) \left| A_i - B_i \right| (A_i + B_i) dz \\
& \leq \gamma (\delta_n a^{-1}_n)^{\beta+1}.2\delta_n a^{-1}_n \left| k \right|_{[-L,L]} \\
& = C_2 (\delta_n a^{-1}_n)^{\beta+2}, \quad 0 < C_2 < \infty,
\end{align*}

where we have set

\begin{align*}
A_i = \int_{x_{i-1}}^{x_i} a^{-1}_n k \left( \frac{x-t}{a_n} \right) dt, \quad \text{and} \\
B_i = \left( \frac{x_i-x_{i-1}}{a_n} \right) k \left( \frac{x-z}{a_n} \right), \quad i = 1, \ldots, n
\end{align*}
Now, according to conditions (ii) and (iii) of Lemma 2.1 and upper bounds (2.7) and (2.3), the Berry–Esseen bound becomes

$$\frac{n}{\text{Var} g_{1n}(x)} \left[ E[g_{1n}(i,x) - E g_{1n}(i,x)]^3 \right]^{3/2}$$

(2.9)

$$\leq \frac{(n\alpha_n)^{-2} \Delta_2^2 \left[ \gamma_k^3(z) dz + C_2 \Delta_2 \alpha_n^{-\beta} \right]}{(n\alpha_n)^{-3/2} \left[ \sigma_1^2 \Delta_1^2 \left[ k_2(z) dz - C_1 \Delta_2 \alpha_n^{-\beta} \right] \right]^{3/2}}$$

$$= 0((n\alpha_n)^{-\delta}), \text{ for sufficiently large } n,$$

and, hence by Liapunov's theorem, assertion (2.4) follows. To prove assertion (2.5), write

(2.10) \ Var g_{2n}(x)

$$= \sigma^2 \left\{ \left[ \int_{x_0}^{x_1} a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz + \int_{x_1}^{x_2} \frac{x^2-z^2}{x_2-x_1} \right] a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \right\}^2$$

$$+ \sum_{i=2}^{n-1} \left[ \int_{x_i}^{x_{i+1}} a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \right] + \sum_{i=2}^{n-1} \left[ \int_{x_i}^{x_{i-1}} a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \right]$$

$$+ \left[ \int_{x_0}^{x_1} \left[ \frac{z-x-1}{x-x-1} \right] a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \right]^2$$

$$= \sigma^2 \left\{ \left[ \int_{x_0}^{x_1} a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \right]^2$$

$$+ 2 \int_{x_1}^{x_2} \frac{x^2-z^2}{x_2-x_1} k \left( \frac{x-z}{a_n} \right) dz \int_{x_0}^{x_1} a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz - \int_{x_1}^{x_2} \frac{z-x}{x_2-x_1} a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \right\}^2$$
+ \ldots + 2 \int_{x_{n-1}}^{x_n} \left( \frac{x - z}{x_{n-2} - x_{n-1}} \right) a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \left[ \int_{x_{n-2}}^{x_{n-1}} \left( \frac{x - z}{x_{n-2} - x_{n-1}} \right) a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \right]

= T_{n1} + T_{n2} \text{ say,}

where $T_{n1} = \text{Var} g_{1n}(x)$, and $T_{n2}$ is the sum of the remaining terms which, after a straightforward simplification, reduces to $0((na_n)^{-2} + (na_n)^{-\beta-1})$.

Consequently, in view of (2.7) with $\delta_n = n^{-1}$, we have

\begin{equation}
(2.11) \quad \left| \text{Var} g_{2n}(x) - \alpha^2 \sum_{i=1}^{n} \left( \frac{x_i - x_{i-1}}{a_n} \right)^2 \int_{x_{i-1}}^{x_i} a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \right| = 0((na_n)^{-2} + (na_n)^{-\beta-1}).
\end{equation}

Similar to (2.6), define $g_{2n}(i, x) = c_{ni} x_i, i = 1, \ldots, n$, where $c_{ni}$ is given by (1.4). Then it can be shown that

\begin{equation}
(2.12) \quad \left| \sum_{i=1}^{n} E[g_{2n}(i, x) - E g_{2n}(i, x)]^2 - \gamma \sum_{i=1}^{n} \left( \frac{x_i - x_{i-1}}{a_n} \right)^2 \int_{x_{i-1}}^{x_i} a_n^{-1} k \left( \frac{x-z}{a_n} \right) dz \right| = 0((na_n)^{-3} + (na_n)^{-\beta-2}),
\end{equation}

and hence assertion (2.5) follows by an application of Liapunov's theorem. □

Remarks. (1) If $g \in \text{Lip}(\alpha)$ and $\frac{1+2\alpha}{n} + \frac{(2\alpha-1)}{a_n} a_n + 0$ as $n \to \infty$,

then, according to the argument of Theorem 2 of Cheng and Lin (1979),

\begin{equation}
(2.13) \quad \frac{||E g_{1n} - g||_{[a,b]}}{[\text{Var} g_{1n}(x)]^{1/2}} = \frac{0((na_n)^{1/2}(a_n^\alpha + n^{-\alpha}))}{[\text{Var} g_{1n}(x)]^{1/2}}
\end{equation}

for $0 < a \leq b < 1$, and hence, for all $x \in (0,1)$,
(2.14) \[
\frac{g_{1n}(x) - g(x)}{[\text{Var } g_{1n}(x)]^{\frac{1}{2}}} \xrightarrow{L} N(0, 1) \quad \text{as } n \to \infty.
\]

(2) The quantity \( \text{Var } g_{1n}(x) \) involves the unknown parameter \( \sigma^2 \). It can be shown that, if \( |k|_{[-L, L]} < \infty \), \( g \in \text{Lip}(\alpha) \) for \( \alpha > 0 \), \( na_n \to \infty \) as \( n \to \infty \), then

\[
(2.15) \quad n^{-1} \sum_{i=1}^{n} \left[ Y_i - g_{1n}(x_i) \right]^2 \xrightarrow{P} \sigma^2 \quad \text{as } n \to \infty.
\]

Thus, in addition to the conditions of Theorem 2.2, if \( g \in \text{Lip}(\alpha) \) for \( \alpha > 0 \) such that \( na_n + n^{-(2\alpha-1)}a_n \to 0 \) as \( n \to \infty \), and \( x_i - x_{i-1} = n^{-1} \) for all \( i \), then

\[
(2.16) \quad \frac{(na_n)^{\frac{1}{2}}[g_{1n}(x) - g(x)]}{\left\{ n^{-1} \sum_{i=1}^{n} \left[ Y_i - g_{1n}(x_i) \right]^2 \right\}^{\frac{1}{2}}} \xrightarrow{L} N(0, 1) \quad \text{as } n \to \infty.
\]

(3) Denote by \( \phi \) the standard normal distribution. Then, under the conditions of Theorem 2.2, we have

\[
(2.17) \quad \sup_{-\infty \leq t \leq \infty} \left| \mathbb{P} \left\{ \frac{g_{1n}(x) - Eg_{1n}(x)}{[\text{Var } g_{1n}(x)]^{\frac{1}{2}}} \leq t \right\} - \phi(t) \right| = O((na_n)^{-\frac{1}{2}})
\]

for \( i = 1, 2 \), and \( x \in (0, 1) \). This result states the rate of convergence in central limit theorem.

(4) The joint limiting distribution of \( (na_n)^{\frac{1}{2}}[g_{1n}(t_i) - Eg_{1n}(t_i)], \ldots, g_{1n}(t_p) - Eg_{1n}(t_p) \) is a \( p \)-variate normal distribution with mean vector \( \mathbf{0} \) and covariance matrix \( \sum = (\sigma_{ij}) \), where if \( x_i - x_{i-1} = n^{-1} \) for all \( i \) and \( na_n \to \infty \) as \( n \to \infty \), then
(2.18) \[ c_{ij} = \begin{cases} \sigma^2 \int k^2(z) dz & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \]

This is also true for the joint limiting distributions of \( g_{2n}(t_i) \) and \( g_{3n}(t_i) \), \( i = 1, \ldots, p \).

3. **Asymptotic distribution of the isotonic estimator.** Based on the initial estimator \( \hat{g}_{1n}(x) \), redefine

\[
\hat{g}_{1n}(x) = g_{1n}(t_{ni}) \quad \text{if } x \in [t_{ni}, t_{ni+1}), \ i = 1, \ldots, n - 1, \tag{3.1}
\]

where \( 0 < a \leq t_{n1} < \ldots < t_{nN} \leq b < 1 \) is a subdivision of \([a, b]\), with

\[
t_{ni} = i \cdot c \cdot n^{-r} \quad \text{and } N = n^r \tag{3.2}
\]

for some \( r > 0 \) and \( 0 < c < \infty \).

It is clear, from the argument of Theorem 2.2, that \( \hat{g}_{1n}(x) \) and \( g_{1n}(x) \) have the same limiting distribution; namely,

\[
\frac{(na_n)^{1/2} [\hat{g}_{1n}(x) - E \hat{g}_{1n}(x)]}{[\sigma^2 \int k^2(z) dz]^{1/2}} \overset{L}{\rightarrow} N(0, 1) \quad \text{as } n \rightarrow \infty. \tag{3.3}
\]

Furthermore, the rate of this convergence is also \( O((na_n)^{-1/2}) \).

The isotonic estimator \( g_{n}^{**} \) of \( g \) is given by (1.6) where \( w \) is a given weight function corresponding to the grid \( S_n = \{t_{ni}\} \) on \([a, b]\). Following the argument of Barlow and Van Zwet (1969), the limiting distribution of \( g_{n}^{**} \) is established in the next theorem.

**Theorem 3.1.** Assume the following conditions:

(i) \( g' \) exists and is nonnegative for \( t \in [a, b] \) with \( g'(t) \geq \varepsilon \) in a neighborhood of \( x \) for some \( \varepsilon > 0 \);
(ii) the weight function \( w \) is bounded by \( b_1 \) and \( b_2 \), \( 0 < b_1 \leq b_2 < \infty \), for all \( n \);

(iii) \( k \in \text{Lip}(1) \) and \( \gamma \equiv \mathbb{E}|e|^3 < \infty \); and

(iv) \( na_n^3 \to 0 \) as \( n \to \infty \) and \( r \) is so chosen that \( n^{-r}(na_n)^{1/4} = n^{\delta} \) for some \( \delta > 0 \).

Then

\[
\frac{(na_n)^{1/2}[g^{**}(x) - g(x)]}{\left\{n^{-1} \sum_{i=1}^{n} [Y_i - g_{ln}(x_i)]^2 k^2(z)dz\right\}^{1/2}} \xrightarrow{L} N(0, 1) \text{ as } n \to \infty.
\]

**Proof.** It suffices to show that

\[
\lim_{n \to \infty} P[g^{**}(x) = \hat{e}_{ln}^-(x)] = 0.
\]

But, according to the construction of \( g^{**}(x) \), this is equivalent to showing

\[
\lim_{n \to \infty} P \left[ \exists \, m > 1 \text{ such that } \sum_{i=1}^{m+1} \frac{\hat{e}_{ln}(t_{nj})w(t_{nj})}{\sum_{j=i}^{m+1} w(t_{nj})} < \hat{e}_{ln}(t_{ni}) \right] = 0
\]

and

\[
\lim_{n \to \infty} P \left[ \exists \, m > 1 \text{ such that } \sum_{i=1-m}^{1} \frac{\hat{e}_{ln}(t_{nj})w(t_{nj})}{\sum_{j=i-m}^{1} w(t_{nj})} > \hat{e}_{ln}(t_{ni}) \right] = 0
\]

where \( t_{ni} \leq x < t_{n i+1} \). We will only show (3.6); the same argument applies to show (3.7). To establish (3.6), it suffices to show that

\[
\lim_{n \to \infty} \sum_{m=1}^{n^{r-1}} P \left\{ \sum_{j=i}^{m+1} [\hat{e}_{ln}(t_{nj}) - \hat{e}_{ln}(t_{ni})]w(t_{nj}) < 0 \right\} = 0.
\]

To this end, define
(3.9) \[ T_j = T_{nj} = \hat{\varepsilon}_{ln}(t_{n \ i+j}) - [E_{ln}(t_{n \ i+j}) - g(t_{n \ i+j})] \]
and

(3.10) \[ \theta_j = g(t_{n \ i+j}) \quad \text{for} \ j = 0, \ldots, m. \]

Then it is clear that

(3.11) \[ (n_{\alpha n})^{1/2}(T_0 - \theta_0, \ldots, T_m - \theta_m) \xrightarrow{L} N_{m+1}(0, \Lambda) \quad \text{as} \ n \to \infty, \]

where \( \Lambda = (\lambda_{ij}) \) with \( \lambda_{ij} = \sigma^2 \int k^2(z)dz \) if \( i = j = 0, \ldots, m; = 0, \text{otherwise} \).

Now define the function

(3.12) \[ h(T_0, \ldots, T_m) = \sum_{j=1}^{m+1} (T_{j-1} - T_0)w(t_{nj}). \]

Then

(3.13) \[ h(\theta_0, \ldots, \theta_m) = \sum_{j=1}^{m+1} (\theta_{j-1} - \theta_0)w(t_{nj}) \]

and

(3.14) \[ (n_{\alpha n})^{1/2}[h(T_0, \ldots, T_m) - h(\theta_0, \ldots, \theta_m)] \xrightarrow{L} N[0, \nu_m^2(\theta)] \]

as \( n \to \infty \), where

(3.15) \[ \nu_m^2(\theta) = \sum_{i=1}^{m} \left( \frac{\partial h}{\partial \theta_i} \right)^2 \sigma^2 \int k^2(z)dz. \]

Thus

(3.16) \[ \left| P\left\{ \sum_{j=1}^{m+1} [\hat{\varepsilon}_{ln}(t_{nj}) - \hat{\varepsilon}_{ln}(t_{ni})]w(t_{nj}) < 0 \right\} \right| \\
\quad - \phi[-(n_{\alpha n})^{1/2}h(\theta_0, \ldots, \theta_m)/\nu_m(\theta)] \\
\quad = 0((n_{\alpha n})^{-1/2}), \quad \text{for all} \ m. \]

Note that
\begin{equation}
V_m^2(\theta) = \sigma^2 \int k^2(z) dz \sum_{j=0}^{m} w(t_{i+j})^2 = o(m^2),
\end{equation}

where we have set

\begin{equation*}
w^*(t_{i+j}) = \begin{cases} 
w(t_{i+j}) & \text{if } j \neq 0 \\
\sum_{j=1}^{m} w(t_{i+j}) & \text{if } j = 0.
\end{cases}
\end{equation*}

Also note that

\begin{equation*}
h(\theta_0, \ldots, \theta_m) = \sum_{j=1}^{m+1} (t_{nj} - t_{ni})g'(t_{nj})w(t_{nj})
\end{equation*}

where $t_{nj} \in (t_{ni}, t_{nj}), j = 1, \ldots, m+1$. Thus

\begin{equation}
(na_n)^{-1/2} h(\theta_0, \ldots, \theta_m)
\end{equation}

\begin{equation*}
= (na_n)^{-1/2} \sum_{j=1}^{m+1} (t_{nj} - t_{ni})g'(t_{nj})w(t_{nj})
\end{equation*}

\begin{equation*}
\geq c \epsilon b_1 n^{-r}(na_n)^{1/2} \text{ for } n \text{ sufficiently large and for each } m,
\end{equation*}

by (3.2) and condition (ii). Therefore, in view of (3.17),

\begin{equation}
\frac{(na_n)^{-1/2} h(\theta_0, \ldots, \theta_m)}{V_m^{1/2}(\theta)} \geq \frac{c \epsilon b_1 n^{-r}(na_n)^{1/2}}{c'm}, \quad 0 < c' < \infty.
\end{equation}

The right hand side of (3.19) tends to $\infty$ as $n \to \infty$ for each $m$, by condition (iv). Now utilizing the approximation to the standard normal distribution given in Feller (1961), p. 166, namely

\begin{equation}
(-x^{-1} + x^{-3})\phi(x) < \phi(x) < -x^{-1}\phi(x) \text{ for all } x < 0,
\end{equation}

where $\phi(x)$ is the standard normal density, we have
\[
\sum_{m=1}^{n^r-1} \phi[-(n a_n)^{-1} h(\theta_1, \ldots, \theta_m)/v_m(\theta)]
\]

\[
< \sum_{m=1}^{n^r-1} \frac{v_m(\theta)}{(2\pi n a_n)^{-1} h(\theta_1, \ldots, \theta_m)} \exp[-i n a_n h^2(\theta_1, \ldots, \theta_m)/v_m^2(\theta)].
\]

Consequently,

\[
\lim_{n \to \infty} \sum_{m=1}^{n^r-1} \phi[-(n a_n)^{-1} h(\theta_1, \ldots, \theta_m)/v_m(\theta)] = 0,
\]

and hence

\[
\sum_{m=1}^{n^r-1} \sum_{j=1}^{m+i} \{ \hat{g}_{1m}(t_{n1}) - \hat{g}_{1m}(t_{ni}) \} w(t_{nj}) < 0,
\]

\[
= \sum_{m=1}^{n^r-1} \phi[-(n a_n)^{-1} h(\theta_1, \ldots, \theta_m)/v_m(\theta)] + (n^r - 1) o(1)(n a_n)^{-1} \]

\[
\to 0 \text{ as } n \to \infty, \text{ uniformly in } i,
\]

by (3.23) and condition (iv). This establishes (3.8) and thus completes the proof of the theorem. \[\square\]
REFERENCES


Nonparametric Estimation of a Regression Function: Limiting Distribution

Kuang-Fu Cheng and Pi-Erh Lin

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Office of Naval Research
Statistics & Probability Program
Arlington, Virginia 22217

Security Classification of this Page

REPORT DOCUMENTATION PAGE

1. REPORT NUMBERS
M518
ONR 148

4. TITLE
Nonparametric Estimation of a Regression Function: Limiting Distribution

5. TYPE OF REPORT & PERIOD COVERED
Technical Report

6. PERFORMING ORG. REPORT NUMBER
FSU Statistics Report M518

7. AUTHOR(s)
Kuang-Fu Cheng and Pi-Erh Lin

8. CONTRACT OR GRANT NUMBER(s)
ONR No. N00014-76-C-0608

9. PERFORMING ORGANIZATION NAME & ADDRESS
The Florida State University
Department of Statistics
Tallahassee, Florida 32306

10. PROGRAM ELEMENT, PROJECT, TASK AREA AND WORK UNIT NOS.

11. CONTROLLING OFFICE NAME & ADDRESS
Office of Naval Research
Statistics & Probability Program
Arlington, Virginia 22217

12. REPORT DATE
September, 1979

13. NUMBER OF PAGES
14

14. MONITORING AGENCY NAME & ADDRESS
(if different from Controlling Office)

15. SECURITY CLASS (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING
SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release: distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS
Asymptotic normality; Berry Esséen bound; isotonic; kernel function; Liapunov's theorem; and Lipschitz

20. ABSTRACT

Consider the regression model \( Y_i = g(x_i) + e_i, \ i = 1, \ldots, n \), where \( g \) is an unknown function defined on \([0, 1]\), \( 0 = x_0 < x_1 < \ldots < x_n \leq 1 \) are chosen so that \( \max_{1 \leq i \leq n} (x_i - x_{i-1}) = o(n^{-1}) \), and where \( \{e_i\} \) are i.i.d. with \( \text{E} e_i = 0 \) and \( \text{Var} e_i = \sigma^2 \).

In a previous paper, Cheng and Lin (1979) study three estimators of \( g \), namely, \( g_{1n} \) of Cheng and Lin (1979), \( g_{2n} \) of Clark (1977), and \( g_{3n} \) of Priestley and Chao (1972). Consistent results are established and rates of strong uniform convergence are obtained. In the current investigation the limiting distribution of \( g_{1n} \), \( i = 1, 2, 3 \), and that of the isotonic estimator \( g^{**} \) are considered.