

AD-A076 378

NORTH CAROLINA UNIV AT CHAPEL HILL INST OF STATISTICS

F/G 12/1

SMOOTH NONPARAMETRIC REGRESSION ANALYSIS.(U)

SEP 79 6 J JOHNSTON

N00014-75-C-0809

UNCLASSIFIED

MIMEO SER-1253

NL

1 OF 1

AD
A076378

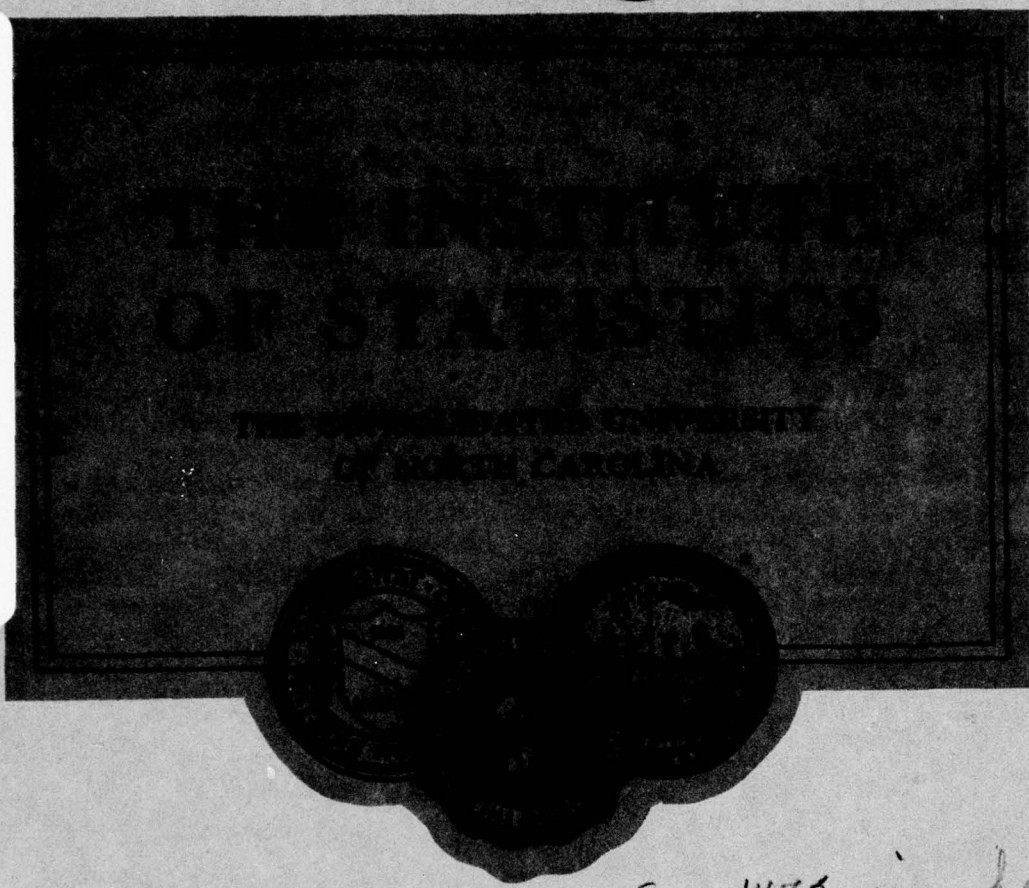
END
DATE
FILMED
12-79
DDC

92

LEVEL

#

AD A 076378



See 1473 in back

DDC FILE COPY

Smooth Nonparametric Regression Analysis

by

Gordon J. Johnston

University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series #1253

September 1979

DDC
RECEIVED
NOV 9 1979
B

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

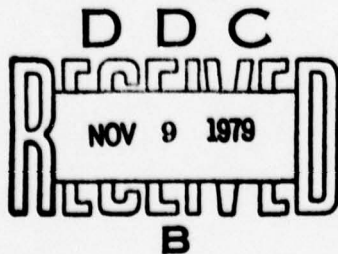
79 11 08 070

SMOOTH NONPARAMETRIC REGRESSION ANALYSIS

Gordon J. Johnston^{*}

Department of Statistics
University of North Carolina, Chapel Hill

^{*} This research was supported in part by the Air Force Office of Scientific Research under Grant AFOSR 75-2796 and by the Office of Naval Research under Grant N00014-75-C-0809.



DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

TABLE OF CONTENTS

CHAPTER 1. Introduction, Summary and Related Work.

1.1 Introduction-----	1
1.2 Summary-----	3
1.3 Related Work-----	4

CHAPTER 2. Consistency, Normality.

2.1 δ -Function Sequences-----	11
2.2 Nonparametric Density Function Estimation-----	16
2.3 Pointwise Consistency Properties of m_n and \bar{m}_n -----	18
2.4 Asymptotic Distribution of m_n -----	24
2.5 Asymptotic Distribution of \bar{m}_n -----	36
2.6 Mean Integrated Square Error-----	37

CHAPTER 3. Asymptotic Properties of Maximum Absolute Deviation.

3.1 Preliminaries-----	42
3.2 Maximum Absolute Deviation of m_n^* -----	47
3.3 Uniform Consistency of m_n and \bar{m}_n -----	71

CHAPTER 4. An Example, Further Research.

4.1 The Estimators m_n and \bar{m}_n -----	74
4.2 An Example-----	75
4.3 Further Research-----	76

FIGURES-----	78
--------------	----

BIBLIOGRAPHY-----	85
-------------------	----

ACCESSION for		
NTIS	White Section	<input checked="" type="checkbox"/>
DDC	Buff Section	<input type="checkbox"/>
UNANNOUNCED		<input type="checkbox"/>
JUSTIFICATION		
BY		
DISTRIBUTION/AVAILABILITY CODES		
Dist.	AVAIL.	and/or SPECIAL
A		

CHAPTER 1. INTRODUCTION, SUMMARY AND RELATED WORK

1.1 Introduction

Regression analysis is concerned with the study of the relationship between a response variable Y and a set of predictor variables $\underline{X} = (X_1, X_2, \dots, X_p)$. An important aspect of regression analysis is the estimation of the regression function, i.e. of the conditional expectation of Y , given \underline{X} . In classical regression analysis, the functional form of the regression function is assumed to be known up to a finite set of unknown parameters, which may be estimated from data.

If no such prior knowledge of the regression function exists, then classical methods do not apply. However, in this case, it may still be desirable to obtain an estimate of the regression function, either for direct analysis or to establish a plausible model for use in the classical regression analysis mentioned above.

Thus there is a need for regression analysis methods which do not assume a specific mathematical form for the regression function, i.e. nonparametric methods.

In this study, a type of nonparametric estimator of the regression function $m(x) = E[Y|X=x]$ will be investigated, where (X,Y) is a bivariate random vector.

Let X and Y be random variables defined on a probability space (Ω, \mathcal{F}, P) with $E|Y| < \infty$. Denote the marginal distribution function of X by F . Then the regression function $m(x)$ is defined by

$m(x) = E[Y|X=x]$, i.e. the (unique a.e. (dF)) Borel measurable function m satisfying

$$(1.1.1) \quad \int_{X^{-1}B} Y dP = \int_B m(x) dF(x)$$

for all Borel sets B . If X and Y have a joint density function f , then it follows that

$$(1.1.2) \quad m(x) = \begin{cases} \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{f(x)} & \text{if } f(x) > 0 \\ 0 & \text{if } f(x) = 0 \end{cases}$$

is a version of the regression function, where f denotes the marginal density of X . Motivated by (1.1.2) and previous work on estimation of density functions by δ -function sequences, Watson (1964) suggested an estimator of $m(x)$ of the form

$$(1.1.3) \quad m_n(x) = \frac{(1/n) \sum_{i=1}^n Y_i \delta_n(x - X_i)}{(1/n) \sum_{j=1}^n \delta_n(x - X_j)}$$

where $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ are independent observations on (X, Y) and $\{\delta_n(x)\}$ is a sequence of weighting functions called a δ -function sequence. The estimator $m_n(x)$ defined in (1.1.3) will be investigated here. By rewriting (1.1.3) as

$$m_n(x) = \sum_{i=1}^n Y_i \left[\frac{\delta_n(x - X_i)}{\sum_{j=1}^n \delta_n(x - X_j)} \right],$$

we have the intuitively appealing interpretation of $m_n(x)$ as a weighted average of the Y -observations, with the weights depending on x through

$\delta_n(x)$. Also, if one desires a smooth estimate of $m(x)$, this can be achieved through the choice of $\delta_n(x)$.

In certain situations, the marginal density f of X is known. For example, suppose in an experiment, we are able to fix the level of the predictor variable X , but we wish to randomize X to reduce sampling bias. Then we would choose X randomly according to a known density f . This situation also arises in certain optimization problems, where the value of the function to be optimized can only be determined up to a random error term (see Devroye (1978)). Since the denominator of (1.1.3) is intended to estimate f , a reasonable way to use the knowledge of f might be to use the modified estimator

$$\bar{m}_n(x) = \frac{(1/n) \sum_{i=1}^n Y_i \delta_n(x - X_i)}{f(x)}$$

We provide some preliminary comparisons of the estimators \bar{m}_n and m_n in the known density case.

1.2 Summary

Since we will assume a specific mathematical form for neither the regression function m nor the underlying probability distribution of (X, Y) , we could not reasonably expect to obtain small sample results for the estimators in question. Thus we shall concern ourselves here almost exclusively with asymptotic results, as the sample size grows larger.

In Chapter 2, we rigorously establish (weak) pointwise consistency of $m_n(x)$, which was proved heuristically by Watson (1964). Asymptotic joint normality of $m_n(x)$, taken at a finite number of points, is demonstrated. In this last result, we significantly weaken a condition of

Schuster (1972) on the δ -function sequence used, at the expense of some mild additional regularity conditions. In the known density case, we establish consistency and asymptotic normality for \bar{m}_n , so as to provide a comparison with the asymptotic normality and consistency results for m_n . We consider the mean integrated square error (MISE) of the numerator of the estimator m_n . An explicit expression for the Fourier transform of the δ -function which minimizes this MISE for each sample size n is derived, much as Watson and Leadbetter (1963) did for density function estimators.

In Chapter 3, we consider the numerator of m_n , and show that the supremum, taken over a finite interval, if properly centered and normalized, converges in distribution to a random variable having an extreme value type distribution. This result is then applied to establish uniform (weak) consistency of m_n , with an associated rate of convergence.

In Chapter 4, we give some examples of calculations of the estimators m_n and \bar{m}_n from simulated data.

1.3 Related Work

Estimators of the form (1.1.3) of the regression function and several other types of nonparametric estimators of the regression function have recently received attention in the literature. Here we survey the recent literature on this subject.

1.3.1 Kernel Type Estimators

Several authors have considered estimators of the form (1.1.3) when the δ -function sequence is of kernel type. Kernel type δ -function sequences (defined rigorously in Lemma 2.1.2) are of the form

$$\delta_n(x) = \epsilon_n^{-1} K(x/\epsilon_n),$$

where K is, e.g. a probability density function and ϵ_n is a positive real

sequence with $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

Schuster (1972) considers the asymptotic normality of this type of estimator. Because of the close relation between Schuster's work and work presented in this dissertation, we defer discussion until Section 2.4.

Schuster and Yokowitz (1978) consider a global error criterion for this type of estimator, and for its derivatives as estimators of the derivatives of the regression function. Let $g^{(r)}$ denote the r -th derivative of the function g . Schuster and Yokowitz give conditions under which, for any $\epsilon > 0$ and n sufficiently large,

$$(1.3.1) \quad P\left[\sup_{a \leq x \leq b} |m_n^{(N)}(x) - m^{(N)}(x)| \geq \epsilon\right] \\ \leq C/(n\epsilon_n^{2N+2}\epsilon^2),$$

where N is a positive integer, $[a, b]$ is a closed, bounded interval and C is a constant not depending on n . If $\{\epsilon_n\}$ is such that $n\epsilon_n^{2N+2} \rightarrow \infty$ as $n \rightarrow \infty$, then (1.3.1) implies that

$$\sup_{a \leq x \leq b} |m_n^{(N)}(x) - m^{(N)}(x)| \xrightarrow{P} 0,$$

as $n \rightarrow \infty$, so that this result is a type of uniform consistency result. It would be of interest to determine a rate of convergence to be associated with the result, i.e., a positive, real sequence $\{b_n\}$ with $b_n \rightarrow \infty$ such that

$$b_n \sup |m_n^{(N)}(x) - m^{(N)}(x)| \xrightarrow{P} 0.$$

This question is addressed in Chapter 3 of this dissertation for the case $N = 0$. Schuster and Yakowitz also consider the case where the X

variable is non-stochastic, i.e., we have $\{F(\cdot; x), x \in [0, 1]\}$ as a family of probability density functions and the object is to estimate

$$w(x) = \int y F(y; x) dy$$

on the basis of an independent sample $Y_i, i = 1, \dots, n$ where Y_i has density $f(\cdot; x_i)$. They give conditions under which a result similar to (1.3.1) holds for the so-called Priestly-Chao estimator

$$(1.3.2) \quad w_n(x) = \epsilon_n^{-1} \sum_{i=1}^n Y_i (x_i - x_{i-1}) K((x - x_i)/\epsilon_n)$$

of $w(x)$ (see Priestly and Chao (1972)).

In the non-stochastic X variable case as described above, Benedetti (1974) shows that both the Watson estimator and the Priestly-Chao estimator are (weakly) consistent and asymptotically normal for appropriate values of x , but he points out some computational advantages of the Watson estimator over the Priestly-Chao.

Konakov (1975) considers a quadratic deviation error criterion for the Watson estimator with kernel type δ -sequences. Define the quadratic deviation to be

$$T_n = n\epsilon_n \int [m_n(x) - m(x)]^2 f_n^2(x) p(x) dx$$

where f_n is a kernel type estimator of the marginal density f of X and p is a bounded integrable weight function. Konakov gives conditions under which T_n , if properly normalized and centered, is asymptotically normal. We do not consider quadratic deviation in this dissertation.

1.3.2 Nearest Neighbor Type Estimators.

Watson (1964) proposed estimating $m(x)$ with the average of the Y values corresponding to the k X values nearest to x , where k is some

integer. This type of estimator is called the k nearest neighbor estimator of the regression function. Earlier work had been done on the classification problem and on estimation of a probability density function using nearest neighbor techniques. (See Fix and Hodges (1951), Cover and Hart (1967), Cover (1968), and Loftgaarden and Quesenberry (1965) for work in these areas.)

Let $k(n)$ be an integer depending on the sample size n and denote by $I_{k(n)}(x)$ the smallest open interval centered at x containing no less than k of the X -observations. Then the k -nearest neighbor estimator \tilde{m}_n , can be written as

$$(1.3.3) \quad \tilde{m}_n(x) = k^{-1} \sum_{\{i: X_i \in I_{k(n)}(x)\}} Y_i .$$

Stone (1977) points out that $\tilde{m}_n(x)$ may be a discontinuous function, and that in some cases, smoothness is a desirable property in a regression function estimator. Lai (1977) proposes a modification of the k nearest neighbor estimator which can have the desired smoothness property. This estimator is very similar to the Watson estimator (1.1.3) with kernel type δ -function sequence. Let W be a probability density with $W(x) = 0$ for $|x| > 1$. Then Lai's estimator is defined by

$$(1.3.4) \quad \hat{m}_n(x) = \frac{\sum_{i=1}^n Y_i W((x-X_i)/R_{k(n)}(x))}{\sum_{i=1}^n W((x-X_i)/R_{k(n)}(x))} ,$$

where $R_{k(n)}$ is the radius of the interval $I_{k(n)}$. This estimator reduces to the form (1.3.3) when $W(x) = 1/2$ for $|x| \leq 1$. Lai proves the following.

1.3.1 Theorem. Assume W is continuous a.e., bounded and $W(x) = 0$ for $|x| > 1$. If there exists an open set U_0 in \mathbb{R} on which

- i) $f(x)$ is continuous, bounded, and $f(x) > 0$,
- ii) $E(|Y| | X=x)$ and $E(\max(Y, 0) | X=x)$ are continuous functions of x ,

- iii) $\limsup_{M \rightarrow \infty} \sup_{x \in U_0} E(|Y| \cdot I_{\{|Y| \geq M\}}(Y) | X=x) = 0$, and if $EY^2 < \infty$ and
- iv) $k(n)/n \rightarrow 0$ and $k(n)/\sqrt{n} \rightarrow \infty$,

then

$$\sup_{z \in A} |\hat{m}_n(z) - m(z)| \rightarrow 0$$

in probability for any compact set $A \subset U_0$. □

A similar result is proved for the estimator (1.1.), with A an interval, in Chapter 3 of this dissertation. There, more regularity conditions are applied to obtain an associated rate of convergence.

Stone (1977) considers the following type of nonparametric regression function estimator:

$$(1.3.5) \quad \hat{m}_n(x) = \sum_{i=1}^n W_{ni}(x) Y_i$$

where $W_{ni}(x) = W_{ni}(x; X_1, \dots, X_n)$ is a weight function. This estimator reduces to the nearest neighbor, modified nearest neighbor and δ -function type estimators discussed above for appropriate choices of the function W_{ni} . Stone gives general conditions on the weight functions W_{ni} for \hat{m}_n to be consistent in L^r , i.e., for

$$E|\hat{m}_n(x) - m(x)|^r \rightarrow 0$$

whenever $E|Y|^r < \infty$. Stone's work applies to give minimal conditions for this type of consistency for some types of estimators, e.g., if $k(n) \rightarrow \infty$ and $k(n)/n \rightarrow 0$, then the k nearest neighbor estimator is consistent in L^r . Stone, however, points out that it is not clear from his results when an estimator of the Watson type is consistent in L^r .

1.3.3 Potential Function Methods.

In this method, introduced by Aizeman, Braverman, and Rozonoer (1970), the regression function m is assumed to belong to a Hilbert space H and have the representation $m(x) = \sum_i C_i \phi_i(x)$, where $\{\phi_i\}$ is a complete system of functions of H . The estimator m_n is calculated recursively by the formula

$$m_n(x) = m_{n-1}(x) + r_n K(x, X_n),$$

where

$$r_n = \gamma_n [Y_n - m_{n-1}(X_n)],$$

and K is a "potential function" of the form

$$K(x, y) = \sum_i \lambda_i^2 \phi_i(x) \phi_i(y),$$

and $\{\gamma_n\}$ is a sequence of real numbers, and m_0 is chosen arbitrarily.

We have the following type of consistency for this set-up. Suppose

$\sum_i (C_i / \lambda_i)^2 < \infty$, $\sum_i \gamma_i = \infty$, $\sum_i \gamma_i^2 < \infty$. Then

$$\int [m_n(x) - m(x)]^2 f(x) dx \rightarrow 0$$

in probability as $n \rightarrow \infty$.

Fisher and Yokowitz (1976) obtain more general results for this type of estimation, but for a more complicated error criterion.

1.3.4 Estimates Based on Ranks.

Let $X_{n1} \leq X_{n2} \leq \dots \leq X_{nn}$ denote the ordered X values and define the concomitant of $X_{ni} = X_j$ to be $Y_{ni} = Y_j$. The set Y_{ni} , $i = 1, \dots, n$, are sometimes called the induced order statistics. Yang (1977) proposes

the following estimator of $m(x)$ based on concomitants:

$$M_n(x) = (n\epsilon_n)^{-1} \sum_{i=1}^n K((\frac{i}{n} - F_n(x))/\epsilon_n) Y_{ni}$$

where $\epsilon_n^{-1}K(x/\epsilon_n)$ is a kernel type δ -function sequence and F_n is the empirical distribution function of the X values. Yang gives conditions under which M_n is (weakly) consistent and asymptotically normal at appropriate points x .

Bhattacharya (1976) discusses estimation of a function related to the regression function based on concomitants. Let F denote the marginal distribution function of X and define

$$h(t) = m \circ F^{-1}(t); H(t) = \int_0^t h(s) ds, \quad 0 \leq t \leq 1.$$

Natural estimators of H are

$$H_n(t) = n^{-1} \sum_{i=1}^{[nt]} Y_{ni}$$

and

$$H_n^*(t) = n^{-1} \sum_{\{i: F(X_{ni}) \leq t\}} Y_{ni}$$

if F is known. Bhattacharya obtains weak convergence in $D[0,1]$ results for these estimators and applies them to estimation and hypothesis testing problems.

2. CONSISTENCY, NORMALITY

2.1 δ -function Sequences

A certain class of δ -function sequences was suggested originally by Rosenblatt (1956) and, under slightly weaker conditions, by Parzen (1962) for use in probability density function estimation. Leadbetter (1963) and Watson and Leadbetter (1964) introduced a more general notion of δ -function sequences, and our approach throughout this study will be to obtain results for the more general type of δ -functions whenever possible.

The following (2.1.1 - 2.1.4) is due to Leadbetter (1963).

2.1.1 Definition. A sequence of integrable functions $\{\delta_n(x)\}$ is called a δ -function sequence if it satisfies the following set of conditions (integrals with no limits of integration are meant to extend over the entire real line):

$$C1. \quad \int |\delta_n(x)| dx < A \quad \text{for all } n \text{ and some fixed } A,$$

$$C2. \quad \int \delta_n(x) dx = 1 \quad \text{for all } n,$$

$$C3. \quad \delta_n(x) \rightarrow 0 \text{ uniformly on } |x| \geq \lambda \text{ for any fixed } \lambda > 0,$$

$$C4. \quad \int_{|x| \geq \lambda} \delta_n(x) dx \rightarrow 0 \quad \text{for any fixed } \lambda > 0$$

□

The next lemma describes the type of δ -function sequence used by Rosenblatt (1956) and Parzen (1962), although the conditions on the function K are slightly different. This type of δ -function sequence will be

referred to as "kernel type" and K as a "kernel function."

2.1.2 Lemma. Let $\{\epsilon_n\}$ be a sequence of non-zero constants with $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$ and let K be an integrable function such that $\int K(x)dx = 1$ and $K(x) = o(x^{-1})$ as $|x| \rightarrow \infty$. Then $\{\epsilon_n^{-1}K(x/\epsilon_n)\}$ is a δ -function sequence. \square

The following lemma demonstrates the similarity of δ -function sequences as defined above and the Dirac δ -function.

2.1.3 Lemma. If $g(x)$ is integrable and continuous at $x = 0$ and $\{\delta_n\}$ is a δ -function sequence, then $g(x)\delta_n(x)$ is integrable for each n and $\int g(x)\delta_n(x)dx \rightarrow g(0)$ as $n \rightarrow \infty$. \square

2.1.4 Lemma. Let $\{\delta_n(x)\}$ be a δ -function sequence such that, for $p \geq 1$, $\alpha_n(p) = \int |\delta_n(u)|^p du < \infty$ for each n . Then $\alpha_n(p) \rightarrow \infty$ and

$$\{\delta_{n,p}(x)\} = \{|\delta_n(x)|^p / \alpha_n(p)\}$$

is a δ -function sequence for $p \geq 1$. \square

Rosenblatt (1971) states the following lemma, which gives a rate of convergence for Lemma 2.1.3, when the δ -function sequence is of kernel type. We include a proof for completeness.

2.1.5 Lemma. Suppose g is an integrable function with bounded, continuous 1st and 2nd derivatives. Let

$$\{\delta_n(x)\} = \{\epsilon_n^{-1}K(x/\epsilon_n)\}$$

be a δ -function sequence of kernel type with

$$\int uK(u)du = 0 \quad \text{and} \quad \int |u^2K(u)|du < \infty .$$

Then

$$\epsilon_n^{-1} \int K((x-u)/\epsilon_n)g(u)du = g(x) + o(\epsilon_n^2) ,$$

where the sequence represented by o does not depend on x .

Proof. Write

$$\epsilon_n^{-1} \int K((x-u)/\epsilon_n)g(u)du = \int K(y)g(x-\epsilon_n y)dy$$

and

$$\begin{aligned} g(x-\epsilon_n y) \\ = g(x) - g'(x)\epsilon_n y + g''(\tau)\epsilon_n^2 y^2/2 \end{aligned}$$

where $\tau = \tau_n(x,y)$ is between x and $x-\epsilon_n y$. Thus

$$\begin{aligned} & \epsilon_n^{-1} \int K((x-u)/\epsilon_n)g(u)du \\ &= g(x) \int K(y)dy - g'(x)\epsilon_n \int yK(y)dy \\ & \quad + (\epsilon_n^2/2) \int g''(\tau)y^2K(y)dy \end{aligned}$$

and hence

$$\begin{aligned} & |\epsilon_n^{-1} \int K((x-u)/\epsilon_n)g(u)du - g(x)| \\ & \leq \epsilon_n^2 \sup_t |g''(t)|/2 \int |y^2K(y)|dy \end{aligned}$$

since $\int K(u)du = 1$ and $\int uK(u)du = 0$. The conclusion follows from the last inequality. \square

The following lemmas will be useful in the sequel.

2.1.6 Lemma. Let $\{\delta_n\}$ be a δ -function sequence and g an integrable function. If g is continuous at x and y and $x \neq y$, then

$\delta_n(x - \cdot)\delta_n(y - \cdot)g(\cdot)$ is integrable and

$$\int \delta_n(x-u)\delta_n(y-u)g(u)du \rightarrow 0$$

as $n \rightarrow \infty$.

Proof. For convenience, assume $x < y$. By Lemma 2.1.3, $\delta_n(y-u)g(u)$ is an integrable function for each n , as is $\delta_n(x-u)g(u)$. Choose λ so that $x < \lambda < y$. Then

$$\begin{aligned} & \left| \int \delta_n(x-u)\delta_n(y-u)g(u)du \right| \\ & \leq \int_{-\infty}^{\lambda} |\delta_n(x-u)\delta_n(y-u)g(u)| du \\ & \quad + \int_{\lambda}^{\infty} |\delta_n(x-u)\delta_n(y-u)g(u)| du \\ & \leq \sup_{u \leq \lambda} |\delta_n(y-u)| \int |\delta_n(x-u)g(u)| du \\ & \quad + \sup_{u \geq \lambda} |\delta_n(x-u)| \int |\delta_n(y-u)g(u)| du . \end{aligned}$$

Now $\sup_{u \leq \lambda} |\delta_n(y-u)|$ and $\sup_{u \geq \lambda} |\delta_n(x-u)|$ converge to zero by C3 of Definition

2.1.1. Further, $\int |\delta_n(x-u)g(u)| du < \infty$ for each n by the preceding remark, and, in fact, by Lemmas 2.1.3 and 2.1.4,

$$(\alpha_n(1))^{-1} \int |\delta_n(x-u)g(u)| du \rightarrow |g(x)|$$

and

$$\alpha_n(1) = \int |\delta_n(u)| du < A$$

for some constant A . Thus $\int |\delta_n(x-u)g(u)|du$ is a bounded sequence, and we have

$$\sup_{u \leq \lambda} |\delta_n(y-u)| \int |\delta_n(x-u)g(u)|du \rightarrow 0$$

as $n \rightarrow \infty$. The same argument applies when x and y are interchanged, and the conclusion follows. \square

2.1.7 Lemma. Let $\{\delta_n\}$ be a δ -function sequence such that δ_n is an even function for each n and g an integrable function such that g has both left and right hand limits at 0. Then $\delta_n(x)g(x)$ is an integrable function for each n and

$$\int \delta_n(x)g(x)dx \rightarrow (g(0+) + g(0-))/2$$

as $n \rightarrow \infty$.

Proof. Define

$$g^+(x) = \begin{cases} g(x), & x > 0 \\ g(0+), & x = 0 \\ g(-x), & x < 0 \end{cases}$$

$$g^-(x) = \begin{cases} g(-x), & x > 0 \\ g(0-), & x = 0 \\ g(x), & x < 0 \end{cases}.$$

Clearly, g^+ and g^- are even functions, continuous at 0. Further, they are both integrable functions, since, e.g.

$$\begin{aligned} \int |g^+(u)|du &= 2 \int_0^\infty |g^+(u)|du \\ &= 2 \int_0^\infty |g(u)|du < \infty. \end{aligned}$$

Thus, by Lemma 2.1.3,

$$\int \delta_n(x) g^+(x) dx \rightarrow g^+(0) = g(0+)$$

as $n \rightarrow \infty$. But

$$\begin{aligned} & \int |\delta_n(x) g(x)| dx \\ &= \int_0^\infty |\delta_n(x) g(x)| dx + \int_{-\infty}^0 |\delta_n(x) g(x)| dx \\ &= (1/2) \int |g^+(x) \delta_n(x)| dx + (1/2) \int |g^-(x) \delta_n(x)| dx < \infty \end{aligned}$$

by Lemma 2.1.3, so that $\delta_n(x)g(x)$ is integrable, and

$$\begin{aligned} & \int \delta_n(x) g(x) dx \\ &= \int_0^\infty \delta_n(x) g(x) dx + \int_{-\infty}^0 \delta_n(x) g(x) dx \\ &= (1/2) \int \delta_n(x) g^+(x) dx + (1/2) \int \delta_n(x) g^-(x) dx \\ &\rightarrow (g(0+) + g(0-))/2 \end{aligned}$$

by the preceding remark. □

2.2 Nonparametric Density Function Estimation.

Nonparametric methods of density function estimation have been studied in great detail (see, e.g., Wegman (1972a) and Wegman (1972b) for a survey and comparison of work in this area). Estimators of a density $f(x)$ of the form

$$(2.2.1) \quad f_n(x) = (1/n) \sum_{i=1}^n \delta_n(x - X_i)$$

are of particular interest here because of the reliance of our proposed regression function estimators on the same type of weighting functions δ_n , and because f_n , defined in (2.2.1) appears in the denominator of m_n as defined in equation (1.1.3).

Since X_i , $i = 1, \dots, n$ are i.i.d. with common density f , we have

$$Ef_n = \int \delta_n(x-u)f(u)du ,$$

which has $f(x)$ as its limiting value as $n \rightarrow \infty$, provided f is continuous at x , by Lemma 2.1.3. That is, f_n is an asymptotically unbiased estimator of f at continuity points of f . Further,

$$\begin{aligned} Ef_n^2(x) &= (1/n)^2 E\left\{ \sum_{i=1}^n \delta_n^2(x-X_i) + \sum_{i \neq j} \delta_n(x-X_i) \delta_n(x-X_j) \right\} \\ &= (1/n) \int \delta_n^2(x-u)f(u)du \\ &\quad + ((n-1)/n) \left[\int \delta_n(x-u)f(u)du \right]^2 , \end{aligned}$$

so that

$$\begin{aligned} \text{Var}[f_n(x)] &= (1/n) \int \delta_n^2(x-u)f(u)du \\ &\quad - (1/n) \left[\int \delta_n(x-u)f(u)du \right]^2 \end{aligned}$$

and we thus have, if $\alpha_n = \int \delta_n^2(u)du < \infty$ for each n , by Lemma 2.1.4,

$$(n/\alpha_n)\text{Var}[f_n(x)] \rightarrow f(x)$$

as $n \rightarrow \infty$ at continuity points x of f . The above calculations (which appear in Watson and Leadbetter (1964)) may be combined to give conditions under which the mean square error of f_n converges to zero, as the following lemma shows.

2.2.1 Lemma. Let $\{\delta_n\}$ be a δ -function sequence for which $\alpha_n = \int \delta_n^2(u) du < \infty$ for each n and $\alpha_n/n \rightarrow 0$ as $n \rightarrow \infty$. Let x be a continuity point of f . Then

$$E[f_n(x) - f(x)]^2 \rightarrow 0 \text{ as } n \rightarrow \infty.$$

□

2.2.2 Remark. By Chebychev's inequality,

$$P\{|f_n(x) - f(x)| > \epsilon\} \leq \epsilon^{-2} E[f_n(x) - f(x)]^2$$

for any $\epsilon > 0$. We thus have $f_n(x) \rightarrow f(x)$ in probability, provided the conditions of Lemma 2.2.1 are satisfied. That is, $f_n(x)$ is a weakly consistent estimator of $f(x)$ for appropriate δ -function sequences and points x .

The preceding discussion on density estimation will suffice for our discussion of pointwise consistency of our proposed regression estimators. We will include other pertinent results on density estimation as they are needed.

2.3 Pointwise Consistency Properties of m_n and \bar{m}_n

We begin our discussion by considering the numerator of the estimators m_n and \bar{m}_n defined in (1.1.3) and (1.1.4), respectively. Denote, for convenience, the numerator by m_n^* , i.e.,

$$m_n^*(x) = (1/n) \sum_{i=1}^n Y_i \delta_n(x - X_i).$$

Then we have the following.

2.3.1 Lemma. Let $\{\delta_n\}$ be a δ -function sequence for which $\alpha_n = \int \delta_n^2(u) du < \infty$ for each n . Suppose $EY^2 < \infty$ and x is a point of continuity of the functions $f(u)$, $m(u) = E[Y|X=u]$ and $s(u) = E[Y^2|X=u]$. Then

$$(i) \quad Em_n^*(x) \rightarrow m(x)f(x)$$

$$(ii) \quad (n/\alpha_n) \text{Var}[m_n^*(x)] \rightarrow s(x)f(x)$$

as $n \rightarrow \infty$.

Proof. We will use the following two well known properties of the regression function:

$$(2.3.1) \quad Eh(Y) = \int E[h(Y)|X=x]f(x)dx$$

for any function h and random variable Y such that $E|h(Y)| < \infty$,

$$(2.3.2) \quad E[g(X)h(Y)|X=x] = g(x)E[h(Y)|X=x]$$

for any functions g and h such that $E|g(X)h(Y)| < \infty$.

Since (X_i, Y_i) , $i = 1, \dots, n$ are i.i.d., we have

$$\begin{aligned} Em_n^*(x) &= EY\delta_n(x-X) \\ &= \int E[Y\delta_n(x-X)|X=u]f(u)du \end{aligned}$$

by (2.3.1),

$$= \int \delta_n(x-u)E[Y|X=u]f(u)du$$

by (2.3.2),

$$= \int \delta_n(x-u)m(u)f(u)du .$$

Now, by assumption, $m(u)f(u)$ is continuous at $u = x$. Thus (i) will follow from Lemma (2.1.3) if we demonstrate that $m(u)f(u)$ is an integrable function. To verify this, note that, by Jensen's inequality

$$|m(x)| = |E[Y|X=x]| \leq E[|Y||X=x] .$$

Thus

$$\int |m(u)f(u)| du \leq \int E[|Y||X=u]f(u)du = E|Y| < \infty$$

by assumption, and (i) follows.

For (ii), note

$$\begin{aligned} E[m_n^*(x)]^2 &= (1/n)^2 E\left\{ \sum_{i=1}^n [Y_i \delta_n(x-X_i)]^2 \right. \\ &\quad \left. + \sum_{i \neq j} Y_i \delta_n(x-X_i) Y_j \delta_n(x-X_j) \right\} \\ &= (1/n) \int s(u)f(u) \delta_n^2(x-u) du \\ &\quad + ((n-1)/n) \left[\int m(u)f(u) \delta_n(x-u) du \right]^2 , \end{aligned}$$

the last step following from (2.3.1) and (2.3.2), as used in the proof of (i). Thus

$$\begin{aligned} \text{Var}[m_n^*(x)] &= (1/n) \int s(u)f(u) \delta_n^2(x-u) du \\ &\quad - (1/n) \left[\int m(u)f(u) \delta_n(x-u) du \right]^2 , \end{aligned}$$

and, since $\alpha_n \rightarrow \infty$ and $\{\delta_n/\alpha_n\}$ is a δ -function sequence, we have

$$(n/\alpha_n) \text{Var}[m_n^*(x)] \rightarrow s(x)f(x) ,$$

as desired. □

We now use the preceding result to demonstrate the consistency of the estimators m_n and \bar{m}_n .

2.3.2 Theorem. Let $\{\delta_n\}$ be a δ -function sequence such that $\alpha_n = \int \delta_n^2(u) du < \infty$ for each n and $\alpha_n = o(n)$. Suppose $EY^2 < \infty$ and x is a continuity point of $f(u)$, $m(u)$ and $s(u)$, and that $f(x) > 0$. Then

- (i) $m_n(x) \xrightarrow{P} m(x)$
- (ii) $\bar{m}_n(x) \xrightarrow{P} m(x)$.

Proof. Since $\alpha_n/n \rightarrow 0$ by assumption, we have $\text{Var}[m_n^*(x)] \rightarrow 0$ by Lemma 2.3.1. Thus

$$E[m_n^*(x) - m(x)f(x)]^2 \rightarrow 0$$

and by applying Chebychev's inequality as in Remark 2.2.2, we see that $m_n^*(x) \xrightarrow{P} m(x)f(x)$. Since, by definition

$$\bar{m}_n(x) = m_n^*(x)/f(x),$$

and $f(x) > 0$, (ii) follows immediately.

For (i), write

$$\begin{aligned} |m_n - m| &= \left| \frac{m_n^* - mf}{f_n} + \frac{m(f - f_n)}{f_n} \right| \\ &\leq \left| \frac{m_n^* - mf}{f_n} \right| + |m| \cdot \left| \frac{f - f_n}{f_n} \right|, \end{aligned}$$

where we have suppressed the argument x for convenience. Now $m_n^* \xrightarrow{P} mf$ by (ii), and since $f_n(x) \xrightarrow{P} f(x) > 0$, we have

$$\frac{m_n^* - mf}{f_n} \xrightarrow{p} 0.$$

Similarly,

$$\frac{f - f_n}{f_n} \xrightarrow{p} 0,$$

and (i) follows. □

We have, then, that both the estimators m_n and \bar{m}_n are weakly consistent estimators of m at continuity points of s , m and f . The following corollary specializes the preceding theorem to kernel type δ -function sequences.

2.3.3 Corollary. Suppose that $\{\delta_n(x)\} = \{\epsilon_n^{-1}K(x/\epsilon_n)\}$ is a δ -function sequence of kernel type with $n\epsilon_n \rightarrow \infty$ as $n \rightarrow \infty$ and $\int K^2(u)du < \infty$. Assume the other conditions of Theorem 2.3.2 are satisfied. Then the conclusions of Theorem 2.3.2 hold.

Proof. We need only verify that $\alpha_n/n \rightarrow 0$. By definition

$$\alpha_n = \int \delta_n^2(u)du = \epsilon_n^{-2} \int K^2(x/\epsilon_n)dx = \epsilon_n^{-1} \int K^2(u)du.$$

Then

$$\alpha_n/n = (n\epsilon_n)^{-1} \int K^2(u)du \rightarrow 0$$

by assumption. □

We have so far been assuming that the density f of X is continuous and positive at points where we wish to estimate the regression function. An important case where these assumptions may not hold is when X is a bounded random variable, i.e. when f has bounded support, and we desire an

estimate at a boundary point of the support of f . We have the following result, which demonstrates that if $m(x+) = m(x)$, $m_n(x)$ is consistent but $\bar{m}_n(x)$ is not.

2.3.4 Theorem. Let $\{\delta_n\}$ be a δ -function sequence such that $\alpha_n < \infty$ for each n and $\alpha_n/n \rightarrow 0$. Suppose $EY^2 < \infty$ and x is a point such that f , m and s have left and right hand limits at x and $f(x+) = f(x) > 0$, $f(x-) = 0$. Then

$$(i) \quad \bar{m}_n(x) \rightarrow m(x+)/2$$

$$(ii) \quad m_n(x) \rightarrow m(x+) .$$

Proof. By Lemma 2.1.7 it follows that

$$Ef_n(x) \rightarrow f(x)/2 ,$$

$$Em_n^*(x) \rightarrow m(x+)f(x)/2 .$$

By an argument similar to the one used in the proof of Lemma 2.3.1, it follows that

$$f_n(x) \xrightarrow{P} f(x)/2 ,$$

$$m_n^*(x) \xrightarrow{P} m(x+)f(x)/2 .$$

Since

$$\bar{m}_n(x) = m_n^*(x)/f_n(x) ,$$

(i) follows. Since

$$m_n(x) = m_n^*(x)/f_n(x) ,$$

(ii) follows by an argument similar to the one used in Theorem 2.3.2.

□

This theorem demonstrates that the estimator \bar{m}_n displays an "end effect" at the boundaries of the range of X which m_n does not display. As we shall see in Chapter 4, this end effect represents a possible disadvantage for \bar{m}_n , depending on how m is defined at the boundaries of its support. We now turn our attention to asymptotic distributional properties of m_n and \bar{m}_n .

2.4 Asymptotic Distribution of m_n .

Nadaraya (1964) and Schuster (1972) have considered the asymptotic normality of the estimator m_n in terms of kernel type δ -function sequences. Nadaraya states that, if Y is a bounded random variable and $n\epsilon_n^2 \rightarrow \infty$ then $(n\epsilon_n)^{1/2}(m_n(x) - Em_n(x))$ has an asymptotically normal distribution with zero mean and variance $s(x) \int K^2(u)du/f(x)$, where

$$s(x) = E[Y^2|X=x] .$$

Schuster (1972) points out that this expression for the variance is incorrect and presents a result with the correct variance which at the same time removes the restriction that Y be bounded and centers at $m(x)$ instead of $Em_n(x)$. We state Schuster's result here for comparison with a new result which represents, in some respects, an improvement over Schuster's.

2.4.1 Theorem. Let $\{\epsilon_n^{-1}K(x/\epsilon_n)\}$ be a δ -function sequence satisfying the condition:

- (i) $K(u)$ and $uK(u)$ are bounded,
- (ii) $\int uK(u)du = 0$, $\int u^2K(u)du < \infty$,

$$(iii) \quad n\epsilon_n^3 \rightarrow \infty, \quad n\epsilon_n^5 \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Suppose x_1, x_2, \dots, x_p are distinct points with $f(x_i) > 0$, $i = 1, \dots, p$. Let $w(u) = m(u)f(u)$ and assume f', w', s', f', w'' exist and are bounded, and that $EY^5 < \infty$. Then

$$(n\epsilon_n)^{1/2}(m_n(x_1) - m(x_1), \dots, m_n(x_p) - m(x_p))'$$

converges in distribution to a multivariate normal random vector with zero mean vector and diagonal covariance matrix with i -th diagonal element given by

$$\sigma^2(x_i) \int K^2(u) du / f(x_i)$$

where

$$\sigma^2(u) = s(u) - m^2(u).$$

Schuster proves this theorem by using the Berry-Esseen theorem to show the joint asymptotic normality of the numerator and denominator of m_n . An application of the Mann-Wald theorem (Billingsley (1968)) then yields the desired result. As we shall see, it is not necessary to consider the joint distribution of the numerator and denominator of m_n in order to establish asymptotic normality. Schuster's proof can thus be simplified. Also, by using the Lindeberg-Feller central limit theorem, instead of the Berry-Esseen theorem, we will be able to require the δ -function sequence to satisfy a less restrictive condition, namely that $n\epsilon_n \rightarrow \infty$ instead of $n\epsilon_n^3 \rightarrow \infty$.

We now present the new asymptotic normality result for m_n . The most important difference between this theorem (when stated in terms of kernel type δ -sequences) and Theorem 2.4.1 is that it only requires $n\epsilon_n \rightarrow \infty$

instead of $n\epsilon_n^3 \rightarrow \infty$. Also, it applies to general δ -function sequences. There are minor differences in the other conditions which will be evident in the statement of the theorem. We first state the main theorem, and then prove a preliminary lemma before returning to the proof of the theorem. We then specialize to kernel type δ -function sequences.

2.4.2 Theorem. Let $\{\delta_n\}$ be a sequence of δ -functions such that

$$\gamma_n = \int |\delta_n(u)|^{2+\eta} du < \infty \quad \text{for each } n,$$

for some $\eta > 0$,

$$\alpha_n = \int \delta_n^2(u) du = o(n) \quad \text{as } n \rightarrow \infty$$

and

$$\gamma_n = o(n^{\eta/2} \alpha_n^{1+\eta/2}), \quad \text{as } n \rightarrow \infty.$$

Suppose $E|Y|^{2+\eta} < \infty$ and the distinct points x_1, x_2, \dots, x_p are continuity points of each of the functions $f(x)$, $m(x)$, $s(x) = E[Y^2|X=x]$ and $E[|Y|^{2+\eta}|X=x]$, and that $f(x_i) > 0$, $i = 1, \dots, p$. Then

$$(Z_n(x_1), Z_n(x_2), \dots, Z_n(x_p))$$

converges in distribution to a multivariate normal random vector with zero mean vector and identity covariance matrix, where

$$Z_n(x) = \frac{m_n(x) - g_n(x)}{[(\alpha_n/n)(\sigma^2(x)/f(x))]^{1/2}},$$

$$g_n(x) = \frac{Em_n^*(x)}{Ef_n(x)},$$

and

$$\sigma^2(x) = s(x) - m^2(x).$$

□

Since $m_n = m_n^*/f_n$ is a ratio of sums of random variables, a direct central limit argument is not possible. However, note that

$$m_n - g_n = [m_n^*/f - (f_n/f)g_n](f/f_n)$$

and $f/f_n \xrightarrow{p} 1$, so that $m_n - g_n$ will have the same asymptotic distribution as the term within square brackets above. The term within square brackets is a sum of random variables, and thus standard arguments may be used to establish its asymptotic normality. This is the outline which the proof of Theorem 2.4.2 will follow, although the notation will be more complicated since the proof will be in a multivariate setting.

The following lemma establishes the asymptotic variance and covariance of $m_n^*/f - (f_n/f)g_n$.

2.4.3 Lemma. Let $\{\delta_n\}$ be a δ -function sequence such that $\alpha_n = \int \delta_n^2(u)du < \infty$ for each n . Suppose $x \neq y$ are continuity points of f , m , and s , and that $f(x) > 0$, $f(y) > 0$, and $EY^2 < \infty$. Define

$$R_n(z) = f(z)/f_n(z)$$

and

$$H_n(z) = m_n^*(z)/f(z) - g_n(z)/R_n(z).$$

Then

$$(i) \quad (n/\alpha_n)\text{Var}[H_n(x)] \rightarrow \sigma^2(x)/f(x)$$

and

$$(ii) \quad n \operatorname{Cov}[H_n(x), H_n(y)] \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

Proof. By definition,

$$H_n(x) = (nf(x))^{-1} \sum_{i=1}^n (Y_i - g_n(x)) \delta_n(x - X_i)$$

and

$$\begin{aligned} EH_n(x) &= Em_n^*(x)/f(x) - g_n(x)Ef_n(x)/f(x) \\ &= 0 \end{aligned}$$

since

$$g_n(x) = Em_n^*(x)/Ef_n(x) ,$$

and we thus have

$$\begin{aligned} \operatorname{Var}[H_n(x)] &= EH_n^2(x) \\ &= (nf(x))^{-2} E \left\{ \sum_{i=1}^n [(Y_i - g_n(x)) \delta_n(x - X_i)]^2 \right. \\ &\quad \left. + \sum_{i \neq j} [(Y_i - g_n(x)) \delta_n(x - X_i)] [(Y_j - g_n(x)) \delta_n(x - X_j)] \right\} . \end{aligned}$$

Now

$$E[(Y - g_n(x)) \delta_n(x - X)] = 0$$

since

$$Em_n^*(x) = EY\delta_n(x - X) ,$$

$$Ef_n(x) = E\delta_n(x - X) .$$

Hence

$$\begin{aligned} (n/\alpha_n) \operatorname{Var}[H_n(x)] \\ = \alpha_n^{-1} (f(x))^{-2} E[(Y - g_n(x)) \delta_n(x - X)]^2 \end{aligned}$$

$$\begin{aligned}
&= \alpha_n^{-1} (f(x))^{-2} \{ \int s(u) f(u) \delta_n^2(x-u) du \\
&\quad - 2g_n(x) \int m(u) f(u) \delta_n^2(x-u) du \\
&\quad + g_n^2(x) \int f(u) \delta_n^2(x-u) du \} \\
&\rightarrow (f(x))^{-2} \{ s(x) f(x) - m^2(x) f(x) \} \\
&= \sigma^2(x) / f(x) ,
\end{aligned}$$

since $\{\delta_n^2(u)/\alpha_n\}$ is a δ -function sequence by Lemma 2.1.4 and $g_n(x) \rightarrow m(x)$.

For (ii), note

$$\begin{aligned}
n \operatorname{Cov}[H_n(x), H_n(y)] &= n E H_n(x) H_n(y) \\
&= (nf(x)F(y))^{-1} E \sum_{i,j=1}^n [(Y_i - g_n(x)) \delta_n(x - X_i)] \cdot \\
&\quad [(Y_j - g_n(y)) \delta_n(y - X_j)] \\
&= (nf(x)f(y))^{-1} \sum_{i=1}^n E [(Y_i - g_n(x)) \delta_n(x - X_i)] \cdot \\
&\quad [(Y_i - g_n(y)) \delta_n(y - X_i)] \\
&= (f(x)f(y))^{-1} \int \delta_n(x-u) \delta_n(y-u) q_n(u) du
\end{aligned}$$

where

$$q_n(u) = f(u) [s(u) - m(u)(g_n(x) + g_n(y)) + g_n(x)g_n(y)]$$

is continuous at $u = x$ and $y = y$ by assumption. Thus

$$n \operatorname{Cov}[H_n(x), H_n(y)] \rightarrow 0$$

by Lemma 2.1.6, and (i) is true. □

We now return to the proof of Theorem 2.4.2. By the Cramér-Wold device (e.g., Billingsley (1968)), it suffices to show

$$\sum_{k=1}^p t_k Z_n(x_k) \xrightarrow{L} N(0, \sum_{k=1}^p t_k^2),$$

or, equivalently,

$$\frac{\sum_{k=1}^p t_k [m_n(x_k) - g_n(x_k)]}{\{(\alpha_n/n) \sum_{k=1}^p t_k^2 (\sigma^2(x_k)/f_1(x_k))\}^{1/2}} \xrightarrow{L} N(0,1),$$

for any real numbers t_1, t_2, \dots, t_p . Write

$$m_n(x) - g_n(x) = H_n(x) R_n(x)$$

where H_n and R_n are as defined in Lemma 2.4.3. Since $f_n(x_k) \xrightarrow{P} f(x_k)$, $k = 1, \dots, p$, it follows that $R_n(x_k) \xrightarrow{P} 1$, $k = 1, \dots, p$, and it thus suffices to show that

$$\frac{\sum_{k=1}^p t_k H_n(x_k)}{\{(\alpha_n/n) \sum_{k=1}^p t_k^2 (\sigma^2(x_k)/f_1(x_k))\}^{1/2}} \xrightarrow{L} N(0,1).$$

Now

$$\gamma_n = \int |\delta_n(u)|^{2+\eta} du < \infty$$

for each n implies

$$\alpha_n = \int \delta_n^2(u) du < \infty$$

for each n , since $\int_{|u| < \lambda} \delta_n^2(u) du < \infty$ for any finite $\lambda > 0$ by Hölder's inequality, and by C1 and C3 of Definition 2.1.1. Thus we have, by Lemma 2.4.3,

$$\begin{aligned} \text{Var} \left[\sum_{k=1}^p t_k H_n(x_k) \right] &= \sum_{k=1}^p t_k^2 \text{Var}[H_n(x_k)] \\ &\quad + \sum_{k \neq j} t_k t_j \text{Cov}[H_n(x_k), H_n(x_j)] \\ &\sim (\alpha_n/n) \sum_{k=1}^p t_k^2 \sigma^2(x_k) / f(x_k) \end{aligned}$$

as $n \rightarrow \infty$. Hence it suffices to prove

$$V_n = \frac{\sum_{k=1}^p t_k H_n(x_k)}{\left\{ \text{Var} \left[\sum_{k=1}^p t_k H_n(x_k) \right] \right\}^{1/2}} \xrightarrow{L} N(0,1) .$$

Since, by definition,

$$H_n(x) = (nf(x))^{-1} \sum_{i=1}^n (Y_i - g_n(x)) \delta_n(x - X_i) ,$$

we may write

$$V_n = \sum_{i=1}^n V_{n,i}$$

where the i.i.d. random variables $V_{n,i}$, $i = 1, \dots, n$, are defined by

$$V_{n,i} = n^{-1/2} \sigma_n^{-1} \sum_{k=1}^p (t_k / f(x_k)) (Y_i - g_n(x_k)) \delta_n(x_k - X_i)$$

where

$$\sigma_n^2 = \text{Var} \left\{ \sum_{k=1}^p (t_k/f(x_k)) (Y - g_n(x_k)) \delta_n(x - X) \right\}.$$

It then follows from the Lindeberg-Feller central limit theorem that if, for some $\eta > 0$,

$$nE|V_{n,1}|^{2+\eta} \rightarrow 0$$

then

$$V_n \xrightarrow{d} N(0,1) \text{ as } n \rightarrow \infty$$

Now by applying the c_r inequality of Loève (1963) repeatedly we have

$$\begin{aligned} nE|V_{n,1}|^{2+\eta} &= nE \left| \sum_{k=1}^p \frac{t_k(Y_1 - g_n(x_k)) \delta_n(x_k - X_1)}{n^{1/2} \sigma_n f(x_k)} \right|^{2+\eta} \\ &\leq \sum_{k=1}^p c_k(\eta) \left\{ \frac{E|Y_1 \delta_n(x_k - X_1)|^{2+\eta}}{n^{\eta/2} \sigma_n^{2+\eta} (f(x_k))^{2+\eta}} \right. \\ &\quad \left. + \frac{E|g_n(x_k) \delta_n(x_k - X_1)|^{2+\eta}}{n^{\eta/2} \sigma_n^{2+\eta} (f(x_k))^{2+\eta}} \right\} \\ &= \sum_{k=1}^p c_k(\eta) [I_{k,n} + J_{k,n}], \end{aligned}$$

where $c_k(\eta)$ depends only on k and η and the constants t_1, \dots, t_k . It is easily seen that

$$\sigma_n^2 = n \text{Var} \left[\sum_{k=1}^p t_k H_n(x_k) \right] \sim \alpha_n \sum_{k=1}^p t_k^2 \sigma^2(x_k)/f(x_k),$$

the last step following by earlier calculations. Further,

$$\begin{aligned}
& E|Y\delta_n(x-X)|^{2+\eta} \\
&= \int |\delta_n(x-u)|^{2+\eta} E[|Y|^{2+\eta}|X=u] f(u) du \\
&= \gamma_n \int \delta_n^{**}(x-u) E[|Y|^{2+\eta}|X=u] f(u) du
\end{aligned}$$

where $\{\delta_n^{**}\} = \{|\delta_n|^{2+\eta}/\gamma_n\}$ is a δ -function sequence by Lemma 2.1.4.

Thus, for $k = 1, 2, \dots, p$, we have

$$I_{k,n} \sim \frac{\gamma_n E[|Y|^{2+\eta}|X=x_k]}{n^{n/2} \alpha_n^{1+n/2} f(x_k) \left[\sum_{i=1}^p t_i^2 \sigma^2(x_i)/f(x_i) \right]^{1+\eta/2}}$$

$\rightarrow 0$ as $n \rightarrow \infty$

since

$$\gamma_n = o(n^{n/2} \alpha_n^{1+n/2})$$

by assumption. Similar calculations yield

$$J_{k,n} \sim \frac{\gamma_n |g_n(x_k)|^{2+\eta}}{n^{n/2} \alpha_n^{1+n/2} (f(x_k))^{2+\eta} \left[\sum_{k=1}^p t_k^2 \sigma^2(x_k)/f(x_k) \right]^{1+\eta/2}}$$

$\rightarrow 0$ as $n \rightarrow \infty$, and the proof is complete. \square

We now give a version of Theorem 2.4.2 for kernel-type δ -function sequences, which may be compared with Schuster's Theorem 2.4.1.

2.4.4 Theorem. Suppose $\{\delta_n(x)\} = \{\epsilon_n^{-1} K(x/\epsilon_n)\}$ is a δ -function sequence of kernel type satisfying

$$(i) \int |K(u)|^{2+\eta} du < \infty \text{ for some } \eta > 0$$

$$(ii) \int uK(u)du = 0, \quad \int u^2K(u)du < \infty$$

$$(iii) \quad n\epsilon_n \rightarrow \infty, \quad n\epsilon_n^5 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Suppose $m(x)$ and $f(x)$ have bounded, continuous 1st and 2nd derivatives, $E|Y|^{2+\eta} < \infty$, the distinct points x_1, x_2, \dots, x_p are continuity points of $s(x)$ and $E[|Y|^{2+\eta}|X=x] < \infty$ and $f(x_k) > 0$, $k = 1, \dots, p$. Then $(Z'_n(x_1), \dots, Z'_n(x_p))$ converges in distribution to a multivariate normal random vector with zero mean vector and identity covariance matrix, where

$$Z'_n(x) = \frac{(n\epsilon_n)^{1/2}(m_n(x) - m(x))}{\{\sigma^2(x) \int K^2(u)du/f(x)\}^{1/2}}.$$

Proof. We first verify that this δ -function sequence satisfies the conditions of Theorem 2.4.2. Now

$$\alpha_n = \int \delta_n^2(u)du = \epsilon_n^{-1} \int K^2(u)du < \infty$$

for each n since $\epsilon_n \neq 0$, $\int K^2(u)du < \infty$. Further,

$$\alpha_n/n = (n\epsilon_n)^{-1} \int K^2(u)du \rightarrow 0$$

since $n\epsilon_n \rightarrow \infty$ by assumption. Similarly,

$$\gamma_n = \int |\delta_n(u)|^{2+\eta}du = (1/\epsilon_n)^{1+\eta} \int |K(u)|^{2+\eta}du < \infty$$

for each n , and

$$\gamma_n/n^{1/2} \alpha_n^{1/2} \propto (n\epsilon_n)^{-\eta/2} \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

by assumption. Thus this type of δ -function sequence satisfies the requirements of Theorem 2.4.2, and since the remaining regularity conditions of Theorem 2.4.2 are clearly satisfied under the present assumptions,

we have that the conclusion holds when $m(x)$ is replaced by $g_n(x) = E m_n^*(x) / E f_n(x)$ in the expression for $Z_n'(x)$. Hence, if we show that

$$(n\epsilon_n)^{\frac{1}{2}}(m(x) - g_n(x)) \rightarrow 0,$$

then the conclusion will follow. Now

$$\begin{aligned} & (n\epsilon_n)^{\frac{1}{2}}(g_n(x) - m(x)) \\ &= (n\epsilon_n)^{\frac{1}{2}} \left\{ \frac{\epsilon_n^{-1} \int K((x-u)/\epsilon_n) m(u) f(u) du}{\epsilon_n^{-1} \int K((x-u)/\epsilon_n) f(u) du} - m(x) \right\} \\ &= (n\epsilon_n)^{\frac{1}{2}} \left\{ \frac{\epsilon_n^{-1} \int K((x-u)/\epsilon_n) n(u) f(u) du - m(x) f(x)}{\epsilon_n^{-1} \int K((x-u)/\epsilon_n) f(u) du} \right. \\ &\quad \left. + \frac{m(x) f(x) - m(x) \epsilon_n^{-1} \int K((x-u)/\epsilon_n) f(u) du}{\epsilon_n^{-1} \int K((x-u)/\epsilon_n) f(u) du} \right\}. \end{aligned}$$

By Lemma 2.1.5, the numerator of the term within the brackets above is $O(\epsilon_n^2)$ and the denominator converges to $f(x) > 0$. Thus

$$\begin{aligned} & (n\epsilon_n)^{\frac{1}{2}}(g_n(x) - m(x)) \\ &= (n\epsilon_n)^{\frac{1}{2}} O(\epsilon_n^2) \\ &= O((n\epsilon_n^5)) \rightarrow 0 \end{aligned}$$

since $n\epsilon_n^5 \rightarrow 0$.

□

2.5 Asymptotic Distribution of \bar{m}_n .

It is evident that, since $\bar{m}_n = m_n^*/f$ is a sum of independent random variables, we may apply the Lindeberg-Feller central limit theorem in much the same way as we did in Theorem 2.4.2 to establish the asymptotic normality of $(\bar{m}_n(x) - E\bar{m}_n(x))/\text{Var}[\bar{m}_n]$. We established in (ii) of Lemma 2.3.1 that

$$\text{Var}[m_n^*(x)] \sim (\alpha_n/n)s(x)f(x)$$

for appropriate points x . Hence we have

$$\text{Var}[\bar{m}_n] \sim (\alpha_n/n)s(x)/f(x) .$$

We therefore have the following theorems, which we state without proof, since the proofs follow those of Theorems 2.4.2 and 2.4.4 very closely. The first theorem concerns the asymptotic normality of \bar{m}_n for general δ -function sequences; the second for kernel type δ -sequences.

2.5.1 Theorem. Under the conditions of Theorem 2.4.2,

$(W_n(x_1), \dots, W_n(x_p))$ converges in distribution to a multivariate normal random vector with zero mean vector and identity covariance matrix, where

$$W_n(x) = \frac{\bar{m}_n(x) - E\bar{m}_n(x)}{\{(\alpha_n/n)s(x)/f(x)\}^{1/2}} .$$

2.5.2 Theorem. Under the conditions of Theorem 2.4.4, $(W'_n(x_1), \dots, W'_n(x_p))$ converges in distribution to a multivariate normal random vector with zero mean and identity covariance matrix, where

$$W'_n(x) = \frac{(n\epsilon_n)^{1/2}(\bar{m}_n(x) - m(x))}{\{s(x) \int K^2(u)du/f(x)\}^{1/2}} .$$

2.6 Mean Integrated Square Error.

The mean integrated square error (MISE) J_n of an estimator

$$f_n(x) = n^{-1} \sum_{i=1}^n \delta_n(x - X_i)$$

of a density f is defined as

$$J_n = E \int (f_n(x) - f(x))^2 dx$$

where δ_n and f are assumed to be square integrable. Watson and Leadbetter (1963) show that J_n is minimized for each n if δ_n is chosen to have a Fourier transform ϕ_{δ_n} expressible as

$$\phi_{\delta_n}(t) = \frac{|\phi_f(t)|^2}{(1/n) + ((n-1)/n)|\phi_f(t)|^2}$$

where ϕ_f is the Fourier transform of f . (Fourier transforms of square integrable functions have the usual interpretation here.) For the regression estimation problem, Watson (1964) considers the error criterion J'_n defined by

$$J'_n = E \int \left[\sum_{i=1}^n \delta_n(x - X_i) m(x) - \sum_{i=1}^n Y_i \delta_n(x - X_i) \right]^2 dx$$

where appropriate assumptions are made on δ_n and m to insure the finiteness of the integral. Watson states that J'_n is minimized for each n if δ_n is chosen so as to have Fourier transform

$$\phi_{\delta_n}^*(t) = \frac{|\phi_{fm}(t)|^2}{n^{-1} E Y^2 + ((n-1)/n) |\phi_{fm}(t)|^2}$$

where ϕ_{fm} is the Fourier transform of fm .

We assume here that δ_n and f_m are square integrable functions and that $EY^2 < \infty$. We define the error criterion I_n by

$$I_n = E \int (m_n^*(x) - f(x)m(x))^2 dx$$

where m_n^* is the numerator of m_n . We will show here that I_n is also minimized for each n by choosing δ_n to have Fourier transform given by $\phi_{\delta_n}^*$, defined above. Note that I_n may be interpreted as the MISE of the numerator of m_n or \bar{m}_n , disregarding the denominator.

By the definition of I_n and Parseval's formula, we have

$$\begin{aligned} (2.6.1) \quad I_n &= E \int (m_n^*(x) - f(x)m(x))^2 dx \\ &= (2\pi)^{-1} E \int |\phi_{m_n}^*(t) - \phi_{fm}(t)|^2 dt \end{aligned}$$

where $\phi_{m_n}^*$ is the Fourier transform of m_n^* , so that I_n may be minimized by minimizing the extreme right hand side of expression (2.6.1) above. Now, by Fubini's theorem for positive functions,

$$\begin{aligned} &E \int |\phi_{m_n}^*(t) - \phi_{fm}(t)|^2 dt \\ &= \int E |\phi_{m_n}^*(t) - \phi_{fm}(t)|^2 dt, \end{aligned}$$

so that I_n may be minimized by minimizing

$$\begin{aligned} (2.6.2) \quad &E |\phi_{m_n}^*(t) - \phi_{fm}(t)|^2 \\ &= E \{ |\phi_{m_n}^*(t)|^2 + |\phi_{fm}(t)|^2 \\ &\quad - [\bar{\phi}_{m_n}^*(t)\phi_{fm}(t) + \phi_{m_n}^*(t)\bar{\phi}_{fm}(t)] \} \end{aligned}$$

for each t , where \bar{g} denotes the conjugate of the complex function g .

Note that since f_m is an integrable function,

$$(2.6.3) \quad \phi_{fm}(t) = \int e^{itu} f(u)m(u)du .$$

Further,

$$\begin{aligned} (2.6.4) \quad \phi_{m_n}^*(t) &= \int [n^{-1} \sum_{j=1}^n Y_j \delta_n(u-X_j)] e^{itu} du \\ &= n^{-1} \sum_{j=1}^n Y_j e^{itX_j} \int \delta_n(u) e^{itu} du \\ &= \phi_{\delta_n}(t) [n^{-1} \sum_{j=1}^n Y_j e^{itX_j}] , \end{aligned}$$

so that

$$\begin{aligned} (2.6.5) \quad E|\phi_{m_n}^*(t)|^2 &= |\phi_{\delta_n}(t)|^2 E|n^{-1} \sum_{j=1}^n Y_j e^{itX_j}|^2 \\ &= |\phi_{\delta_n}(t)|^2 \cdot n^{-2} E \sum_{j,k=1}^n Y_j Y_k e^{itX_j} e^{-itX_k} \\ &= |\phi_{\delta_n}(t)|^2 n^{-2} \{ E \sum_{j=1}^n Y_j^2 \\ &\quad + E \sum_{j \neq k} Y_j Y_k e^{itX_j} e^{-itX_k} \} . \end{aligned}$$

Now

$$\begin{aligned} EY e^{itX} &= \int m(u)f(u)e^{itu} du \\ &= \phi_{fm}(t) , \end{aligned}$$

and thus

$$\begin{aligned}
 (2.6.5) \quad E|\phi_{m_n}^*(t)|^2 &= |\phi_{\delta_n}(t)|^2 [(1/n)EY^2 + ((n-1)/n)|\phi_{fm}(t)|^2] .
 \end{aligned}$$

Finally, from (2.6.3) and (2.6.4), we have

$$\begin{aligned}
 (2.6.6) \quad E[\phi_{fm}(t)\bar{\phi}_{m_n}^*(t) + \bar{\phi}_{fm}(t)\phi_{m_n}^*(t)] &= \phi_{fm}(t)\bar{\phi}_{\delta_n}(t)E[n^{-1}\sum_{j=1}^n Y_j e^{-itX_j}] \\
 &\quad + \bar{\phi}_{fm}(t)\phi_{\delta_n}(t)E[n^{-1}\sum_{j=1}^n Y_j e^{itX_j}] \\
 &= |\phi_{fm}(t)|^2 [\bar{\phi}_{\delta_n}(t) + \phi_{\delta_n}(t)] \\
 &= 2 \operatorname{Re}[\phi_{\delta_n}(t)] |\phi_{fm}(t)|^2 .
 \end{aligned}$$

Combining (2.6.3), (2.6.5) and (2.6.6) yields

$$\begin{aligned}
 (2.6.7) \quad E|\phi_{m_n}^*(t) - \phi_{fm}(t)|^2 &= |\phi_{fm}(t)|^2 - 2 \operatorname{Re}[\phi_{\delta_n}(t)] |\phi_{fm}(t)|^2 \\
 &\quad + |\phi_{\delta_n}(t)|^2 [(1/n)EY^2 + ((n-1)/n)|\phi_{fm}(t)|^2] \\
 &= [(1/n)EY^2 + ((n-1)/n)|\phi_{fm}(t)|^2] \cdot \\
 &\quad \left| \phi_{\delta_n}(t) - \frac{|\phi_{fm}(t)|^2}{(1/n)EY^2 + ((n-1)/n)|\phi_{fm}(t)|^2} \right|^2
 \end{aligned}$$

$$+ \frac{|\phi_{fm}(t)|^2 [EY^2 - |\phi_{fm}(t)|^2]}{EY^2 + (n-1)|\phi_{fm}(t)|^2},$$

the last equality following by completing the square and rearranging terms. Now

$$\begin{aligned} |\phi_{fm}(t)| &= \left| \int m(u) f(u) e^{itu} du \right| \\ &\leq \int |m(u)| f(u) du \\ &\leq \int E[|Y| | X=u] f(u) du = E|Y|, \end{aligned}$$

so that

$$|\phi_{fm}(t)|^2 \leq (E|Y|)^2 \leq EY^2,$$

for all t , and hence (2.6.7) is minimized for each t by choosing

$$\phi_{\delta_n} = \phi_{\delta_n}^*.$$

3. ASYMPTOTIC PROPERTIES OF MAXIMUM ABSOLUTE DEVIATION

3.1 Preliminaries

Since our goal in many cases is the estimation of the regression function over the entire real line, or some subset of the real line, it is natural to investigate the behavior of our estimators under some global error criterion. An attempt at this direction was made in Section 2.6, where we considered mean integrated square error. This was not entirely satisfactory, however, since we were only able to determine the δ -function sequence which minimized the MISE of the numerator of the estimators in question, disregarding the denominator. In this chapter, we consider a different global error criterion, the maximum absolute deviation, defined as $\sup_{x \in I} |m_n(x) - m(x)|$ where I is a closed, bounded interval of the real line, which we will take without loss of generality to be $[0,1]$. We shall mainly be concerned here with conditions under which the maximum absolute deviation converges to zero in probability (in this case we say that the estimator in question is *uniformly consistent* over I). We will also be able to find a large sample confidence bound for the regression function, based on the estimator \bar{m}_n .

Our method of analysis will follow the one briefly outlined below used by Bickel and Rosenblatt (1973) and Rosenblatt (1976) for probability density estimators. For a density function estimator

$$f_n(u) = (n\epsilon_n)^{-1} \sum_{i=1}^n K((u-X_i)/\epsilon_n) ,$$

the deviation about the mean $f_n(u) - Ef_n(u)$, normalized so as to have

non-zero asymptotic standard deviation, may be written as

$$\begin{aligned}
 (3.1.1) \quad & \frac{(n\epsilon_n)^{1/2}(f_n(u) - Ef_n(u))}{[f(u)]^{1/2}} \\
 &= [f(u)\epsilon_n]^{-1/2} \int K((u-s)/\epsilon_n) dZ_n(s) \\
 &= Y_n(u) \text{ , say,}
 \end{aligned}$$

where Z_n is the empirical process defined by

$$Z_n(s) = n^{1/2}[F_n(s) - F(s)] \text{ ,}$$

where F_n is the empirical distribution function (EDF) of X_i ,

$i = 1, \dots, n$, and F is the cumulative distribution function of X_1 .

Koml s, Major and Tusn dy (1975) have shown that a sequence of Brownian bridges $\{B_n\}$ on $[0,1]$ may be constructed such that

$$(3.1.2) \quad \sup_{-\infty < u < \infty} |Z_n(u) - B_n(F(u))| = O(n^{-1/2} \log n)$$

a.s. This fact is exploited, using integration by parts in (3.1.1), to show that

$$\begin{aligned}
 & (\log n)^{1/2} \sup_{0 \leq u \leq 1} |Y_n(u)| \\
 &= (\log n)^{1/2} \sup_{0 \leq u \leq 1} |Y_{1,n}(u)| + o_p(1)
 \end{aligned}$$

where $Y_{1,n}$ is the stochastic process obtained by replacing $Z_n(s)$ with $B_n(F(s))$ in the defining expression for Y_n . Further stages of approximation finally yield

$$\begin{aligned}
 (3.1.3) \quad & (\log n)^{\frac{1}{2}} \sup_{0 \leq u \leq 1} |Y_n(u)| \\
 & = (\log n)^{\frac{1}{2}} \sup_{0 \leq u \leq 1} |Y_{2,n}(u)| + o_p(1)
 \end{aligned}$$

where $Y_{2,n}$ is the Gaussian process on $[0,1]$ defined by

$$Y_{2,n}(u) = \epsilon_n^{-\frac{1}{2}} \int K((u-s)/\epsilon_n) dW(s)$$

where W is a Wiener process on \mathbb{R} . The asymptotic distribution of $(\log n)^{\frac{1}{2}} \sup_{0 \leq u \leq 1} |Y_{2,n}(u)|$ with proper centering constants, is determined, and, in light of (3.1.3), $(\log n)^{\frac{1}{2}} \sup_{0 \leq u \leq 1} |Y_n(u)|$ has the same asymptotic distribution.

We will employ this method to determine the asymptotic distribution of the maximum absolute deviation of the numerator m_n^* of the estimators m_n and \bar{m}_n , properly normalized and centered. Algebraic manipulation and elementary analysis will then yield uniform consistency of the estimators, with an associated rate of convergence. Since the denominator of \bar{m}_n is non-stochastic, an asymptotic confidence band for m , based on \bar{m}_n , may also be specified.

In the forthcoming development, we will need to use integrals of the form

$$(3.1.4) \quad Y_n(t) = \iint y k\left(\frac{t-x}{\epsilon_n}\right) dW(T(x,y)) ,$$

where $T: \mathbb{R}^2 \rightarrow [0,1]^2$ is the transformation defined by

$$T(x,y) = (F_{X|Y}(x,y), F_Y(y)) ,$$

and $W(\cdot, \cdot)$ is the Wiener process on $[0,1]^2$. In this section, we will give conditions for the existence of (3.1.4) and prove some useful properties.

If $H(s,t)$ is a real, measurable function on $[0,1]^2$, then it is well-known that the L_2 integral

$$\iint H(s,t) dW(s,t)$$

exists if

$$\iint H^2(s,t) ds dt < \infty$$

(see Masani (1968), Chap. 5).

Suppose that $f(x,y) > 0$ for all real x and y so that T is one-to-one and hence T^{-1} is a well-defined function on $[0,1]^2$ to \mathbb{R}^2 . Denote, for fixed n and t

$$G_t(x,y) = yK((t-x)/\epsilon_n) .$$

Then, by Theorem 5.19 of Masani (1968), we have

$$\begin{aligned} (3.1.5) \quad & \iint_{\mathbb{R}^2} yK((t-x)/\epsilon_n) dW(T(x,y)) \\ &= \iint_{[0,1]^2} G_t(T^{-1}(s,u)) dW(s,u) \end{aligned}$$

in the sense that if either integral exists, then so does the other and they are equal. By the previous remark, the integral on the right hand side of (3.1.5) exists if

$$\iint_{[0,1]^2} G_t^2(T^{-1}(s,u)) ds du < \infty .$$

Now

$$\begin{aligned} (3.1.6) \quad & \iint_{[0,1]^2} G_t^2(T^{-1}(s,u)) ds du \\ &= \iint_{\mathbb{R}^2} G_t^2(x,y) |J(x,y)| dx dy \end{aligned}$$

where $J(x,y)$ is the Jacobian of T (see, e.g. Buck (1965), Sec. 6.1, Thm. 4), if $|J(x,y)| > 0$ for all real x and y and $G_t(x,y)$, $f(x,y)$ and $f(y)$ are continuous. By definition,

$$\begin{aligned} J(x,y) &= \begin{vmatrix} \frac{\partial}{\partial x} F_{X|Y}(x|y) & \frac{\partial}{\partial y} F_{X|Y}(x|y) \\ \frac{\partial}{\partial x} F_Y(y) & \frac{\partial}{\partial y} F_Y(y) \end{vmatrix} \\ &= f_{X|Y}(x|y) f_Y(y) \\ &= f(x,y) > 0 \end{aligned}$$

by assumption, using the obvious notation for conditional and marginal densities. Thus

$$\begin{aligned} &\iint_{[0,1]^2} G_t^2(T^{-1}(s,u)) ds du \\ &= \iint_{\mathbb{R}^2} y^2 K((t-x)/\epsilon_n) f(x,y) dx dy \\ &= EY^2 K^2((t-X)/\epsilon_n) < \infty \end{aligned}$$

if, e.g., $EY^2 < \infty$ and K is bounded. We note that the above development holds if, instead of having $f(x,y) > 0$ for all real x and y , we have $f(x,y) > 0$ for x and y in some rectangle of \mathbb{R}^2 , and the range of integration is appropriately adjusted. We will henceforth assume this to be true without comment.

We will now give properties of the integral (3.1.4) which will be useful in the future development. We will show

$$(3.1.7) \quad EY_n(t) = 0 ,$$

$$\begin{aligned}
 (3.1.8) \quad EY_n(t_1)Y_n(t_2) \\
 = \iint y^2 K((t_1-x)/\epsilon_n) K((t_2-x)/\epsilon_n) f(x,y) dx dy,
 \end{aligned}$$

for $t_1 \neq t_2$. In view of (3.1.5) and the definition of the stochastic integral, (3.1.7) follows. For (3.1.8), we note, by (3.1.5) and (5.2) of Masani (1968)

$$\begin{aligned}
 EY_n(t_1)Y_n(t_2) \\
 = \iint G_{t_1}(T^{-1}(s,u)) G_{t_2}(T^{-1}(s,u)) ds du \\
 = \iint y^2 K((t_1-x)/\epsilon_n) K((t_2-x)/\epsilon_n) f(x,y) dx dy
 \end{aligned}$$

as in (3.1.8).

We finally note to close this section that, since W is a Gaussian process on $[0,1]^2$ and since $Y_n(t)$ is an L_2 limit of linear combinations of $W(\cdot, \cdot)$, we have that $Y_n(t)$ is itself a (one-parameter) Gaussian stochastic process for each n , with mean given by (3.1.7) and covariance function given by (3.1.8).

3.2 Maximum Absolute Deviation of m_n^*

It is convenient to introduce certain assumptions at this point which will be in force in our main theorem. Let $f(x,y)$ denote the joint density of (X,Y) , $f_Y(y)$ the marginal density of Y , and let $\{a_n\}$ be a real sequence with $a_n \rightarrow \infty$ as $n \rightarrow \infty$. We make the following assumptions:

$$(A1) \quad (\log n) \epsilon_n^{-3} \int_{|y| \geq a_n} y^2 f_Y(y) dy \leq c$$

for all n and some constant c ,

$$(A2) \quad a_n \epsilon_n^{-1/2} n^{-1/6} (\log n)^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

$$(A3) \quad (\log n) \sup_{0 \leq x \leq 1} \int_{|y| > a_n} y^2 f(x, y) dy \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

(A4) There exists a constant $\eta > 0$ such that

$$g_n(x) = \int_{-a_n}^{a_n} y^2 f(x, y) dy$$

satisfies

$$g_n(x) > \eta \quad \forall x \in [0, 1] \text{ and some } n,$$

and $g_n^{1/2}$ has a continuous 1st derivative on some interval $[-A, A]$.

Further, the functions

$$s(x)f(x) = \int y^2 f(x, y) dy,$$

$$E[|Y| | X=x] f(x) = \int |y| f(x, y) dy$$

are uniformly bounded.

If Y is a bounded random variable, then clearly any sequence $\{a_n\}$ with $a_n \rightarrow \infty$ satisfies assumptions A1 and A3. If the marginal distribution of Y is normal and $\epsilon_n = n^{-\delta}$ as in Theorem 3.2.1, then it is readily checked that $\{a_n\} = \{\log n\}$ satisfies A1 and A2.

We normalize $m_n^*(t) - Em_n^*(t)$ by $(n\epsilon_n)^{-1/2} [s(t)f(t)]^{1/2}$, which is proportional to its asymptotic standard deviation, thus defining the following stochastic process on $[0, 1]$:

$$(3.2.1) \quad Y_n(t) = \frac{(n\epsilon_n)^{1/2} [m_n^*(t) - Em_n^*(t)]}{[s(t)f(t)]^{1/2}}.$$

Then we have the following theorem.

3.2.1 Theorem. Suppose the kernel function K vanishes outside a finite interval $[-A, A]$ and is absolutely continuous and has a bounded derivative on $[-A, A]$ and that the marginal density of X is positive on an interval containing $[0, 1]$. Suppose A1-A4 hold. Then, for $\epsilon_n = n^{-\delta}$, $0 < \delta < \frac{1}{2}$,

$$P\left\{(2\delta \log n)^{\frac{1}{2}} \left[\frac{\sup_{0 \leq t \leq 1} |Y_n(t)|}{[\lambda(K)]^{\frac{1}{2}}} - d_n \right] < x \right\} \rightarrow e^{-2e^{-x}}$$

as $n \rightarrow \infty$, where

$$\lambda(K) = \int K^2(u) du ,$$

$$d_n = (2\delta \log n)^{\frac{1}{2}} + (2\delta \log n)^{-\frac{1}{2}} \left\{ \log\left(\frac{c_1(K)}{\pi^{\frac{1}{2}}}\right) + \frac{1}{2}[\log \delta + \log \log n] \right\}$$

if

$$c_1(K) = \frac{K^2(A) + K^2(-A)}{2\lambda(K)} > 0 ,$$

and otherwise

$$d_n = (2\delta \log n)^{\frac{1}{2}} + (2\delta \log n)^{-\frac{1}{2}} \left[\log\left(\frac{c_3(K)}{2\pi}\right) \right]$$

where

$$c_2(K) = \frac{\int [K'(u)]^2 du}{2\lambda(K)} .$$

The proof of Theorem 3.2.1 is based on Theorems 3.2.2 and 3.3.3, which follow. Theorem 3.2.2 is due to Bickel and Rosenblatt (1973), who used it in proving a result similar to Theorem 3.2.1 for probability density estimators. We will here denote by $\int K(t) dW(t)$ the L_2 integral

of k with respect to the Wiener process W (see e.g. Doob (1953), Chap. IX, Sec. 2).

3.2.2 Theorem. Suppose $K(\cdot)$ is a kernel function which vanishes outside $[-A, A]$ and is absolutely continuous on $[-A, A]$. Define on $[0, 1]$ the stochastic process

$$Z_n(t) = \varepsilon_n^{-1/2} \int K\left(\frac{t-x}{\varepsilon_n}\right) dW(x)$$

where

$$\varepsilon_n = n^{-\delta}$$

with $0 < \delta < 1/2$ and $W(x)$ is a Wiener process on $(-\infty, \infty)$. Then

$$P\left\{(2\delta \log n)^{1/2} \left[\frac{\sup_{0 \leq t \leq 1} |Z_n(t)|}{[\lambda(K)]^{1/2}} - d_n \right] < x\right\} \rightarrow e^{-2e^{-x}}$$

as $n \rightarrow \infty$, where d_n and $\lambda(K)$ are as in Theorem 1.

Theorem 3.2.3 is a special case of Theorem E of Révész (1976).

3.2.3 Theorem. Let X_1 and X_2 be independent random variables, each uniformly distributed over $[0, 1]$. Define the empirical process of (X_1, X_2) by

$$Z_n(x_1, x_2) = n^{1/2} [F_n(x_1, x_2) - x_1 x_2]$$

on $[0, 1]^2$, where F_n denotes the empirical distribution function of (X_1, X_2) . Then one can define a sequence $\{B_n\}$ of independent Brownian bridges on $[0, 1]^2$ such that

$$\begin{aligned}
 (3.2.2) \quad & \sup_{0 \leq x_1, x_2 \leq 1} |Z_n(x_1, x_2) - B_n(x_1, x_2)| \\
 &= O(n^{-1/6} (\log n)^{3/2}) \quad \text{a.s.}
 \end{aligned}$$

Proof of Theorem 3.2.1: For convenience, denote $\sup_{0 \leq t \leq 1} |g(t)|$ by $\|g\|$ and note that, for any sequence of processes $\{Z_n(t)\}$ defined on $[0,1]$,

$$\begin{aligned}
 & (\log n)^{1/2} [\|Y_n\| - d_n] \\
 &= (\log n)^{1/2} [\|Z_n\| - d_n] \\
 &+ (\log n)^{1/2} [\|Y_n\| - \|Z_n\|] .
 \end{aligned}$$

Thus, if we show that

$$(\log n)^{1/2} \|Z_n - Y_n\| \xrightarrow{P} 0$$

and

$$(\log n)^{1/2} [\|Z_n\| - d_n]$$

converges in law, then

$$(\log n)^{1/2} [\|Y_n\| - d_n]$$

also converges in law, and has the same limiting distribution.

We will apply the preceding remark to eventually "approximate" the process Y_n with the process Z_n of Theorem 3.2.2, thus obtaining the desired result. We will proceed through several stages of such approximation, and the details will be given in the sequence of lemmas which immediately follows the proof of this theorem.

We first note that Y_n may be written as

$$(3.2.3) \quad Y_n(t) = [s(t)f(t)]^{-\frac{1}{2}} \epsilon_n^{-\frac{1}{2}} \iint y K\left(\frac{t-x}{\epsilon_n}\right) dZ_n(x,y)$$

where Z_n is the empirical process defined by

$$Z_n(x,y) = n^{\frac{1}{2}} [F_n(x,y) - F(x,y)] .$$

Now define the following processes on $[0,1]$:

$$(3.2.4) \quad Y_{0,n}(t) = [s(t)f(t)]^{-\frac{1}{2}} \epsilon_n^{-\frac{1}{2}} \iint_{|y| \leq a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dZ_n(x,y) .$$

$$(3.2.5) \quad Y_{1,n}(t) = [s_n(t)f(t)]^{-\frac{1}{2}} \epsilon_n^{-\frac{1}{2}} \iint_{|y| \leq a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dZ_n(x,y)$$

where

$$s_n(t) = E[Y^2 I_{\{|y| \leq a_n\}}(Y) | X=t] ,$$

$$(3.2.6) \quad Y_{2,n}(t) = [s_n(t)f(t)]^{-\frac{1}{2}} \epsilon_n^{-\frac{1}{2}} \iint_{|y| \leq a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dB_n(T(x,y))$$

where $\{B_n\}$ is a sequence of Brownian bridges as in Theorem 3.2.3 and $T: \mathbb{R}^2 \rightarrow [0,1]^2$ is the transformation defined by

$$(3.2.7) \quad T(x,y) = (F_{X|Y}(x|y), F_Y(y)) ,$$

$$(3.2.8) \quad Y_{3,n}(t) = [s_n(t)f(t)]^{-\frac{1}{2}} \epsilon_n^{-\frac{1}{2}} \iint_{|y| \leq a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dW_n(T(x,y))$$

where $\{W_n\}$ is a sequence of independent Wiener processes used in constructing $\{B_n\}$ as

$$B_n(u, s) = W_n(u, s) - us W_n(1, 1)$$

(Révész (1976)), $0 \leq u, s \leq 1$,

$$(3.2.9) \quad Y_{4,n}(t) = [s_n(t)f(t)]^{-\frac{1}{2}} \epsilon_n^{-\frac{1}{2}} \int [s_n(x)f(x)]^{\frac{1}{2}} K\left(\frac{t-x}{\epsilon_n}\right) dW(x),$$

$$(3.2.10) \quad Y_{5,n}(t) = \epsilon_n^{-\frac{1}{2}} \int K\left(\frac{t-x}{\epsilon_n}\right) dW(x),$$

where W is a Wiener process on $(-\infty, \infty)$.

We have, by Lemma 3.2.4,

$$||Y_n - Y_{0,n}|| = o_p((\log n)^{-\frac{1}{2}}),$$

where $o_p(a_n)$ refers to a sequence of random variables A_n such that $A_n/a_n \rightarrow 0$ in probability. Lemma 3.2.8 gives

$$||Y_{0,n} - Y_{1,n}|| = o_p((\log n)^{-\frac{1}{2}}).$$

By Lemma 3.2.5

$$||Y_{1,n} - Y_{2,n}|| = O(a_n \epsilon_n^{-1/2} n^{-1/6} (\log n)^{3/2}) \text{ a.s.},$$

and by A2,

$$a_n \epsilon_n^{-1/2} n^{-1/6} (\log n)^2 \rightarrow 0,$$

so that

$$||Y_{1,n} - Y_{2,n}|| = o_p((\log n)^{-\frac{1}{2}}).$$

By Lemma 3.2.6,

$$\begin{aligned} ||Y_{2,n} - Y_{3,n}|| &= o_p(\epsilon_n^{\frac{1}{2}}) \\ &= o_p((\log n)^{-\frac{1}{2}}) \end{aligned}$$

since $\epsilon_n = n^{-\delta}$ and hence $\epsilon_n \log n \rightarrow 0$.

Now $Y_{3,n}$ is zero mean Gaussian process on $[0,1]$ with covariance function

$$\begin{aligned} (3.2.11) \quad r(t_1, t_2) &= EY_{3,n}(t_1)Y_{3,n}(t_2) \\ &= [s_n(t_1)f(t_1)]^{-\frac{1}{2}}[s_n(t_2)f(t_2)]^{-\frac{1}{2}} \cdot \\ &\quad \epsilon_n^{-1} \iint_{|y| \leq a_n} y^2 K\left(\frac{t_1-x}{\epsilon_n}\right) K\left(\frac{t_2-x}{\epsilon_n}\right) f(x,y) dx dy \end{aligned}$$

(cf. Section 3.1). The integral on the right hand side of (3.2.11) may be written as

$$\begin{aligned} &\epsilon_n^{-1} E \left\{ Y^2 I_{\{|y| \leq a_n\}}(Y) K\left(\frac{t_1-x}{\epsilon_n}\right) K\left(\frac{t_2-x}{\epsilon_n}\right) \right\} \\ &= \epsilon_n^{-1} \int s_n(x) f(x) K\left(\frac{t_1-x}{\epsilon_n}\right) K\left(\frac{t_2-x}{\epsilon_n}\right) dx . \end{aligned}$$

Thus the process $Y_{4,n}$ is a Gaussian process with the same covariance function as $Y_{3,n}$, i.e., they have the same finite dimensional distributions. Hence the asymptotic distribution of

$$(2\delta \log n)^{\frac{1}{2}} \left[\frac{||Y_{3,n}||}{[\lambda(K)]^{\frac{1}{2}}} - d_n \right]$$

is the same as that of

$$(2\delta \log n)^{\frac{1}{2}} \left[\frac{||Y_{4,n}||}{[\lambda(K)]^{\frac{1}{2}}} - d_n \right].$$

Further, by Lemma 3.2.7

$$\begin{aligned} ||Y_{4,5} - Y_{5,n}|| &= o_p(\epsilon_n^{\frac{1}{2}}) \\ &= o_p((\log n)^{-\frac{1}{2}}). \end{aligned}$$

By Theorem 3.2.2

$$(2\delta \log n)^{\frac{1}{2}} \left[\frac{||Y_{5,n}||}{[\lambda(K)]^{\frac{1}{2}}} - d_n \right]$$

has the desired limit distribution, and the theorem is proved. \square

3.2.4 Lemma. If A1 is satisfied and

$$g(x) - s(x)f(x) = \int y^2 f(x,y) dy$$

is bounded away from zero on $[0,1]$, then

$$||Y_n - Y_{0,n}|| = o_p((\log n)^{-\frac{1}{2}}).$$

Proof. Note

$$Y_n(t) - Y_{0,n}(t) = [g(t)]^{-\frac{1}{2}} \epsilon_n^{-\frac{1}{2}} \iint_{|y| > a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dZ_n(x,y)$$

so that

$$||Y_n - Y_{0,n}|| \leq \epsilon_n^{-\frac{1}{2}} ||g^{-\frac{1}{2}}|| \cdot \left| \iint_{|y| > a_n} y K\left(\frac{\cdot-x}{\epsilon_n}\right) dZ_n(x,y) \right|.$$

By assumption,

$$||g^{-1/2}|| < \infty ,$$

and thus it suffices to prove that

$$(\log n)^{1/2} \sup_{0 \leq t \leq 1} \left| \epsilon_n^{-1/2} \iint_{|y| > a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dZ_n(x,y) \right| \xrightarrow{P} 0 .$$

Now

$$\begin{aligned} (3.2.12) \quad & (\log n)^{1/2} \epsilon_n^{-1/2} \iint_{|y| > a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dZ_n(x,y) \\ &= (\log n)^{1/2} (n\epsilon_n)^{-1/2} \sum_{i=1}^n \left\{ Y_i I_{\{|y| > a_n\}} (Y_i) K\left(\frac{t-X_i}{\epsilon_n}\right) \right. \\ & \quad \left. - E Y_i I_{\{|y| > a_n\}} K\left(\frac{t-X_i}{\epsilon_n}\right) \right\} \\ &= \sum_{i=1}^n X_{n,i}(t) = U_n(t) , \end{aligned}$$

say, where $X_{n,i}(t)$, $i = 1, \dots, n$ are i.i.d. with

$$EX_{n,i}(t) = 0$$

for each $t \in [0,1]$. Thus

$$(3.2.13) \quad EU_n^2(t) = \sum_{i=1}^n EX_{n,i}^2(t)$$

and

$$\begin{aligned} (3.2.14) \quad & EX_{n,i}^2(t) \\ &\leq (\log n) (n\epsilon_n^{-1}) E Y_i^2 I_{\{|y| > a_n\}} (Y_i) K^2\left(\frac{t-X_i}{\epsilon_n}\right) \\ &\leq \bar{k} (\log n) (n\epsilon_n^{-1}) E Y_i^2 I_{\{|y| > a_n\}} (Y_i) \end{aligned}$$

where

$$\bar{k} = \sup_{-A \leq u \leq A} k^2(u),$$

and A is as defined in Theorem 3.2.1.

Combining (3.2.13) and (3.2.14) yields

$$\begin{aligned} E \left\{ \sum_{i=1}^n X_{n,i}(t) \right\}^2 &\leq \bar{k} (\log n) \epsilon_n^{-1} E Y_i^2 I_{\{|Y| > a_n\}}(Y_i) \\ &= \bar{k} (\log n) \epsilon_n^{-1} \int_{|y| > a_n} y^2 f_Y(y) dy \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$ by A1. This implies that

$$(3.2.15) \quad U_n(t) \xrightarrow{P} 0$$

for $0 \leq t \leq 1$.

In order to show that

$$(3.2.16) \quad ||U_n|| = \sup_{0 \leq t \leq 1} U_n(t) \xrightarrow{P} 0$$

we note that $U_n(t)$ is an element of the space $D[0,1]$ of right continuous functions with left hand limits for each n , and that, if we show that U_n converges weakly to the zero element of $D[0,1]$, then (3.2.16) will follow, since $||\cdot||$ is a continuous functional on $D[0,1]$. Since (3.2.15) implies

$$(U_n(t_1), U_n(t_2), \dots, U_n(t_k)) \xrightarrow{D} 0$$

in R^k for distinct points t_1, t_2, \dots, t_k of $[0,1]$, it suffices to verify the

following moment condition to show weak convergence of U_n (Billingsley, (1968), Th. 15.6):

$$\begin{aligned} E\{|U_n(t) - U_n(t_1)| |U_n(t_2) - U_n(t)|\} \\ \leq B(t_2 - t_1)^2 \end{aligned}$$

where B is a constant.

By the Schwarz inequality,

$$\begin{aligned} E\{|U_n(t) - U_n(t_1)| \cdot |U_n(t_2) - U_n(t)|\} \\ \leq \{E[U_n(t) - U_n(t_1)]^2 \cdot E[U_n(t_2) - U_n(t)]^2\}^{1/2}. \end{aligned}$$

Defining

$$G_n(u, s, X) = K\left(\frac{u-X}{\epsilon_n}\right) - K\left(\frac{s-X}{\epsilon_n}\right),$$

we have

$$\begin{aligned} & \{E[U_n(t) - U_n(t_1)]^2\} \\ &= (\log n) (n\epsilon_n)^{-1} \{E[\sum_{i=1}^n [Y_i I_{\{|y| > a_n\}}(Y_i) G_n(t, t_1, X_i) \\ & \quad - EY_i I_{\{|y| > a_n\}}(Y_i) G_n(t, t_1, X_i)]^2\} \\ &= (\log n) (n\epsilon_n)^{-1} \{ \sum_{i=1}^n E[Y_i I_{\{|y| > a_n\}}(Y_i) G_n(t, t_1, X_i) \\ & \quad - EY_i I_{\{|y| > a_n\}}(Y_i) G_n(t, t_1, X_i)]^2\} \\ &\leq (\log n) (n\epsilon_n)^{-1} \{ \sum_{i=1}^n EY_i^2 I_{\{|y| > a_n\}}(Y_i) G_n^2(t, t_1, X_i) \}. \end{aligned}$$

Since K has a bounded derivative on $[-A, A]$, it satisfies a Lipschitz condition:

$$|K(u) - K(s)| \leq B_1 |u-s|$$

where B_1 is a constant. Thus

$$\begin{aligned} & \{E[U_n(t) - U_n(t_1)]^2\}^{1/2} \\ & \leq B_1^{1/2} (\log n)^{1/2} \epsilon_n^{-3/2} |t-t_1| \{EY_1^2 I_{\{|y|>a_n\}}(Y_1)\}^{1/2} \\ & = B_1 (\log n)^{1/2} \epsilon_n^{-3/2} |t-t_1| \left\{ \int_{|y|>a_n} y^2 f_Y(y) dy \right\}^{1/2}. \end{aligned}$$

Applying the same argument to

$$E[U_n(t_2) - U_n(t)]^2$$

yields

$$\begin{aligned} & E\{|U_n(t) - U_n(t_1)| |U_n(t_2) - U_n(t)|\} \\ & \leq B_1^2 \log n \epsilon_n^{-3} |t-t_1| \cdot |t_2-t| \int_{|y|>a_n} y^2 f_Y(y) dy \\ & \leq c(t_1-t_2)^2 \end{aligned}$$

by A1 and using the fact that $t_1 \leq t \leq t_2$. The moment condition is therefore satisfied, and the result follows. \square

Before going on to Lemma 3.2.5, we state the useful integration by parts formula for Riemann-Stieltjes integrals on rectangles of \mathbb{R}^2 .

Let f and g be two functions defined on $[0,1]^2$. If all of the integrals below exist and are finite, then we have

$$\begin{aligned} (3.2.17) \quad & \int_0^1 \int_0^1 f(x,y) dg(x,y) = \int_0^1 \int_0^1 g(x,y) df(x,y) \\ & + \int_0^1 f(1,y) dg(1,y) - \int_0^1 f(0,y) dg(0,y) \end{aligned}$$

$$+ \int_0^1 g(x,1) df(x,1) - \int_0^1 g(x,0) df(x,0) .$$

We note that if $g(x,y)$ is a Wiener process on $[0,1]^2$ and $f(x,y)$ is a measurable function on $[0,1]^2$ such that $\int_0^1 \int_0^1 f(x,y) dg(x,y)$ exists, then (3.2.17) remains valid provided the integrals on the right hand side of (3.2.17) also exist.

3.2.5 Lemma. If K is absolutely continuous on $[-A,A]$ and zero outside $[-A,A]$, then

$$||Y_{1,n} - Y_{2,n}|| = O(a_n \epsilon_n^{-1/2} n^{-1/6} (\log n)^{3/2})$$

a.s.

Proof. First we note that the random pair

$$(3.2.18) \quad (X', Y') = T(X, Y) ,$$

where $T: \mathbb{R}^2 \rightarrow [0,1]^2$ is defined by (3.2.7), is jointly uniformly distributed on $[0,1]^2$, X' and Y' are independent, and $Z_n(T^{-1}(x', y'))$, $0 \leq x', y' \leq 1$, is the empirical process of (X', Y') (Rosenblatt (1952)). Theorem 3.2.3 thus applies to (X', Y') , and we may conclude that

$$\begin{aligned} \sup_{0 \leq x', y' \leq 1} |B_n(x', y') - Z_n(T^{-1}(x', y'))| \\ = O(n^{-1/6} (\log n)^{3/2}) \quad \text{a.s.}, \end{aligned}$$

or, equivalently,

$$\begin{aligned} (3.2.19) \quad \sup_{x, y \in \mathbb{R}} |B_n(T(x, y)) - Z_n(x, y)| \\ = O(n^{-1/6} (\log n)^{3/2}) \quad \text{a.s.} \end{aligned}$$

Applying the integration by parts formula (3.2.17), we have

$$\begin{aligned}
 (3.2.20) \quad & \int_{|y| \leq a_n} \int y K\left(\frac{t-x}{\epsilon_n}\right) dZ_n(x, y) \\
 & \leq \int_{u=-A}^A \int_{y=-a_n}^{a_n} y K(u) dZ_n(t-\epsilon_n u, y) \\
 & = \int_{-A}^A \int_{-a_n}^{a_n} Z_n(t-\epsilon_n u, y) d[yK(u)] \\
 & \quad + a_n \int_{u=-A}^A Z_n(t-\epsilon_n u, a_n) dK(u) \\
 & \quad + a_n \int_{u=-A}^A Z_n(t-\epsilon_n u, -a_n) dK(u) \\
 & \quad + k(A) \int_{y=-a_n}^{a_n} y dZ_n(t-\epsilon_n A, y) \\
 & \quad - k(-A) \int_{y=-a_n}^{a_n} y dZ_n(t+\epsilon_n A, y) .
 \end{aligned}$$

The second to last integral above may be written, using ordinary one-dimensional) integration by parts,

$$\begin{aligned}
 & \int_{-a_n}^{a_n} y dZ_n(t-\epsilon_n A, y) \\
 & = \int_{-a_n}^{a_n} Z_n(t-\epsilon_n A, y) dy + a_n Z_n(t-\epsilon_n A, a_n) - a_n Z_n(t-\epsilon_n A, -a_n) ,
 \end{aligned}$$

and similarly for the last integral on the right hand side of (3.2.20).

By using a similar argument, we obtain (where the integrals are defined in the L_2 sense)

$$\begin{aligned}
(3.2.21) \quad & \int_{|y| \leq a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dB_n(T(x, y)) \\
&= \int_{u=-A}^A \int_{y=-a_n}^{a_n} B_n(T(t-\epsilon_n u, y)) d[yK(u)] \\
&\quad + a_n \int_{u=-A}^A B_n(T(t-\epsilon_n u, a_n)) dK(u) \\
&\quad + a_n \int_{u=-A}^A B_n(T(t-\epsilon_n u, -a_n)) dK(u) \\
&\quad + k(a) \left\{ \int_{y=-a_n}^{a_n} B_n(T(t-\epsilon_n A, y)) dy \right. \\
&\quad + a_n B_n(T(t-\epsilon_n A, a_n)) \\
&\quad + a_n B_n(T(t-\epsilon_n A, -a_n)) \} \\
&\quad - K(1-A) \left\{ \int_{y=-a_n}^{a_n} B_n(T(t+\epsilon_n A, y)) dy \right. \\
&\quad + a_n B_n(T(t+\epsilon_n A, a_n)) \\
&\quad + a_n B_n(T(t+\epsilon_n A, -a_n)) \}
\end{aligned}$$

Subtracting (3.2.21) from (3.2.20) and using (3.2.19) and the assumption that K is absolutely continuous on $[-A, A]$, we obtain

$$\begin{aligned}
(3.2.22) \quad & \epsilon_n^{1/2} |g_n(t)|^{1/2} |Y_{1,n}(t) - Y_{2,n}(t)| \\
&= O(n^{-1/6} (\log n)^{3/2}) \\
&\quad \{ 4a_n \int_{-A}^A |K'(u)| du + 4a_n [K(A) + K(-A)] \} \quad \text{a.s.} \\
&= O(a_n n^{-1/6} (\log n)^{3/2})
\end{aligned}$$

since

$$\int_{-A}^A |K'(u)| du < \infty .$$

Thus, since $\|g_n^{-1/2}\|$ is a bounded sequence by assumption,

$$\|Y_{1,n} - Y_{2,n}\| = O_p(a_n \epsilon_n^{-1/2} n^{-1/6} (\log n)^{3/2}) ,$$

and the proof is complete. \square

We may write the sequence of Brownian bridges $\{B_n\}$ of Theorem 3.2.3 as

$$(3.2.23) \quad B_n(x,y) = W_n(x,y) - xyW_n(1,1) ,$$

$0 \leq x, y \leq 1$, where $\{W_n\}$ is a sequence of independent Wiener processes on $[0,1]^2$ (Révész (1976)). The next lemma shows that, for our purposes, the only significant part of (3.2.23) is $W_n(x,y)$.

3.2.6 Lemma. If A4 holds, then

$$\|Y_{2,n} - Y_{3,n}\| = O_p(\epsilon_n^{1/2}) .$$

Proof. By definition of $Y_{2,n}$ and $Y_{3,n}$, we have

$$\begin{aligned} |Y_{2,n}(t) - Y_{3,n}(t)| \\ = |[g_n(t)]^{-1/2} \epsilon_n^{-1/2} \iint_{|y| \leq a_n} y K\left(\frac{t-x}{\epsilon_n}\right) f(x,y) dx dy| \cdot |W_n(1,1)| , \end{aligned}$$

since the Jacobian of the transformation T is $f(x,y)$. Thus

$$\begin{aligned} \epsilon_n^{-1/2} \|Y_{2,n} - Y_{3,n}\| \\ \leq |W_n(1,1)| \|g_n^{-1/2}\| . \end{aligned}$$

$$\begin{aligned}
& \cdot \sup_{u \leq t \leq 1} \epsilon_n^{-1} \iint_{|y| \leq a_n} |y K(\frac{t-x}{\epsilon_n})| f(x,y) dx dy \\
& \leq |W_n(1,1)| \cdot ||g_n^{-\frac{1}{2}}|| \\
& \cdot \sup_{u \leq t \leq 1} \epsilon_n^{-1} \int [\int |y| f(x,y) dy] |K(\frac{t-x}{\epsilon_n})| dx .
\end{aligned}$$

By A4,

$$h(x) = \int |y| f(x,y) dy$$

is a bounded function and $||g_n^{-\frac{1}{2}}||$ is a bounded sequence, so that for some constant M we have

$$\begin{aligned}
& \epsilon_n^{-\frac{1}{2}} ||Y_{2,n} - Y_{3,n}|| \\
& \leq |W_n(1,1)| M \epsilon_n^{-1} \int |K(\frac{t-x}{\epsilon_n})| dx \\
& = |W_n(1,1)| M \int |K(u)| du \\
& = O_p(1) .
\end{aligned}$$

Thus

$$||Y_{2,n} - Y_{3,n}|| = O_p(\epsilon_n^{\frac{1}{2}})$$

and the proof is complete. \square

3.2.7 Lemma. Under the assumptions of Theorem 3.2.1,

$$||Y_{4,n} - Y_{5,n}|| = O_p(\epsilon_n)^{\frac{1}{2}}.$$

Proof. By definition,

$$\begin{aligned}
(3.2.24) \quad & |Y_{4,n}(t) - Y_{5,n}(t)| \\
&= \epsilon_n^{-1/2} \left| \int \left\{ \left[\frac{g_n(x)}{g_n(t)} \right]^{1/2} - 1 \right\} K\left(\frac{t-x}{\epsilon_n}\right) dW(x) \right| \\
&= \epsilon_n^{-1/2} \left| \int_{-A}^A \left\{ \left[\frac{g_n(t-u\epsilon_n)}{g_n(t)} \right]^{1/2} - 1 \right\} k(u) dW(t-u\epsilon_n) \right|.
\end{aligned}$$

By using integration by parts and the assumptions that $g_n^{1/2}$ and K are absolutely continuous, we may bound the integral on the extreme right hand side of (3.2.24) with

$$\begin{aligned}
(3.2.25) \quad & \left| \epsilon_n^{-1/2} \int_{-A}^A W(t-u\epsilon_n) \frac{\partial}{\partial u} \left\{ \left[\left(\frac{g_n(t-u\epsilon_n)}{g_n(t)} \right)^{1/2} - 1 \right] K(u) \right\} du \right| \\
&+ \left| \epsilon_n^{-1/2} K(A) W(t-A\epsilon_n) \left\{ \left[\frac{g_n(t-A\epsilon_n)}{g_n(t)} \right]^{1/2} - 1 \right\} \right| \\
&+ \left| \epsilon_n^{-1/2} K(-A) W(t+A\epsilon_n) \left\{ \left[\frac{g_n(t+A\epsilon_n)}{g_n(t)} \right]^{1/2} - 1 \right\} \right| \\
&= J_{1,n}(t) + J_{2,n}(t) + J_{3,n}(t),
\end{aligned}$$

say. We will show that the supremum over $[0,1]$ of each of these three terms is in $O_p(\epsilon_n^{1/2})$, thus completing the proof.

First of all, note

$$\begin{aligned}
&\epsilon_n^{-1/2} |J_{2,n}| \\
&\leq K(A) \sup_{0 \leq t \leq 1} |W(t-A\epsilon_n)| \sup_{0 \leq t \leq 1} \epsilon_n^{-1} \left| \left[\frac{g_n(t-A\epsilon_n)}{g_n(t)} \right]^{1/2} - 1 \right|
\end{aligned}$$

and

$$\sup_{0 \leq t \leq 1} |W(t-A\epsilon_n)| = O_p(1).$$

Now

$$\begin{aligned}
 & \sup_{0 \leq t \leq 1} \epsilon_n^{-1} \left| \left[\frac{g_n(t - A\epsilon_n)}{g_n(t)} \right]^{\frac{1}{2}} - 1 \right| \\
 &= \sup_{0 \leq t \leq 1} \epsilon_n^{-1} \left\{ \frac{|[g_n(t - A\epsilon_n)]^{\frac{1}{2}} - [g_n(t)]^{\frac{1}{2}}|}{|g_n(t)|^{\frac{1}{2}}} \right\} \\
 &\leq ||g_n^{-\frac{1}{2}}|| \epsilon_n^{-1} \sup_{0 \leq t \leq 1} |[g_n(t - A\epsilon_n)]^{\frac{1}{2}} - [g_n(t)]^{\frac{1}{2}}|.
 \end{aligned}$$

By assumption, $||g_n^{-\frac{1}{2}}||$ is a bounded sequence, and by the mean value theorem,

$$\begin{aligned}
 & \epsilon_n^{-1} |[g_n(t - A\epsilon_n)]^{\frac{1}{2}} - [g_n(t)]^{\frac{1}{2}}| \\
 & \leq \frac{1}{2} \epsilon_n^{-1} |g_n(t - A\epsilon_n) - g_n(t)| \cdot |x_n(t, A)|^{-\frac{1}{2}}
 \end{aligned}$$

where $x_n(t, A)$ is between $g_n(t - A\epsilon_n)$ and $g_n(t)$. Applying the mean value theorem to g_n yields

$$\begin{aligned}
 & g_n(t - A\epsilon_n) - g_n(t) \\
 &= A\epsilon_n g'_n(t_n(t, A)),
 \end{aligned}$$

where $t_n(t, A)$ is between $t - A\epsilon_n$ and t . Thus

$$\begin{aligned}
 & \epsilon_n^{-\frac{1}{2}} ||J_{2,n}|| \\
 & \leq K(A) \sup_{-A \leq t \leq 1} |W(t)| \cdot \frac{A}{2} \sup_{0 \leq t \leq 1} \frac{|g'_n(t_n(t, A))|}{|x_n(t, A)|^{\frac{1}{2}}} \\
 &= O_p(1)
 \end{aligned}$$

since g'_n is uniformly bounded and g_n is bounded away from zero, by

assumption, and thus

$$||J_{2,n}|| = O_p(\epsilon_n^{1/2}).$$

A similar argument shows that

$$||J_{3,n}|| = O_p(\epsilon_n^{1/2}),$$

so we now consider $J_{1,n}$.

Carrying out the differentiation in the integrand of $J_{1,n}$, we have

$$\begin{aligned} \epsilon_n^{-1/2} J_{1,n}(t) &= \left| \epsilon_n^{-1} \int_{-A}^A W(t-u\epsilon_n) \left\{ K'(u) \left[\left(\frac{g_n(t-u\epsilon_n)}{g_n(t)} \right)^{1/2} - 1 \right] \right\} du \right. \\ &\quad \left. - \frac{1}{2} \int_{-A}^A W(t-u\epsilon_n) K(u) \left(\frac{g_n(t-u\epsilon_n)}{g_n(t)} \right)^{-1/2} \left(\frac{g'_n(t-u\epsilon_n)}{g_n(t)} \right) du \right| \\ &= |C_{1,n}(t) - C_{2,n}(t)|, \end{aligned}$$

say. Now the non-stochastic terms in the integrand of $C_{2,n}$ are uniformly bounded in their arguments and in n , by assumption. We therefore have

$$||C_{2,n}|| \leq C_2 \int_{-A}^A |W(t-u\epsilon_n)| du = O_p(1)$$

where C_2 is a constant. For $C_{1,n}$, apply the same argument used in considering $J_{2,n}$ to conclude that

$$\sup_{0 \leq t \leq 1} \epsilon_n^{-1} \left| \left(\frac{g_n(t-u\epsilon_n)}{g_n(t)} \right)^{1/2} - 1 \right| \leq C_1 |u|$$

where C_1 is a constant. Then

$$||C_{1,n}|| \leq C_1 \sup_{0 \leq t \leq 1} \int_{-A}^A |W(t-u\epsilon_n)K'(u)u| du$$

$$= O_p(1),$$

and the proof is complete. \square

We now use the results proved thus far in showing that $Y_{0,n}$ and $Y_{1,n}$ are sufficiently close to one another.

3.2.8 Lemma. Under the assumptions of Theorem 3.2.1,

$$||Y_{0,n} - Y_{1,n}|| = O_p((\log n)^{-\frac{1}{2}}).$$

Proof. We must show that

$$\sup_{0 \leq t \leq 1} \left\{ \left| [g(t)]^{-\frac{1}{2}} - [g_n(t)]^{-\frac{1}{2}} \right| \epsilon_n^{-\frac{1}{2}} \left| \iint_{|y| \leq a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dZ_n(x,y) \right| \right\}$$

$$= O_p((\log n)^{-\frac{1}{2}}).$$

By the preceding four lemmas and Theorem 3.2.2,

$$(\log n)^{\frac{1}{2}} [||Y_{1,n}|| [\lambda(K)]^{-\frac{1}{2}} - d_n]$$

converges in distribution to some random variable, and is therefore a $O_p(1)$ sequence. Since, by definition,

$$d_n = O((\log n)^{\frac{1}{2}}),$$

we have

$$||Y_{1,n}|| = O_p((\log n)^{\frac{1}{2}}),$$

and since $|g_n^{-\frac{1}{2}}|$ is a bounded sequence, we have

$$\sup_{0 \leq t \leq 1} \epsilon_n^{-1/2} \left| \iint_{|g| \leq a_n} y K\left(\frac{t-x}{\epsilon_n}\right) dZ_n(x, y) \right|$$

$$= O_p((\log n)^{1/2}) .$$

Thus it suffices to prove

$$(\log n) \|g_n^{-1/2} - g^{-1/2}\| \rightarrow 0$$

as $n \rightarrow \infty$. By the mean value theorem,

$$|g_n^{-1/2} - g^{-1/2}| = |g_n - g| \cdot |h_n^{-3/2}|$$

where h_n is between g_n and g . Since g_n and g are bounded away from zero, $\|h_n^{-3/2}\|$ is a bounded sequence, and since, by A3,

$$(\log n) \|g_n - g\| \rightarrow 0 ,$$

the result is proved. \square

Since $m_n^*(t)$ is an asymptotically unbiased estimator of $m^*(t) = m(t)f(t)$, it is natural to seek conditions under which $Em_n^*(t)$ may be replaced by $m^*(t)$ in Theorem 3.2.1. Define the process

$$Y'_n(t) = \frac{(n\epsilon_n)^{1/2} [m_n^*(t) - m^*(t)]}{[s(t)f(t)]^{1/2}} .$$

Then we have the following corollary to Theorem 3.2.1.

3.2.9 Corollary. Suppose all the conditions of Theorem 3.2.1 hold and in addition

$$\epsilon_n = n^{-\delta} , \quad 1/5 < \delta < 1/2 ,$$

K satisfies

$$\int uK(u)du = 0 ,$$

$$\int u^2K(u)du < \infty$$

and the function

$$m^*(t) = m(t)f(t) = \int yf(t,y)dy$$

has bounded, continuous 1st and 2nd derivatives. Then the conclusion of Theorem 3.2.1 holds, with Y'_n replacing Y_n .

Proof. According to the remark at the beginning of the proof of Theorem 3.2.1, it suffices to show

$$||Y'_n - Y_n|| = O_p((\log n)^{-\frac{1}{2}}) .$$

But

$$||Y'_n - Y_n|| \leq (n\epsilon_n)^{\frac{1}{2}} ||m^* - Em_n^*|| \cdot ||g^{-\frac{1}{2}}|| .$$

By assumption,

$$||g^{-\frac{1}{2}}|| < \infty$$

and we know that, under the assumptions on m^* and K ,

$$||m^* - Em_n^*|| = O(\epsilon_n^2) .$$

Since

$$\epsilon_n = n^{-\delta} , \delta > 1/5 ,$$

then

$$\epsilon_n^2 (n\epsilon_n)^{\frac{1}{2}} (\log n)^{\frac{1}{2}} = (n\epsilon_n^5 \log n)^{\frac{1}{2}} \rightarrow 0 ,$$

and the proof is complete. □

Based on this corollary, we may construct a confidence band for $m(t)$, $0 \leq t \leq 1$ as follows. Using the asymptotic distribution, we have

$$P\left\{(2\delta \log n)^{\frac{1}{2}} \left[\frac{\sup |Y'_n(t)|}{[\lambda(K)]^{\frac{1}{2}}} - d_n \right] < C(\alpha) \right\} \approx 1 - \alpha$$

where

$$C(\alpha) = \log 2 - \log |\log (1-\alpha)|.$$

Inverting the above expression in the usual way, we obtain as a $(1-\alpha) \times 100\%$ confidence band for $m(t)$:

$$(3.2.26) \quad \bar{m}_n(t) \pm (n\epsilon_n)^{-\frac{1}{2}} \left[\frac{s(t)}{f(t)} \right]^{\frac{1}{2}} \left[\frac{c(\lambda)}{(2\delta \log n)^{\frac{1}{2}}} + d_n \right] [\lambda(K)]^{\frac{1}{2}},$$

$$0 \leq t \leq 1.$$

3.3 Uniform Consistency of m_n and \bar{m}_n .

We saw in Corollary 3.2.9 that the sequence of random variables

$$(\log n)^{\frac{1}{2}} [(\epsilon_n)^{\frac{1}{2}} \sup_{0 \leq t \leq 1} \left| \frac{m_n^*(t) - m(t)f(t)}{[s(t)f(t)]^{\frac{1}{2}}} \right| - d_n]$$

converges in distribution, and is thus a $O_p(1)$ sequence. We employ this fact to show the uniform consistency of m_n^* , and specify a rate of convergence.

3.3.1 Lemma. Under the conditions of Corollary 3.2.9,

$$(3.3.1) \quad \sup_{0 \leq t \leq 1} |m_n^*(t) - m(t)f(t)| = O_p[(\log n)^{\frac{1}{2}} (n\epsilon_n)^{-\frac{1}{2}}].$$

Proof. By definition,

$$d_n = O((\log n)^{\frac{1}{2}}),$$

and thus

$$\begin{aligned}
 (n\epsilon_n)^{\frac{1}{2}} \sup_{0 \leq t \leq 1} \left| \frac{m_n^*(t) - m(t)f(t)}{[s(t)f(t)]^{\frac{1}{2}}} \right| \\
 = O_p((\log n)^{-\frac{1}{2}}) + O((\log n)^{\frac{1}{2}}) \\
 = O_p((\log n)^{\frac{1}{2}}) .
 \end{aligned}$$

Now, using the assumption that $g(t) = s(t)f(t)$ is bounded away from zero, the conclusion follows. \square

We now use the preceding lemma to show uniform consistency of m_n and \bar{m}_n .

3.3.2 Theorem. Under the conditions of Corollary 3.2.9, we have

$$(3.3.2) \quad ||\bar{m}_n - m|| = O_p[(\log n)^{\frac{1}{2}}(n\epsilon_n)^{-\frac{1}{2}}] ,$$

$$(3.3.3) \quad ||m_n - m|| = O_p[(\log n)^{\frac{1}{2}}(n\epsilon_n)^{-\frac{1}{2}}] .$$

Proof. Note that

$$||\bar{m}_n - m|| \leq ||f_1^{-1}|| \cdot ||m_n^* - m^*||$$

where

$$m^*(t) = f(t)m(t) .$$

By assumption, f is bounded away from zero on $[0,1]$, and thus $||f^{-1}|| < \infty$. An application of Lemma 3.3.1 thus proves (3.3.2).

For (3.3.3), note

$$||m_n - m|| = \left| \left| \frac{m_n^* f - f_n m_n^*}{f_n f} \right| \right|$$

$$\leq \left| \left| \frac{m_n^* f - m_n^* f_n}{f_n f} \right| \right| + \left| \left| \frac{m_n^* f_n - m_n^* f_n}{f_n f} \right| \right|$$

$$= A + B ,$$

say. Now

$$B = \left| \left| \bar{m}_n - m \right| \right| = O_p[(\log n)^{1/2} (n\epsilon_n)^{-1/2}]$$

by (3.3.2). Further,

$$\begin{aligned} A &\leq \left| \left| \frac{m_n^*}{f_n} \right| \right| \cdot \left| \left| \frac{f_n - f}{f} \right| \right| \\ &= \left| \left| \frac{m_n^*}{f_n} \right| \right| O_p[(\log n)^{1/2} (n\epsilon_n)^{-1/2}] \end{aligned}$$

(Bickel and Rosenblatt (1973)). Since

$$\left| \left| \frac{m_n^*}{f_n} \right| \right| \leq \left| m_n^* \right| \cdot \left[\inf_{0 \leq t \leq 1} |f_n(t)| \right]^{-1}$$

and it is easily verified that $\left| m_n^* \right| \xrightarrow{P} \left| m^* \right|$, $\inf_{0 \leq t \leq 1} |f_n(t)| \rightarrow \inf_{0 \leq t \leq 1} |f(t)| > 0$, (3.3.3) follows. □

4. AN EXAMPLE, FURTHER RESEARCH

As we noted in the introductory chapter, if the density of X is known, then either the estimator m_n or \bar{m}_n may be used to estimate the regression function. Here we will summarize some results given in Chapter 2 which relate to the relative performance of m_n and \bar{m}_n in this case. We then present an example in which m_n and \bar{m}_n are computed from a set of simulated data.

4.1 The Estimators m_n and \bar{m}_n .

We first note that, according to Theorem 2.3.4, if the density function of X has, say, an interval for its support and is non-zero at the endpoints of the interval, then m_n is a consistent estimator at the endpoints, whereas \bar{m}_n is not. The implication of this for finite sample sizes is that \bar{m}_n is likely to display a bias near the endpoints of the X variable which m_n will not have.

According to Theorems 2.4.4 and 2.5.2, under appropriate conditions, both $m_n(x)$ and $\bar{m}_n(x)$ have asymptotic normal distributions with mean $m(x)$ (for kernel type estimators). However, the sequence of scaling constants required for unit asymptotic variance differs for the two estimators; for $m_n(x)$ it is $\{\sigma^2(x) \int K^2(u)du / (n\epsilon_n)f_1(x)\}^{1/2}$ and for $\bar{m}_n(x)$ it is $\{s(x) \int K^2(u)du / (n\epsilon_n)f_1(x)\}^{1/2}$. Since

$$\sigma^2(x) = s(x) - m^2(x) \leq s(x) ,$$

this indicates that \bar{m}_n may display more dispersion about m for finite sample sizes than m_n .

4.2 An Example.

In order to illustrate the behavior of the estimators in one specific case, we have computed m_n and \bar{m}_n for a set of artificial data. We have also computed the approximate confidence intervals given by (3.2.26) for m , based on \bar{m}_n . The results of the computations are depicted in Figures 1-6, and we have also shown a scatterplot of the data and the true regression function on each figure. The data consists of $n = 200$ points (X_i, Y_i) chosen independently with $X_i \sim U(-3, 2)$ and

$$Y_i = X_i^3/3 + X_i^2 + Z_i$$

where Z_i is a standard normal variable independent of X_i . Thus, for this data

$$m(x) = x^3/3 + x^2.$$

All calculations are for kernel type estimators with kernel function given by a standard normal density function, truncated at ± 3 and normalized so as to be a probability density.

Figures 1 and 2 show the estimators m_n and \bar{m}_n , respectively, with $\epsilon_n = n^{-.21}$ and Figures 3 and 4 show m_n and \bar{m}_n with slightly less smoothing, $\epsilon_n = n^{-.4}$. The previously discussed bias of \bar{m}_n is evident at the upper endpoint on Figures 2 and 4, although m_n and \bar{m}_n do not differ by very much at the lower endpoint. The difference in the asymptotic variances of m_n and \bar{m}_n does not manifest itself in this example, although \bar{m}_n in Figure 4 has a slightly more variable appearance than m_n in Figure 3.

Figures 5 and 6 show the approximate confidence bands given by (3.2.26) for $\alpha = .1$, and $\epsilon_n = n^{-.21}$ and $\epsilon_n = n^{-.4}$, respectively. The confidence bands (3.2.26) are asymptotically valid for any subinterval of $[-3, 2]$. In practice, however, one should consider these confidence

bands to be approximately valid only for intervals well within the support of X , since the earlier remarks on the endpoint bias of \bar{m}_n apply to the confidence bands also. These confidence bands were calculated using the true conditional second moment

$$s(t) = 1 + [t^3/3 + t^2]^2.$$

In practice, one would use an estimator of $s(t)$, e.g. the consistent estimator

$$s_n(t) = (n\epsilon_n)^{-1} \sum_{i=1}^n Y_i^2 K((t-X_i)/\epsilon_n).$$

4.3 Further Research.

Theorem 3.2.1 was proved for the process

$$Y_n(t) = \frac{(n\epsilon_n)^{1/2} [m_n^*(t) - Em_n^*(t)]}{[s(t)f(t)]^{1/2}}.$$

It should be possible to carry out a similar program for the process

$$V_n(t) = \frac{(n\epsilon_n)^{1/2} [m_n(t) - \frac{Em_n^*(t)}{Ef_n(t)}]}{[\sigma^2(t)/f(t)]^{1/2}}.$$

A first step in such a proof might be to show the equivalence of V_n to the process

$$V'_n(t) = [f(t)/f_n(t)] V_n(t)$$

(in the sense of $||V_n - V'_n|| = o_p((\log n)^{-1/2})$). Successive approximations, as in Theorem 3.2.1 would lead eventually to the equivalence of V'_n to the Wiener process of Theorem 3.2.2, and thus to the asymptotic distribution of the maximum absolute deviation of V_n .

We have not been able to carry out the technical details of the proof of such a theorem. However, if it were to be proved, one application would be a confidence band such as (3.2.26), but based on m_n instead of \bar{m}_n , and therefore narrower since m_n is asymptotically less variable than \bar{m}_n .

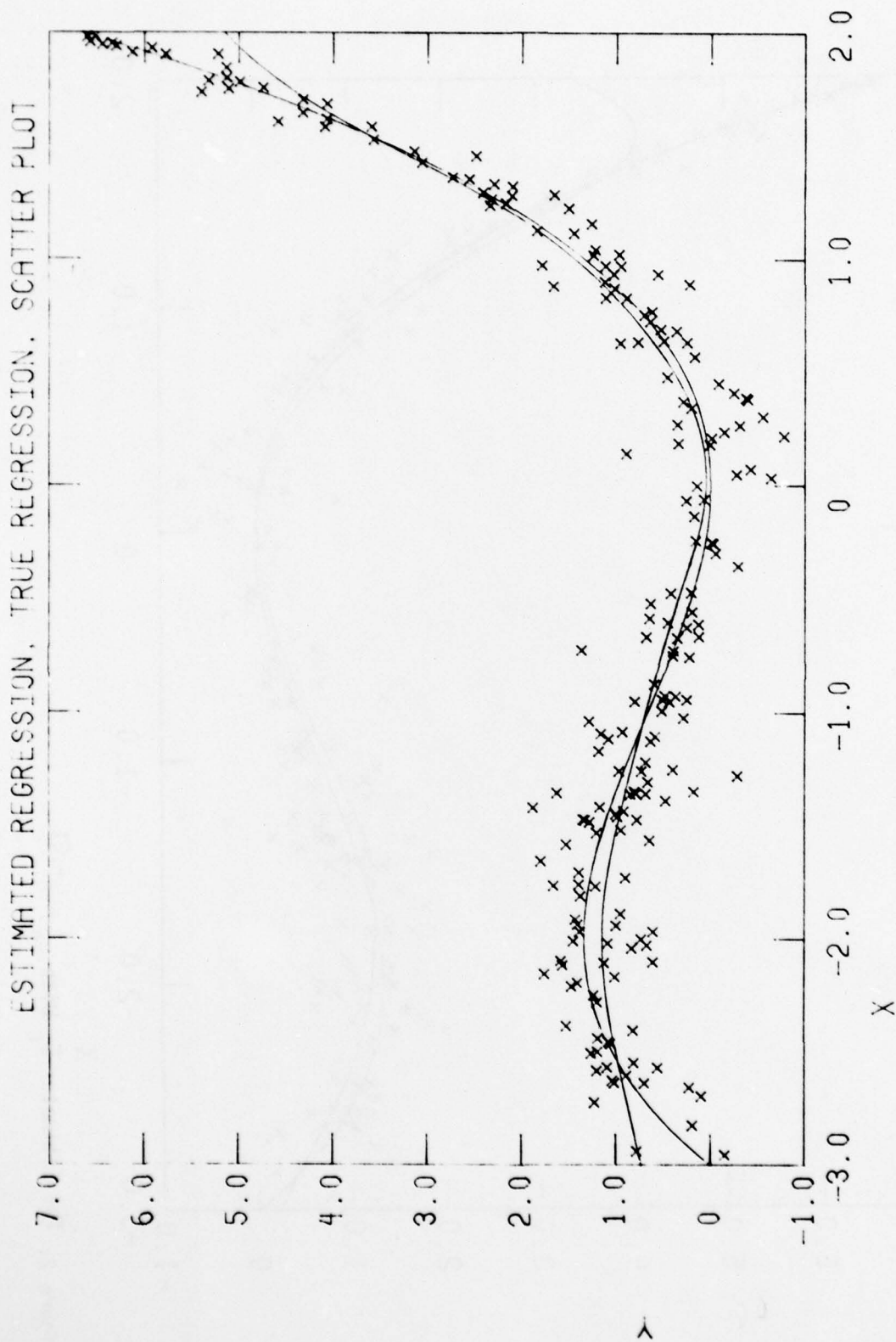


Figure 1. The Estimator m_n With $\epsilon_n = n^{-.21}$

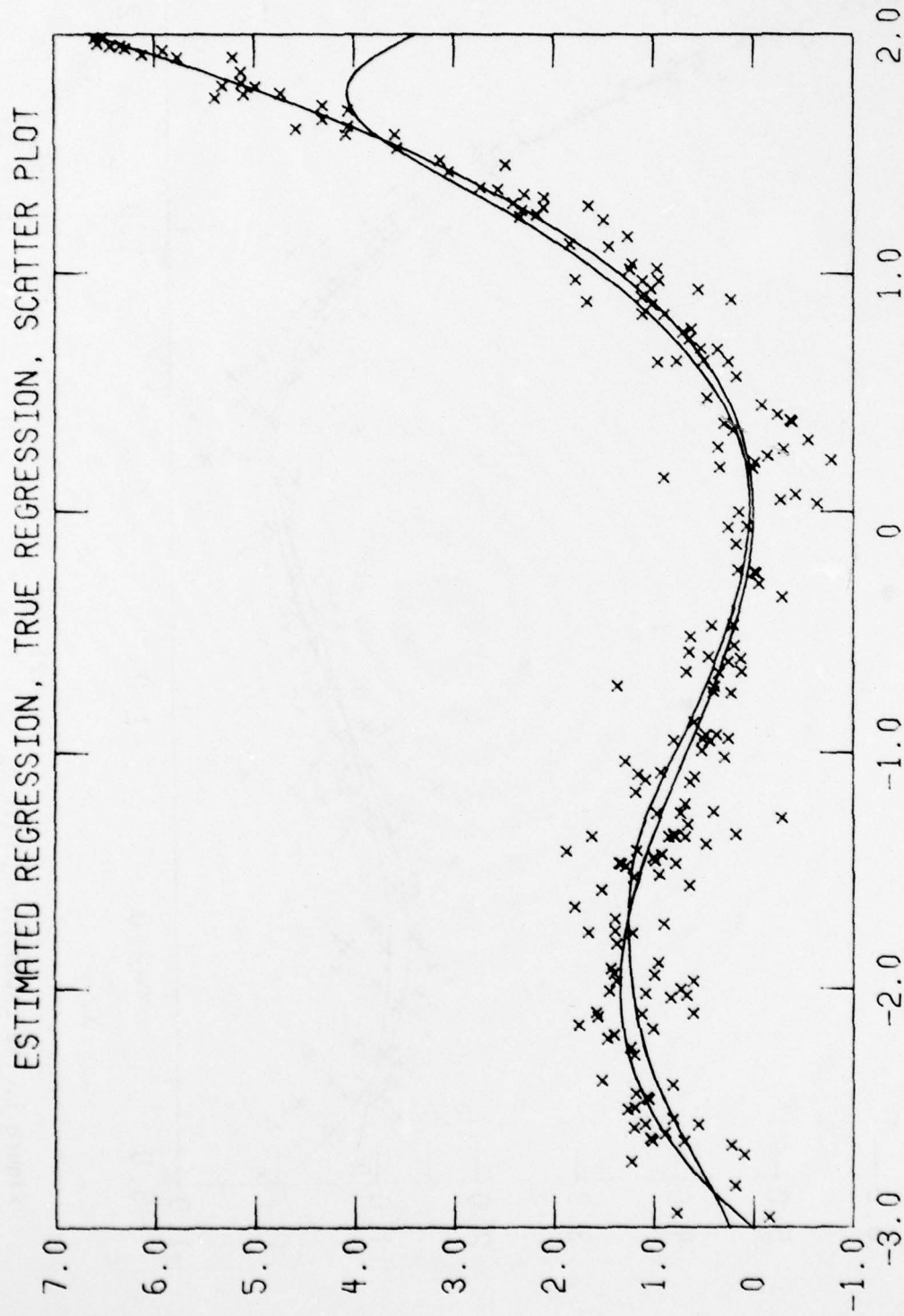


Figure 2. The Estimator \bar{m}_n With $\epsilon_n = n^{-.21}$

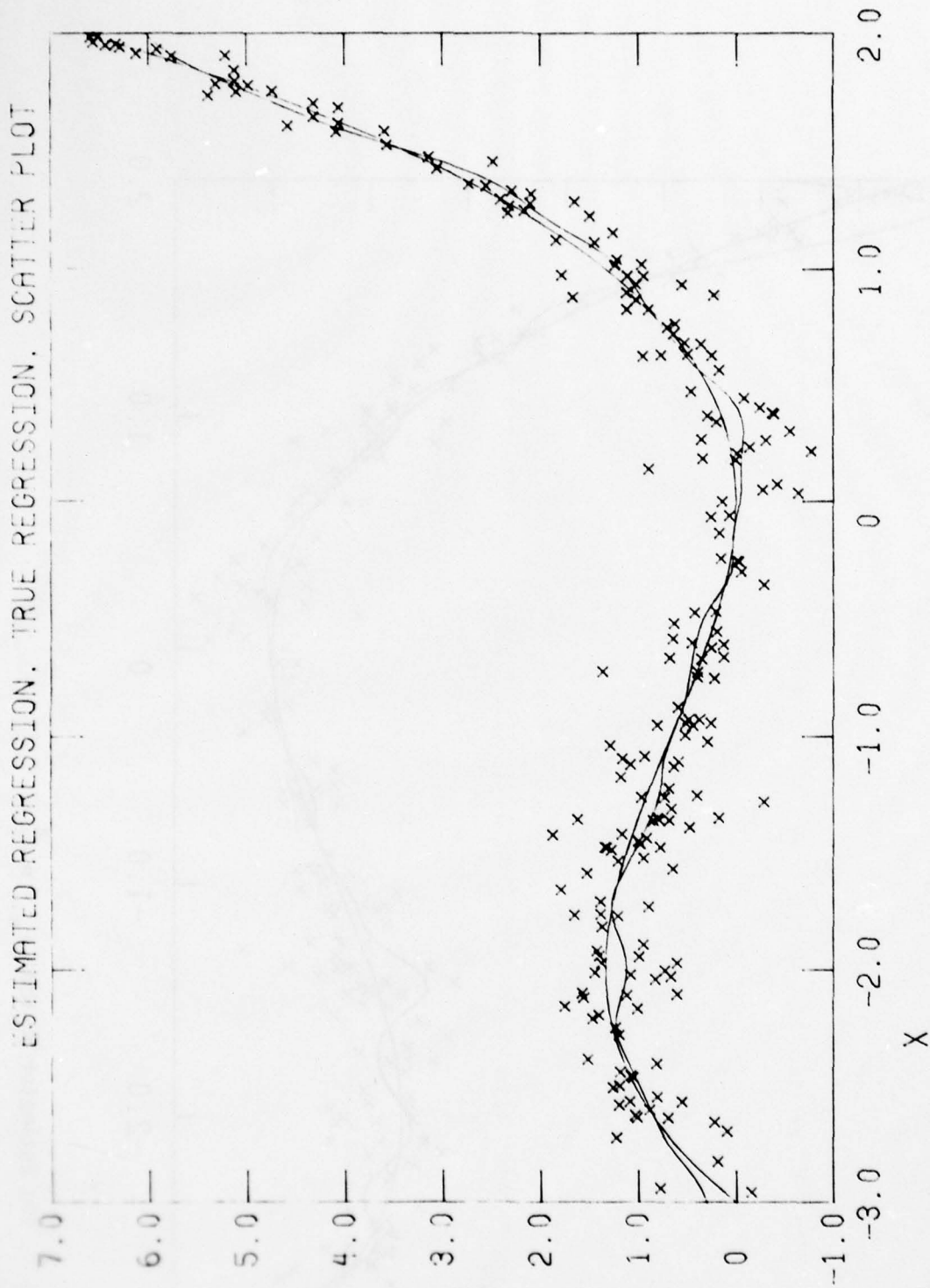


Figure 3. The Estimator m_n With $\epsilon_n = n^{-.4}$

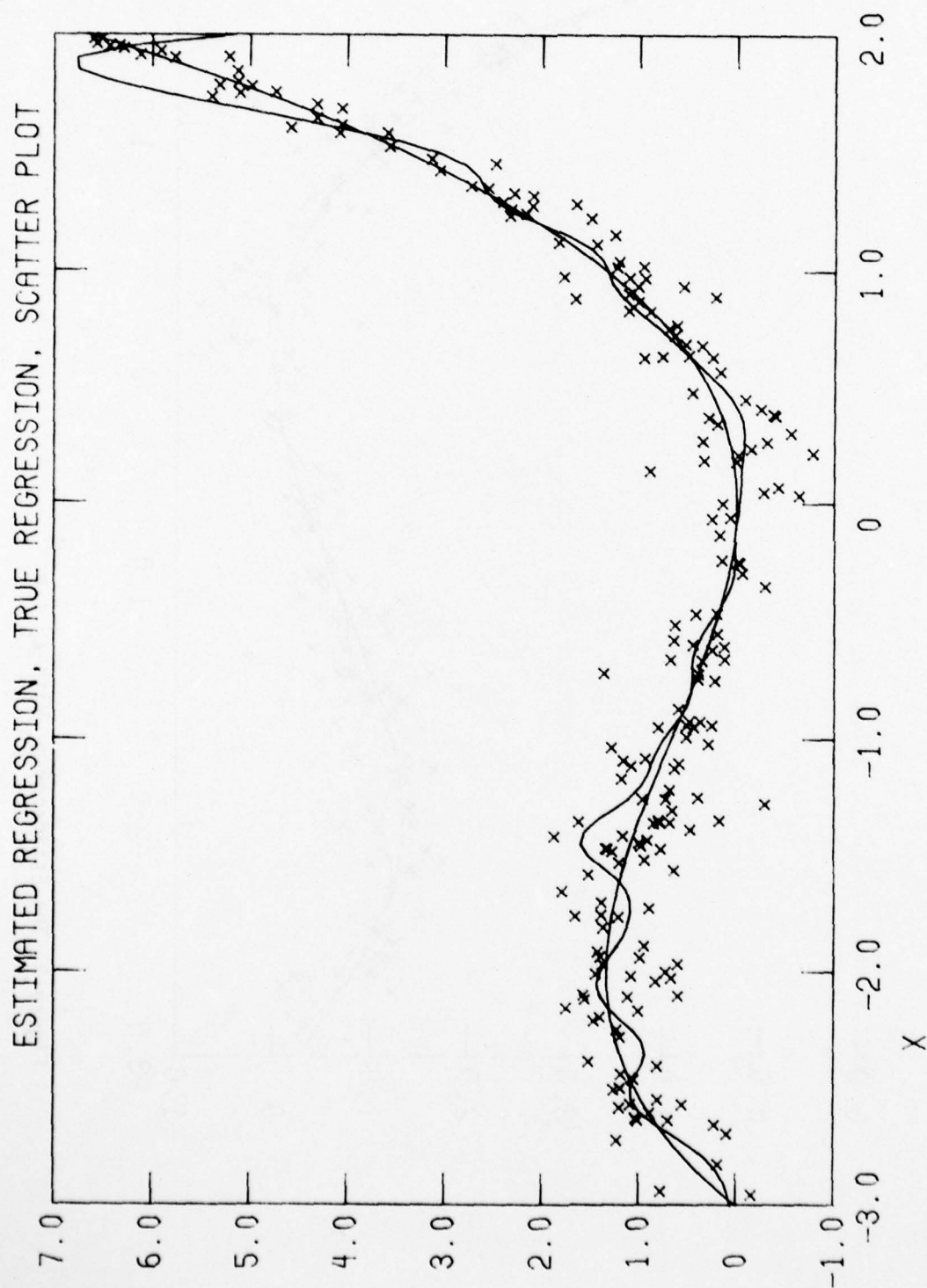


Figure 4. The Estimator \bar{m}_n With $\varepsilon_n = n^{-.4}$

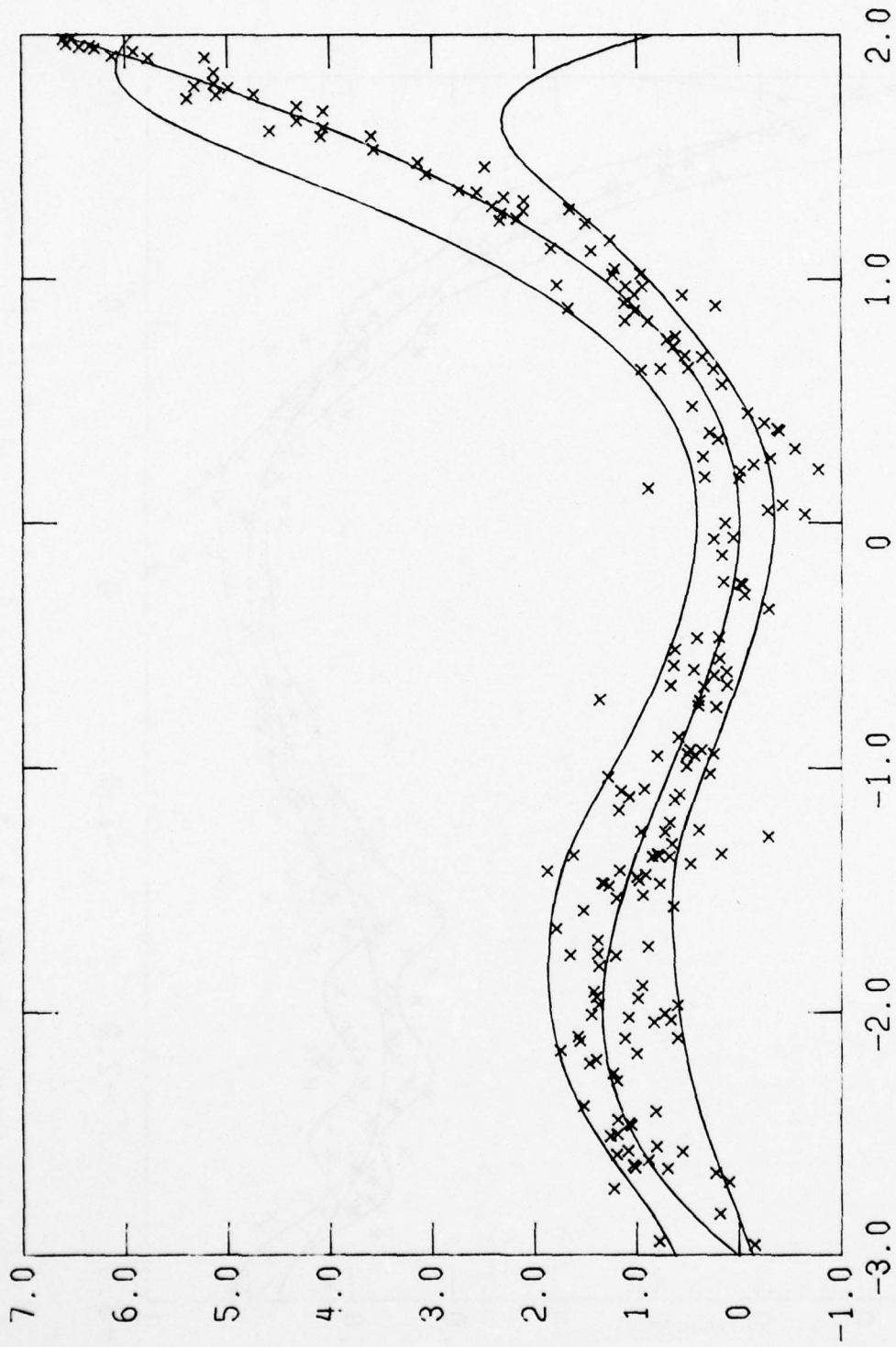


Figure 5. Confidence Band Based on \bar{m}_n With $\epsilon_n = n^{-.21}$

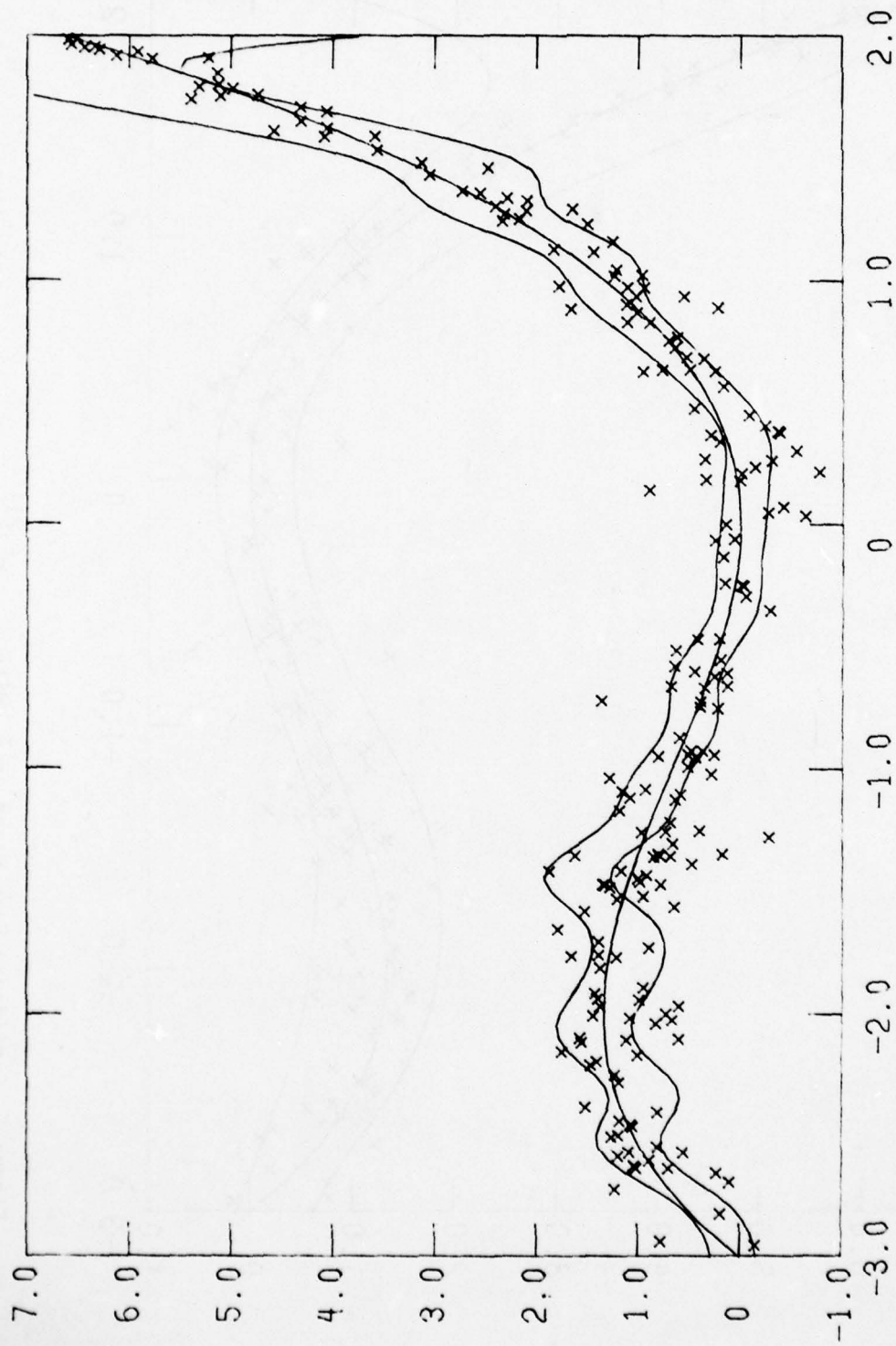


Figure 6. Confidence Band Based on \bar{m}_n With $\varepsilon_n = n^{-.4}$

BIBLIOGRAPHY

1. Aizerman, M.A., Braverman, E.M. and Rozonoer, J.T. (1970). Extrapolation problems in automatic control and the method of potential functions. *American Mathematical Society Translations* (2), Vol. 87.
2. Benedetti, J.K. (1974). Kernel estimation of regression functions. Ph.D. Dissertation, Department of Biostatistics, University of Washington.
3. Bhattacharya, P.K. (1976). An invariance principal in regression analysis. *Annals of Statistics*, 4, no. 3, 621-624.
4. Bickel, P.J. and Rosenblatt, M. (1973). On some global measures of the deviation of density function estimators. *Annals of Mathematical Statistics*, 1, 1071-1095.
5. Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley.
6. Buck, R.C. (1965). *Advanced Calculus*, Mc-Graw-Hill.
7. Cover, T.M. (1968). Estimation by the nearest neighbor rule. *IEEE Transactions on Information Theory*, IT-14, 50-55.
8. Cover, T.M. and Hart, P.E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, IT-13, 21-27.
9. Devroye, L.P. (1978). The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Transactions on Information Theory*, IT-24, no. 2.
10. Fisher, L. and Yakowitz, S. (1976). Uniform convergence of the potential function algorithm. *SIAM Journal on Control*, 14, 95-103.
11. Fix, E. and Hodges, J.L., Jr. (1951). Discriminatory analysis, nonparametric discrimination, consistency properties. Randolph Field, Texas, Project 21-49-004, Report no. 4.
12. Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent R.V.'s and the sample DF, I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32, 111-131.
13. Konakov, V.D. (1977). On a global measure of deviation for an estimate of the regression line. *Theory of Probability and Applications*, 22, no. 4, 858-868.
14. Lai, S.L. (1977). Large sample properties of K-nearest neighbor procedures. Ph.D. dissertation, Department of Mathematics, UCLA.

15. Leadbetter, M.R. (1963). On the nonparametric estimation of probability densities. Ph.D. dissertation, Department of Statistics, University of North Carolina at Chapel Hill.
16. Loève, M. (1963). *Probability Theory*, 3rd ed., D. Van Nostrand Co.
17. Loftsgaarden, D.O. and Quesenberry, C.P. (1965). A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, 36, 1049-1051.
18. Masani, P. (1968). *Orthogonally Scattered Measures*, Advanced Mathematics.
19. Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and Applications*, 9, 141-142.
20. Parzen, E. (1962). On estimation of a probability density function. *Annals of Mathematical Statistics*, 33, 1065-1076.
21. Priestly, M.B. and Chao, M.T. (1972). Nonparametric function fitting. *Journal of the Royal Statistical Society, Series B*, 34, 385-392.
22. Révész, P. (1976). On strong approximation of the multidimensional empirical process. *Annals of Probability*, 4, 729-743.
23. Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23, 470-472.
24. Rosenblatt, M. (1956). Remarks on some nonparametric estimators of a density function. *Annals of Mathematical Statistics*, 27, 832-837.
25. Rosenblatt, M. (1971). Curve estimates. *Annals of Mathematical Statistics*, 42, 1815-1842.
26. Rosenblatt, M. (1976). On the maximal deviation of a k-dimensional density estimator. *Annals of Probability*, 4, 1009-1015.
27. Schuster, E.F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Annals of Mathematical Statistics*, 43, 84-88.
28. Schuster, E.F. and Yakowitz, S. (1979). Contributions to the theory of nonparametric regression, with applications to system identification. *Annals of Statistics*, 7, 139-149.
29. Stone, C. (1977). Consistent nonparametric regression. *Annals of Statistics*, 5, 595-645.
30. Watson, G.S. (1964). Smooth regression analysis. *Sankhyā, Series A*, 26, 359-372.

31. Watson, G.S. and Leadbetter, M.R. (1963). On the estimation of the probability density, I. *Annals of Mathematical Statistics*, 34, 480-491.
32. Watson, G.S. and Leadbetter, M.R. (1964). Hazard analysis, II. *Sankhyā, Series A*, 26, 101-116.
33. Wegman, E.J. (1972a). Nonparametric probability density estimation I: A summary of available methods. *Technometrics*, 14, 533-546.
34. Wegman, E.J. (1972b). Nonparametric probability density estimation II: A comparison of density estimation methods. *Journal of Statistical Computation and Simulation*, 1, 225-245.
35. Yang, S.S. (1977). Linear functions of concomitants of order statistics. Technical Report No. 7, Department of Mathematics, M.I.T.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Smooth Nonparametric Regression Analysis		5. TYPE OF REPORT & PERIOD COVERED (9) TECHNICAL Repts.
7. AUTHOR(s) (10) Gordon Johnston		6. PERFORMING ORG. REPORT NUMBER Mimeo Series No.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina Chapel Hill, North Carolina 27514		8. CONTRACT OR GRANT NUMBER(s) AFOSR 75-2796 (15) N00014-75-C-0809 AFOSR 75-2796
11. CONTROLLING OFFICE NAME AND ADDRESS 1. Air Force Office of Scientific Research Bolling AFB, DC 2. Office of Naval Research		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (12) 89
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Arlington, VA 22217		12. REPORT DATE (11) September 79
		13. NUMBER OF PAGES 86
		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 entered in Block 20, if different from Report) (14) MIMEO SER-1253		
18. SUPPLEMENTARY NOTES Ph.D. dissertation under the joint direction of Professors R.J. Carroll and M.R. Leadbetter		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Regression function, Watson-type estimator, Pointwise weak consistency, Asymptotic joint normality. $E[Y X=x] \neq \bar{Y} \neq X=x$		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This work investigates properties of the Watson type estimator of the re- gression function $E[Y X=x]$ of a bivariate random vector (X,Y) . Pointwise weak consistency of the estimator is demonstrated. Sufficient conditions are given for asymptotic joint normality of the estimator taken at a finite number of distinct points, and for the uniform consistency of the estimator over a bounded interval. A large sample confidence band for the regression function, based on the estimator, is derived.		

DD FORM 1473

1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

420 064

xcl