THE UNIVERSITY OF MANITOBA

APPLICATIONS OF ESTIMATES OF A PROBABILITY DENSITY FUNCTION AND ITS DERIVATIVES IN NONPARAMETRIC INFERENCE

by

©

RUJAGAATA XAVIER MUGISHA

A THESIS

SUBMITTED TO THE FACULTY OF

GRADUATE STUDIES IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Department of Statistics

WINNIPEG, MANITOBA

JUNE, 1981

APPLICATIONS OF ESTIMATES OF A PROBABILITY DENSITY FUNCTION AND ITS DERIVATIVES IN NONPARAMETRIC INFERENCE

BY

RUJAGAATA XAVIER MUGISHA

A thesis submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

DOCTOR OF PHILOSOPHY

TABLE OF CONTENTS

	Page
ABSTRACT	iv
ACKNOWLEDGEMENT	vi
CHAPTER I: INTRODUCTION	1
1.1 Kernel Estimates of a Probability Density Function and Its Derivatives	1
1.2 Applications of Kernel Estimates to Statistical Problems	4
1.3 Estimation of the Shift Between Two Densities	5
1.4 Estimation of a Multivariate Multiple Regression Function	10
1.5 Outline of the Dissertation	12
CHAPTER II: ESTIMATION OF A PROBABILITY DENSITY FUNCTION AND MODE WHEN SAMPLE SIZE IS RANDOM	13
2.1 Introduction and Summary	13
2.2 Asymptotic Normality of $f_n(x)$ and $f_n^{(1)}(x)$	15
2.3 Uniform Strong Consistency of the Estimates $f_n^{(s)}(x)$, $s = 0,1,\ldots,r$	25
2.4 Uniform Consistency of the Estimates $f_{N_t}^{(s)}(x)$, $s = 0,1,\ldots,r$ and the Sample Mode θ_{N_t}	26

Page

2.5 Asymptotic Normality of $f_{N_t}^{(r)}(x)$, $r = 0,1$ 29

2.6 Asymptotic Normality of the Sample Mode 37

CHAPTER III: EFFICIENT NONPARAMETRIC ESTIMATION OF A SHIFT PARAMETER 41

3.1 The Empirical Likelihood Method for Estimating a Shift Parameter 41

3.2 Conditions A.3 45

3.3 Convergence Properties of Kernel Estimates of a Density Function and Its Derivatives 47

3.4 Consistency and Asymptotic Efficiency of the Nonparametric Estimator of the Shift Parameter 51

3.5 A Computational Method for an Asymptotically Efficient Estimate of the Shift Parameter 78

3.6 Miscellaneous Remarks 80

CHAPTER IV: NONPARAMETRIC ESTIMATION OF A MULTIVARIATE MULTIPLE REGRESSION FUNCTION 84

4.1 Introduction and Summary 84

4.2 Uniform Strong Consistency of $m_{in}(x_1,x_2)$, $i = 1,2$ 85

4.3 Asymptotic Joint Normality of the Estimates $m_{in}(x_1,x_2)$, $i = 1,2$ 95

	Page
REFERENCES	115
BIOGRAPHICAL DATA	123

ABSTRACT

In this dissertation we have considered three problems of nonparametric inference, namely, (i) the estimation of a probability density function and the mode when the sample size N_t is a random variable depending on a positive parameter t, (ii) the estimation of the shift between two probability density functions, and (iii) the estimation of a multivariate multiple regression function. In solving all the three problems we have employed the so-called kernel method of estimating a probability density function and its derivatives.

In the first problem we have considered a situation where the sample size N_t is distributed as an integer valued random variable depending on a positive parameter t. By making suitable assumptions on the convergence of N_t as t tends to infinity we have proved, under certain regularity conditions, the uniform strong consistency and the asymptotic normality of the estimates of a probability density function and the mode as t approaches infinity.

In the second problem we have considered the case when we have two independent random samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ coming from two distributions with probability density functions f and g respectively. It is

assumed that these two densities differ only in location and their common form is unknown. Motivated by the likelihood principle, we have then defined an empirical likelihood equation based on the kernel estimates of a probability density function and its derivatives. We have found that under certain regularity conditions, the solution of this empirical likelihood equation gives rise to an estimate of the shift between the densities f and g, and that the large-sample properties of this estimate are the same as the large-sample properties of the corresponding maximum likelihood estimate.

Finally, we have used the kernel method to estimate a multivariate multiple regression function. We have proved the uniform strong consistency of the estimated regression function, and the joint asymptotic normality of the estimate when this is computed at two or more distinct points.

ACKNOWLEDGEMENT

I would like to thank my supervisor and chairman of my advisory committee, Professor M. Samanta, for suggesting the problems and for guiding me throughout the entire preparation of this dissertation. I would also like to thank the rest of the advisory committee, Professor J.C. Fu, Professor K. Kocherlakota, Professor G.I. Paul, and Professor H.C. Finlayson, for their helpful suggestions.

Financial support was kindly granted to me by the Canadian Commonwealth Scholarship and Fellowship Committee, and to them I also wish to express my sincere gratitude.

I am very thankful to my wife, Stella, for her encouragement and moral support, and to my parents for putting in all that they had in order to get me through my earlier education.

Finally, I wish to express my appreciation to Mrs. Eva Loewen for the fine work she did in typing this dissertation.

CHAPTER I

INTRODUCTION

1.1 Kernel Estimates of a Probability Density Function and Its Derivatives

Let $X_1, X_2, \ldots, X_n$ be independent random varibles having a common probability density function f and let ϕ be a real valued Borel measurable function such that

$$\int_{-\infty}^{\infty} \phi(w)\,dw = 1.$$

We now define $\hat{f}_n(x)$ by

$$\hat{f}_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} \phi\left(\frac{x-X_i}{a_n}\right) \qquad \ldots(1.1)$$

where $\{a_n\}$ is a sequence of positive numbers converging to zero as n tends to infinity. The function $\hat{f}_n(x)$ in (1.1) is said to be a kernel estimate of f(x).

Convergence properties of this estimate have been extensively studied. We only mention here the works of Rosenblatt [32, 34], Parzen [29], Nadaraya [26, 27], Bhattacharya [5], Schuster [41] and Singh [45].

Multivariate analogues of the estimate in (1.1) for a multivariate density have been considered among others by Cacoullos [7], Van Ryzin [53], and Epanechnikov [14].

For the one-dimensional case, let $p \geq 0$ be an integer and denote by $f^{(p)}$ the p-th order derivative of f where

$f^{(0)} = f$. Suppose that the kernel ϕ and its first (p+1) derivatives satisfy certain regularity conditions. Then for estimating $f^{(r)}(x)$, $r = 0,1,...,p$ Bhattacharya [5] suggested the estimates given by

$$\hat{f}_n^{(r)}(x) = \frac{1}{na_n^{r+1}} \sum_{i=1}^{n} \phi^{(r)}\left(\frac{x-X_i}{a_n}\right), \quad r = 0,1,...,p \qquad ...(1.2)$$

and he studied their asymptotic properties. Schuster [41] also studied the asymptotic properties of the estimates in (1.2) and obtained their rates of convergence.

Singh [45] suggested different estimates for $f^{(r)}(x)$ of the form given by

$$\tilde{f}_n^{(r)}(x) = \frac{1}{na_n^{r+1}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{a_n}\right), \quad r = 0,1,... \qquad ...(1.3)$$

where the kernel K is chosen such that

$$\frac{1}{j!} \int_{-\infty}^{\infty} y^j K(y)\,dy = \begin{cases} 1 & \text{if } j = r \\ 0 & \text{if } j \neq r,\ j = 0,1,...,r-1 \end{cases}$$

and later in Singh [47], he studied further asymptotic properties of these estimates.

Yamato [59] and Davies [12] considered a kernel estimate of f given by

$$f_n(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{a_j} \phi\left(\frac{x-X_j}{a_j}\right) \qquad ...(1.4)$$

which has the following property:

$$f_n(x) = \frac{n-1}{n} f_{n-1}(x) + \frac{1}{na_n} \phi\left(\frac{x-X_n}{a_n}\right).$$

Because of this recursive property, Yamato [59] called the estimate in (1.4) a "sequential" estimate for a probability density function. However, it was not until the works of Davies and Wegman [13] and Caroll [8] that sequential density estimation, in its proper sense, was investigated.

Now, suppose that the sample size is a random variable distributed as a positive integer valued random variable N_t depending on a positive parameter t. On the basis of $X_1, X_2, \ldots, X_{N_t}$, Srivastava [48] considered an estimate of f(x) similar to that in (1.1) given by

$$\hat{f}_{N_t}(x) = \frac{1}{N_t a_{N_t}} \sum_{i=1}^{N_t} \phi\left(\frac{x-X_i}{a_{N_t}}\right) \qquad \ldots(1.5)$$

and he studied its asymptotic properties as t approaches infinity. He also attempted to prove the asymptotic normality of this estimate when N_t and the sequence $\{X_i\}$ are not independent. However, his proof is incorrect as Caroll [8] has pointed out.

In Chapter II we have proposed an estimate of f(x) based on $X_1, X_2, \ldots, X_{N_t}$ and given by

$$f_{N_t}(x) = \frac{1}{N_t} \sum_{j=1}^{N_t} \frac{1}{a_j} \phi\left(\frac{x-X_j}{a_j}\right). \qquad \ldots(1.6)$$

We have established conditions under which our estimate given in (1.6) is uniformly strongly consistent and asymptotically normal. Following Parzen [29] we have also proposed an estimate θ_{N_t} of the mode θ of f(x) and we have proved its strong consistency and asymptotic normality as t tends to infinity.

1.2 Applications of Kernel Estimates to Statistical Problems

In this section we shall briefly mention some instances where kernel estimates for density functions and their derivatives have been applied to statistical problems.

Parzen [29] and Nadaraya [26] considered the problem of estimating a mode of a univariate probability density function. Van Ryzin [53] and Samanta [37] considered the estimation of a mode of a multivariate density. Murthy [24] applied the kernel method to the estimation of jumps, reliability, and hazard rate. Bhattacharya [5] gave a solution to the estimation of the Fisher information. Nadaraya [25, 27], Rosenblatt [33], Schuster [42], Schuster and Yakowitz [43] all considered the estimation of regression curves and Singh and Tracy [44] considered the same problem for a special model where the conditional density of Y given X belongs to a Lebesgue exponential family. In Singh [46], a wide range of problems including those in econometrics were suggested which could be solved by using the kernel method. Rosenblatt [35] considered among other things a test of independence using functionals of kernel estimates for density functions. Ahmad and Lin [2] suggested estimates for a vector valued bivariate failure rate. Aitken and MacDonald [3] applied kernel-based density estimates to categorical data and Titterington [51] explored their use in this area. Copas and Fryer [10] used kernel estimation techniques to determine the suicide risks among psychiatric patients.

In this dissertation we shall apply the kernel method in estimating the shift between two densities and also in estimating a multivariate multiple regression function.

In the next section we shall briefly discuss the problem of estimating the shift between two densities.

1.3 Estimation of the Shift Between Two Densities

Let $X_1, X_2, \ldots, X_m$ be independent random variables having a common probability density function f and $Y_1, Y_2, \ldots, Y_n$ be independent random variables having a common probability density function g. Suppose that there exists a function h and two real numbers θ_1 and θ_2 such that

$$f(x) = h(x-\theta_1) \quad \text{and} \quad g(x) = h(x-\theta_2) \qquad \ldots(1.7)$$

for all real numbers x. The number $\theta = \theta_2 - \theta_1$ is said to be the shift between the two densities f and g.

Let $\{m_N\}$ and $\{n_N\}$ be two sequences of positive integers such that $m_N + n_N = N$ for each N and $\lim_{N\to\infty} \frac{m_N}{N} = \lambda$ where $0 < \lambda < 1$. If h were known, then θ_1 and θ_2 could be estimated separately from $X_1, X_2, \ldots, X_{m_N}$ and $Y_1, Y_2, \ldots, Y_{n_N}$ using the method of maximum likelihood by solving for t_1 and t_2 in the following equations:

$$\sum_{i=1}^{m_N} \frac{h^{(1)}(X_i - t_1)}{h(X_i - t_1)} = 0 \qquad \ldots(1.8)$$

and

$$\sum_{j=1}^{n_N} \frac{h^{(1)}(Y_j-t_2)}{h(Y_j-t_2)} = 0. \qquad \ldots(1.9)$$

Under certain regularity conditions on h (see Cramer [11], p.500), the above equations have solutions t_{1N} and t_{2N} which converge in probability to θ_1 and θ_2 so that $t_N = t_{2N} - t_{1N}$ converges in probability to the shift θ. Moreover, the estimate t_N is asymptotically normally distributed with mean θ and variance $\frac{1}{N\lambda(1-\lambda)J}$ where

$$J = \int_{-\infty}^{\infty} \frac{\{f^{(1)}(x)\}^2}{f(x)}\,dx = \int_{-\infty}^{\infty} \frac{\{h^{(1)}(x)\}^2}{h(x)}\,dx. \qquad \ldots(1.10)$$

Further properties of the estimate t_N can be found in Wald [54], LeCam [18], and Chernoff [9].

The nonparametric counterpart of this problem arises when h is unknown. Many nonparametric estimates of θ have been suggested in the case when h is unknown. For testing the hypothesis H_o: $\theta = 0$ against H_a: $\theta > 0$, Wilcoxon [57] proposed the statistic W given by

$$W = \sum_{j=1}^{n_N} R_j \qquad \ldots(1.11)$$

where R_j, $j = 1,\ldots,n_N$ are the ranks of the observations $Y_1, Y_2, \ldots, Y_{n_N}$ in the combined ordered arrangement of the two samples $X_1, X_2, \ldots, X_{m_N}$ and $Y_1, Y_2, \ldots, Y_{n_N}$. A large-sample confidence interval for θ based on the Wilcoxon's statistic is discussed in Lehmann [20]. Hodges and Lehmann [15] proposed a point estimate of θ based on the statistic U of Mann and Whitney [23] (which is equivalent to the statistic W) where U is given by

$$U = \sum_{i=1}^{m_N} \sum_{j=1}^{n_N} I(X_i, Y_j) \qquad \ldots(1.12)$$

and where

$$I(a,b) = \begin{cases} 1 & \text{if } b > a \\ 0 & \text{otherwise.} \end{cases}$$

The Hodges-Lehmann point estimate $\hat{\theta}$ of θ based on the Mann-Whitney statistic is given by

$$\hat{\theta} = \text{median of } \{Y_j - X_i,\ i = 1,\ldots,m_N,\ j = 1,\ldots,n_N\}.$$

The small-sample and large-sample properties of this estimate are discussed in Hodges and Lehmann [15].

Asymptotically efficient nonparametric estimation of θ has been the subject of intense research in recent years. An estimate θ_N of θ is said to be asymptotically efficient if the limiting distribution of $\sqrt{N\lambda(1-\lambda)}\ \{\theta_N - \theta\}$ is a normal distribution with a variance which attains the Cramer-Rao lower bound given by $1/J$ where J is the Fisher information as defined in (1.10). The existence of nonparametric estimates of θ having this property was first indicated by Stein [49].

van Eeden [52] proposed asymptotically efficient rank estimates of θ based on a fraction of the N observations $X_1,\ldots,X_{m_N}$, $Y_1,\ldots,Y_{n_N}$. Beran [4] used all the N observations in constructing an estimate of θ that is asymptotically efficient. Weiss and Wolfowitz [56] considered simultaneous estimation of the shift and scale parameters in the two-sample problem and Wolfowitz [58] continued this work for scale parameters.

Analogous estimates for the location parameter in the one-sample problem have been considered among others by Stone [50] and Sacks [36].

Motivated by the maximum likelihood method for estimating the shift θ in the two-sample problem, Bhattacharya [6] proposed a nonparametric estimate of θ by first reducing one sample to a frequency distribution over a fixed set of class intervals. He then defined an "empirical likelihood inequality" and proved that this empirical likelihood inequality has a solution (in some sense) which converges in probability to θ. He also showed that the asymptotic efficiency of his estimate relative to the maximum likelihood estimate is equal to the ratio of the Fisher information in a grouped observation to the Fisher information in an ungrouped observation.

Samanta [38] improved the method of Bhattacharya by defining an empirical likelihood equation using kernel estimates of a probability density function and its derivatives. He proved that with arbitrarily high probability this equation has a solution $\hat{T}_N^*$ which is consistent for θ. He also proved that $\sqrt{N\lambda(1-\lambda)}\ \{\hat{T}_N^*-\theta\}$ has an asymptotic normal distribution with a variance equal to $\frac{1}{J_{a_1,a_2}(f)}$ where a_1 and a_2 $(a_1 < a_2)$ are two arbitrarily fixed real numbers chosen in advance,

$$J_{a_1,a_2}(f) = \frac{\{f(a_1)\}^2}{F(a_1)} + \int_{a_1}^{a_2} \frac{\{f^{(1)}(x)\}^2}{f(x)}\, dx + \frac{\{f(a_2)\}^2}{1-F(a_2)} \qquad \ldots(1.13)$$

and F(y) is the distribution function of f(•). The quantity $J_{a_1,a_2}(f)$ has the property that although it is less than J it converges to J as a_1 and a_2 approach $-\infty$ and $+\infty$ respectively.

In this dissertation we have further improved the method of Samanta [38]. Using kernel estimates of a probability density function and its derivative, we have devised a nonparametric estimate of θ which we have shown to be asymptotically efficient. In Chapter III we shall first describe our method, then obtain an estimate T_N of θ and proceed to study its asymptotic distribution using the theory of a two-sample U-statistic. We shall also describe a simple computational procedure which leads to an estimate $\hat{T}_N$ of θ having the asymptotic properties possessed by T_N. We shall then apply our method to the estimation of a location parameter in the one-sample problem.

In the next section we are going to discuss the problem of estimating a multivariate multiple regression function using the kernel method.

1.4 Estimation of a Multivariate Multiple Regression Function

Let $(X_{1i},\ldots,X_{p_1 i}, Y_{1i},\ldots,Y_{p_2 i})'$, $i = 1,\ldots,n$ be n independent observations of a vector random variable $\begin{pmatrix}\underline{X}\\ \underline{Y}\end{pmatrix} = (X_1,\ldots,X_{p_1}, Y_1,\ldots,Y_{p_2})'$ having the joint distribution function $F(x_1,\ldots,x_{p_1}, y_1,\ldots,y_{p_2})$ and the joint probability density function $f(x_1,\ldots,x_{p_1}, y_1,\ldots,y_{p_2})$. We denote the joint distribution function and the joint probability density function of $(X_1,\ldots,X_{p_1})'$ by $G(x_1,\ldots,x_{p_1})$ and $g(x_1,\ldots,x_{p_1})$ respectively. The expectation of $\underline{Y}$ given $\underline{X}' = (x_1,\ldots,x_{p_1})$, denoted by $E(\underline{Y}|\underline{X} = \underline{x})$ defines the regression of $\underline{Y}$ on $\underline{X}$ and is given by the p_2-dimensional vector $(m_1(x_1,\ldots,x_{p_1}),\ldots,m_{p_2}(x_1,\ldots,x_{p_2}))'$ where for $i = 1,\ldots,p_2$, $m_i(x_1,\ldots,x_{p_1})$ is given by

$$m_i(x_1,\ldots,x_{p_1}) = \frac{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} y_i f(x_1,\ldots,x_{p_1}, y_1,\ldots,y_{p_2})\,dy_1,\ldots,dy_{p_2}}{g(x_1,\ldots,x_{p_1})}.$$

Let $\phi_i(y)$, $i = 1,\ldots,p_1$ be p_1 univariate probability density functions and let $\{a_n\}$ be a sequence of positive numbers converging to zero as n tends to infinity. Following Nadaraya [25, 27] and Watson [55] we shall propose a nonparametric kernel estimate for the population regression function given by $(m_{1n}(x_1,\ldots,x_{p_1}),\ldots,m_{p_2 n}(x_1,\ldots,x_{p_1}))'$ where

$$m_{in}(x_1,\ldots,x_{p_1})$$

$$= \frac{\sum_{j=1}^{n} Y_{ij} \prod_{\ell=1}^{p_1} \phi_\ell\left(\frac{x_\ell - X_{\ell j}}{a_n}\right)}{\sum_{j=1}^{n} \prod_{\ell=1}^{p_1} \phi_\ell\left(\frac{x_\ell - X_{\ell j}}{a_n}\right)}, \quad i = 1,\ldots,p_2 \qquad \ldots(1.14)$$

Let S be a closed set in the p_1-dimensional Euclidean space such that $\inf_{\underline{x}\in S} g(\underline{x}) = \mu > 0$.

The problem of estimating the regression function when $p_1 = p_2 = 1$ was considered among others by Nadaraya [25, 27], Rosenblatt [33], Schuster [42], and Schuster and Yakowitz [43]. Schuster [42] proved the asymptotic joint normality of the estimates of the regression function at q distinct points. For arbitrary p_1 and $p_2 = 1$, Ahmad and Lin [1] considered the estimation of a multiple regression function using a recursive-type of kernel estimate for a probability density function given in (1.4) and they proved the asymptotic joint normality of the estimates when these estimates are computed at q distinct points in S. One of the regularity conditions assumed by Ahmad and Lin [1] is that $E(|Y|^3 | \underline{X} = \underline{x})$ is a bounded function over the entire p_1-dimensional Euclidean space. This assumption, however, excludes very important classes of regression functions.

In this dissertation we have considered the problem of estimating a multivariate multiple regression function for the special case when $p_1 = p_2 = 2$. By assuming that for arbitrarily small $\eta > 0$, $E[|Y_i|^{3+\eta}] < \infty$, $i = 1,2$, we have also studied the asymptotic joint distribution of the regression estimates when these estimates are computed at two distinct points in S. We have discussed how these results can be extended to the case when p_1 and p_2 are arbitrary and the regression estimates are computed at q ($q \geq 2$) distinct points.

1.5 Outline of the Dissertation

In the next chapter (Chapter II) we discuss the problem of estimating a probability density function and mode when the sample size is random. In Chapter III we consider the problem of estimating the shift between two densities and finally, in Chapter IV, we examine the problem of estimating a multivariate multiple regression function.

All integrals in this dissertation will be understood to be Lebesgue integrals.

CHAPTER II

ESTIMATION OF A PROBABILITY DENSITY FUNCTION AND MODE WHEN SAMPLE SIZE IS RANDOM

2.1 Introduction and Summary

Let $X_1, X_2, \ldots, X_n$ be n independent observations of a random variable X having the probability density function f(x). Yamato [59] and Davies [12] have studied the asymptotic properties of the estimate $f_n(x)$ of f(x) given by

$$f_n(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{a_j} \phi\left(\frac{x-X_j}{a_j}\right) \qquad \ldots(2.1)$$

where $\phi(u)$ is a continuous probability density function and $\{a_n\}$ is a monotonically decreasing sequence of positive numbers converging to zero.

In many practical situations the number of observations N_t which we observe in time $(0,t]$ is a random variable. For example, we consider the problem of estimating the probability density function of the waiting times of customers at a service station and we may assume that the number of customers N_t arriving in time $(0,t]$ is a Poisson random variable with parameter πt $(\pi > 0)$.

Considering such problems we assume that for any $t > 0$, N_t is an integer valued random variable. Let

$X_1, X_2, X_3, \ldots,$ be independent observations of X having the probability density function $f(x)$. The random variables $X_1, X_2, X_3, \ldots$ need not be independent of the random variable N_t. In this chapter we consider the estimate $f_{N_t}^{(r)}(x)$ of the r-th derivative $f^{(r)}(x)$ $(r = 0,1,2,\ldots)$ based on $X_1, X_2, \ldots, X_{N_t}$ and given by

$$f_{N_t}^{(r)}(x) = \frac{1}{N_t} \sum_{j=1}^{N_t} \frac{1}{a_j^{r+1}} \phi^{(r)}\left(\frac{x-X_j}{a_j}\right) \qquad \ldots(2.2)$$

where $\phi^{(r)}(u)$ is the r-th derivative of $\phi(u)$.

If $\phi(u)$ is so chosen that $\phi(u)$ tends to 0 as u tends to $\pm\infty$, then for every sample sequence $f_{N_t}(x)$ is continuous and tends to 0 as x tends to $\pm\infty$. Consequently, there is a random variable θ_{N_t} such that

$$f_{N_t}(\theta_{N_t}) = \max_{-\infty<x<\infty} f_{N_t}(x). \qquad \ldots(2.3)$$

Similarly, if the probability density function $f(x)$ is uniformly continuous, then $f(x)$ possesses a mode θ defined by

$$f(\theta) = \max_{-\infty<x<\infty} f(x). \qquad \ldots(2.4)$$

We consider θ_{N_t} as an estimate for θ and we shall assume that θ is unique.

In this chapter we have proved that under certain regularity conditions the estimates $f_{N_t}^{(r)}(x)$ in (2.2) are uniformly (uniform in x) strongly consistent and the sample mode θ_{N_t} in (2.3) is strongly consistent (Theorems 2.1 and

2.2). We have also shown that under certain conditions the estimates $f_{N_t}^{(r)}(x)$, $r = 0,1$ and θ_{N_t} are asymptotically normally distributed (Theorems 2.3 and 2.4). These four theorems can be regarded as appropriate extensions of the earlier results due to Parzen [29], Yamato [59], and Davies [12].

Some allied work in this area has been done by Srivastava [48], Davies and Wegman [13], and Caroll [8]. However, our methods are fairly standard, and working with the derivative of the density function, we have been able to derive the asymptotic normality of the estimated mode as well.

2.2 Asymptotic Normality of $f_n(x)$ and $f_n^{(1)}(x)$

We assume that the function $\phi(u)$ and its first derivative $\phi^{(1)}(u)$ satsify the following conditions:

$$\lim_{|u|\to\infty} |u\phi(u)| = 0 \qquad \ldots(2.5)$$

$$\sup_{-\infty<u<\infty} |\phi^{(1)}(u)| < \infty \text{ and } \int_{-\infty}^{\infty} |\phi^{(1)}(u)|du < \infty \qquad \ldots(2.6)$$

$$\lim_{|u|\to\infty} |u\phi^{(1)}(u)| = 0. \qquad \ldots(2.7)$$

We note that Condition (2.5) and the existence of $\phi^{(1)}(u)$ imply that $\sup_{-\infty<u<\infty} |\phi(u)| < \infty$. Let $C(f)$ denote the set of continuity points of $f(x)$.

Lemma 2.1. Let $K(u)$ be a real-valued Borel measurable function. If $x \in C(f)$ and if

$$\sup_{-\infty<u<\infty} |K(u)| < \infty$$

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

and

$$\lim_{|u|\to\infty} |uK(u)| = 0$$

then for any $\eta \geq 0$

$$\lim_{n\to\infty} \{|\frac{1}{a_n} \int_{-\infty}^{\infty} |K(\frac{y}{a_n})|^{1+\eta} f(x-y)dy$$

$$- f(x) \int_{-\infty}^{\infty} |K(y)|^{1+\eta} dy|\} = 0.$$

Proof. We note that if

$$\sup_{\infty<u<\infty} |K(u)| < \infty$$

and

$$\int_{-\infty}^{\infty} |K(u)|du < \infty,$$

then for any $\eta \geq 0$

$$\int_{-\infty}^{\infty} |K(u)|^{1+\eta} du < \infty.$$

The rest of the proof follows along the lines of the proof of Theorem 1A of Parzen [29].

Lemma 2.2. Let $\{g_n\}$ be a sequence of functions converging to a function g at a point y as n tends to infinity. Then

$$\lim_{n\to\infty} \{\frac{1}{n} \sum_{j=1}^{n} g_j(y)\} = g(y).$$

Proof. This is a well known result in calculus.

Lemma 2.3 Let $\{\nu_{nj}, j = 1,2,\ldots,n; n = 1,2,3,\ldots\}$ be a uniformly bounded double array of numbers such that

$$\lim_{n\to\infty} \{\frac{1}{n} \sum_{j=1}^{n} \nu_{nj}\} = \nu.$$

Let $\{g_j\}$ be a sequence of functions converging to a function g at a point y as j tends to infinity. Then

$$\lim_{n\to\infty} \{\frac{1}{n} \sum_{j=1}^{n} \nu_{nj} g_j(y)\} = \nu g(y).$$

Proof. The proof of this lemma follows from Lemma 2.2 in conjunction with the inequality

$$\left|\frac{1}{n} \sum_{j=1}^{n} \nu_{nj} g_j(y) - \nu g(y)\right|$$

$$\leq \frac{1}{n} \sum_{j=1}^{n} |\nu_{nj}| |g_j(y) - g(y)|$$

$$+ |g(y)| \left|\frac{1}{n} \sum_{j=1}^{n} \nu_{nj} - \nu\right|.$$

We now assume that for some real numbers $s \geq 1$ the sequence $\{a_n\}$ satisfies the following condition:

$$\lim_{n\to\infty} \{\frac{1}{n} \sum_{j=1}^{n} (\frac{a_n}{a_j})^s\} = \gamma_s < \infty. \qquad \ldots(2.8)$$

Lemma 2.4. Let $\phi(u)$ and $\phi^{(1)}(u)$ satisfy Conditions (2.5), (2.6), and (2.7) and the sequence $\{a_n\}$ satisfy Condition (2.8). If $x \in C(f)$, then for $r = 0,1$ and for any $\eta \geq 0$

$$\lim_{n\to\infty} \{\frac{1}{n} \sum_{j=1}^{n} (\frac{a_n}{a_j})^s \int_{-\infty}^{\infty} \frac{1}{a_j} |\phi^{(r)}(\frac{y}{a_j})|^{1+\eta} f(x-y)dy\}$$

$$= \gamma_s f(x) \int_{-\infty}^{\infty} |\phi^{(r)}(y)|^{1+\eta} dy$$

where

$$\phi^{(o)}(u) = \phi(u).$$

Proof. We let

$$\gamma_{nj} = (\frac{a_n}{a_j})^s$$

$$g_j(x) = \int_{-\infty}^{\infty} \frac{1}{a_j} |\phi^{(r)}(\frac{y}{a_j})|^{1+\eta} f(x-y)dy$$

and

$$g(x) = f(x) \int_{-\infty}^{\infty} |\phi^{(r)}(y)|^{1+\eta} dy.$$

We note that $|\gamma_{nj}| \leq 1$ for all $j = 1,2,\ldots,n$; $n = 1,2,3,\ldots$ and by Lemma 2.1

$$\lim_{j\to\infty} g_j(x) = g(x).$$

An application of Lemma 2.3 completes the proof.

We assume that the estimate $f_n^{(r)}(x)$ is defined in a similar manner as in (2.2), i.e.,

$$f_n^{(r)}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{a_j^{r+1}} \phi^{(r)}(\frac{x-X_j}{a_j}).$$

Lemma 2.5. Let $\phi(u)$ and $\phi^{(1)}(u)$ satisfy Conditions (2.5), (2.6), and (2.7), and the sequence $\{a_n\}$ satisfy Condition (2.8) for $s = 1,2,3$. If $x \in C(f)$, then for $r = 0,1$

$$\lim_{n\to\infty} [na_n^{1+2r}\{\mathrm{Var}(f_n^{(r)}(x))\}] = \gamma_{1+2r} f(x) \int_{-\infty}^{\infty} |\phi^{(r)}(u)|^2 du$$

Proof. We have

$$na_n^{1+2r}[\mathrm{Var}\{f_n^{(r)}(x)\}]$$

$$= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{a_n}{a_j}\right)^{1+2r}\int_{-\infty}^{\infty}\frac{1}{a_j}\left|\phi^{(r)}\left(\frac{u}{a_j}\right)\right|^2 f(x-u)\,du$$

$$- \frac{a_n}{n}\sum_{j=1}^{n}\left(\frac{a_n}{a_j}\right)^{2r}\left\{\int_{-\infty}^{\infty}\frac{1}{a_j}\phi^{(r)}\left(\frac{u}{a_j}\right)f(x-u)\,du\right\}^2.$$

By Lemma 2.1 we get for $r = 0,1$

$$\lim_{n\to\infty}\left[\frac{1}{a_n}\int_{-\infty}^{\infty}\left|\phi^{(r)}\left(\frac{u}{a_n}\right)\right|^2 f(x-u)\,du\right] = f(x)\int_{-\infty}^{\infty}|\phi^{(r)}(u)|^2 du$$

and

$$\lim_{n\to\infty}\left[\frac{1}{a_n}\int_{-\infty}^{\infty}\phi^{(r)}\left(\frac{u}{a_n}\right)f(x-u)\,du\right]^2 = \left[f(x)\int_{-\infty}^{\infty}\phi^{(r)}(u)\,du\right]^2.$$

An application of Lemma 2.4 completes the proof.

Remark 2.1. If $a_n = n^{-\delta}$, $\delta > 0$, then for $s > 0$

$$\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{j=1}^{n}\left(\frac{a_n}{a_j}\right)^s\right\} = \lim_{n\to\infty}\left\{\frac{1^{s\delta}+2^{s\delta}+\ldots+n^{s\delta}}{n^{1+s\delta}}\right\} = \frac{1}{1+s\delta}.$$

For a proof of this result see page 46 of Korovkin [17]. Since $\frac{1}{1+\delta} < 1$ we note that for this particular choice of $\{a_n\}$ the asymptotic variance of $f_n(x)$ is smaller than the asymptotic variance of the estimate given in Parzen [29]. This fact was first observed by Yamato [59] who proved Lemma 2.5 for $r = 0$. However, his proof is based on the incorrect assumption that a probability density function continuous on $(-\infty,\infty)$ is necessarily bounded.

We now assume that the sequence $\{a_n\}$ satisfies the condition

$$\lim_{n\to\infty} na_n = \infty. \qquad \ldots(2.9)$$

Lemma 2.6. Let $\phi(u)$ and $\phi^{(1)}(u)$ satisfy Conditions (2.5), (2.6), and (2.7), and the sequence $\{a_n\}$ satisfy Condition (2.8) for $s = 1,2,3$ and Condition (2.9). If $x \in C(f)$, then for $r = 0,1$

$$(na_n^{1+2r})^{\frac{1}{2}}[f_n^{(r)}(x) - E\{f_n^{(r)}(x)\}]$$

converges in distribution to a normal random variable with mean zero and variance

$$\gamma_{1+2r} f(x)\{\int_{-\infty}^{\infty} |\phi^{(r)}(u)|^2 du\}.$$

Proof. We note that Yamato [59] has proved this lemma for $r = 0$. The proof for $r = 1$ can be accomplished in a similar way. We define for $r = 0,1$

$$W_j^{(r)}(x) = \frac{1}{a_j^{1+r}} \phi^{(r)}(\frac{x-X_j}{a_j}), \quad j = 1,2,\ldots,n.$$

Then

$$f_n^{(r)}(x) = \frac{1}{n} \sum_{j=1}^{n} W_j^{(r)}(x).$$

To prove the lemma it suffices to show that for some $\delta > 0$ (see Loeve [21], p.275)

$$\frac{\sum_{j=1}^{n} E\left|W_j^{(r)}(x) - E\{W_j^{(r)}(x)\}\right|^{2+\delta}}{\left\{\sum_{j=1}^{n} \operatorname{Var}(W_j^{(r)}(x))\right\}^{1+\frac{\delta}{2}}} \qquad \ldots(2.10)$$

converges to zero as n tends to infinity. Now, the expression in (2.10) is equal to

$$\frac{\left(\frac{a_n^{1+2r}}{n}\right)^{1+\frac{\delta}{2}} \left[\sum_{j=1}^{n} E\left|W_j^{(r)}(x) - E\{W_j^{(r)}(x)\}\right|^{2+\delta}\right]}{\left[n a_n^{1+2r}\{\operatorname{Var}(f_n^{(r)}(x))\}\right]^{1+\frac{\delta}{2}}}$$

By Lemma 2.5, the denominator of the above expression converges to

$$\left[\gamma_{1+2r} f(x) \int_{-\infty}^{\infty} |\phi^{(r)}(u)|^2 du\right]^{1+\frac{\delta}{2}}$$

as n tends to infinity. Hence, to prove the lemma it suffices to show that the numerator of the above expression converges to zero as n approaches infinity. Using the C_r-inequality (see Loeve [21], p.155) we have

$$\left(\frac{a_n^{1+2r}}{n}\right)^{1+\frac{\delta}{2}} \left[\sum_{j=1}^{n} E\left|W_j^{(r)}(x) - E\{W_j^{(r)}(x)\}\right|^{2+\delta}\right]$$

$$\leq 2^{1+\delta}\left(\frac{a_n^{1+2r}}{n}\right)^{1+\frac{\delta}{2}} \left[\sum_{j=1}^{n} E\left|W_j^{(r)}(x)\right|^{2+\delta}\right]$$

$$+ 2^{1+\delta}\left(\frac{a_n^{1+2r}}{n}\right)^{1+\frac{\delta}{2}} \left[\sum_{j=1}^{n} \left|E\{W_j^{(r)}(x)\}\right|^{2+\delta}\right]. \qquad \ldots(2.11)$$

Now,

$$\left(\frac{a_n^{1+2r}}{n}\right)^{1+\frac{\delta}{2}} [\sum_{j=1}^{n} E|W_j^{(r)}(x)|^{2+\delta}]$$

$$= \frac{1}{n}\sum_{j=1}^{n} \frac{1}{(na_j)^{\delta/2}} \left(\frac{a_n}{a_j}\right)^{1+\frac{\delta}{2}+r(2+\delta)}$$

$$\cdot \int_{-\infty}^{\infty} \frac{1}{a_j} |\phi^{(r)}(\frac{y}{a_j})|^{2+\delta} f(x-y)dy$$

$$\leq \frac{1}{(na_n)^{\delta/2}} [\frac{1}{n}\sum_{j=1}^{n} \int_{-\infty}^{\infty} \frac{1}{a_j} |\phi^{(r)}(\frac{y}{a_j})|^{2+\delta} f(x-y)dy].$$

From this inequality, using Lemma 2.1 and Lemma 2.2 and Condition (2.9) we conclude that

$$\lim_{n\to\infty} \left[\left(\frac{a_n^{1+2r}}{n}\right)^{1+\frac{\delta}{2}} \{\sum_{j=1}^{n} E|W_j^{(r)}(x)|^{2+\delta}\}\right] = 0.$$

For the other term in the right hand side of (2.11), we have

$$\left(\frac{a_n^{1+2r}}{n}\right)^{1+\frac{\delta}{2}} [\sum_{j=1}^{n} |E\{W_j^{(r)}(x)\}|^{2+\delta}]$$

$$= \frac{a_n^{1+\frac{\delta}{2}}}{n^{\delta/2}} [\frac{1}{n}\sum_{j=1}^{n} \left(\frac{a_n}{a_j}\right)^{r(2+\delta)} |\int_{-\infty}^{\infty} \frac{1}{a_j}\phi^{(r)}(\frac{y}{a_j}) f(x-y)dy|^{2+\delta}]$$

$$\leq \frac{a_n^{1+\frac{\delta}{2}}}{n^{\delta/2}} [\frac{1}{n}\sum_{j=1}^{n} |\int_{-\infty}^{\infty} \frac{1}{a_j}\phi^{(r)}(\frac{y}{a_j}) f(x-y)dy|^{2+\delta}].$$

From this inequality we conclude in a similar manner that

$$\lim_{n\to\infty} \left(\frac{a_n^{1+2r}}{n}\right)^{1+\frac{\delta}{2}} [\sum_{j=1}^{n} |E\{W_j^{(r)}(x)\}|^{2+\delta}] = 0.$$

This completes the proof of the lemma.

We now assume that the function $\phi(u)$ and the sequence $\{a_n\}$ satisfy the following conditions:

$$\int_{-\infty}^{\infty} |u|\phi(u)du < \infty \qquad \text{...(2.12)}$$

$$\frac{1}{n}\sum_{j=1}^{n} a_j \leq Ca_n \quad (C > 0),\ n = 1,2,3,... \qquad \text{...(2.13)}$$

Lemma 2.7. Let $\phi(u)$ and $\phi^{(1)}(u)$ satisfy Conditons (2.6) and (2.12) and the sequence $\{a_n\}$ satisfy Condition (2.13). If for $r = 0,1$ the first $(r+1)$ derivatives of $f(x)$ exist and are bounded, then

$$\sqrt{na_n^{1+2r}} \sup_{-\infty<x<\infty} [E\{f_n^{(r)}(x)\} - f^{(r)}(x)] = 0(na_n^{3+2r})^{\frac{1}{2}}.$$

Proof. For $r = 0$ we have

$$|E\{f_n(x)\} - f(x)|$$

$$= |\frac{1}{n}\sum_{j=1}^{n} \frac{1}{a_j}\int_{-\infty}^{\infty} \phi(\frac{x-u}{a_j})f(u)du - f(x)|$$

$$= |\frac{1}{n}\sum_{j=1}^{n} \int_{-\infty}^{\infty} f(x-a_ju)\phi(u)du - f(x)|. \qquad \text{...(2.14)}$$

For $r = 1$, we integrate by parts to get

$$|E\{f_n^{(1)}(x)\} - f^{(1)}(x)|$$

$$= |\frac{1}{n}\sum_{j=1}^{n} \frac{1}{a_j^2}\int_{-\infty}^{\infty} \phi^{(1)}(\frac{x-u}{a_j})f(u)du - f^{(1)}(x)|$$

$$= |\frac{1}{n}\sum_{j=1}^{n} \int_{-\infty}^{\infty} f^{(1)}(x-a_ju)\phi(u)du - f^{(1)}(x)|. \qquad \text{...(2.15)}$$

In (2.14) and (2.15) expanding $f(x-a_j u)$ and $f^{(1)}(x-a_j u)$ around x to the order of a_j we get for r = 0,1

$$\sup_{-\infty<x<\infty} |E\{f_n^{(r)}(x)\} - f^{(r)}(x)|$$

$$\leq \sup_{-\infty<x<\infty} |f^{(r+1)}(x)| \int_{-\infty}^{\infty} |u|\phi(u)du\{\frac{1}{n}\sum_{j=1}^{n} a_j\}$$

$$\leq Ca_n$$

where C is a positive constant. The proof of the lemma follows from the above observation.

Remark 2.2. If $a_n = n^{-\delta}$, $0 < \delta < 1$, then

$$\frac{1}{n}\sum_{j=1}^{n} a_j \leq \frac{1}{(1-\delta)} a_n, \quad \text{(see Korovkin [17], p.26).}$$

Thus, Condition (2.13) is satisfied for this choice of $\{a_n\}$.

Lemma 2.8. Under the conditions of Lemmas 2.6 and 2.7, if for r = 0,1, $\sqrt{na_n^{3+2r}} = o(1)$, then $\sqrt{na_n^{1+2r}}\{f_n^{(r)}(x) - f^{(r)}(x)\}$ converges in distribution to a normal random variable with mean zero and variance $\gamma_{1+2r} f(x)\{\int_{-\infty}^{\infty} |\phi^{(r)}(u)|^2 du\}$.

Proof. We have

$$\sqrt{na_n^{1+2r}}\,[f_n^{(r)}(x) - f^{(r)}(x)]$$

$$= \sqrt{na_n^{1+2r}}\,[f_n^{(r)}(x) - E\{f_n^{(r)}(x)\}]$$

$$+ \sqrt{na_n^{1+2r}}\,[E\{f_n^{(r)}(x)\} - f^{(r)}(x)]$$

$$= \sqrt{na_n^{1+2r}}\,[f_n^{(r)}(x) - E\{f_n^{(r)}(x)\}] + o(1),$$

by Lemma 2.7. The desired conclusion now follows from Lemma 2.6.

2.3 Uniform Strong Consistency of the Estimates $f_n^{(s)}(x)$, $s = 0,1,\ldots,r$

We define

$$\alpha(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx$$

and

$$k(t) = \int_{-\infty}^{\infty} e^{itx} \phi(x)\,dx.$$

We now assume that for some integer $r \geq 0$ the following conditions on the functions $\alpha(u)$, $\phi(u)$ and $k(u)$ and the sequence $\{a_n\}$ are satisfied.

$$\int_{-\infty}^{\infty} |u^{r+1}|\,|\alpha(u)|\,du < \infty \qquad \ldots(2.16)$$

$$\int_{-\infty}^{\infty} |u^{r}|\,|k(u)|\,du < \infty \qquad \ldots(2.17)$$

$$|k(w)| \leq |k(u)| \quad \text{if} \quad |u| \leq |w| \qquad \ldots(2.18)$$

$$\int_{-\infty}^{\infty} |\phi^{(s)}(u)|\,du < \infty, \quad s = 1,2,\ldots,r \qquad \ldots(2.19)$$

$$\lim_{n\to\infty} \left\{\frac{a_n}{a_{n+1}}\right\} = 1 \qquad \ldots(2.20)$$

$$\sum_{n=1}^{\infty} \frac{1}{(na_n^{1+r})^2} < \infty \qquad \ldots(2.21)$$

$$\sum_{n=1}^{\infty} \frac{1}{na_n^{2(1+r)-1}} \left|\frac{1}{a_{n+1}} - \frac{1}{a_n}\right| < \infty \qquad \ldots(2.22)$$

$$\lim_{n\to\infty} na_n^{2(1+r)} = \infty. \qquad \ldots(2.23)$$

We note that Condition (2.16) implies that the first (r+1) derivatives of f(x) are bounded and Conditon (2.19) means that $\phi(u)$ and its first (r-1) derivatives are functions of bounded variation on $(-\infty,\infty)$.

Lemma 2.9. If Condition (2.12) and for some integer $r \geq 0$ Conditions (2.16) through (2.23) are satisfied, then for $s = 0,1,...,r$

$$\sup_{-\infty<x<\infty} |f_n^{(s)}(x) - f^{(s)}(x)|$$

converges to zero with probability one as n tends to infinity.

Proof. For r = 0, the proof has been given by Davies [12]. The proof for $r \geq 1$ can be accomplished along the same lines as in Davies [12] and is omitted.

Remark 2.3. If we chose $\phi(y)$ to be the standard normal probability density function and $a_n = n^{-\delta}$, $0 < \delta < \frac{1}{2(1+r)}$, then Conditions (2.17) through (2.23) are satisfied.

2.4 Uniform Consistency of the Estimates $f_{N_t}^{(s)}(x)$, $s = 0,1,...,r$, and the Sample Mode θ_{N_t}

We now prove the following theorem.

Theorem 2.1. Suppose that Condition (2.12) and for some positive integer $r \geq 0$ Conditions (2.16) through (2.23) are satisfied.

(a) If for every $\varepsilon > 0$

$$\lim_{t\to\infty} P\{|\frac{N_t}{t} - \pi| > \varepsilon\} = 0, \quad 0 < \pi < \infty, \qquad ...(2.24)$$

then for $s = 0,1,...,r$

$$\sup_{-\infty<x<\infty} |f_{N_t}^{(s)}(x) - f^{(s)}(x)|$$

converges in probability to zero as t tends to infinity.

(b) If

$$P\{\lim_{t\to\infty} \frac{N_t}{t} = \pi\} = 1, \quad 0 < \pi < \infty, \qquad \ldots(2.25)$$

then for s = 0,1,...,r

$$\sup_{-\infty<x<\infty} |f_{N_t}^{(s)}(x) - f^{(s)}(x)|$$

converges to zero with probability one as t tends to infinity.

Proof. For any s = 0,1,...,r let

$$W_n = \sup_{-\infty<x<\infty} |f_n^{(s)}(x) - f^{(s)}(x)|$$

and

$$W_{N_t} = \sup_{-\infty<x<\infty} |f_{N_t}^{(s)}(x) - f^{(s)}(x)|.$$

To prove part (a) of the theorem, let ε and η be arbitrarily small positive numbers. By Lemma 2.9 we can find an integer $n_o = n_o(\varepsilon,\eta)$ such that

$$P\{W_n < \frac{\varepsilon}{2} \text{ for all n greater than } n_o\} \geq 1 - \eta$$

and

$$P\{|W_{m+n} - W_n| < \frac{\varepsilon}{2} \text{ for all m whenever } n > n_o\} \geq 1 - \eta.$$

From (2.24) we can also find $t_o = t_o(\varepsilon,\eta)$ such that for $t > t_o$ we have

$$P\{|N_t - \pi t| < \varepsilon t\} \geq 1 - \eta.$$

Let $N_1' = [t(\pi-\varepsilon)]$ and $N_2' = [t(\pi+\varepsilon)] + 1$ where $[x]$ denotes the integral part of x. We now choose $t > t_o$ such that $N_1' > n_o$. Then we get

$$
\begin{aligned}
& P\{W_{N_t} < \varepsilon\} \\
& \quad \geq P\{W_{N_t} < \varepsilon;\ |N_t-\pi t| < \varepsilon t;\ |W_{N_t}-W_n| < \tfrac{\varepsilon}{2} \text{ for some} \\
& \qquad \text{integer } n \text{ lying between } N_1' \text{ and } N_2'\} \\
& \quad \geq P\{W_n < \tfrac{\varepsilon}{2};\ |N_t-\pi t| < \varepsilon t;\ |W_m-W_n| < \tfrac{\varepsilon}{2} \text{ for all} \\
& \qquad \text{integers } m \text{ and } n \text{ lying between } N_1' \text{ and } N_2'\} \\
& \quad \geq P\{W_n < \tfrac{\varepsilon}{2} \text{ for all integers } n \text{ lying between } N_1' \\
& \qquad \text{and } N_2'\} \\
& \quad - P\{|N_t-\pi t| \geq \varepsilon t\} \\
& \quad - P\{|W_m-W_n| \geq \tfrac{\varepsilon}{2} \text{ for all integers } m \text{ and } n \text{ lying} \\
& \qquad \text{between } N_1' \text{ and } N_2'\} \\
& \quad > 1 - 3\eta.
\end{aligned}
$$

For a proof of part (b) of the theorem we refer to Srivastava [48], p.80.

We now recall the definitions of the sample mode θ_{N_t} and the population mode θ given in (2.3) and (2.4) respectively.

Theorem 2.2. If the Conditions of part (b) of Theorem 2.1 are satisfied for $r = 0$, then the sample mode θ_{N_t} converges to θ with probability one as t tends to infinity.

Proof. The proof follows from the fact that since $f(x)$ is uniformly continuous with a unique mode θ, for every $\varepsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(\theta)| \geq \delta$ whenever $|x - \theta| \geq \varepsilon$.

2.5 Asymptotic Normality of $f_{N_t}^{(r)}(x)$, $r = 0,1$

We assume that the random variable N_t satisfies Condition (2.24). Then for every $\varepsilon > 0$, there exists $t_o = t_o(\varepsilon)$ such that for $t \geq t_o$ we have

$$P\{|N_t - \pi t| \geq \pi t\varepsilon\} < \varepsilon.$$

We define

$$N_1 = [\pi t(1-\varepsilon)]$$

and

$$N_2 = [\pi t(1+\varepsilon)]$$

where $[\cdot]$ denotes the greatest integer function.

We note that for any $0 < \varepsilon < \frac{1}{2}$ and $t > \frac{1}{\pi\varepsilon}$ the numbers N_1 and N_2 defined above satisfy the following inequalities:

$$\frac{N_2}{N_1} < \frac{1+\varepsilon}{1-2\varepsilon} \quad \text{and} \quad \frac{N_2 - N_1}{N_1} < \frac{3\varepsilon}{1-2\varepsilon}. \qquad \ldots(2.26)$$

We define for $r = 0,1$

$$W_j^{(r)}(x) = \frac{1}{a_j^{1+r}} \phi^{(r)}\left(\frac{x-X_j}{a_j}\right) \qquad \ldots(2.27)$$

$$S_{m,r}(x) = \sum_{j=1}^{m} \{W_j^{(r)}(x) - f^{(r)}(x)\} \qquad \ldots(2.28)$$

and

$$Z_r(x) = \max_{N_1 < n \leq N_2} \left| \sum_{j=N_1+1}^{n} [W_j^{(r)}(x) - E\{W_j^{(r)}(x)\}]\right| \ldots(2.29)$$

In order to study the asymptotic distribution of $\sqrt{N_t a_{N_t}^{1+2r}}\,[f_{N_t}^{(r)}(x) - f^{(r)}(x)]$, we find it convenient to choose a specific sequence $\{a_n = n^{-\delta},\ n = 1,2,3,\ldots\}$, where δ is some positive number. With this choice of $\{a_n\}$ we have the following lemma.

Lemma 2.10. Let $\phi(u)$ and $\phi^{(1)}(u)$ satisfy Conditions (2.5), (2.6), and (2.7), and $\{a_n = n^{-\delta}\}$, $\delta > 0$. If $x \in C(f)$, then for any $0 < \varepsilon < \frac{1}{2}$, $t > \frac{1}{\pi\varepsilon}$ and $r = 0,1$, we have

$$P\left[Z_r(x) \geq \varepsilon^{\frac{1}{3}} \sqrt{\frac{N_1}{a_{N_1}^{1+2r}}}\right] < \frac{C\varepsilon^{\frac{1}{3}}}{1-2\varepsilon} \left(\frac{1+\varepsilon}{1-2\varepsilon}\right)^{\delta(1+2r)}$$

where C is a positive constant.

Proof. By Kolmogorov's inequality,

$$P\left[Z_r(x) \geq \varepsilon^{\frac{1}{3}} \sqrt{\frac{N_1}{a_{N_1}^{1+2r}}}\right]$$

$$\leq \frac{\sum_{j=N_1+1}^{N_2} \operatorname{Var}\{W_j^{(r)}(x)\}}{\varepsilon^{\frac{2}{3}}\left(\frac{N_1}{a_{N_1}^{1+2r}}\right)}$$

$$\leq \frac{\sum_{j=N_1+1}^{N_2} E\left|W_j^{(r)}(x)\right|^2}{\varepsilon^{\frac{2}{3}}\left(\frac{N_1}{a_{N_1}^{1+2r}}\right)}$$

$$= \frac{\sum_{j=N_1+1}^{N_2} \frac{1}{a_j^{1+2r}} \int_{-\infty}^{\infty} \frac{1}{a_j} \left|\phi^{(r)}\left(\frac{y}{a_j}\right)\right|^2 f(x-y)\,dy}{\varepsilon^{\frac{2}{3}}\left(\frac{N_1}{a_{N_1}^{1+2r}}\right)}.$$

By Lemma 2.1 for any fixed $x \in C(f)$ and $r = 0,1$, the sequence $\{\int_{-\infty}^{\infty} \frac{1}{a_j} |\phi^{(r)}(\frac{y}{a_j})|^2 f(x-y)dy,\ j = 1,2,\ldots\}$ is bounded. Using this result and the inequalities in (2.26), we now conclude that for any $0 < \varepsilon < \frac{1}{2}$ and $t > \frac{1}{\pi\varepsilon}$ and $r = 0,1$

$$P\left[Z_r(x) \geq \varepsilon^{\frac{1}{3}} \sqrt{\frac{N_1}{a_{N_1}^{1+2r}}}\right]$$

$$< \frac{C(N_2-N_1)}{3\varepsilon^{\frac{2}{3}} N_1} \left(\frac{a_{N_1}}{a_{N_2}}\right)^{1+2r}$$

$$\leq \frac{C\varepsilon^{\frac{1}{3}}}{1-2\varepsilon} \left(\frac{1+\varepsilon}{1-2\varepsilon}\right)^{\delta(1+2r)}$$

where C is a positive constant.

Lemma 2.11. Let $\phi(u)$ and $\phi^{(1)}(u)$ satisfy Conditions (2.6), (2.12) and $\{a_n = n^{-\delta}\}$, $\delta > 0$. If for $r = 0,1$ the first $(r+1)$ derivatives of $f(x)$ exist and are bounded and if $0 < \varepsilon < \frac{1}{2}$, $t > \frac{1}{\pi\varepsilon}$ and $\delta > \frac{1}{3+2r}$, then for all n such that $N_1 < n \le N_2$

$$\sup_{-\infty<x<\infty} \left| \sqrt{\frac{a_{N_1}^{1+2r}}{N_1}} \sum_{j=N_1+1}^{n} [E\{W_j^{(r)}(x)\} - f^{(r)}(x)] \right| < \frac{C\varepsilon}{1-2\varepsilon}$$

where C is a positive constant.

Proof. For $r = 0,1$ we obtain from the proof of Lemma 2.7 that for all n such that $N_1 < n \le N_2$

$$\sum_{j=N_1+1}^{n} [E\{W_j^{(r)}(x)\} - f^{(r)}(x)]$$

$$\le \{\sup_{-\infty<y<\infty} |f^{(r+1)}(y)| \int_{-\infty}^{\infty} |u|\phi(u)du\}\{\sum_{j=N_1+1}^{n} a_j\}$$

$$\le \frac{C}{3}(N_2 - N_1)a_{N_1}.$$

Hence,

$$\sup_{-\infty<x<\infty} \left| \sqrt{\frac{a_{N_1}^{1+2r}}{N_1}} \sum_{j=N_1+1}^{n} [E\{W_j^{(r)}(x)\} - f^{(r)}(x)] \right|$$

$$\le \frac{C(N_2-N_1)}{3N_1}(N_1 a_{N_1}^{3+2r})^{\frac{1}{2}}$$

$$= \frac{C(N_2-N_1)}{3N_1}(N_1^{1-\delta(3+2r)})^{\frac{1}{2}}$$

$$\le \frac{C\varepsilon}{1-2\varepsilon}.$$

<u>Theorem 2.3</u>. Let $\phi(u)$ and $\phi^{(1)}(u)$ satisfy Conditions (2.5), (2.6), (2.7) and (2.12) and $\{a_n = n^{-\delta}\}$, $\delta > 0$. If for $r = 0,1$, the first $(r+1)$ derivatives of $f(x)$ exist and are bounded and if $\frac{1}{3+2r} < \delta < 1$, then $\sqrt{N_t a_{N_t}^{1+2r}}\ \{f_{N_t}^{(r)}(x) - f^{(r)}(x)\}$ converges in distribution to a normal random variable with mean zero and variance $\gamma_{1+2r} f(x) \int_{-\infty}^{\infty} |\phi^{(r)}(u)|^2 du$ as t tends to infinity.

<u>Proof</u>. The proof resembles that of Theorem 1 in Renyi [31]. let ε ($\varepsilon < \frac{1}{2}$) be an arbitrarily small positive number. Let $t \geq t_o$ where $t_o = t_o(\varepsilon) > \frac{1}{\pi\varepsilon}$ and let N_1 and N_2 be chosen as before.

We have for $r = 0,1$ and $y > 0$

$$P[\sqrt{N_t a_{N_t}^{1+2r}}\ \{f_{N_t}^{(r)}(x) - f^{(r)}(x)\} < y]$$

$$= \sum_{n=1}^{\infty} P[\sqrt{n a_n^{1+2r}}\ \{f_n^{(r)}(x) - f^{(r)}(x)\} < y;\ N_t = n]$$

$$= \sum_{|n-\pi t| < \pi t\varepsilon} P[\sqrt{n a_n^{1+2r}}\ \{f_n^{(r)}(x) - f^{(r)}(x)\} < y;$$
$$N_t = n]$$
$$+ \sum_{|n-\pi t| \geq \pi t\varepsilon} P[\sqrt{n a_n^{1+2r}}\ \{f_n^{(r)}(x) - f^{(r)}(x)\} < y;$$
$$N_t = n]$$

$$\leq \sum_{|n-\pi t| < \pi t\varepsilon} P[\sqrt{n a_n^{1+2r}}\ \{f_n^{(r)}(x) - f^{(r)}(x)\} < y;$$

$$N_t = n]$$

$$+ \varepsilon.$$

Hence,

$$\Big| P[\sqrt{N_t a_{N_t}^{1+2r}}\ \{f_{N_t}^{(r)}(x) - f^{(r)}(x)\} < y]$$

$$- \sum_{|n-\pi t|<\pi t\varepsilon} P[\sqrt{n a_n^{1+2r}}\ \{f_n^{(r)}(x) - f^{(r)}(x)\} < y;$$

$$N_t = n]\Big| < \varepsilon. \qquad \ldots(2.30)$$

Introducing the random variables $W_j^{(r)}(x)$, $S_{m,r}(x)$, and $Z_r(x)$ as defined in (2.27), (2.28), and (2.29) respectively, we have for any n such that $N_1 < n \leq N_2$

$$P[\sqrt{n a_n^{1+2r}}\ \{f_n^{(r)}(x) - f^{(r)}(x)\} < y;\ N_t = n]$$

$$= P\Bigg[\sqrt{\frac{a_n^{1+2r}}{n}}\ \{\sum_{j=1}^{N_1} (W_j^{(r)}(x) - f^{(r)}(x))$$

$$+ \sum_{j=N_1+1}^{n} (W_j^{(r)}(x) - E\{W_j^{(r)}(x)\}) \qquad \ldots(2.31)$$

$$+ \sum_{j=N_1+1}^{n} (E\{W_j^{(r)}(x)\} - f^{(r)}(x))\} < y;\ N_t = n\Bigg]$$

$$\leq P[S_{N_1,r}(x) < y\sqrt{\frac{N_2}{a_{N_2}^{1+2r}}} - \sum_{j=N_1+1}^{n} \{E(W_j^{(r)}(x))$$

$$- f^{(r)}(x)\} + Z_r(x);\ N_t = n]$$

$$= P\Bigg[\sqrt{\frac{a_{N_1}^{1+2r}}{N_1}}\ S_{N_1,r}(x) < y\sqrt{\frac{N_2}{N_1}\left(\frac{a_{N_1}}{a_{N_2}}\right)^{1+2r}}$$

$$- \sqrt{\frac{a_{N_1}^{1+2r}}{N_1}} \sum_{j=N_1+1}^{n} (E\{W_j^{(r)}(x)\} - f^{(r)}(x))$$

$$+ \sqrt{\frac{a_{N_1}^{1+2r}}{N_1}}\; Z_r(x);\; N_t = n\Bigg]$$

$$\leq P\Big[\sqrt{N_1 a_{N_1}^{1+2r}}\,\{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y\sqrt{\left(\frac{N_2}{N_1}\right)^{1+\delta(1+2r)}}$$

$$+ \frac{C\varepsilon}{1-2\varepsilon} + \sqrt{\frac{a_{N_1}^{1+2r}}{N_1}}\; Z_r(x);\; N_t = n\Big], \quad \text{by Lemma 2.11.}$$

Hence,

$$\sum_{|n-\pi t|<\pi t\varepsilon} P\Big[\sqrt{n a_n^{1+2r}}\,\{f_n^{(r)}(x) - f^{(r)}(x)\} < y;\; N_t = n\Big]$$

$$\leq P\Big[\sqrt{N_1 a_{N_1}^{1+2r}}\,\{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y\sqrt{\left(\frac{N_2}{N_1}\right)^{1+\delta(1+2r)}}$$

$$+ \frac{C\varepsilon}{1-2\varepsilon} + \sqrt{\frac{a_{N_1}^{1+2r}}{N_1}}\; Z_r(x);\; |N_t - \pi t| < \pi t\varepsilon\Big]$$

$$\leq P\Big[\sqrt{N_1 a_{N_1}^{1+2r}}\,\{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y\left(\frac{1+\varepsilon}{1-2\varepsilon}\right)^{\frac{1+\delta(1+2r)}{2}}$$

$$+ \frac{C\varepsilon}{1-2\varepsilon} + \varepsilon^{\frac{1}{3}};\; Z_r(x) < \varepsilon^{\frac{1}{3}}\sqrt{\frac{N_1}{a_{N_1}^{1+2r}}};\; |N_t - \pi t| < \pi t\varepsilon\Big]$$

$$+ P\left[Z_r(x) \geq \varepsilon^{\frac{1}{3}}\sqrt{\frac{N_1}{a_{N_1}^{1+2r}}}\right]$$

$$\leq P\Big[\sqrt{N_1 a_{N_1}^{1+2r}}\,\{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y\left(\frac{1+\varepsilon}{1-2\varepsilon}\right)^{\frac{1+\delta(1+2r)}{2}}$$

$$+ \frac{C\varepsilon}{1-2\varepsilon} + \varepsilon^{\frac{1}{3}}\Big] + \frac{C\varepsilon^{\frac{1}{3}}}{1-2\varepsilon}\left(\frac{1+\varepsilon}{1-2\varepsilon}\right)^{\delta(1+2r)}, \qquad \ldots(2.32)$$

by Lemma 2.10.

From (2.31) we get in a similar manner

$$\sum_{|n-\pi t|<\pi t\varepsilon} P[\sqrt{na_n^{1+2r}}\,\{f_n^{(r)}(x) - f^{(r)}(x)\} < y;\ N_t = n]$$

$$\geq P[\sqrt{N_1 a_{N_1}^{1+2r}}\,\{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y$$

$$- \frac{C\varepsilon}{1-2\varepsilon} - \sqrt{\frac{a_{N_1}^{1+2r}}{N_1}}\, Z_r(x);\ |N_t - \pi t| < \pi t\varepsilon]$$

$$\geq P[\sqrt{N_1 a_{N_1}^{1+2r}}\,\{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y - \frac{C\varepsilon}{1-2\varepsilon} - \varepsilon^{\frac{1}{3}};$$

$$Z_r(x) < \varepsilon^{\frac{1}{3}} \sqrt{\frac{N_1}{a_{N_1}^{1+2r}}}\ ;\ |N_t - \pi t| < \pi t\varepsilon]$$

$$\geq P[\sqrt{N_1 a_{N_1}^{1+2r}}\,\{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y - \frac{C\varepsilon}{1-2\varepsilon} - \varepsilon^{\frac{1}{3}}]$$

$$- P\left[Z_r(x) \geq \varepsilon^{\frac{1}{3}} \sqrt{\frac{N_1}{a_{N_1}^{1+2r}}}\right] - P[|N_t - \pi t| \geq \pi t\varepsilon]$$

$$\geq P[\sqrt{N_1 a_{N_1}^{1+2r}}\,\{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y - \frac{C\varepsilon}{1-2\varepsilon} - \varepsilon^{\frac{1}{3}}]$$

$$- \frac{C\varepsilon^{\frac{1}{3}}}{1-2\varepsilon} \left(\frac{1+\varepsilon}{1-2\varepsilon}\right)^{\delta(1+2r)} - \varepsilon. \qquad \ldots(2.33)$$

From (2.30), (2.32), and (2.33) we conclude that for $t \geq t_o$,

$$P[\sqrt{N_t a_{N_t}^{1+2r}}\,\{f_{N_t}^{(r)}(x) - f^{(r)}(x)\} < y]$$

$$\leq P[\sqrt{N_1 a_{N_1}^{1+2r}}\,\{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y\left(\frac{1+\varepsilon}{1-2\varepsilon}\right)^{\frac{1+\delta(1+2r)}{2}}$$

$$+ \frac{C\varepsilon}{1-2\varepsilon} + \varepsilon^{\frac{1}{3}}] + \frac{C\varepsilon^{\frac{1}{3}}}{1-2\varepsilon} \left(\frac{1+\varepsilon}{1-2\varepsilon}\right)^{\delta(1+2r)} + \varepsilon$$

and

$$P[\sqrt{N_t a_{N_t}^{1+2r}} \{f_{N_t}^{(r)}(x) - f^{(r)}(x)\} < y]$$

$$\geq P[\sqrt{N_1 a_{N_1}^{1+2r}} \{f_{N_1}^{(r)}(x) - f^{(r)}(x)\} < y - \frac{C\varepsilon}{1-2\varepsilon} - \varepsilon^{\frac{1}{3}}]$$

$$- \frac{C\varepsilon^{\frac{1}{3}}}{1-2\varepsilon} \left(\frac{1+\varepsilon}{1-2\varepsilon}\right)^{\delta(1+2r)} - 2\varepsilon.$$

Similar statements hold for $y < 0$. We now invoke Lemma 2.8 and the continuity of the distribution function of a normal random variable to complete the proof of the theorem.

Remark 2.4. We note that Caroll [8] has given a proof for the asymptotic normality of $f_{N_t}^{(r)}(x)$, $r = 0$, under conditions different from ours.

2.6 Asymptotic Normality of the Sample Mode

In this section we study conditions under which the distribution of the sample mode θ_{N_t} is asymptotically normal.

Lemma 2.12. If for $r = 2$ the conditions of part (a) of Theorem 2.1 are satisfied and if the random variable $\hat{\theta}_{N_t}$ converges in probability to θ as t tends to infinity, then $f_{N_t}^{(2)}(\hat{\theta}_{N_t})$ converges in probability to $f^{(2)}(\theta)$ as t tends to infinity.

Proof. We have with probability one

$$|f_{N_t}^{(2)}(\hat{\theta}_{N_t}) - f^{(2)}(\theta)|$$

$$\leq |f_{N_t}^{(2)}(\hat{\theta}_{N_t}) - f^{(2)}(\hat{\theta}_{N_t})| + |f^{(2)}(\hat{\theta}_{N_t}) - f^{(2)}(\theta)|$$

$$\leq \sup_{-\infty<x<\infty} |f_{N_t}^{(2)}(x) - f^{(2)}(x)| + |f^{(2)}(\hat{\theta}_{N_t}) - f^{(2)}(\theta)|.$$

The proof follows from part (a) of Theorem 2.1 and the hypothesis.

We assume that $\{a_n = n^{-\delta}\}$, $\delta > 0$.

Expanding $f_{N_t}^{(1)}(\theta_{N_t})$ around θ we get

$$f_{N_t}^{(1)}(\theta_{N_t}) = f_{N_t}^{(1)}(\theta) + (\theta_{N_t}-\theta)f_{N_t}^{(2)}(\theta_{N_t}^*) = 0 \qquad \ldots(2.34)$$

where $|\theta_{N_t}^*-\theta| < |\theta_{N_t}-\theta|$. Hence, if $\theta_{N_t}-\theta = o_p(1)$, then $\theta_{N_t}^*-\theta = o_p(1)$. From (2.34) we obtain

$$\sqrt{N_t a_{N_t}^3}\,(\theta_{N_t}-\theta) = \frac{-\sqrt{N_t a_{N_t}^3}\; f_{N_t}^{(1)}(\theta)}{f_{N_t}^{(2)}(\theta_{N_t}^*)} . \qquad \ldots(2.35)$$

If $f^{(1)}(\theta) = 0$, then we can conclude from Theorem 2.3 that $\sqrt{N_t a_{N_t}^3}\; f_{N_t}^{(1)}(\theta)$ converges in distribution to a normal random variable with mean zero and variance $\gamma_3 f(\theta) \int_{-\infty}^{\infty} |\phi^{(1)}(u)|^2 du$ as t tends to infinity, provided $\{a_n = n^{-\delta}\}$, $\frac{1}{5} < \delta < 1$. Similarly from Lemma 2.12 we can conclude that

$$f_{N_t}^{(2)}(\theta_{N_t}^*) = f^{(2)}(\theta) + o_p(1)$$

provided $\{a_n = n^{-\delta}\}$, $0 < \delta < \frac{1}{6}$. We now note that there does not exist a sequence $\{a_n = n^{-\delta}\}$ which can be used to prove the convergence results for the numerator and the denominator of (2.35). Due to this difficulty we shall use additional conditions in Theorem 2.3. We state the following conditions:

$$\int_{-\infty}^{\infty} u\phi(u)du = 0 \qquad \ldots(2.36)$$

$$\int_{-\infty}^{\infty} u^2\phi(u)du < \infty, \qquad \ldots(2.37)$$

$$f^{(1)}(\theta) = 0; \quad f^{(2)}(\theta) < 0. \qquad \ldots(2.38)$$

We prove the following theorem.

Theorem 2.4. Let $\{a_n = n^{-\delta}\}$, $\frac{1}{7} < \delta < \frac{1}{6}$. If Conditions (2.5), (2.6), (2.7), (2.18), (2.24), (2.36), (2.37), and (2.38) are satisfied, and if for $r = 2$ Conditions (2.16), (2.17) and (2.19) are also satisfied, then $\sqrt{N_t a_{N_t}^3}\,(\theta_{N_t} - \theta)$ converges in distribution to a normal random variable with mean zero and variance

$$\gamma_3 f(\theta)\,\frac{\int_{-\infty}^{\infty} |\phi^{(1)}(u)|^2 du}{\{f^{(2)}(\theta)\}^2}$$ as t tends to infinity.

Proof. We first note that according to the hypothesis, the conditions of part (a) of Theorem 2.1 with $r = 0,1,2$ are satisfied. We also note that from the uniqueness of the mode θ, the uniform continuity of $f(x)$, and part (a) of Theorem 2.1, the sample mode θ_{N_t} converges in probability to θ as t tends to infinity. Using Lemma 2.12, we can write relation (2.35) as:

$$\sqrt{N_t a_{N_t}^3}\ (\theta_{N_t} - \theta) = \frac{-\sqrt{N_t a_{N_t}^3}\ f_{N_t}^{(1)}(\theta)}{f^{(2)}(\theta) + o_p(1)} . \qquad \ldots (2.39)$$

We now show that under the hypothesis $\sqrt{N_t a_{N_t}^3}\ f_{N_t}^{(1)}(\theta)$ converges in distribution to a normal random variable with mean zero and variance $\gamma_3 f(\theta) \int_{-\infty}^{\infty} |\phi^{(1)}(u)|^2 du$ as t tends to infinity. We note that Condition (2.16) with r = 2 implies that the third derivative of f(x) is bounded. In relation (2.15) expanding $f^{(1)}(x-a_j u)$ around x to the order of $a_j^2 = j^{-2\delta}$ $(\frac{1}{7} < \delta < \frac{1}{6})$ we get

$$\sqrt{n a_n^3} \sup_{-\infty<x<\infty} [E\{f_n^{(1)}(x)\} - f^{(1)}(x)] = O(n^{1-7\delta})^{\frac{1}{2}}$$

$$= o(1).$$

Hence, we note that the conclusions of Lemma 2.8 and Theorem 2.3 hold with r = 1. From relation (2.39) we have

$$\sqrt{N_t a_{N_t}^3}\ (\theta_{N_t} - \theta) = \frac{-\sqrt{N_t a_{N_t}^3}\ f_{N_t}^{(1)}(\theta)}{f^{(2)}(\theta)} \{1 + o_p(1)\}$$

$$= \frac{-\sqrt{N_t a_{N_t}^3}\ f_{N_t}^{(1)}(\theta)}{f^{(2)}(\theta)} + O_p(1) \cdot o_p(1)$$

$$= \frac{-\sqrt{N_t a_{N_t}^3}\ f_{N_t}^{(1)}(\theta)}{f^{(2)}(\theta)} + o_p(1).$$

The proof of the theorem now follows from this computation.

CHAPTER III

EFFICIENT NONPARAMETRIC ESTIMATION OF A SHIFT PARAMETER

3.1 The Empirical Likelihood Method for Estimating a Shift Parameter

Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be independent random variables with probability density functions f and g respectively. Suppose there exists a probability density function h and two real numbers θ_1 and θ_2 such that

$$f(x) = h(x-\theta_1) \quad \text{and} \quad g(x) = h(x-\theta_2)$$

for all real numbers x. In such a case, the number $\theta = \theta_2 - \theta_1$ is said to be the shift between the densities f and g. Let $\{m_N\}$ and $\{n_N\}$ be sequences of positive integers such that $m_N + n_N = N$ for each N and $\lim_{N\to\infty} \frac{m_N}{N} = \lambda$, $0 < \lambda < 1$. The classical maximum likelihood method of estimating θ on the basis of $X_1, X_2, \ldots, X_{m_N}$, $Y_1, Y_2, \ldots, Y_{n_N}$ when h is known consists of solving the equations

$$\sum_{i=1}^{m_N} \frac{h^{(1)}(X_i - t_1)}{h(X_i - t_1)} = 0 \qquad \ldots(3.1)$$

$$\sum_{j=1}^{n_N} \frac{h^{(1)}(Y_j - t_2)}{h(Y_j - t_2)} = 0. \qquad \ldots(3.2)$$

If h satisfies certain regularity conditions (see for example, Cramer [11], p.500), then equations (3.1) and (3.2) have solutions t_{1N} and t_{2N} converging in probability to θ_1 and θ_2 respectively so that $t_N = t_{2N}-t_{1N}$ converges in probability to the shift θ. Furthermore, $\sqrt{N}\ (t_N-\theta)$ converges in distribution to a normal random variable with mean zero and variance $\frac{1}{\lambda\mu J(f)}$ where $\mu = 1-\lambda$ and $J(f)$ is the Fisher information defined by

$$J(f) = \int_{-\infty}^{\infty} \frac{\{f^{(1)}(x)\}^2}{f(x)}\, dx. \qquad \ldots(3.3)$$

The nonparametric counterpart of this problem arises when f is unknown. In this chapter we propose a method for estimating θ in the nonparametric set-up which first obtains a kernel estimate of the common density of $Y_1, Y_2, \ldots, Y_{N\mu}$ and then the maximum likelihood principle is invoked to estimate the shift required on $X_1, X_2, \ldots, X_{N\lambda}$ in order to match them with this estimated density. In what follows, we shall take $m_N = N\lambda$ and $n_N = N\mu$ and proceed as if $N\lambda$ and $N\mu$ were integers. In this way we shall avoid unnecessary complications without affecting our analysis.

By means of a Borel measurable function ϕ and a sequence $\{a_N\}$ of positive numbers converging to zero, we define a kernel estimate of the r-th derivative of the common density function of $Y_1, Y_2, \ldots, Y_{N\mu}$ as

$$f_N^{(r)}(x) = \frac{1}{N\mu a_N^{1+r}} \sum_{j=1}^{N\mu} \phi^{(r)}\left(\frac{x-Y_j}{a_N}\right), \quad r = 0,1,2,.. \qquad \ldots(3.4)$$

for all real x. Let $\{b_N\}$ be a sequence of positive numbers converging to infinity and I_N be the indicator function of the interval $(-b_N, b_N)$. For every t, we define

$$L_N(t) = f_N(-b_N+t) + \frac{1}{N\lambda}\sum_{i=1}^{N\lambda} I_N(X_i)\,\frac{f_N^{(1)}(X_i+t)}{f_N(X_i+t)}$$

$$- f_N(b_N+t) \qquad \ldots(3.5)$$

and call

$$L_N(t) = 0 \qquad \ldots(3.6)$$

the empirical likelihood equation for estimating θ.

We examine the maximum likelihood estimate of θ in a different way to explain the motivation behind equation (3.6). We have

$$\sum_{i=1}^{N\lambda} \log h(X_i - t_{1N}) = \sum_{i=1}^{N\lambda} \log h(X_i + t_N - t_{2N})$$

$$= \sum_{i=1}^{N\lambda} \log \hat{g}_N(X_i + t_N)$$

where for every t, $\hat{g}_N(z+t) = h(z+t-t_{2N})$ is an estimate of $h(z+t-\theta_2)$ which is the density function of the random variables $Y_j - t$, $j = 1,2,\ldots,N\mu$. Hence we propose a method of estimating θ in the nonparametric set-up which first obtains an estimate $f_N(z+t)$ of $f(z+t-\theta)$ based on the Y-sample and then uses the X-sample to find the value of t for which $\sum_{i=1}^{N\lambda} \log f_N(X_i+t)$ is maximized. This gives the following equation:

$$\frac{1}{N\lambda}\sum_{i=1}^{N\lambda} \frac{f_N^{(1)}(X_i+t)}{f_N(X_i+t)} = 0.$$

However, for technical reasons we have modified the above equation. For those X_i's lying in the interval $(-b_N, b_N)$ we have used the above form of the left hand side of the equation. We have then added $f_N(-b_N+t)$ and $-f_N(b_N+t)$ to this expression to give rise to equation (3.5).

It will be shown in Theorem 3.1 that under certain regularity conditions, with probability approaching one as N tends to infinity the empirical likelihood equation (3.6) has a solution T_N which is consistent for θ. Furthermore, $\sqrt{N}\ (T_N-\theta)$ converges in distribution (Theorem 3.2) to a normal random variable with mean zero and variance $\frac{1}{\lambda\mu J(f)}$, where J(f) is the Fisher information as defined in (3.3).

The estimate T_N of the shift proposed here is very difficult to compute. However, in Theorem 3.3 we shall give a simple computational procedure which starts with an easily computable first approximation and after exactly one iteration leads to an estimate $\hat{T}_N$ of θ having the asymptotic properties possessed by T_N.

The method proposed above is an improvement upon a method developed by Samanta [38].

We now state and discuss the regularity conditions on f, ϕ, and the sequence $\{a_N\}$ which we shall refer to as Conditions A.3.

3.2 Conditions A.3

Condition 3.1: f and its first five derivatives are bounded.

Condition 3.2: $J(f) = \int_{-\infty}^{\infty} \{\frac{d}{dx} \log f(x)\}^2 f(x)dx < \infty.$

Condition 3.3: $\frac{\partial^2}{\partial x^2} \log f(x)$ is uniformly continuous.

Condition 3.4: $\int_{-\infty}^{\infty} f^{(1)}(x)dx = \int_{-\infty}^{\infty} f^{(2)}(x)dx = 0.$

Condition 3.5: There exists a strictly monotone increasing function H such that $\sup_{|x| \leq y} \frac{1}{f(x)} \leq H(y)$ for all y and $H(b_N+1) = N^{1/25}$.

Condition 3.6: ϕ and its first two derivatives are continuous functions of bounded variation.

Condition 3.7: $\int_{-\infty}^{\infty} \phi(u)du = 1, \int_{-\infty}^{\infty} u^r \phi(u)du = 0,$ $r = 1,2,3,4,$ and $\int_{-\infty}^{\infty} |u^5 \phi(u)|du < \infty.$

Condition 3.8: $a_N = N^{-\delta}, \quad \frac{1}{8} < \delta < \frac{19}{150}$.

Probability density functions which satisfy Conditions 3.1, 3.2, 3.3, and 3.4 include among others the normal, the Cauchy, the contaminated normal, and mixtures of the Cauchy and normal probability density functions.

Condition 3.5 is similar to that in Bhattacharya [5], p.381. If we let $H(y) = \sqrt{2\pi}\, e^{y^2/2}$, then Condition 3.5 is satisfied if f is the standard normal density function or any other density function having flatter tails than the standard normal probability density curve.

The kernel ϕ satisfying Conditions 3.6 and 3.7 can be obtained in the following manner. For some integer $\ell \geq 0$ let $P_{2\ell}(x) = c(1+\beta_1 x^2 + \ldots + \beta_\ell x^{2\ell})$ be a polynomial of degree 2ℓ. We now define another kernel $K_{2\ell}(x)$ given by

$$K_{2\ell}(x) = P_{2\ell}(x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$= c(1+\beta_1 x^2 + \ldots + \beta_\ell x^{2\ell}) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad \ldots(3.7)$$

where the constants $c, \beta_1, \ldots, \beta_\ell$ are to be chosen such that

$$\int_{-\infty}^{\infty} K_{2\ell}(x)dx = 1 \qquad \ldots(3.8)$$

and

$$\int_{-\infty}^{\infty} x^{2r} K_{2\ell}(x)dx = 0, \text{ for } r = 1,\ldots,\ell. \qquad \ldots(3.9)$$

Since $\int_{-\infty}^{\infty} x^{2r} e^{-x^2/2} dx = 2^{\frac{2r+1}{2}} \Gamma(\frac{2r+1}{2})$, we can use the condition in (3.9) to determine the constants $\beta_1, \ldots, \beta_\ell$ by solving a set of ℓ simultaneous equations. Then c can be chosen such that (3.8) is satisfied. The kernel $K_{2\ell}(x)$ so obtained, together with all its derivatives, is a continuous function of bounded variation. We note that the kernels obtained in this manner possess similar properties as those suggested by Schucany and Sommers [40]. For $\ell = 2$, the kernel $K_{2\ell}(x)$ suggested in (3.7) is given by

$$K_4(x) = \frac{15}{8}\left(1 - \frac{2}{3}x^2 + \frac{1}{15}x^4\right)\frac{1}{\sqrt{2\pi}}\, e^{-x^2/2},$$

and this is the kernel given in Nadaraya [28]. By setting $\phi(x) = K_4(x)$, we get a kernel function which satisfies Conditions 3.6 and 3.7 of Conditions A.3.

3.3 Convergence Properties of Kernel Estimates of a Density Function and Its Derivatives

We now consider the following lemmas.

Lemma 3.1. Under Conditions 3.1, 3.6, and 3.7,

$$\sup_{-\infty<z<\infty} \left|E[f_N^{(r)}(z)] - f^{(r)}(z-\theta)\right| \leq Ca_N^{5-r}, \quad r = 0,1,2,$$

where C is a positive constant.

Proof. The proof is very similar to that in Samanta [38] and will be omitted.

Lemma 3.2. Under Conditions 3.6 there exists a universal constant C such that for any $N > 0$ and $\varepsilon_N > 0$

$$P\left[\sup_{-\infty<z<\infty} \left|f_N^{(r)}(z) - E\{f_N^{(r)}(z)\}\right| \geq \varepsilon_N\right]$$

$$\leq C \exp(-2N\mu\varepsilon_N^2 a_N^{2r+2}/\mu_r^2)$$

for $r = 0,1,2$ and where $\mu_r = \int_{-\infty}^{\infty} |\phi^{(r+1)}(u)|du.$

Proof. See Lemma 2.2 of Schuster [41].

Lemma 3.3. Under Conditions 3.1, 3.6, and 3.7 for sufficiently large N and for $r = 0,1,2$, if $a_N^{5-r} = o(\varepsilon_N)$, then

$$P[\sup_{-\infty<z<\infty} |f_N^{(r)}(z) - f^{(r)}(z-\theta)| \geq \varepsilon_N]$$

$$\leq C_1 \exp(-C_2 N \varepsilon_N^2 a_N^{2r+2})$$

where C_1 and C_2 are positive constants independent of N.

Proof. We have for $r = 0,1,2$

$$\sup_{-\infty<z<\infty} |f_N^{(r)}(z) - f^{(r)}(z-\theta)|$$

$$\leq \sup_{-\infty<z<\infty} |f_N^{(r)}(z) - E[f_N^{(r)}(z)]|$$

$$+ \sup_{-\infty<z<\infty} |E[f_N^{(r)}(z)] - f^{(r)}(z-\theta)|$$

$$\leq \sup_{-\infty<z<\infty} |f_N^{(r)}(z) - E[f_N^{(r)}(z)]| + Ca_N^{5-r}, \text{ by Lemma 3.1.}$$

Hence, if $a_N^{5-r} = o(\varepsilon_N)$, then we obtain for all sufficiently large N,

$$P[\sup_{-\infty<z<\infty} |f_N^{(r)}(z) - f^{(r)}(z-\theta)| \geq \varepsilon_N]$$

$$\leq P[\sup_{-\infty<z<\infty} |f_N^{(r)}(z) - E\{f_N^{(r)}(z)\}| \geq \frac{\varepsilon_N}{2}]$$

$$\leq C_1 \exp(-C_2 N \varepsilon_N^2 a_N^{2r+2}), \text{ by Lemma 3.2.}$$

This completes the proof.

We now consider the function

$$L_N^*(t) = \frac{1}{N\lambda} \sum_{i=1}^{N\lambda} I_N(X_i) \frac{f^{(1)}(X_i+t-\theta)}{f(X_i+t-\theta)}.$$

We also define $L_N^{(1)}(t) = \frac{\partial}{\partial t} L_N(t)$ and $L_N^{*(1)}(t) = \frac{\partial}{\partial t} L_N^*(t)$.

We have the following lemma on the convergence to zero of $L_N(t) - L_N^*(t)$ and $L_N^{(1)}(t) - L_N^{*(1)}(t)$.

<u>Lemma 3.4</u>. Under Conditions A.3 for every $\varepsilon > 0$,

(a) $\lim_{N\to\infty} P[\sup_{|t-\theta|\le 1} |L_N(t) - L_N^*(t)| > \varepsilon] = 0$

(b) $\lim_{N\to\infty} P[\sup_{|t-\theta|\le 1} |L_N^{(1)}(t) - L_N^{*(1)}(t)| > \varepsilon] = 0.$

<u>Proof</u>. We shall only demonstrate the proof of part (a) of the lemma. The proof of part (b) is similar and is omitted. We have

$$\sup_{|t-\theta|\le 1} |L_N(t) - L_N^*(t)|$$

$$\le \sup_{|t-\theta|\le 1} |f_N(-b_N+t) - f(-b_N+t-\theta)| + \sup_{|t-\theta|\le 1} |f(-b_N+t-\theta)|$$

$$+ \sup_{|t-\theta|\le 1} |f_N(b_N+t) - f(b_N+t-\theta)| + \sup_{|t-\theta|\le 1} |f(b_N+t-\theta)|$$

$$+ \sup_{|t-\theta|\le 1} \frac{1}{N\lambda} \sum_{i=1}^{N\lambda} I_N(X_i) \left| \frac{f_N^{(1)}(X_i+t)}{f(X_i+t)} - \frac{f^{(1)}(X_i+t-\theta)}{f(X_i+t-\theta)} \right| .$$

In view of Lemma 3.3 it now suffices to show that for any $\varepsilon > 0$,

$$\lim_{N\to\infty} N\lambda P[\sup_{|z|\le b_N+1} \left| \frac{f_N^{(1)}(z+\theta)}{f_N(z+\theta)} - \frac{f^{(1)}(z)}{f(z)} \right| > \varepsilon] = 0.$$

If

$$\sup_{|z|\le b_N+1} |f_N(z+\theta) - f(z)| < \varepsilon_N,$$

$$\sup_{|z|\le b_N+1} |f_N^{(1)}(z+\theta) - f^{(1)}(z)| < \varepsilon_N,$$

$$\sup_{|z|\le b_N+1} \frac{1}{f(z)} \le H(b_N+1),$$

and

$$\lim_{N\to\infty} \varepsilon_N\{H(b_N+1)\} = 0,$$

then for all sufficiently large N,

$$\sup_{|z|\le b_N+1} \left| \frac{f_N^{(1)}(z+\theta)}{f_N(z+\theta)} - \frac{f^{(1)}(z)}{f(z)} \right|$$

$$\le \sup_{|z|\le b_N+1} \frac{|f_N^{(1)}(z+\theta)-f^{(1)}(z)|}{|f_N(z+\theta)|}$$

$$+ \sup_{|z|\le b_N+1} \frac{|f^{(1)}(z)|\,|f_N(z+\theta)-f(z)|}{|f(z)|\,|f_N(z+\theta)|}$$

$$\le C\ \varepsilon_N\{H(b_N+1)\}^2.$$

If we define $\varepsilon_N = \dfrac{\varepsilon}{C\{H(b_N+1)\}^2}$, then

$$\lim_{N\to\infty} \varepsilon_N\{H(b_N+1)\} = \lim_{N\to\infty} \frac{\varepsilon}{C\{H(b_N+1)\}} = 0.$$

Hence, for all sufficiently large N,

$$P\left[\sup_{|z|\le b_N+1} \left| \frac{f_N^{(1)}(z+\theta)}{f_N(z+\theta)} - \frac{f^{(1)}(z)}{f(z)} \right| > \varepsilon\right]$$

$$\le P\left[\sup_{|z|\le b_N+1} |f_N(z+\theta) - f(z)| \ge \frac{\varepsilon}{C\{H(b_N+1)\}^2}\right]$$

$$+ P\left[\sup_{|z|\le b_N+1} |f_N^{(1)}(z+\theta) - f^{(1)}(z)| \ge \frac{\varepsilon}{C\{H(b_N+1)\}^2}\right].$$

The proof of part (a) now follows from Lemma 3.3 and the hypothesis.

3.4 Consistency and Asymptotic Efficiency of the Nonparametric Estimator of the Shift Parameter

We now prove the following lemma.

Lemma 3.5. Under Conditions 3.2 and 3.4,

(a) $L_N^*(\theta) = o_p(1)$

(b) $L_N^{*(1)}(\theta) = -J(f) + o_p(1)$.

Proof. We first give the proof of part (b) of the lemma. Let us introduce the random variable η_N defined by

$$\eta_N = \frac{1}{N\lambda}\sum_{i=1}^{N\lambda}\left[\frac{f^{(2)}(X_i)}{f(X_i)} - \left\{\frac{f^{(1)}(X_i)}{f(X_i)}\right\}^2\right].$$

Using Conditions 3.2 and 3.4 we have

$$E\left[\frac{f^{(2)}(X_1)}{f(X_1)} - \left\{\frac{f^{(1)}(X_1)}{f(X_1)}\right\}^2\right] = -J(f).$$

Hence, by Khintchne's theorem, $\eta_N = -J(f) + o_p(1)$. Using this fact, it now suffices to show that

$\lim_{N\to\infty} E[|L_N^{*(1)}(\theta) - \eta_N|] = 0$. Now,

$$\begin{aligned} &E[|L_N^{*(1)}(\theta) - \eta_N|] \\ &\le E\Big[\frac{1}{N\lambda}\sum_{i=1}^{N\lambda}|I_N(X_i) - 1|\left|\frac{f^{(2)}(X_i)}{f(X_i)}\right| \\ &\quad + \frac{1}{N\lambda}\sum_{i=1}^{N\lambda}|1 - I_N(X_i)|\left\{\frac{f^{(1)}(X_i)}{f(X_i)}\right\}^2\Big] \\ &= \int_{|u|\ge b_N}|f^{(2)}(u)|\,du + \int_{|u|\ge b_N}\frac{\{f^{(1)}(u)\}^2}{f(u)}\,du. \end{aligned}$$

The right hand side of the last equality converges to zero as N tends to infinity. This completes the proof of part (b). The proof of part (a) can be accomplished in a similar manner by letting

$$\eta_N = \frac{1}{N\lambda} \sum_{i=1}^{N\lambda} \frac{f^{(1)}(X_i)}{f(X_i)}.$$

<u>Lemma 3.6</u>. Under Condition 3.3 for every $\varepsilon_1 > 0$, there is a $\delta_1 > 0$ such that

$$P[\sup_{|t-\theta|\leq\delta_1} |L_N^{*(1)}(t) - L_N^{*(1)}(\theta)| < \varepsilon_1] = 1.$$

<u>Proof</u>. We have

$$L_N^{*(1)}(t) - L_N^{*(1)}(\theta)$$

$$= \frac{1}{N\lambda} \sum_{i=1}^{N\lambda} I_N(X_i)[\frac{\partial^2}{\partial t^2} \log f(X_i+t-\theta)$$

$$- \frac{\partial^2}{\partial t^2} \log f(X_i+t-\theta)\Big|_{t=\theta}].$$

Due to the uniform continuity of $\frac{\partial^2}{\partial x^2} \log f(x)$, we conclude that for every $\varepsilon_1 > 0$ there is a $\delta_1 > 0$ such that

$$\left|\frac{\partial^2}{\partial t^2} \log f(x-t-\theta) - \frac{\partial^2}{\partial t^2} \log f(x+t-\theta)\Big|_{t=\theta}\right| < \varepsilon_1$$

for all x whenever $|t-\theta| \leq \delta_1$.

Using this fact we have

$$P[\sup_{|t-\theta|\leq\delta_1} |L_N^{*(1)}(t) - L_N^{*(1)}(\theta)| < \varepsilon_1] = 1.$$

Lemma 3.7. Under Conditions 3.2, 3.3, and 3.4 for arbitrarily small $\varepsilon > 0$ and $\delta > 0$, there exists $N_o = N_o(\varepsilon, \delta)$ such that for $N > N_o$, $P[\{L_N^*(\theta-\delta) > 0\} \cap \{L_N^*(\theta+\delta) < 0\}] > 1-\varepsilon$. Hence, with probability approaching one as N tends to infinity the equation $L_N^*(t) = 0$ has a solution which is consistent for θ.

Proof. Expanding $L_N^*(t)$ around θ gives

$$L_N^*(t) = L_N^*(\theta) + (t-\theta)L_N^{*(1)}(\theta_1)$$

where $|\theta_1-\theta| < |t-\theta|$. Let ε_1 in Lemma 3.6 be less than $\frac{J(f)}{4}$ and let the corresponding δ_1 be chosen. We select arbitrarily small positive numbers ε, δ where $\delta < \delta_1$.

If $|t-\theta| < \delta$, then $|\theta_1-\theta| < \delta$. Hence, by Lemma 3.6,

$$P[|L_N^{*(1)}(\theta_1) - L_N^{*(1)}(\theta)| < \varepsilon_1] = 1.$$

Therefore,

$$P[L_N^{*(1)}(\theta) - \frac{J(f)}{4} < L_N^{*(1)}(\theta_1) < L_N^{*(1)}(\theta) + \frac{J(f)}{4}] = 1. \quad \text{...(3.10)}$$

By Lemma 3.5, $L_N^*(\theta)$ and $L_N^{*(1)}(\theta)$ converge in probability to zero and to $-J(f)$ respectively as N tends to infinity. Using these facts and the result in (3.10), we conclude that there exists $N_o = N_o(\varepsilon, \delta)$ such that for $N > N_o$,

$$P[|L_N^*(\theta)| \geq \delta^2] < \frac{\varepsilon}{2} \quad \text{and} \quad P[L_N^{*(1)}(\theta_1) \geq \frac{-J(f)}{2}] < \frac{\varepsilon}{2}.$$

We have

$$P[\{|L_N^*(\theta)| \geq \delta^2\} \cup \{L_N^{*(1)}(\theta_1) \geq \frac{-J(f)}{2}\}]$$

$$\leq P[|L_N^*(\theta)| \geq \delta^2] + P[L_N^{*(1)}(\theta_1) \geq \frac{-J(f)}{2}] < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence, for $N > N_o$,

$$P[\{|L_N^*(\theta)| < \delta^2\} \cap \{L_N^{*(1)}(\theta_1) < \frac{-J(f)}{2}\}] > 1-\varepsilon.$$

Suppose $t = \theta \pm \delta$. For any sample point satisfying the inequalities

$$|L_N^*(\theta)| < \delta^2 \quad \text{and} \quad L_N^{*(1)}(\theta_1) < \frac{-J(f)}{2},$$

the expression $L_N^*(t) = L_N^*(\theta) + (t-\theta)L_N^{*(1)}(\theta_1)$ is less than $\delta^2 + \delta(\frac{-J(f)}{2})$ if $t = \theta+\delta$. Hence, $L_N^*(t)$ assumes a negative value if $t = \theta+\delta$ and if δ is chosen less than $\frac{J(f)}{2}$. Similarly for $|L_N^*(\theta)| < \delta^2$ and $L_N^{*(1)}(\theta_1) < \frac{-J(f)}{2}$, the expression $L_N^*(t) = L_N^*(\theta) + (t-\theta)L_N^{*(1)}(\theta_1)$ assumes a positive value if $t = \theta-\delta$ and if δ is chosen less than $\frac{J(f)}{2}$.

We now appeal to the fact that $L_N^*(t)$ is a continuous function of t for every sample point. Thus, for arbitrarily small positive numbers ε and δ, the equation $L_N^*(t) = 0$ will, with a probability exceeding $1-\varepsilon$, have a root between the limits $\theta \pm \delta$ as soon as $N > N_o(\varepsilon,\delta)$.

We now prove the following theorem.

Theorem 3.1. Under Conditions A.3 with probability approaching one as N tends to infinity the empirical likelihood equation $L_N(t) = 0$ has a solution T_N which is consistent for θ.

Proof. By Lemma 3.7 for arbitrarily small positive numbers ε and δ there exists $N_o = N_o(\varepsilon,\delta)$ such that

$$P[\{L_N^*(\theta-\delta) > \varepsilon\} \cap \{L_N^*(\theta+\delta) < -\varepsilon\}] > 1-\varepsilon$$

whenever $N > N_o$. And by part (a) of Lemma 3.4 for arbitrarily small $\varepsilon_1 > 0$ there exists $N_1 = N_1(\varepsilon_1)$ such that

$$P[\sup_{|t-\theta|\leq 1} |L_N(t) - L_N^*(t)| < \varepsilon_1] > 1-\varepsilon_1$$

whenever $N > N_1$. Combining these two results, we conclude that

$$P[\{L_N^*(t) - \varepsilon_1 < L_N(t) < L_N^*(t) + \varepsilon_1 \text{ for all } |t-\theta| \leq 1\}$$

$$\cap \{L_N^*(\theta-\delta) > \varepsilon\} \cap \{L_N^*(\theta+\delta) < -\varepsilon\}] > 1-\varepsilon-\varepsilon_1$$

whenever $N > N^*$ where $N^* = \max(N_o, N_1)$.

If we choose $\delta < 1$ and $\varepsilon_1-\varepsilon < 0$ (hence, $\varepsilon-\varepsilon_1 > 0$), then

$$P[\{L_N(\theta-\delta) > 0\} \cap \{L_N(\theta+\delta) < 0\}] > 1-\varepsilon-\varepsilon_1$$

whenever $N > N^*$.

We now appeal to the fact that $L_N(t)$ is a continuous function of t for every sample point. Hence, with probability greater than $1-\varepsilon-\varepsilon_1$, the equation $L_N(t) = 0$ has a root between the limits $\theta \pm \delta$ as soon as N exceeds N^*.

For any given sample size N there may be sample points for which the empirical likelihood equation $L_N(t) = 0$ has no solution. In such a case we define T_N in an arbitrary manner. But we note that the probability of such sample points can be made arbitrarily small for all sufficiently large N. With this convention, let $T_N = T_N(X_1, X_2, \ldots, X_{N\lambda}, Y_1, Y_2, \ldots, Y_{N\mu})$ be a consistent solution of the equation $L_N(t) = 0$. The following lemma will be used in the derivation of the asymptotic distribution of $\sqrt{N}\,(T_N - \theta)$.

Lemma 3.8. Under Conditions A.3, if $\{\hat{\theta}_N\}$ is a sequence of random variables such that $\hat{\theta}_N = \theta + o_p(1)$, then $-L_N^{(1)}(\hat{\theta}_N) = J(f) + o_p(1)$.

Proof. We have with probability one

$$
\begin{aligned}
&|L_N^{(1)}(\hat{\theta}_N) + J(f)| \\
&\quad \leq |L_N^{(1)}(\hat{\theta}_N) - L_N^{*(1)}(\hat{\theta}_N)| + |L_N^{*(1)}(\hat{\theta}_N) - L_N^{*(1)}(\theta)| \\
&\qquad + |L_N^{*(1)}(\theta) + J(f)|.
\end{aligned}
$$

Hence, for any $\varepsilon > 0$

$$
\begin{aligned}
&P[|L_N^{(1)}(\hat{\theta}_N) + J(f)| > \varepsilon] \\
&\quad \leq P[|L_N^{(1)}(\hat{\theta}_N) - L_N^{*(1)}(\hat{\theta}_N)| > \tfrac{\varepsilon}{3}] \\
&\qquad + P[|L_N^{*(1)}(\hat{\theta}_N) - L_N^{*(1)}(\theta)| > \tfrac{\varepsilon}{3}] \\
&\qquad + P[|L_N^{*(1)}(\theta) + J(f)| > \tfrac{\varepsilon}{3}].
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\lim_{N\to\infty} P[|L_N^{(1)}(\hat{\theta}_N) - L_N^{*(1)}(\hat{\theta}_N)| < \tfrac{\varepsilon}{3}] \\
&\quad \geq \lim_{N\to\infty} P[\{|L_N^{(1)}(\hat{\theta}_N) - L_N^{*(1)}(\hat{\theta}_N)| < \tfrac{\varepsilon}{3}\} \cap \{|\hat{\theta}_N - \theta| \leq 1]\\
&\quad \geq \lim_{N\to\infty} P[\{\sup_{|t-\theta|\leq 1} |L_N^{(1)}(t) - L_N^{*(1)}(t)| < \tfrac{\varepsilon}{3}\} \\
&\qquad\quad \cap \{|\hat{\theta}_N - \theta| \leq 1\}] = 1, \text{ by Lemma 3.4(b) and the hypothesis.}
\end{aligned}
$$

Hence,

$$L_N^{(1)}(\hat{\theta}_N) - L_N^{*(1)}(\hat{\theta}_N) = o_p(1).$$

Using the uniform continuity of $\frac{\partial^2}{\partial x^2} \log f(x)$, it follows from Theorem 2 of Mann and Wald [22] that

$$L_N^{*(1)}(\hat{\theta}_N) - L_N^{*(1)}(\theta) = o_p(1).$$

Finally, by part (b) of Lemma 3.5, we have

$$L_N^{*(1)}(\theta) + J(f) = o_p(1).$$

The proof of the lemma now follows from the above observations.

Expanding $L_N(T_N)$ around θ we obtain

$$L_N(T_N) = L_N(\theta) + (T_N - \theta) L_N^{(1)}(T_N^*) \qquad \ldots (3.11)$$

where

$$|T_N^* - \theta| < |T_N - \theta|.$$

Hence, $T_N^* = \theta + o_p(1)$, by Theorem 3.1. Using Theorem 3.1 and relation (3.11) we have for any real number C

$$\lim_{N\to\infty} P[\sqrt{N}\,(T_N - \theta) \le C] = \lim_{N\to\infty} P[\sqrt{N}\,(T_N - \theta) \le C;\ L_N(T_N) = 0]$$

$$= \lim_{N\to\infty} P\left[\frac{\sqrt{N}\, L_N(\theta)}{-L_N^{(1)}(T_N^*)} \le C\right]. \qquad \ldots (3.12)$$

We conclude that $\sqrt{N}\,(T_N - \theta)$ and $\frac{\sqrt{N}\, L_N(\theta)}{-L_N^{(1)}(T_N^*)}$ have the same limiting distribution. By Lemma 3.8 we have

$$-L_N^{(1)}(T_N^*) = J(f) + o_p(1).$$

It is now evident that the asymptotic distribution of

$\sqrt{N}\ (T_N-\theta)$ depends on the asymptotic distribution of $\sqrt{N}\ L_N(\theta)$. We have

$$L_N(\theta) = V_{1N} + V_{2N}$$

where

$$V_{1N} = V_{1N}(X_1, X_2, \ldots, X_{N\lambda},\ Y_1, Y_2, \ldots, Y_{N\mu})$$

$$= f_N(-b_N+\theta) + \frac{1}{N\lambda} \sum_{i=1}^{N\lambda} I_N(X_i) \frac{f_N^{(1)}(X_i+\theta)}{f(X_i)}$$

$$\cdot\{2 - \frac{f_N(X_i+\theta)}{f(X_i)}\} - f_N(b_N+\theta) \qquad \ldots(3.14)$$

and

$$V_{2N} = V_{2N}(X_1, X_2, \ldots, X_{N\lambda},\ Y_1, Y_2, \ldots, Y_{N\mu})$$

$$= \frac{1}{N\lambda} \sum_{i=1}^{N\lambda} I_N(X_i) \frac{f_N^{(1)}(X_i+\theta)}{f_N(X_i+\theta)} \{\frac{f_N(X_i+\theta)-f(X_i)}{f(X_i)}\}^2 \qquad \ldots(3.15)$$

<u>Lemma 3.9</u>. Under Conditions A.3 for every $\varepsilon > 0$,

$$\lim_{N\to\infty} P[\sqrt{N}\ |V_{2N}| > \varepsilon] = 0.$$

<u>Proof</u>. To prove the lemma is suffices to show that

$$\lim_{N\to\infty} N\lambda P[\sup_{|x|\le b_N} \frac{\{|f_N^{(1)}(x+\theta)|\}^{\frac{1}{2}}\{N^{\frac{1}{4}}|f_N(x+\theta)-f(x)|\}}{\{|f_N(x+\theta)|\}^{\frac{1}{2}}f(x)} > \sqrt{\varepsilon}] = 0.$$

Now, if

$$\sup_{|x|\le b_N} |f_N^{(1)}(x+\theta) - f^{(1)}(x)| < 1,$$

$$\sup_{|x|\le b_N} |f_N(x+\theta) - f(x)| < \varepsilon_N \quad \text{for } \varepsilon_N > 0,$$

$$\sup_{|x|\le b_N} \frac{1}{f(x)} \le H(b_N), \quad \text{and} \quad \lim_{N\to\infty} \varepsilon_N H(b_N) = 0,$$

then

$$\sup_{|x|\le b_N} \frac{\{|f_N^{(1)}(x+\theta)|\}^{\frac{1}{2}}\{N^{\frac{1}{4}}|f_N(x+\theta)-f(x)|\}}{\{|f_N(x+\theta)|\}^{\frac{1}{2}}f(x)} \le C\varepsilon_N\{H(b_N)\}^{\frac{3}{2}}N^{\frac{1}{4}}.$$

The proof can now be completed in a similar manner as in Lemma 3.4.

From this lemma we conclude that $\sqrt{N}\ L_N(\theta)$ and $\sqrt{N}\ V_{1N}$ have the same limiting distribution. We shall show that $\sqrt{N}\ V_{1N}$ can be expressed as a sum of a function of a two sample U-statistic and a random variable which converges in probability to zero. We shall now briefly outline the theory of a two sample U-statistic.

We let $Z_i = Y_i-\theta$, $i = 1,2,\ldots$ Then $X_1,X_2,\ldots$ and $Z_1,Z_2,\ldots$ are independent random variables with a common probability density function f. For each N let $g_N(x_1,x_2,\ldots,x_r,\ z_1,z_2,\ldots,z_s)$ be a Borel measurable function of r+s real variables $x_1,x_2,\ldots,x_r,\ z_1,z_2,\ldots,z_s$. Define for each N for which $m = m_N = N\lambda \ge r$, $n = n_N = N\mu \ge s$,

$$U_N(X_1,X_2,\ldots,X_m,\ Z_1,Z_2,\ldots,Z_n)$$

$$= \frac{1}{m_{(r)}n_{(s)}} \sum_P g_N(X_{i_1},X_{i_2},\ldots,X_{i_r},\ Z_{j_1},Z_{j_2},\ldots,Z_{j_s}) \qquad \ldots(3.16)$$

where $m_{(r)} = \frac{m!}{(m-r)!}$, $n_{(s)} = \frac{n!}{(n-s)!}$ and the summation $\sum_P$ is over all permutations $(i_1,i_2,\ldots,i_r)$ of r distinct integers selected from $(1,2,\ldots,m)$ and over all permutations $(j_1,j_2,\ldots,j_s)$ of s distinct integers selected from $(1,2,\ldots,n)$. Let us define for each N another Borel measurable function

$g_N^*(x_1,x_2,\ldots,x_r,\ z_1,z_2,\ldots,z_s)$ which is symmetric in the x's and symmetric in the z's by

$$g_N^*(x_1,x_2,\ldots,x_r,\ z_1,z_2,\ldots,z_s)$$
$$= \frac{1}{r!s!} \sum_P g_N(x_{i_1},x_{i_2},\ldots,x_{i_r},\ z_{j_1},z_{j_2},\ldots,z_{j_s}),$$

the summation $\sum_P$ being over all permutations $(i_1,i_2,\ldots,i_r)$ of the integers $(1,2,\ldots,r)$ and over all permutations $(j_1,j_2,\ldots,j_s)$ of the integers $(1,2,\ldots,s)$. The random variable $U_N(X_1,X_2,\ldots,X_m,\ Z_1,Z_2,\ldots,Z_n)$ defined by (3.16) can then be written as

$$U_N(X_1,X_2,\ldots,X_m,\ Z_1,Z_2,\ldots,Z_n)$$
$$= \frac{1}{\binom{m}{r}\binom{n}{s}} \sum_C g_N^*(x_{i_1},x_{i_2},\ldots,x_{i_r},\ z_{j_1},z_{j_2},\ldots,z_{j_s})$$

where $\binom{m}{r}$ and $\binom{n}{s}$ are the binomial coefficients and the summation $\sum_C$ extends over all combinations $(i_1,i_2,\ldots,i_r)$ of r distinct integers chosen from $(1,2,\ldots,m)$ and all combinations $(j_1,j_2,\ldots,j_s)$ of s distinct integers chosen from $(1,2,\ldots,n)$. For notational simplicity we shall write U_N for $U_N(X_1,X_2,\ldots,X_m,\ Z_1,Z_2,\ldots,Z_n)$. The statistic U_N defined in (3.16) is called a two sample U-statistic first studied by Lehmann [19] where the function g_N does not depend on N.

We assume that $E[g_N^*(X_1,X_2,\ldots,X_r,\ Z_1,Z_2,\ldots,Z_s)]^2$ exists for each N. This implies that $E[g_N^*(X_1,X_2,\ldots,X_r,$ $Z_1,Z_2,\ldots,Z_s)]$ also exists for each N and we let

$$\theta_N = E[g_N^*(X_1, X_2, \ldots, X_r, Z_1, Z_2, \ldots, Z_s)].$$

It then follows that

$$E[U_N] = \theta_N = E[g_N(X_1, X_2, \ldots, X_r, Z_1, Z_2, \ldots, Z_s)].$$

Let us define

$$h_N(x_1, x_2, \ldots, x_r, z_1, z_2, \ldots, z_s)$$

$$= g_N^*(x_1, x_2, \ldots, x_r, z_1, z_2, \ldots, z_s) - \theta_N$$

and for $c = 0, 1, \ldots, r$ and $d = 0, 1, \ldots, s$ let $h_{N,cd}(x_1, x_2, \ldots, x_c, z_1, z_2, \ldots, z_d)$ be the conditional expectation of $h_N(X_1, X_2, \ldots, X_r, Z_1, Z_2, \ldots, Z_s)$ given $X_1 = x_1$, $X_2 = x_2, \ldots$ $\ldots, X_c = x_c$, $Z_1 = z_1$, $Z_2 = z_2, \ldots, Z_d = z_d$. We define $\xi_{N,cd} = E[h_{N,cd}(X_1, X_2, \ldots, X_c, Z_1, Z_2, \ldots, Z_d)]^2$. In particular, $\xi_{N,rs} = V[g_N^*(X_1, X_2, \ldots, X_r, Z_1, Z_2, \ldots, Z_s)]$ where V stands for variance.

We now assume the following conditions:

Condition 3.9(a): $\xi_{10} = \lim_{N\to\infty} \xi_{N,10}$ exists.

Condition 3.9(b): $\xi_{01} = \lim_{N\to\infty} \xi_{N,01}$ exists.

Condition 3.9(c): $\frac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} h_{N,10}(X_i)$ converges in distribution to a normal random variable with mean zero and variance ξ_{10}.

Condition 3.9(d): $\frac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} h_{N,01}(Z_j)$ converges in distribution to a normal random variable with mean zero and variance ξ_{01}.

The following result on the asymptotic distribution of U_N is a direct generalization of Lehmann's theorem and its proof is omitted.

Lemma 3.10. (a) If for all (c,d) with the exception of $(1,0)$ and $(0,1)$, $\xi_{N,cd} = o(N^{c+d-1})$, then

$$V(\sqrt{N}\, U_N) = \{\frac{r^2}{\lambda}\, \xi_{N,10} + \frac{s^2}{\mu}\, \xi_{N,01} + o(1)\}\{1 + o(1)\}.$$

(b) If further, Conditions 3.9(a) to 3.9(d) hold with at least one of the numbers ξ_{10} and ξ_{01} positive, then $\sqrt{N}\,(U_N-\theta_N)$ converges in distribution to a normal random variable with mean 0 and variance $\frac{r^2}{\lambda}\, \xi_{10} + \frac{s^2}{\mu}\, \xi_{01}$.

From relation (3.14) we obtain

$$V_{1N} = \frac{N\mu-1}{N\mu}\, U_N + V_{3N} \qquad \text{...(3.17)}$$

where $U_N = U_N(X_1, X_2, ..., X_{N\lambda}, Z_1, Z_2, ..., Z_{N\mu})$ is a two-sample U-statistic defined by

$$U_N = \frac{1}{N\lambda N\mu(N\mu-1)} \sum_{i=1}^{N\lambda} \sum_{\substack{j,k=1 \\ j\neq k}}^{N\mu} g_N(X_i, Z_j, Z_k) \qquad \text{...(3.18)}$$

and

$$g_N(x_1, z_1, z_2) = \frac{1}{a_N}\, \phi(\frac{-b_N-z_1}{a_N})$$
$$+ I_N(x_1)\, \frac{\frac{1}{a_N^2}\, \phi^{(1)}(\frac{x_1-z_1}{a_N})}{f(x_1)} \left\{2 - \frac{\frac{1}{a_N}\, \phi(\frac{x_1-z_2}{a_N})}{f(x_1)}\right\}$$
$$- \frac{1}{a_N}\, \phi(\frac{b_N-z_1}{a_N}) \qquad \text{...(3.19)}$$

and where

$$V_{3N} = \frac{1}{N\mu} \{f_N(-b_N+\theta) - f_N(b_N+\theta)\}$$

$$+ \frac{1}{N\lambda(N\mu)^2} \sum_{i=1}^{N\lambda} \sum_{j=1}^{N\mu} I_N(X_i) \frac{\frac{1}{a_N^2} \phi^{(1)}(\frac{X_i-Z_j}{a_N})}{f(X_i)}$$

$$\cdot\left\{2 - \frac{\frac{1}{a_N} \phi(\frac{X_i-Z_j}{a_N})}{f(X_i)}\right\}. \qquad \ldots(3.20)$$

We now prove the following lemma on the convergence to zero of the random variable V_{3N}.

Lemma 3.11. Under Conditions A.3

$$\sqrt{N}\; V_{3N} = o_p(1).$$

Proof. By Conditions 3.1 and 3.7, we have

$$\frac{1}{\sqrt{N}} E|f_N(z)| \le \frac{1}{\sqrt{N}} f(z-\theta) \int_{-\infty}^{\infty} |\phi(u)|du + C \frac{a_N}{\sqrt{N}}$$

where C is a positive constant. Hence

$$\frac{1}{\sqrt{N}} E|f_N(z)| = o(1). \qquad \ldots(3.21)$$

Similarly, using Conditions 3.5 and 3.6, we get

$$E\left[\left| I_N(X_i) \frac{\frac{1}{a_N^2} \phi^{(1)}(\frac{X_i-Z_j}{a_N})}{f(X_i)} \left\{2 - \frac{\frac{1}{a_N} \phi(\frac{X_i-Z_j}{a_N})}{f(X_i)}\right\} \right|\right]$$

$$\le E\left[\left| I_N(X_i) \frac{\frac{1}{a_N^2} \phi^{(1)}(\frac{X_i-Z_j}{a_N})}{f(X_i)} \right| \left\{ 2 + \left| \frac{\frac{1}{a_N} \phi(\frac{X_i-Z_j}{a_N})}{f(X_i)} \right| \right\}\right]$$

$$\le [2 + \frac{C' H(b_N)}{a_N}]\; E\left[\left| I_N'(X_i) \frac{\frac{1}{a_N^2} \phi^{(1)}(\frac{X_i-Z_j}{a_N})}{f(X_i)} \right|\right]$$

where C' is a positive generic constant. Hence,

$$E\left[\left|\frac{1}{N^{5/2}}\sum_{i=1}^{N\lambda}\sum_{j=1}^{N\mu} I_N(X_i)\frac{\frac{1}{a_N^2}\phi^{(1)}\left(\frac{X_i-Z_j}{a_N}\right)}{f(X_i)}\left\{2-\frac{\frac{1}{a_N}\phi\left(\frac{X_i-Z_j}{a_N}\right)}{f(X_i)}\right\}\right|\right]$$

$$\leq C'\left[\frac{N^{-\frac{1}{2}}\{H(b_N)\}}{a_N^2}\right]$$

$$= O\left(N^{-\frac{31}{150}}\right), \quad \text{by Condition 3.8}$$

$$= o(1). \qquad \ldots(3.22)$$

The proof of the lemma now follows from the results in (3.21) and (3.22).

It is now evident that $\sqrt{N}\, V_{1N}$ and $\frac{N\mu-1}{N\mu}\{\sqrt{N}\, U_N\}$ have the same limiting distribution.

In the next six lemmas we shall study the asymptotic distribution of the U-statistic defined by (3.18) and (3.19).

<u>Lemma 3.12</u>. Under Conditions 3.1, 3.5, 3.6, and 3.7

$$\theta_N = O[a_N^4 + a_N^5\{H(b_N)\}].$$

<u>Proof</u>. By Fubini's theorem and Lemma 3.1, we have

$$\theta_N = E[g_N(X_1, Z_1, Z_2)]$$

$$= f(-b_N) + O(a_N^5)$$

$$+ \int_{-b_N}^{b_N}\left\{\int_{-\infty}^{\infty} f^{(1)}(x-a_N u)\phi(u)\,du\right\}\left\{2-\frac{f(x)+O(a_N^5)}{f(x)}\right\}dx$$

$$- f(b_N) + O(a_N^5)$$

$$= f(-b_N) + O(a_N^5)$$

$$+ \int_{-b_N}^{b_N} \{ \int_{-\infty}^{\infty} f^{(1)}(x-a_N u)\phi(u)du\}\{\frac{0(a_N^5)}{f(x)}\}dx$$

$$+ \int_{-b_N}^{b_N} \{f^{(1)}(x) + a_N^4 K_N(x)\}dx - f(b_N) + 0(a_N^5)$$

where

$$K_N(x) = \frac{1}{4!} \int_{-\infty}^{\infty} u^4 f^{(5)}(x-\alpha a_N u)\phi(u)du \quad \text{and} \quad 0 < \alpha < 1.$$

We have

$$\left| \int_{-b_N}^{b_N} \{ \int_{-\infty}^{\infty} f^{(1)}(x-a_N u)\phi(u)du\}\{\frac{0(a_N^5)}{f(x)}\}dx \right|$$

$$\leq \{ \int_{-\infty}^{\infty} |f^{(1)}(x)|dx\} 0[a_N^5\{H(b_N)\}]$$

$$\leq C[a_N^5\{H(b_N)\}],$$

since by the Cauchy-Schwarz inequality,

$$\int_{-\infty}^{\infty} |f^{(1)}(x)|dx \leq \{J(f)\}^{\frac{1}{2}} < \infty.$$

Similarly,

$$\left| \int_{-b_N}^{b_N} K_N(x)dx \right| = \frac{1}{4!} \left| \int_{-b_N}^{b_N} \int_{-\infty}^{\infty} u^4 f^{(5)}(x-\alpha a_N u)\phi(u)dudx \right| < \infty.$$

The proof of the lemma now follows from the above computations.

<u>Corollary</u>. If further, Condition 3.8 is satisfied, then

$$\sqrt{N}\ \theta_N = o(1).$$

For each N let the function $t_N(z)$ be defined by

$$t_N(z) = \frac{1}{a_N} \int_{-b_N}^{b_N} \frac{f^{(1)}(u)}{f(u)} \phi(\frac{u-z}{a_N})du.$$

Lemma 3.13. Under Conditions 3.1, 3.5, and 3.7 there exists a constant C independent of N such that

$$\left| t_N(z) - \frac{f^{(1)}(z)}{f(z)} \int_{\frac{-b_N-z}{a_N}}^{\frac{b_N-z}{a_N}} \phi(t)\,dt \right| \leq C a_N \{H(b_N)\}^2$$

for all $-b_N < z < b_N$.

Proof. We have

$$t_N(z) = \frac{1}{a_N} \int_{-b_N}^{b_N} \frac{f^{(1)}(u)}{f(u)} \phi\left(\frac{u-z}{a_N}\right) du.$$

Put $t = \frac{u-z}{a_N}$. Then

$$t_N(z) = \int_{\frac{-b_N-z}{a_N}}^{\frac{b_N-z}{a_N}} \frac{f^{(1)}(z+a_N t)}{f(z+a_N t)} \phi(t)\,dt.$$

Now, expanding $\frac{f^{(1)}(z+a_N t)}{f(z+a_N t)}$ around z we get

$$t_N(z) = \frac{f^{(1)}(z)}{f(z)} \int_{\frac{-b_N-z}{a_N}}^{\frac{b_N-z}{a_N}} \phi(t)\,dt$$

$$+ a_N \int_{\frac{-b_N-z}{a_N}}^{\frac{b_N-z}{a_N}} \left[\frac{f^{(2)}(z+\beta a_N t)}{f(z+\beta a_N t)} - \left\{ \frac{f^{(1)}(z+\beta a_N t)}{f(z+\beta a_N t)} \right\}^2 \right] t\phi(t)\,dt$$

where $0 < \beta < 1$. Since

$$\frac{-b_N - z}{a_N} \le t \le \frac{b_N - z}{a_N} ,$$

we conclude that for $0 < \beta < 1$, $-b_N < z + \beta a_N t < b_N$. The proof of the lemma now follows from the hypothesis and the above computation.

Corollary. Under Conditions 3.1, 3.5, 3.6, 3.7, and 3.8,

$$\lim_{N\to\infty} t_N(z) = \frac{f^{(1)}(z)}{f(z)} \quad \text{for all } -\infty < z < \infty.$$

The proof follows immediately since, for any z, N can be chosen sufficiently large to satisfy the condition $-b_N < z < b_N$, and the limits of integration in Lemma 3.13, $\frac{b_N - z}{a_N}$ and $\frac{-b_N - z}{a_N}$ approach $+\infty$ and $-\infty$ respectively as N tends to infinity.

Lemma 3.14. Under Conditions 3.1, 3.5, 3.6, and 3.7

(a) $$h_{N,10}(x) = I_N(x) \frac{f^{(1)}(x)}{f(x)} + f(-b_N) - f(b_N) + R_N(x)$$

where $|R_N(x)| \le C[a_N^5\{H(b_N)\}^2 + a_N^4\{H(b_N)\}]$ and where C is independent of x.

(b) $$h_{N,01}(z) = -\frac{1}{2} t_N(z) + \frac{1}{2}\{f(b_N) - f(-b_N)\} + R_N'(z)$$

where $|R_N'(z)| \le C'[a_N^3\{H(b_N)\}]$ and C' is independent of z.

Proof. By Fubini's theorem and Lemmas 3.1 and 3.12 we have

$$\begin{aligned} h_{N,10}(x) &= E[g_N^*(x, Z_1, Z_2)] - \theta_N \\ &= E[\tfrac{1}{2}\{g_N(x, Z_1, Z_2) + g_N(x, Z_2, Z_1)\}] - \theta_N \\ &= f(-b_N) + O(a_N^5) \end{aligned}$$

$$+ \ I_N(x) \ \frac{f^{(1)}(x)+O(a_N^4)}{f(x)} \ \{2 - \frac{f(x)+O(a_N^5)}{f(x)}\}$$

$$- \ f(b_N) + O(a_N^5) + O[a_N^4 + a_N^5\{H(b_N)\}]$$

$$= I_N(x) \ \frac{f^{(1)}(x)}{f(x)} + O[a_N^5\{H(b_N)\}^2 + a_N^4\{H(b_N)\}]$$

$$+ \ f(-b_N) - f(b_N) + O(a_N^5) + O[a_N^4 + a_N^5\{H(b_N)\}]$$

$$= I_N(x) \ \frac{f^{(1)}(x)}{f(x)} + f(-b_N) - f(b_N)$$

$$+ \ O[a_N^5\{H(b_N)\}^2 + a_N^4\{H(b_N)\}].$$

The proof of part (a) of the lemma is now complete.

(b) By Fubini's theorem and Lemma 3.1 we get

$E[g_N(X_1,z,Z_2)]$

$$= \frac{1}{a_N} \phi(\frac{-b_N-z}{a_N}) + \int_{-b_N}^{b_N} \frac{1}{a_N^2} \phi^{(1)}(\frac{x-z}{a_N})\{2 - \frac{f(x)+O(a_N^5)}{f(x)}\}dx$$

$$- \ \frac{1}{a_N} \phi(\frac{b_N-z}{a_N})$$

$$= \frac{1}{a_N} \phi(\frac{-b_N-z}{a_N}) + \int_{-b_N}^{b_N} \frac{1}{a_N^2} \phi^{(1)}(\frac{x-z}{a_N}) dx$$

$$+ \ O[a_N^4\{H(b_N)\}] - \frac{1}{a_N} \phi(\frac{b_N-z}{a_N})$$

$$= O[a_N^4\{H(b_N)\}]. \qquad \ldots(3.23)$$

Similarly,

$E[g_N(X_1,Z_2,z)]$

$$= f(-b_N) + O(a_N^5)$$

$$+ \int_{-b_N}^{b_N} \{f^{(1)}(x) + a_N^4 K_N(x)\} \left\{ 2 - \frac{\frac{1}{a_N}\phi(\frac{x-z}{a_N})}{f(x)} \right\} dx$$

$$- f(b_N) + O(a_N^5),$$

where

$$K_N(x) = \frac{1}{4!} \int_{-\infty}^{\infty} u^4 f^{(5)}(x - \alpha a_N u)\phi(u)du \quad \text{and} \quad 0 < \alpha < 1.$$

Hence,

$$E[g_N(X_1, Z_2, z)]$$

$$= f(b_N) - f(-b_N) + O(a_N^5) - \int_{-b_N}^{b_N} \frac{f^{(1)}(x)}{f(x)} \frac{1}{a_N} \phi(\frac{x-z}{a_N})dx$$

$$+ O(a_N^4) + O[a_N^3\{H(b_N)\}]$$

$$= -t_N(z) + f(b_N) - f(-b_N) + O[a_N^4 + a_N^3\{H(b_N)\}]$$

$$= -t_N(z) + f(b_N) - f(-b_N) + O[a_N^3\{H(b_N)\}]. \quad \ldots(3.24)$$

Hence, using relation (3.23), (3.24), and Lemma 3.12, we get

$$h_{N,01}(z) = \frac{1}{2} E[g_N(X_1, z, Z_2) + g_N(X_1, Z_2, z)] - \theta_N$$

$$= -\frac{1}{2} t_N(z) + \frac{1}{2} f(b_N) - \frac{1}{2} f(-b_N)$$

$$+ O[a_N^3\{H(b_N)\}]. \quad \ldots(3.25)$$

The proof of part (b) now follows from (3.25).

Lemma 3.15. Under Conditions A.3

(a) $\lim_{N\to\infty} \xi_{N,10} = J(f)$

(b) $\lim_{N\to\infty} \xi_{N,01} = \frac{1}{4} J(f)$

(c) $\lim_{N\to\infty} E\{h_{N,10}(X_1) \frac{f^{(1)}(X_1)}{f(X_1)}\} = J(f)$

(d) $\lim_{N\to\infty} E\{h_{N,01}(Z_1)\ \dfrac{f^{(1)}(Z_1)}{f(Z_1)}\} = -\frac{1}{2}\ J(f)$

(e) $\dfrac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} h_{N,10}(X_i)$ converges in distribution to a normal random variable with mean 0 and variance $J(f)$.

(f) $\dfrac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} h_{N,01}(Z_j)$ converges in distribution to a normal random variable with mean 0 and variance $\frac{1}{4}\ J(f)$.

<u>Proof</u>. (a) By part (a) of Lemma 3.14 we have

$$\lim_{N\to\infty} |h_{N,10}(x)|^k = \left|\frac{f^{(1)}(x)}{f(x)}\right|^k \quad \text{for } k = 1,2.$$

We then note that for all sufficiently large N, $|h_{N,10}(x)|^k$ is bounded above by $\{\left|\frac{f^{(1)}(x)}{f(x)}\right| + 1\}^k$ and $E\{\left|\frac{f^{(1)}(X_1)}{f(X_1)}\right| + 1\}^k$ is finite for $k = 1,2$.

An application of the Lebesgue dominated convergence theorem completes the proof.

(b) By the corollary of Lemma 3.13 and part (b) of Lemma 3.14 we have

$$\lim_{N\to\infty} |h_{N,01}(z)|^k = (\tfrac{1}{2})^k \left|\frac{f^{(1)}(z)}{f(z)}\right|^k \quad \text{for } k = 1,2.$$

We then note that for all sufficiently large N, $|h_{N,01}(z)|^k$ is bounded above by

$$\{(\tfrac{1}{2})\cdot \left|\frac{f^{(1)}(z)}{f(z)}\right| + 1\}^k \quad \text{for } k = 1,2.$$

The proof is complete by the Lebesgue dominated convergence theorem.

(c) We note that for all sufficiently large N,

$$|h_{N,10}(x)\ \frac{f^{(1)}(x)}{f(x)}| \le \{|\frac{f^{(1)}(x)}{f(x)}| + 1\}\{|\frac{f^{(1)}(x)}{f(x)}|\}$$

and

$$\lim_{N\to\infty} [h_{N,10}(x)\ \frac{f^{(1)}(x)}{f(x)}] = \{\frac{f^{(1)}(x)}{f(x)}\}^2.$$

An application of the Lebesgue dominated convergence theorem completes the proof.

(d) For all sufficiently large N,

$$|h_{N,01}(z)\ \frac{f^{(1)}(z)}{f(z)}| \le \{\frac{1}{2}\ |\frac{f^{(1)}(z)}{f(z)}| + 1\}|\frac{f^{(1)}(z)}{f(z)}|,$$

and

$$\lim_{N\to\infty} [h_{N,01}(z)\ \frac{f^{(1)}(z)}{f(z)}] = -\ \frac{1}{2}\ \{\frac{f^{(1)}(z)}{f(z)}\}^2.$$

The proof is now complete by the Lebesgue dominated convergence theorem.

(e) We have

$$E\left[\frac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} h_{N,10}(X_i) - \frac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} \frac{f^{(1)}(X_i)}{f(X_i)}\right]^2$$

$$= E\left[\frac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} h_{N,10}(X_i)\right]^2 + E\left[\frac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} \frac{f^{(1)}(X_i)}{f(X_i)}\right]^2$$

$$- \frac{1}{N\lambda} \sum_{i=1}^{N\lambda} 2E\left[h_{N,10}(X_i)\ \frac{f^{(1)}(X_i)}{f(X_i)}\right]$$

$$= \xi_{N,10} + J(f) - 2E\left[h_{N,10}(X_1)\ \frac{f^{(1)}(X_1)}{f(X_1)}\right].$$

Hence,

$$\lim_{N\to\infty} E\left[\frac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} h_{N,10}(X_i) - \frac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} \frac{f^{(1)}(X_i)}{f(X_i)}\right]^2$$

$$= J(f) + \lim_{N\to\infty} \xi_{N,10} - 2\left[\lim_{N\to\infty} E\{h_{N,10}(X_1) \frac{f^{(1)}(X_1)}{f(X_1)}\}\right]$$

$$= J(f) + J(f) - 2J(f), \text{ by part (a) and part (c)}$$

$$= 0.$$

We conclude that

$$\frac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} h_{N,10}(X_i) = \frac{1}{\sqrt{N\lambda}} \sum_{i=1}^{N\lambda} \frac{f^{(1)}(X_i)}{f(X_i)} + o_p(1).$$

The proof of part (e) is complete by an application of the central limit theorem.

(f) We have

$$E\left[\frac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} h_{N,01}(Z_j) + \frac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} \frac{1}{2} \frac{f^{(1)}(Z_j)}{f(Z_j)}\right]^2$$

$$= E\left[\frac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} h_{N,01}(Z_j)\right]^2 + E\left[\frac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} \frac{1}{2} \frac{f^{(1)}(Z_j)}{f(Z_j)}\right]^2$$

$$+ \frac{1}{N\mu} \sum_{j=1}^{N\mu} E\left[h_{N,01}(Z_j) \frac{f^{(1)}(Z_j)}{f(Z_j)}\right]$$

$$= \xi_{N,01} + \frac{1}{4} J(f) + E\left[h_{N,01}(Z_1) \frac{f^{(1)}(Z_1)}{f(Z_1)}\right]$$

Hence,

$$\lim_{N\to\infty} E\left[\frac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} h_{N,01}(Z_j) + \frac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} \frac{1}{2} \frac{f^{(1)}(Z_j)}{f(Z_j)}\right]^2$$

$$= \frac{1}{4} J(f) + \frac{1}{4} J(f) - \frac{1}{2} J(f), \text{ by part (b) and part (d)}$$

$$= 0.$$

We conclude that

$$\frac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} h_{N,01}(Z_j) = -\frac{1}{\sqrt{N\mu}} \sum_{j=1}^{N\mu} \frac{1}{2} \frac{f^{(1)}(Z_j)}{f(Z_j)} + o_p(1).$$

An application of the central limit theorem completes the proof.

Lemma 3.16. Under Conditions A.3

(a) $\xi_{N,11} = O\left(N^{\frac{44}{75}}\right)$

(b) $\xi_{N,02} = O\left(N^{\frac{44}{75}}\right)$

(c) $\xi_{N,12} = O\left(N^{\frac{23}{25}}\right)$.

Proof. (a) We have $\xi_{N,11} = E[h_{N,11}(X_1,Z_1)]^2$, where

$$\begin{aligned} h_{N,11}(x_1,z_1) &= E[g_N^*(x_1,z_1,Z_2)] - \theta_N \\ &= \frac{1}{2} E[g_N(x_1,z_1,Z_2) + g_N(x_1,Z_2,z_1)] - \theta_N. \end{aligned}$$

Now

$$\begin{aligned} &E[g_N(x_1,z_1,Z_2)] \\ &= \frac{1}{a_N} \phi\left(\frac{-b_N-z_1}{a_N}\right) \\ &\quad + I_N(x_1) \frac{\frac{1}{a_N^2} \phi^{(1)}\left(\frac{x_1-z_1}{a_N}\right)}{f(x_1)} \left\{ 2 - \frac{f(x_1)+O(a_N^5)}{f(x_1)} \right\} \\ &\quad - \frac{1}{a_N} \phi\left(\frac{b_N-z_1}{a_N}\right) \\ &= O\left[\frac{1}{a_N} + \frac{H(b_N)}{a_N^2} + a_N^3\{H(b_N)\}^2\right] \end{aligned}$$

$$= 0\left[\frac{H(b_N)}{a_N^2} + a_N^3\{H(b_N)\}^2\right]$$

$$= 0\left(N^{\frac{22}{75}}\right).$$

Similarly,

$$E[g_N(x_1,Z_2,z_1)]$$

$$= f(-b_N) + 0(a_N^5)$$

$$+ I_N(x_1)\ \frac{f^{(1)}(x_1)+0(a_N^4)}{f(x_1)}\left\{2 - \frac{\frac{1}{a_N}\ \phi(\frac{x_1-z_1}{a_N})}{f(x_1)}\right\}$$

$$- f(b_N) + 0(a_N^5)$$

$$= f(-b_N) - f(b_N) + 0(a_N^5)$$

$$+ 0[H(b_N) + \frac{\{H(b_N)\}^2}{a_N} + a_N^4\{H(b_N)\} + a_N^3\{H(b_N)\}^2]$$

$$= 0\left[\frac{\{H(b_N)\}^2}{a_N}\right]$$

$$= 0\left(N^{\frac{31}{150}}\right).$$

Hence,

$$\xi_{N,11} = 0\left(N^{\frac{44}{75}}\right).$$

(b) We have $\xi_{N,02} = E[h_{N,02}(Z_1,Z_2)]^2$, where

$$h_{N,02}(z_1,z_2) = E[g_N^*(X_1,z_1,z_2)] - \theta_N$$

$$= \frac{1}{2}\ E[g_N(X_1,z_1,z_2) + g_N(X_1,z_2,z_1)] - \theta_N.$$

Now,

$$E[g_N(X_1,z_1,z_2)]$$

$$= \frac{1}{a_N}\,\phi\left(\frac{-b_N-z_1}{a_N}\right) + \int_{-b_N}^{b_N} \frac{1}{a_N^2}\,\phi^{(1)}\left(\frac{x-z_1}{a_N}\right)\left\{2 - \frac{\frac{1}{a_N}\,\phi\left(\frac{x-z_2}{a_N}\right)}{f(x)}\right\} dx$$

$$- \frac{1}{a_N}\,\phi\left(\frac{b_N-z_1}{a_N}\right)$$

$$= 0\left[\frac{1}{a_N} + \frac{H(b_N)}{a_N^2}\right]$$

$$= 0\left[\frac{H(b_N)}{a_N^2}\right]$$

$$= 0\left(N^{\frac{22}{75}}\right).$$

Similarly,

$$E[g_N(X_1,z_2,z_1)] = 0\left(N^{\frac{22}{75}}\right).$$

Hence,

$$\xi_{N,02} = 0\left(N^{\frac{44}{75}}\right).$$

(c) We have

$$\xi_{N,12} = E[h_{N,12}(X_1,Z_1,Z_2)]^2, \text{ where}$$

$$h_{N,12}(x_1,z_1,z_2) = \frac{1}{2}\,[g_N(x_1,z_1,z_2) + g_N(x_1,z_2,z_1)] - \theta_N.$$

Now,

$$g_N(x_1,z_1,z_2) = \frac{1}{a_N}\,\phi\left(\frac{-b_N-z_1}{a_N}\right)$$

$$+ I_N(x_1)\,\frac{\frac{1}{a_N^2}\,\phi\left(\frac{x_1-z_1}{a_N}\right)}{f(x_1)}\left\{2 - \frac{\frac{1}{a_N}\,\phi\left(\frac{x_1-z_2}{a_N}\right)}{f(x_1)}\right\}$$

$$- \frac{1}{a_N} \phi\left(\frac{b_N - z_1}{a_N}\right)$$

$$= 0\left[\frac{1}{a_N} + \frac{H(b_N)}{a_N^2} + \frac{\{H(b_N)\}^2}{a_N^3}\right]$$

$$= 0\left[\frac{\{H(b_N)\}^2}{a_N^3}\right]$$

$$= 0\left(N^{\frac{23}{50}}\right).$$

Similarly,

$$g_N(x_1, z_2, z_1) = 0\left(N^{\frac{23}{50}}\right).$$

Hence,

$$\xi_{N,12} = 0\left(N^{\frac{23}{25}}\right).$$

Lemma 3.17. Under Conditions A.3, $\sqrt{N}\, U_N$ converges in distribution to a normal random variable with mean 0 and variance $\frac{J(f)}{\lambda\mu}$.

Proof. By Lemmas 3.15 and 3.16 the hypotheses of Lemma 3.10 are satisfied. Hence, $\sqrt{N}\,(U_N - \theta_N)$ converges in distribution to a normal random variable with mean zero and variance $\frac{\xi_{10}}{\lambda} + \frac{4\xi_{01}}{\mu}$. By the corollary of Lemma 3.12 and part (a) and part (b) of Lemma 3.15 the proof now follows.

Lemma 3.18. Under Conditions A.3, $\sqrt{N}\, L_N(\theta)$ converges in distribution to a normal random variable with mean 0 and variance $\frac{J(f)}{\lambda\mu}$.

Proof. Using relation (3.13) we have

$$\sqrt{N}\ L_N(\theta) = \sqrt{N}\ V_{1N} + \sqrt{N}\ V_{2N}$$

$$= \sqrt{N}\ V_{1N} + o_p(1), \text{ by Lemma 3.9}$$

$$= \sqrt{N}\ \left(\frac{N\mu-1}{N\mu} U_N + V_{3N}\right) + o_p(1), \text{ by relation (3.17)}$$

$$= \sqrt{N}\ U_N + \sqrt{N}\ U_N \cdot o(1) + \sqrt{N}\ V_{3N} + o_p(1)$$

$$= \sqrt{N}\ U_N + O_p(1)o(1) + o_p(1) + o_p(1), \text{ by Lemmas 3.11 and 3.17}$$

$$= \sqrt{N}\ U_N + o_p(1).$$

Another application of Lemma 3.17 now completes the proof.

Theorem 3.2. Under Conditions A.3 with probability approaching one as N tends to infinity the likelihood equation $L_N(t) = 0$ has a solution T_N which is consistent for θ. Furthermore, the solution is asymptotically normally distributed with mean θ and variance $\frac{1}{N\lambda\mu J(f)}$.

Proof. The first part of the theorem is a restatement of Theorem 3.1. For the second part we note that in view of relation (3.12) it suffices to obtain the asymptotic distribution of $\frac{\sqrt{N}\ L_N(\theta)}{-L_N^{(1)}(T_N^*)}$. We have

$$\frac{\sqrt{N}\ L_N(\theta)}{-L_N^{(1)}(T_N^*)} = \frac{\sqrt{N}\ L_N(\theta)}{J(f)+o_p(1)}, \text{ by Lemma 3.8}$$

$$= \frac{\sqrt{N}\ L_N(\theta)}{J(f)} \{1 + o_p(1)\}$$

$$= \frac{\sqrt{N}\ L_N(\theta)}{J(f)} + O_p(1)o_p(1), \text{ by Lemma 3.18}$$

$$= \frac{\sqrt{N}\ L_N(\theta)}{J(f)} + o_p(1). \qquad \ldots(3.26)$$

Another application of Lemma 3.18 now completes the proof.

3.5 A Computational Method for an Asymptotically Efficient Estimate of the Shift Parameter

The estimate T_N of the shift parameter studied in this chapter is computationally formidable. We now outline a simple procedure leading to an estimate having the asymptotic properties possessed by a consistent sequence of solutions of the empirical likelihood equation.

Let T_N^* be any estimate of the shift θ such that $T_N^* - \theta = O_p(\frac{1}{\sqrt{N}})$. For example, we can take for T_N^* the median of the first sample subtracted from the median of the second sample. With T_N^* as a first approximation for a solution of the equation $L_N(t) = 0$ we get the next approximation $\hat{T}_N$ going through exactly one iteration by the Newton-Raphson method, i.e.,

$$\hat{T}_N = T_N^* + \frac{L_N(T_N^*)}{-L_N^{(1)}(T_N^*)} \qquad \ldots(3.27)$$

For the estimate $\hat{T}_N$ we have the following theorem.

<u>Theorem 3.3</u>. Under Conditions A.3, $\hat{T}_N$ is asymptotically normally distributed with mean θ and variance $\frac{1}{N\mu\lambda J(f)}$.

<u>Proof</u>. We first note that $T_N^* - \theta = o_p(1)$. Hence, by Lemma 3.8, we have

$$-L_N^{(1)}(T_N^*) = J(f) + o_p(1). \qquad \ldots(3.28)$$

From (3.27) and (3.28) we get

$$\sqrt{N}\ (\hat{T}_N-\theta) = \sqrt{N}\ (T_N^*-\theta) + \frac{\sqrt{N}\ L_N(T_N^*)}{-L_N^{(1)}(T_N^*)}$$

$$= \sqrt{N}\ (T_N^*-\theta) + \frac{\sqrt{N}\ L_N(\theta)+\sqrt{N}\ (T_N^*-\theta)L_N^{(1)}(T_N')}{J(f) + o_p(1)}$$

where $|T_N'-\theta| < |T_N^*-\theta|$. This implies that $T_N' = \theta + o_p(1)$. Hence, by another application of Lemma 3.8 we have

$$\sqrt{N}\ (\hat{T}_N-\theta) = \frac{\sqrt{N}\ L_N(\theta)}{J(f)+o_p(1)} + \sqrt{N}\ (T_N^*-\theta)\{1 - \frac{J(f)+o_p(1)}{J(f)+o_p(1)}\}$$

$$= \frac{\sqrt{N}\ L_N(\theta)}{J(f)}\ \{1 + o_p(1)\}$$

$$+ O_p(1)[1 - \{1+o_p(1)\}\{1+o_p(1)\}]$$

$$= \frac{\sqrt{N}\ L_N(\theta)}{J(f)} + O_p(1)o_p(1) + O_p(1)o_p(1),$$

by Lemma 3.18

$$= \frac{\sqrt{N}\ L_N(\theta)}{J(f)} + o_p(1).$$

Another application of Lemma 3.18 now completes the proof.

3.6 Miscellaneous Remarks

(i) Choice of the sequence $\{b_N\}$.

The choice of the sequence $\{b_N\}$ requires a knowledge of the tails of the density function. If the density function has flatter tails than the standard normal probability density function, then b_N may be chosen to be equal to $\sqrt{\frac{2}{25} \ln N}$ for all sufficiently large N. The density functions which possess flatter tails than the standard normal density function include among others the Cauchy density function, the mixture of the standard normal and the Cauchy density functions, and the density function of the random variable t with degrees of freedom $1 < \nu < \infty$.

(ii) Estimation of J(f).

Following Bhattacharya [5] we propose the following estimate of J(f) given by

$$\hat{J}(f) = \int_{-b_N}^{b_N} \frac{\{f_N^{(1)}(x)\}^2}{f_N(x)} dx.$$

The consistency of $\hat{J}(f)$ can be proved in a similar manner as in Bhattacharya [5].

(iii) Large sample tests of hypothesis about θ.

Suppose we want to test the null hypothesis $H_o: \theta = \theta_o$ against alternatives on both sides. We have shown that under H_o, $\sqrt{N}\ (T_N-\theta_o)$ is asymptotically normally distributed with mean 0 and variance $\frac{1}{\lambda\mu J(f)}$. Using a consistent estimate $\hat{J}(f)$ of J(f) (see Miscellaneous Remarks (ii)), we now define the statistic

$$\tau_N = \{N\lambda\mu\hat{J}(f)\}^{\frac{1}{2}}(T_N - \theta_o)$$

which is asymptotically standard normal under H_o. τ_N can be used as a test statistic for testing H_o and the critical region $|\tau_N| \geq R^{-1}(1 - \frac{\alpha}{2})$ will have an approximate level of significance α for large N, where R is the standard normal probability distribution function. In an analogous manner we can construct a large sample confidence interval for the unknown parameter for a prescribed confidence coefficient. The test statistic τ_N can also be used for testing the hypothesis H_o: $\theta \leq \theta_o$ (or $\theta \geq \theta_o$) against H_1: $\theta > \theta_o$ (or $\theta < \theta_o$). It can be shown that the asymptotic efficiency (ARE) of these tests relative to the tests based on the maximum likelihood estimate is 1.

(iv) Estimation of the location parameter in the one sample problem.

Let $W_1, W_2, \ldots, W_N$ be independent and identically distributed random variables with probability density function $h(\cdot - \nu)$ where ν is a location parameter and $h(\cdot)$ is symmetric about zero, i.e., $h(w) = h(-w)$ for all w. If h were known, then ν could be estimated from $W_1, W_2, \ldots, W_N$ using the method of maximum likelihood by solving the following likelihood equation:

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial\nu}\{\log h(W_i - \nu)\} = 0.$$

It is well known (see Cramer [11], p.500) that under certain regularity conditions on h, with probability approaching one as N tends to infinity, the above likelihood equation has a solution ν_N^* which is consistent for ν. Furthermore, $\sqrt{N}\,(\nu_N^*-\nu)$ converges in distribution to a normal random variable with mean zero and variance $\frac{1}{J(h)}$ where

$$J(h) = \int_{-\infty}^{\infty} \frac{\{h^{(1)}(w)\}^2}{h(w)}\, dw.$$

The nonparametric counterpart of this problem arises when h is unknown. We define $m_N = [\frac{N}{2}]$ and $n_N = N - [\frac{N}{2}]$ where $[\frac{N}{2}]$ is the largest integer contained in $\frac{N}{2}$. Since h is symmetric it follows that the probability density function of $-W_1$ is $h(\cdot+\nu)$. Then $X_i = -W_i$, $i = 1,2,\ldots,m_N$ and $Y_j = W_j$, $j = m_N+1, m_N+2,\ldots,N$ are independent random variables. The common probability density function of the random variables X_i, $i = 1,2,\ldots,m_N$, is $h(\cdot+\nu)$ and the common probability density function of the random variables Y_j, $j = 1,2,\ldots,n_N$ is $h(\cdot-\nu)$. The shift between these two density functions is $\theta = 2\nu$. Using the method developed in this chapter, we define the sample empirical likelihood equation as given in (3.5) and conclude (see Theorems 3.1 and 3.2) that under Conditions A.3, with probability approaching one as N tends to infinity, this equation has a solution θ_N^{**} which is consistent for θ. Furthermore,

$\sqrt{N}\ (\theta_N^{**}-\theta)$ converges in distribution to a normal random variable with mean zero and variance $\frac{4}{J(h)}$. From this result, it follows that the estimate $\frac{\theta_N^{**}}{2}$ converges in probability to the location parameter ν and $\sqrt{N}\ (\frac{\theta_N^{**}}{2} - \nu)$ converges in distribution to a normal random variable with mean 0 and variance $\frac{1}{J(h)}$. We also note that using the Newton-Raphson technique one can easily compute an estimate θ_N^{***} which, according to Theorem 3.3, has the same asymptotic properties possessed by θ_N^{**}.

CHAPTER IV

NONPARAMETRIC ESTIMATION OF A MULTIVARIATE MULTIPLE REGRESSION FUNCTION

4.1 Introduction and Summary

Let $\binom{\underline{X}}{\underline{Y}}$ be a vector random variable where $\underline{X} = (X_1, X_2, \ldots, X_{p_1})'$ and $\underline{Y} = (Y_1, Y_2, \ldots, Y_{p_2})'$. In this situation the expectation of $\underline{Y}$ given $\underline{X} = \underline{x}$ defines the regression of $\underline{Y}$ on $\underline{X}$. We note that this regression is a vector function having p_2 components. In this chapter we consider the problem of estimating this regression function on the basis of a random sample of size n. For notational simplicity we shall consider the situation when $p_1 = p_2 = 2$. The asymptotic properties of our estimate can be proved under similar conditions for arbitrary p_1 and p_2. Let $(X_{11}, X_{21}, Y_{11}, Y_{21})$, $(X_{12}, X_{22}, Y_{12}, Y_{22}), \ldots, (X_{1n}, X_{2n}, Y_{1n}, Y_{2n})$ be independent vector random variables identically distributed as the vector random variable (X_1, X_2, Y_1, Y_2) having the joint distribution function $F(x_1, x_2, y_1, y_2)$ and the joint probability density function $f(x_1, x_2, y_1, y_2)$. We also denote the joint distribution function and the joint probability density function of the vector random variable (X_1, X_2) by $G(x_1, x_2)$ and $g(x_1, x_2)$ respectively, and the regression of Y_i on (X_1, X_2) by $m_i(x_1, x_2)$, $i = 1,2$. Then we have

$$m_i(x_1,x_2) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_i f(x_1,x_2,y_1,y_2)\,dy_1 dy_2}{g(x_1,x_2)}, \quad i = 1,2. \qquad \ldots(4.1)$$

Let $\phi_i(y)$, $i = 1,2$ be two univariate probability density functions and let $\{a_n\}$ be a sequence of positive numbers converging to zero as n tends to infinity. Following Watson [55] and Nadaraya [25, 27] we propose to estimate the population regression function $m_i(x_1,x_2)$ by the statistic $m_{in}(x_1,x_2)$ defined by

$$m_{in}(x_1,x_2) = \frac{\sum_{j=1}^{n} Y_{ij}\phi_1\left(\frac{x_1-X_{1j}}{a_n}\right)\phi_2\left(\frac{x_2-X_{2j}}{a_n}\right)}{\sum_{j=1}^{n} \phi_1\left(\frac{x_1-X_{1j}}{a_n}\right)\phi_2\left(\frac{x_2-X_{2j}}{a_n}\right)}, \quad i = 1,2. \qquad \ldots(4.2)$$

Theorems 4.1 and 4.2 state the asymptotic properties of the estimates $m_{in}(x_1,x_2)$, $i = 1,2$. In Theorem 4.1 we have shown that under a set of regularity conditions the estimates are uniformly (uniform over a closed finite rectangle) strongly consistent. In Theorem 4.2 we have proved that under a second set of regularity conditions the estimates are asymptotically jointly normally distributed. These two theorems can be regarded as appropriate generalizations of the earlier results due to Nadaraya [25, 27] and Schuster [42].

4.2 Uniform Strong Consistency of $m_{in}(x_1,x_2)$, $i = 1,2$

In this section the following conditions will be referred to as Conditions (A.4.2).

Condition 4.2.1: $m_i(x_1,x_2)$, $i = 1,2$ and $g(x_1,x_2)$ are continuous over the entire two dimensional Euclidean plane.

Condition 4.2.2: $g(x_1,x_2)$ is bounded away from zero over every closed finite rectangle of the Euclidean plane.

Condition 4.2.3: $\lim_{|y|\to\infty} y^2\phi_i(y) = 0$, $i = 1,2$.

Condition 4.2.4: $\phi_i(y)$, $i = 1,2$ are continuous functions of bounded variation on $(-\infty,\infty)$.

Condition 4.2.5: The random variables Y_i, $i = 1,2$ are bounded with probability one, i.e., there exist real constants, A_i, B_i, $i = 1,2$ such that $P[A_i \leq Y_i \leq B_i] = 1$, $i = 1,2$.

Condition 4.2.6: The infinite series $\sum_{n=1}^{\infty} \exp(-\gamma n a_n^4)$ is convergent for any $\gamma > 0$.

From Conditions 4.2.3 and 4.2.4 we conclude that the functions $\phi_i(y)$ and $y^2\phi_i(y)$, $i = 1,2$ are bounded.

Let $R_i = \{(x_1,x_2): a_{ij} \leq x_j \leq b_{ij},\ j = 1,2\}$, $i = 1,2$ be two closed finite rectangles in the Euclidean plane.

We first prove five lemmas. Without any loss of generality we shall assume that $A_i = 0$, $B_i = 1$, $i = 1,2$. We also define

$$\psi_i(x_1,x_2) = m_i(x_1,x_2)g(x_1,x_2)$$

$$\psi_{in}(x_1,x_2) = \frac{1}{na_n^2}\sum_{j=1}^{n} Y_{ij}\phi_1\left(\frac{x_1-X_{1j}}{a_n}\right)\phi_2\left(\frac{x_2-X_{2j}}{a_n}\right) \quad \ldots(4.3)$$

$$g_n(x_1,x_2) = \frac{1}{na_n^2}\sum_{j=1}^{n} \phi_1\left(\frac{x_1-X_{1j}}{a_n}\right)\phi_2\left(\frac{x_2-X_{2j}}{a_n}\right).$$

Lemma 4.1. Under Conditions 4.2.1, 4.2.3, 4.2.4, and 4.2.5,

$$\lim_{n\to\infty} \sup_{(x_1,x_2)\in R_1} \left|E\{\psi_{1n}(x_1,x_2)\} - \psi_1(x_1,x_2)\right| = 0.$$

Proof. We have

$$E\{\psi_{1n}(x_1,x_2)\} = \frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^1\int_0^1 y_1\phi_1\left(\frac{x_1-t_1}{a_n}\right)\phi_2\left(\frac{x_2-t_2}{a_n}\right)$$

$$\cdot f(t_1,t_2,y_1,y_2)\,dy_1dy_2dt_1dt_2$$

$$= \frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[Y_1\phi_1\left(\frac{x_1-X_1}{a_n}\right)\phi_2\left(\frac{x_2-X_2}{a_n}\right)$$

$$\cdot|X_1 = t_1,\ X_2 = t_2]g(t_1,t_2)\,dt_1dt_2$$

$$= \frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1\left(\frac{x_1-t_1}{a_n}\right)\phi_2\left(\frac{x_2-t_2}{a_n}\right)E[Y_1|X_1 = t_1,$$

$$X_2 = t_2]g(t_1,t_2)\,dt_1dt_2 \quad \ldots(4.4)$$

$$= \frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1\left(\frac{x_1-t_1}{a_n}\right)\phi_2\left(\frac{x_2-t_2}{a_n}\right)\psi_1(x_1,x_2)\,dt_1dt_2.$$

Let $\delta > 0$. We define $M = \sup_{(x_1,x_2)\in R_1} |\psi_1(x_1,x_2)|$. Using relation (4.4) we get

$$\sup_{(x_1,x_2)\in R_1} \left| E\{\psi_{1n}(x_1,x_2)\} - \psi_1(x_1,x_2)\right|$$

$$\leq \sup_{(x_1,x_2)\in R_1} \sup_{\{(t_1^2+t_2^2)^{\frac{1}{2}}<\delta\}} \left|\psi_1(x_1-t_1,x_2-t_2)\right.$$

$$\left. - \psi_1(x_1,x_2)\right| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\phi_1(u)\phi_2(v)|dudv$$

$$+ \sup_{(x_1,x_2)\in R_1} \iint_{\{(t_1^2+t_2^2)^{\frac{1}{2}}\geq\delta\}} \frac{|\psi_1(x_1-t_1,x_2-t_2)|}{(t_1^2+t_2^2)}$$

$$\cdot \frac{(t_1^2+t_2^2)}{a_n^2} \left|\phi_1(\frac{t_1}{a_n})\phi_2(\frac{t_2}{a_n})\right| dt_1dt_2$$

$$+ M \iint_{\{(t_1^2+t_2^2)^{\frac{1}{2}}\geq\delta/a_n\}} |\phi_1(t_1)\phi_2(t_2)|dt_1dt_2$$

$$\leq \sup_{(x_1,x_2)\in R_1} \sup_{\{(t_1^2+t_2^2)^{\frac{1}{2}}<\delta\}} \left|\psi_1(x_1-t_1,x_2-t_2)\right.$$

$$\left. - \psi_1(x_1,x_2)\right| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\phi_1(u)\phi_2(v)|dudv$$

$$+ \frac{1}{\delta^2} \sup_{\{(u_1^2+u_2^2)^{\frac{1}{2}}\geq\delta/a_n\}} (u_1^2+u_2^2)|\phi_1(u_1)\phi_2(u_2)| \qquad \ldots(4.5)$$

$$+ M \iint_{\{(t_1^2+t_2^2)^{\frac{1}{2}}\geq\delta/a_n\}} |\phi_1(t_1)\phi_2(t_2)|dt_1dt_2.$$

Let η be an arbitrary positive number. By letting n tend to infinity we can make the sum of the last two terms in (4.5) less than $\eta/2$. Then by choosing δ sufficiently small we can make the first term less than $\eta/2$. This proves the lemma.

Lemma 4.2. Under Conditions 4.2.4 and 4.2.5 we have for any $\varepsilon > 0$

$$P[\sup_{(x_1,x_2)\in R_1} |\psi_{1n}(x_1,x_2) - E\{\psi_{1n}(x_1,x_2)\}| > \varepsilon]$$

$$\leq \lambda_1 e^{-\lambda_2 n a_n^4} + \lambda_3 e^{-\lambda_4 n a_n^4}$$

where λ_i, $i = 1,2,3,4$ are positive constants.

Proof. We have

$$\psi_{1n}(x_1,x_2) = \frac{1}{a_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{0}^{1} y\phi_1(\frac{x_1-t_1}{a_n})\phi_2(\frac{x_2-t_2}{a_n}) dH_n(t_1,t_2,y)$$

where $H_n(u,v,w)$ is the empirical distribution function of the sample $(X_{11},X_{21},Y_{11}),(X_{12},X_{22},Y_{12}),\ldots,(X_{1n},X_{2n},Y_{1n})$ defined by

$$H_n(u,v,w) = \frac{1}{n} \sum_{j=1}^{n} \Phi(u-X_{1j})\Phi(v-X_{2j})\Phi(w-Y_{1j})$$

where $\Phi(x-y) = 1$ if $y \leq x$ and vanishes otherwise. Let $H(u,v,w)$ be the joint distribution function of (X_1,X_2,Y_1), $H(u|v,w)$ be the conditional distribution function of X_1 given (X_2,Y_1) and $H_1(v,w)$ be the joint distribution function of (X_2,Y_1).

We have

$$E\{\psi_{1n}(x,y)\} = \frac{1}{a_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{0}^{1} w\phi_1(\frac{x-u}{a_n})\phi_2(\frac{y-v}{a_n}) dH(u,v,w)$$

$$= \frac{1}{a_n^2} \int_{-\infty}^{\infty} \int_{0}^{1} w\phi_2(\frac{y-v}{a_n})\{\int_{-\infty}^{\infty} \phi_1(\frac{x-u}{a_n}) dH(u|v,w)\} dH_1(v,w).$$

Integrating by parts the integral in the brackets and interchanging the order of integration we get

$$E\{\psi_{1n}(x,y)\} = \frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_0^1 w\phi_2\left(\frac{y-v}{a_n}\right)\{-\int_{-\infty}^{\infty} H(u|v,w)$$

$$\cdot d\phi_1\left(\frac{x-u}{a_n}\right)\}dH_1(v,w)$$

$$= -\frac{1}{a_n^2}\int_{-\infty}^{\infty}\{\int_{-\infty}^{\infty}\int_0^1 w\phi_2\left(\frac{y-v}{a_n}\right)H(u|v,w) \qquad \ldots(4.6)$$

$$\cdot dH_1(v,w)\}d\phi_1\left(\frac{x-u}{a_n}\right)$$

$$= -\frac{1}{a_n^2}\int_{-\infty}^{\infty}\{\int_{-\infty}^{\infty}\int_0^1 w\phi_2\left(\frac{y-v}{a_n}\right)dH_u^*(v,w)\}d\phi_1\left(\frac{x-u}{a_n}\right)$$

where for fixed u_o the function $H_{u_o}^*(v,w)$ is defined by

$$H_{u_o}^*(v,w) = \int_{-\infty}^{v}\int_0^w H(u_o|t_1,t_2)dH_1(t_1,t_2) = H(u_o,v,w)$$

Repeating twice this operation with the integral in the bracket in (4.6) we get

$$E\{\psi_{1n}(x,y)\} = \frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(u,v)d\phi_2\left(\frac{y-v}{a_n}\right)d\phi_1\left(\frac{x-u}{a_n}\right)$$
$$- \frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^1 H(u,v,w)dwd\phi_2\left(\frac{y-v}{a_n}\right)d\phi_1\left(\frac{x-u}{a_n}\right). \qquad \ldots(4.7)$$

Similarly,

$$\psi_{1n}(x,y) = \frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_n(u,v)d\phi_2\left(\frac{y-v}{a_n}\right)d\phi_1\left(\frac{x-u}{a_n}\right)$$
$$- \frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^1 H_n(u,v,w)dwd\phi_2\left(\frac{y-v}{a_n}\right)d\phi_1\left(\frac{x-u}{a_n}\right) \qquad \ldots(4.8)$$

where $G_n(u,v)$ is the empirical distribution function of the random sample $(X_{11},X_{21}),(X_{12},X_{22}),\ldots,(X_{1n},X_{2n})$ and can be defined in a similar manner.

Let E_ℓ be the ℓ-dimensional Euclidean space and μ_1 and μ_2 be the variations of ϕ_1 and ϕ_2 defined by

$$\mu_1 = \int_{-\infty}^{\infty} |d\phi_1(u)|$$

$$\mu_2 = \int_{-\infty}^{\infty} |d\phi_2(v)|.$$

Then we have

$$\sup_{(x_1,x_2)\in R_1} |\psi_{1n}(x_1,x_2) - E\{\psi_{1n}(x_1,x_2)\}|$$

$$= \sup_{(x_1,x_2)\in R_1} \Big| \frac{1}{a_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G_n(t_1,t_2) - G(t_1,t_2)]$$

$$\cdot d\phi_2(\frac{x_2-t_2}{a_n}) d\phi_1(\frac{x_1-t_1}{a_n})$$

$$- \frac{1}{a_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^1 [H_n(t_1,t_2,y) - H(t_1,t_2,y)]$$

$$\cdot dy d\phi_2(\frac{x_2-t_2}{a_n}) d\phi_1(\frac{x_1-t_1}{a_n}) \Big|$$

$$\leq \frac{\mu_1\mu_2}{a_n^2} [\sup_{(u,v)\in E_2} |G_n(u,v) - G(u,v)|$$

$$+ \sup_{(u,v,w)\in E_3} |H_n(u,v,w) - H(u,v,w)|].$$

By invoking the result of Keifer and Wolfowitz [16] we get for any $\varepsilon > 0$

$$P[\sup_{(x_1,x_2)\in R_1} |\psi_{1n}(x_1,x_2) - E\{\psi_{1n}(x_1,x_2)\}| > \varepsilon]$$

$$\leq P[\sup_{(u,v)\in E_2} |G_n(u,v) - G(u,v)| > \frac{\varepsilon a_n^2}{2\mu_1\mu_2}]$$

$$+ P[\sup_{(u,v,w)\epsilon E_3} |H_n(u,v,w) - H(u,v,w)| > \frac{\epsilon a_n^2}{2\mu_1\mu_2}]$$

$$\leq \lambda_1 e^{-\lambda_2 n a_n^4} + \lambda_3 e^{-\lambda_4 n a_n^4}$$

where λ_i, $i = 1,2,3,4$ are positive constants.

Lemma 4.3. Under Conditions 4.2.1, 4.2.3, 4.2.4, and 4.2.5 for all sufficiently large n and for any $\epsilon > 0$,

$$P[\sup_{(x_1,x_2)\epsilon R_1} |\psi_{1n}(x_1,x_2) - \psi_1(x_1,x_2)| > \epsilon]$$

$$\leq \lambda_1 e^{-\lambda_2 n a_n^4} + \lambda_3 e^{-\lambda_4 n a_n^4}$$

where λ_i, $i = 1,2,3,4$ are positive constants.

Proof. For any $\epsilon > 0$ we have for all sufficiently large n

$$P[\sup_{(x_1,x_2)\epsilon R_1} |\psi_{1n}(x_1,x_2) - \psi_1(x_1,x_2)| > \epsilon]$$

$$\leq P[\sup_{(x_1,x_2)\epsilon R_1} |\psi_{1n}(x_1,x_2) - E\{\psi_{1n}(x_1,x_2)\}| > \frac{\epsilon}{2}]$$

$$\leq \lambda_1 e^{-\lambda_2 n a_n^4} + \lambda_3 e^{-\lambda_4 n a_n^4}, \quad \text{by Lemmas 4.1 and 4.2.}$$

Lemma 4.4. Under Conditions 4.2.1, 4.2.3, and 4.2.4

$$\lim_{n\to\infty} \sup_{(x_1,x_2)\epsilon R_1} |E\{g_n(x_1,x_2)\} - g(x_1,x_2)| = 0.$$

We omit the proof which can be accomplished using a similar method to that used in Lemma 4.1.

Lemma 4.5. Under Conditions 4.2.1, 4.2.3, and 4.2.4 for all sufficiently large n we have for any $\epsilon > 0$

$$P[\sup_{(x_1,x_2)\epsilon R_1} |g_n(x_1,x_2) - g(x_1,x_2)| > \epsilon]$$

$$\leq \lambda_1 \exp\left(-\frac{\lambda_2 \varepsilon^2 n a_n^4}{\mu_1^2 \mu_2^2}\right)$$

where λ_1 and λ_2 are positive constants.

<u>Proof.</u> Using Lemma 4.4 we get for all sufficiently large n

$$P[\sup_{(x_1,x_2)\in R_1} |g_n(x_1,x_2) - g(x_1,x_2)| > \varepsilon]$$

$$\leq P[\sup_{(x_1,x_2)\in R_1} |g_n(x_1,x_2) - E\{g_n(x_1,x_2)\}| > \frac{\varepsilon}{2}]$$

$$\leq \lambda_1 \exp\left(-\frac{\lambda_2 \varepsilon^2 n a_n^4}{\mu_1^2 \mu_2^2}\right).$$

The last inequality follows from Lemma 1 of Samanta [39].

We are now in a position to prove the following theorem.

<u>Theorem 4.1.</u> Under Conditions (A.4.2) we have

$$P[\lim_{n\to\infty} \sup_{(x_1,x_2)\in R_i} |m_{in}(x_1,x_2) - m_i(x_1,x_2)|$$

$$= 0,\ i = 1,2] = 1.$$

<u>Proof.</u> To prove the theorem it suffices to show that

$$P[\lim_{n\to\infty} \sup_{(x_1,x_2)\in R_1} |m_{1n}(x_1,x_2) - m_1(x_1,x_2)| = 0] = 1$$

and

$$P[\lim_{n\to\infty} \sup_{(x_1,x_2)\in R_2} |m_{2n}(x_1,x_2) - m_2(x_1,x_2)| = 0] = 1.$$

Since the definitions of $m_{1n}(x_1,x_2)$ and $m_{2n}(x_1,x_2)$ and the corresponding parametric functions are similar it is enough to prove the first result.

We define

$$\ell_1 = \min_{(x_1,x_2)\in R_1} |g(x_1,x_2)|,$$

and

$$\ell_2 = \max_{(x_1,x_2)\in R_1} |m_1(x_1,x_2)|.$$

Let $\varepsilon > 0$ ($\varepsilon < \ell_1$) be arbitrarily chosen. Using standard computations in probability theory we get

$$P[\sup_{(x_1,x_2)\in R_1} |m_{1n}(x_1,x_2) - m_1(x_1,x_2)| > \varepsilon]$$

$$= P[\{\sup_{(x_1,x_2)\in R_1} \left|\frac{\psi_{1n}(x_1,x_2)}{g_n(x_1,x_2)} - m_1(x_1,x_2)\right| > \varepsilon\}$$

$$\cap \{\sup_{(x_1,x_2)\in R_1} |g_n(x_1,x_2) - g(x_1,x_2)| \leq \varepsilon\}]$$

$$+ P[\{\sup_{(x_1,x_2)\in R_1} \left|\frac{\psi_{1n}(x_1,x_2)}{g_n(x_1,x_2)} - m_1(x_1,x_2)\right| > \varepsilon\}$$

$$\cap \{\sup_{(x_1,x_2)\in R_1} |g_n(x_1,x_2) - g(x_1,x_2)| > \varepsilon\}]$$

$$\leq P[\sup_{(x_1,x_2)\in R_1} |\psi_{1n}(x_1,x_2) - m_1(x_1,x_2)g_n(x_1,x_2)|$$

$$> \varepsilon(\ell_1-\varepsilon)]$$

$$+ P[\sup_{(x_1,x_2)\in R_1} |g_n(x_1,x_2) - g(x_1,x_2)| > \varepsilon]$$

$$\leq P[\sup_{(x_1,x_2)\in R_1} |\psi_{1n}(x_1,x_2) - \psi_1(x_1,x_2)| > \frac{\varepsilon(\ell_1-\varepsilon)}{2}]$$

$$+ P[\sup_{(x_1,x_2)\in R_1} |g_n(x_1,x_2) - g(x_1,x_2)| > \frac{\varepsilon(\ell_1-\varepsilon)}{2\ell_2}]$$

$$+ P[\sup_{(x_1,x_2)\in R_1} |g_n(x_1,x_2) - g(x_1,x_2)| > \varepsilon].$$

We now invoke Lemmas 4.3 and 4.5 to conclude that there exist positive constants λ_i, $i = 1,2,\ldots,8$ such that for sufficiently large n

$$P[\sup_{(x_1,x_2)\in R_1} |m_{1n}(x_1,x_2) - m_1(x_1,x_2)| > \varepsilon]$$

$$\leq \lambda_1 e^{-\lambda_2 n a_n^4} + \lambda_3 e^{-\lambda_4 n a_n^4} + \lambda_5 e^{-\lambda_6 n a_n^4} + \lambda_7 e^{-\lambda_8 n a_n^4}.$$

This inequality in conjunction with the Borel-Cantelli lemma and Condition 4.2.6 completes the proof of the theorem.

Remark 4.2.1. Theorem 4.1 can be generalized to the situation where $\underline{X} = (X_1, X_2, \ldots, X_{p_1})'$ and $\underline{Y} = (Y_1, Y_2, \ldots, Y_{p_2})'$. When the regression functions and their estimates are defined in a similar manner, the theorem remains true under conditions similar to those mentioned before. For example, Conditions 4.2.3 and 4.2.6 should be replaced by:

Condition 4.2.7: $\lim_{|y|\to\infty} y^{p_1} \phi_i(y) = 0$, $i = 1,2,\ldots,p_1$

Condition 4.2.8: The infinite series $\sum_{n=1}^{\infty} \exp(-\gamma n a_n^{2p_1})$ is convergent for any $\gamma > 0$.

4.3 Asymptotic Joint Normality of the Estimates $m_{in}(x_1,x_2)$, $i = 1,2$

We shall investigate the asymptotic joint distribution of our estimates for two distinct points.

We note that for $i = 1,2$

$$E(Y_i | X_1 = x_1, X_2 = x_2) = \frac{w_i(x_1,x_2)}{g(x_1,x_2)} \qquad \ldots(4.9)$$

where

$$w_i(x_1,x_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_i f(x_1,x_2,y_1,y_2)\,dy_1 dy_2.$$

We now define for $i,j = 1,2$

$$E(Y_iY_j | X_1 = x_1,\ X_2 = x_2) = \frac{v_{ij}(x_1,x_2)}{g(x_1,x_2)}$$

where

$$v_{ij}(x_1,x_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_i y_j f(x_1,x_2,y_1,y_2)\,dy_1 dy_2.$$

The following conditions will be referred to as Conditions (A.4.3).

Condition 4.3.1: (x_{11},x_{21}) and (x_{12},x_{22}) are two distinct points such that $g(x_{11},x_{21}) > 0$ and $g(x_{12},x_{22}) > 0$.

Condition 4.3.2: For arbitrarily small positive η, $E[|Y_i|^{3+\eta}] < \infty$, $i = 1,2$.

Condition 4.3.3: $\frac{\partial g(x_1,x_2)}{\partial x_r}$, $\frac{\partial^2 g(x_1,x_2)}{\partial x_r \partial x_s}$, $\frac{\partial w_i(x_1,x_2)}{\partial x_r}$, $\frac{\partial^2 w_i(x_1,x_2)}{\partial x_r \partial x_s}$, $\frac{\partial v_{ij}(x_1,x_2)}{\partial x_r}$, $i,j,r,s = 1,2$ exist and are bounded.

Condition 4.3.4: $\phi_i(y)$ and $|y\phi_i(y)|$ are bounded and $\lim_{|y|\to\infty} |y\phi_i(y)| = 0$, $i = 1,2$.

Condition 4.3.5: $\int_{-\infty}^{\infty} y\phi_i(y)\,dy = 0$, $i = 1,2$.

Condition 4.3.6: $\int_{-\infty}^{\infty} y^2\phi_i(y)\,dy < \infty$, $i = 1,2$.

Condition 4.3.7: $\lim_{n\to\infty} na_n^{\delta} = \infty$ for $\delta < 6$ and $\lim_{n\to\infty} na_n^6 = 0$.

We define for $i,s = 1,2$ and $j = 1,2,...,n$ the following random variables.

$$U_{jn}^{*}(x_{1s},x_{2s}) = \frac{1}{a_n^2}\phi_1\left(\frac{x_{1s}-X_{1j}}{a_n}\right)\phi_2\left(\frac{x_{2s}-X_{2j}}{a_n}\right)$$

$$U_{jn}(x_{1s},x_{2s}) = a_n[U_{jn}^{*}(x_{1s},x_{2s}) - E\{U_{jn}^{*}(x_{1s},x_{2s})\}]$$

$$V_{ijn}^{*}(x_{1s},x_{2s}) = Y_{ij}U_{jn}^{*}(x_{1s},x_{2s})$$

$$V_{ijn}(x_{1s},x_{2s}) = a_n[V_{ijn}^{*}(x_{1s},x_{2s}) - E\{V_{ijn}^{*}(x_{1s},x_{2s})\}]$$

$$U_n(x_{1s},x_{2s}) = \sum_{j=1}^{n} U_{jn}(x_{1s},x_{2s})$$

$$V_{in}(x_{1s},x_{2s}) = \sum_{j=1}^{n} V_{ijn}(x_{1s},x_{2s}) \qquad ...(4.10)$$

$$\underline{W}_{jn} = (U_{jn}(x_{11},x_{21}),\ V_{1jn}(x_{11},x_{21}),\ V_{2jn}(x_{11},x_{21}),$$
$$U_{jn}(x_{12},x_{22}),\ V_{1jn}(x_{12},x_{22}),\ V_{2jn}(x_{12},x_{22}))'$$

$$n^{\frac{1}{2}}\underline{Z}_n = (U_n(x_{11},x_{21}),\ V_{1n}(x_{11},x_{21}),\ V_{2n}(x_{11},x_{21}),$$
$$U_n(x_{12},x_{22}),\ V_{1n}(x_{12},x_{22}),\ V_{2n}(x_{12},x_{22}))'$$

$$n^{\frac{1}{2}}\underline{Z}_n^{*} = a_n\sum_{j=1}^{n}(U_{jn}^{*}(x_{11},x_{21}) - g(x_{11},x_{21}),\ V_{1jn}^{*}(x_{11},x_{21})$$
$$- w_1(x_{11},x_{21}),\ V_{2jn}^{*}(x_{11},x_{21}) - w_2(x_{11},x_{21})$$
$$U_{jn}^{*}(x_{12},x_{22}) - g(x_{12},x_{22}),\ V_{1jn}^{*}(x_{12},x_{22})$$
$$- w_1(x_{12},x_{22}),\ V_{2jn}^{*}(x_{12},x_{22}) - w_2(x_{12},x_{22}))'.$$

Then we can write

$$\underline{Z}_n = \underline{Z}_n^{*} + \underline{F}_n$$

where

$$\underline{F}_n = \sum_{j=1}^{n} \underline{F}_{jn}$$

with

$$\underline{F}_{jn} = \frac{a_n}{n^{\frac{1}{2}}} (g(x_{11},x_{21}) - E\{U^*_{jn}(x_{11},x_{21})\},\ w_1(x_{11},x_{21})$$

$$- E\{V^*_{1jn}(x_{11},x_{21})\},\ w_2(x_{11},x_{21}) - E\{V^*_{2jn}(x_{11},x_{21})\},$$

$$g(x_{12},x_{22}) - E\{U^*_{jn}(x_{12},x_{22})\},\ w_1(x_{12},x_{22})$$

$$- E\{V^*_{1jn}(x_{12},x_{22})\},\ w_2(x_{12},x_{22})$$

$$- E\{V^*_{2jn}(x_{12},x_{22})\})'.$$

We also define the following matrices:

$$\Lambda = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1^2(t_1)\phi_2^2(t_2)dt_1dt_2 \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}$$

$$\Lambda_1 = ((\lambda_{1ij}))_{2\times 2}, \quad \Lambda_2 = ((\lambda_{2ij}))_{2\times 2}$$

$$\lambda_{1ij} = \mathrm{Cov}(Y_i,Y_j | X_1 = x_{11},\ X_2 = x_{21})/g(x_{11},x_{21})$$

$$\lambda_{2ij} = \mathrm{Cov}(Y_i,Y_j | X_1 = x_{12},\ X_2 = x_{22})/g(x_{12},x_{22})$$

$$C = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1^2(t_1)\phi_2^2(t_2)dt_1dt_2 \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \qquad \ldots(4.11)$$

$$D_s = \begin{bmatrix} g(x_{1s},x_{2s}) & w_1(x_{1s},x_{2s}) & w_2(x_{1s},x_{2s}) \\ w_1(x_{1s},x_{2s}) & v_{11}(x_{1s},x_{2s}) & v_{12}(x_{1s},x_{2s}) \\ w_2(x_{1s},x_{2s}) & v_{12}(x_{1s},x_{2s}) & v_{22}(x_{1s},x_{2s}) \end{bmatrix} \quad s = 1,2.$$

We prove eight basic lemmas.

Lemma 4.6. Under Conditions 4.3.3, 4.3.5, and 4.3.6 we have for $i,s = 1,2$ and $j = 1,2,\dots,n$

(i) $\quad |E\{U^*_{jn}(x_{1s},x_{2s})\} - g(x_{1s},x_{2s})| = 0(a_n^2)$

(ii) $\quad |E\{V^*_{ijn}(x_{1s},x_{2s})\} - w_i(x_{1s},x_{2s})| = 0(a_n^2)$.

Proof. We shall prove part (ii) of this lemma. The proof of part (i) is similar and is omitted.

We have for $i,s = 1,2$

$$|E\{V^*_{ijn}(x_{1s},x_{2s}) - w_i(x_{1s},x_{2s})|$$

$$= \Big|\frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_i\phi_1\Big(\frac{x_{1s}-u}{a_n}\Big)\phi_2\Big(\frac{x_{2s}-t}{a_n}\Big)$$

$$\cdot f(u,t,y_1,y_2)dy_1dy_2dudt - w_i(x_{1s},x_{2s})\Big|$$

$$= \Big|\frac{1}{a_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_1\Big(\frac{x_{1s}-u}{a_n}\Big)\phi_2\Big(\frac{x_{2s}-t}{a_n}\Big)w_i(u,t)dudt - w_i(x_{1s},x_{2s})\Big|$$

$$= \Big|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_1(u)\phi_2(t)\{w_i(x_{1s}-a_nu,\ x_{2s}-a_nt)$$

$$- w_i(x_{1s},x_{2s})\}dudt\Big|.$$

Expanding $w_i(x_{1s}-a_nu,\ x_{2s}-a_nt)$ around (x_{1s},x_{2s}) up to the order of a_n^2 and using the hypothesis we get the desired result.

Lemma 4.7. Under Conditions 4.3.3, 4.3.5, 4.3.6, and 4.3.7, $\underline{F}_n$ converges to the null vector in E_6 as n tends to infinity.

Proof. By Lemma 4.6 we have

$$\underline{F}_{jn} = \frac{O(a_n^3)}{n^{1/2}} (1,1,1,1,1,1)'.$$

The proof now follows from Condition 4.3.7 and the fact that

$$\underline{F}_n = \sum_{j=1}^{n} \underline{F}_{jn}.$$

Lemma 4.8. Under Conditions 4.3.3, 4.3.4, 4.3.5, and 4.3.6 the following results hold for $i,\ell,r,s = 1,2$ and $j = 1,2,\ldots,n$.

(i) $E\{U_{jn}^2(x_{1s},x_{2s})\}$

$$= g(x_{1s},x_{2s}) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1^2(u)\phi_2^2(t)\,du\,dt + O(a_n)$$

(ii) $E\{V_{ijn}(x_{1s},x_{2s})V_{\ell jn}(x_{1s},x_{2s})\} = v_{i\ell}(x_{1s},x_{2s})$

$$\cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1^2(u)\phi_2^2(t)\,du\,dt + O(a_n)$$

(iii) $E\{U_{jn}(x_{1s},x_{2s})V_{ijn}(x_{1s},x_{2s})\} = w_i(x_{1s},x_{2s})$

$$\cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1^2(u)\phi_2^2(t)\,du\,dt + O(a_n)$$

(iv) $E\{U_{jn}(x_{1s},x_{2s})U_{jn}(x_{1r},x_{2r})\} = O(a_n), \quad r \neq s$

(v) $E\{V_{ijn}(x_{1s},x_{2s})V_{\ell jn}(x_{1r},x_{2r})\} = O(a_n), \quad r \neq s$

(vi) $E\{U_{jn}(x_{1s},x_{2s})V_{ijn}(x_{1r},x_{2r})\} = O(a_n), \quad r \neq s.$

Proof. We shall prove part (ii) and part (v) of the lemma. The proofs of the other parts are similar and are omitted.

(ii) $E\{V_{ijn}(x_{1s},x_{2s})V_{\ell jn}(x_{1s},x_{2s})\}$

$$= \frac{1}{a_n^2} E\left[Y_i Y_\ell \phi_1^2\left(\frac{x_{1s}-X_1}{a_n}\right)\phi_2^2\left(\frac{x_{2s}-X_2}{a_n}\right)\right]$$

$$- a_n^2 E\left[\frac{1}{a_n^2} Y_i \phi_1\left(\frac{x_{1s}-X_1}{a_n}\right)\phi_2\left(\frac{x_{2s}-X_2}{a_n}\right)\right]$$

$$\cdot E\left[\frac{1}{a_n^2} Y_\ell \phi_1\left(\frac{x_{1s}-X_1}{a_n}\right)\phi_2\left(\frac{x_{2s}-X_2}{a_n}\right)\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1^2(t_1)\phi_2^2(t_2)v_{i\ell}(x_{1s}-a_nt_1,\ x_{2s}-a_nt_2)dt_1dt_2$$

$$- a_n^2 \{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1(t_1)\phi_2(t_2)w_i(x_{1s}-a_nt_1,$$

$$x_{2s}-a_nt_2)dt_1dt_2\}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1(t_1)\phi_2(t_2)w_\ell(x_{1s}-a_nt_1,$$

$$x_{2s}-a_nt_2)dt_1dt_2\}.$$

Expanding $v_{i\ell}(x_{1s}-a_nt_1,\ x_{2s}-a_nt_2)$ and $w_i(x_{1s}-a_nt_1,\ x_{2s}-a_nt_2)$ around (x_{1s},x_{2s}) to the order of a_n and using the hypothesis, the desired conclusion follows.

(v) Let $\delta_{1n} = \dfrac{x_{12}-x_{11}}{a_n}$ and $\delta_{2n} = \dfrac{x_{22}-x_{21}}{a_n}$. Since the points (x_{11},x_{21}) and (x_{12},x_{22}) are distinct at least one of the quantities δ_{1n} and δ_{2n} is nonzero. We shall assume without any loss of generality that $\delta_{1n} > 0$.

We have

$$E[V_{ijn}(x_{11},x_{21})V_{\ell jn}(x_{12},x_{22})]$$

$$= \frac{1}{a_n^2} E\left[Y_i Y_\ell \phi_1(\frac{x_{11}-X_1}{a_n})\phi_2(\frac{x_{21}-X_2}{a_n})\phi_1(\frac{x_{12}-X_1}{a_n})\phi_2(\frac{x_{22}-X_2}{a_n})\right]$$

$$- a_n^2 E\left[\frac{1}{a_n^2} Y_i \phi_1(\frac{x_{11}-X_1}{a_n})\phi_2(\frac{x_{21}-X_2}{a_n})\right] E\left[\frac{1}{a_n^2} Y_\ell \phi_1(\frac{x_{12}-X_1}{a_n})\right.$$

$$\left.\cdot \phi_2(\frac{x_{22}-X_2}{a_n})\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1(t_1)\phi_2(t_2)\phi_1(\delta_{1n}+t_1)\phi_2(\delta_{2n}+t_2)$$

$$\cdot v_{i\ell}(x_{11}-a_n t_1, x_{21}-a_n t_2)dt_1 dt_2$$

$$- a_n^2 \{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1(t_1)\phi_2(t_2)w_i(x_{11}-a_n t_1, x_{21}-a_n t_2)dt_1 dt_2\}$$

$$\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1(t_1)\phi_2(t_2)w_\ell(x_{12}-a_n t_1, x_{22}-a_n t_2)dt_1 dt_2\}.$$

Expanding $v_{i\ell}(x_{11}-a_n t_1, x_{21}-a_n t_2)$ around (x_{11},x_{21}) and $w_\ell(x_{12}-a_n t_1, x_{22}-a_n t_2)$ around (x_{12},x_{22}) to the order of a_n and using the hypothesis, we get

$$|E[V_{ijn}(x_{11},x_{21})V_{\ell jn}(x_{12},x_{22})]|$$

$$\leq |v_{i\ell}(x_{11},x_{21})| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1(t_1)\phi_2(t_2)\phi_1(\delta_{1n}+t_1)$$

$$\cdot \phi_2(\delta_{2n}+t_2)dt_1 dt_2 + O(a_n)$$

$$= |v_{i\ell}(x_{11},x_{21})| \int_{-\infty}^{\infty}\{\int_{|t_1|<\frac{\delta_{1n}}{2}} \phi_1(t_1)\phi_2(t_2)$$

$$\cdot \phi_1(\delta_{1n}+t_1)\phi_2(\delta_{2n}+t_2)dt_1\}dt_2$$

$$+ |v_{i\ell}(x_{11},x_{21})| \int_{-\infty}^{\infty}\{ \int_{|t_1| \geq \frac{\delta_{1n}}{2}} \phi_1(t_1)\phi_2(t_2)$$

$$\cdot \phi_1(\delta_{1n}+t_1)\phi_2(\delta_{2n}+t_2)dt_1\}dt_2 + O(a_n)$$

$$\leq |v_{i\ell}(x_{11},x_{21})| \{ \sup_{|t_1| < \frac{\delta_{1n}}{2}} \phi_1(\delta_{1n}+t_1)\}$$

$$\cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_1(t_1)\phi_2(t_2)\phi_2(\delta_{2n}+t_2)dt_1dt_2$$

$$+ |v_{i\ell}(x_{11},x_{21})| \{ \sup_{|t_1| \geq \frac{\delta_{1n}}{2}} \phi_1(t_1)\}$$

$$\cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_2(t_2)\phi_1(\delta_{1n}+t_1)\phi_2(\delta_{2n}+t_2)dt_1dt_2 + O(a_n)$$

$$= \{ \sup_{|t_1| < \delta_{1n/2}} \phi_1(\delta_{1n}+t_1)\} \cdot O(1)$$

$$+ \{ \sup_{|t_1| \geq \delta_{1n/2}} \phi_1(t_1)\} \cdot O(1) + O(a_n)$$

$$\leq 2\{ \sup_{|z| \geq \delta_{1n/2}} \phi_1(z)\} \cdot O(1) + O(a_n)$$

$$= 2 \sup_{|z| \geq \delta_{1n/2}} \frac{1}{|z|} |z\phi_1(z)| \cdot O(1) + O(a_n)$$

$$\leq 4\delta_{1n}^{-1} \sup_{|z| \geq \delta_{1n/2}} |z\phi_1(z)| \cdot O(1) + O(a_n)$$

$$= \frac{4a_n}{(x_{12}-x_{11})} \sup_{|z| \geq \delta_{1n/2}} |z\phi_1(z)| \cdot O(1) + O(a_n)$$

$$= O(a_n) \cdot o(1) + O(a_n)$$

$$= O(a_n).$$

Lemma 4.9. Under Conditions 4.3.3, 4.3.4, and 4.3.6 for any $\delta > 0$ the following results hold for $i,s = 1,2$ and $j = 1,2,...,n$

(i) $E\{|U_{jn}(x_{1s},x_{2s})|^{2+\delta}\} = O\left(\frac{1}{a_n^{\delta}}\right)$

(ii) $E\{|V_{ijn}(x_{1s},x_{2s})|^{2+\delta}\} = O\left(\frac{1}{a_n^{2+\delta-\frac{2\varepsilon}{1+\varepsilon}}}\right)$

where ε is any positive number for which $E[|Y_i|^{(2+\delta)(1+\varepsilon)}]$ is finite.

Proof. (i) We have

$$E\{|U_{jn}(x_{1s},x_{2s})|^{2+\delta}\} = a_n^{2+\delta}E[|U_{jn}^*(x_{1s},x_{2s})$$
$$- E\{U_{jn}^*(x_{1s},x_{2s})\}|^{2+\delta}].$$

By Minkowski's inequality and by an analysis similar to that in Lemma 4.8 we have

$$[E\{|U_{jn}(x_{1s},x_{2s})|^{2+\delta}]^{\frac{1}{2+\delta}}$$

$$\leq a_n[\{E(|U_{jn}^*(x_{1s},x_{2s})|)^{2+\delta}\}^{\frac{1}{2+\delta}}$$

$$+ \{|E(U_{jn}^*(x_{1s},x_{2s}))|^{2+\delta}\}^{\frac{1}{2+\delta}}]$$

$$= O\left(\frac{1}{a_n^{\delta/(2+\delta)}}\right) + O(a_n)$$

$$= O\left(\frac{1}{a_n^{\delta/(2+\delta)}}\right).$$

The desired conclusion for (i) now follows.

(ii) We have

$$E\{|V_{ijn}(x_{1s},x_{2s})|^{2+\delta}\} = a_n^{2+\delta}E[|V^*_{ijn}(x_{1s},x_{2s})$$
$$- E\{V^*_{ijn}(x_{1s},x_{2s})\}|^{2+\delta}].$$

By Minkowski's inequality and the results in Lemma 4.8 we get

$$\{E|V_{ijn}(x_{1s},x_{2s})|^{2+\delta}\}^{\frac{1}{2+\delta}}$$
$$\leq a_n[\{E(|V^*_{ijn}(x_{1s},x_{2s})|)^{2+\delta}\}^{\frac{1}{2+\delta}}$$
$$+ \{|E(V^*_{ijn}(x_{1s},x_{2s}))|^{2+\delta}\}^{\frac{1}{2+\delta}}]$$
$$= a_n[E\{|V^*_{ijn}(x_{1s},x_{2s})|^{2+\delta}\}]^{\frac{1}{2+\delta}} + O(a_n).$$

Using Holder's inequality we get for any $\varepsilon > 0$

$$a_n^{2+\delta}[E\{|V^*_{ijn}(x_{1s},x_{2s})|^{2+\delta}\}]$$
$$= a_n^{-(2+\delta)}\left[E\{\left|Y_i\phi_1(\frac{x_{1s}-X_1}{a_n})\phi_2(\frac{x_{2s}-X_2}{a_n})\right|^{2+\delta}\}\right]$$
$$\leq a_n^{-(2+\delta)}[E\{|Y_i|^{(2+\delta)(1+\varepsilon)}\}]^{\frac{1}{1+\varepsilon}}$$
$$\cdot\left[E\{\phi_1(\frac{x_{1s}-X_1}{a_n})\phi_2(\frac{x_{2s}-X_2}{a_n})\}^{(2+\delta)(\frac{1+\varepsilon}{\varepsilon})}\right]^{\frac{\varepsilon}{1+\varepsilon}}$$
$$= a_n^{-(2+\delta)+2\varepsilon/(1+\varepsilon)}[E\{|Y_i|^{(2+\delta)(1+\varepsilon)}\}]^{\frac{1}{1+\varepsilon}}$$
$$\cdot\left[\frac{1}{a_n^2}E\{\phi_1(\frac{x_{1s}-X_1}{a_n})\phi_2(\frac{x_{2s}-X_2}{a_n})\}^{(2+\delta)(\frac{1+\varepsilon}{\varepsilon})}\right]^{\frac{\varepsilon}{1+\varepsilon}}.$$

Using an analysis similar to that in Lemma 4.8 and the hypothesis we conclude that

$$a_n^{2+\delta}[E\{|V_{ijn}^*(x_{1s},x_{2s})|^{2+\delta}\}] = O\left(\frac{1}{a_n^{2+\delta-2\varepsilon/(1+\varepsilon)}}\right) .$$

Hence,

$$\{E|V_{ijn}(x_{1s},x_{2s})|^{2+\delta}\}^{\frac{1}{2+\delta}} = O\left(\frac{1}{a_n^{1-\frac{2\varepsilon}{(1+\varepsilon)(2+\delta)}}}\right) + O(a_n)$$

$$= O\left(\frac{1}{a_n^{1-\frac{2\varepsilon}{(1+\varepsilon)(2+\delta)}}}\right) .$$

The conclusion now follows from the above observation.

Let $\underline{d} = (d_1,d_2,d_3,d_4,d_5,d_6)$ be any nonnull row vector in E_6. We define for $j = 1,2,\ldots,n$

$$\sigma_{jn}^2 = \mathrm{Var}\{\frac{\underline{d}\underline{W}_{jn}}{n^{\frac{1}{2}}}\}$$

and

$$\rho_{jn}^{2+\delta} = E\{\left|\frac{\underline{d}\underline{W}_{jn}}{n^{\frac{1}{2}}}\right|^{2+\delta}\}.$$

<u>Lemma 4.10</u>. Under Conditions 4.3.3, 4.3.4, 4.3.5, and 4.3.6 for $j = 1,2,\ldots,n$

(i) $\sigma_{jn}^2 = \{\underline{d}C\underline{d}'\}/n + O(\frac{a_n}{n})$

where the matrix C is as in (4.11).

(ii) $\rho_{jn}^{2+\delta} = O\left(\frac{1}{n^{1+\frac{\delta}{2}}\, a_n^{2+\delta-\frac{2\varepsilon}{1+\varepsilon}}}\right)$

where δ and ε are positive numbers for which

$E\{|Y_i|^{(2+\delta)(1+\varepsilon)}\} < \infty, \; i = 1,2.$

Proof.

(i) Using the results in Lemma 4.8 we get

$$
\begin{aligned}
n\sigma_{jn}^2 &= \mathrm{Var}\{\underline{d}\underline{W}_{jn}\} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_1^2(u)\phi_2^2(t)\,du\,dt[d_1^2 g(x_{11},x_{21}) \\
&\quad + d_2^2 v_{11}(x_{11},x_{21}) + d_3^2 v_{22}(x_{11},x_{21}) + d_4^2 g(x_{12},x_{22}) \\
&\quad + d_5^2 v_{11}(x_{12},x_{22}) + d_6^2 v_{22}(x_{12},x_{22}) \qquad \ldots(4.12) \\
&\quad + 2d_1d_2w_1(x_{11},x_{21}) + 2d_1d_3w_2(x_{11},x_{21}) \\
&\quad + 2d_2d_3v_{12}(x_{11},x_{21}) + 2d_4d_5w_1(x_{12},x_{22}) \\
&\quad + 2d_4d_6w_2(x_{12},x_{22}) + 2d_5d_6v_{12}(x_{12},x_{22})] + O(a_n) \\
&= \underline{d}C\underline{d}' + O(a_n).
\end{aligned}
$$

The desired conclusion now follows.

(ii) We have

$$
\begin{aligned}
n^{1+\frac{\delta}{2}}\cdot\rho_{jn}^{2+\delta} &= E\{|\underline{d}\underline{W}_{jn}|^{2+\delta}\} \\
&\leq |\underline{d}|^{2+\delta}E\{|\underline{W}_{jn}|^{2+\delta}\} \\
&= |\underline{d}|^{2+\delta}E\{U_{jn}^2(x_{11},x_{21}) + V_{1jn}^2(x_{11},x_{21}) \\
&\quad + V_{2jn}^2(x_{11},x_{21}) + U_{jn}^2(x_{12},x_{22}) \\
&\quad + V_{1jn}^2(x_{12},x_{22}) + V_{2jn}^2(x_{12},x_{22})\}^{\frac{2+\delta}{2}} \\
&\leq |\underline{d}|^{2+\delta}\cdot 6^{\frac{2+\delta}{2}}\; E[\max\{U_{jn}^2(x_{11},x_{21}),\; V_{1jn}^2(x_{11},x_{21}),
\end{aligned}
$$

$$V_{2jn}^2(x_{11},x_{21}),\ U_{jn}^2(x_{12},x_{22}),\ V_{1jn}^2(x_{12},x_{22}),$$

$$V_{2jn}^2(x_{12},x_{22})\}]^{\frac{2+\delta}{2}}$$

$$= |\underline{d}|^{2+\delta}\cdot 6^{\frac{2+\delta}{2}}\ E[\max\{|U_{jn}(x_{11},x_{21})|^{2+\delta},$$

$$|V_{1jn}(x_{11},x_{21})|^{2+\delta},\ |V_{2jn}(x_{11},x_{21})|^{2+\delta},$$

$$|U_{jn}(x_{12},x_{22})|^{2+\delta} + |V_{1jn}(x_{12},x_{22})|^{2+\delta},$$

$$|V_{2jn}(x_{12},x_{22})|^{2+\delta}\}]$$

$$\leq |\underline{d}|^{2+\delta}\cdot 6^{\frac{2+\delta}{2}}\ E[|U_{jn}(x_{11},x_{21})|^{2+\delta}$$

$$+ |V_{1jn}(x_{11},x_{21})|^{2+\delta} + |V_{2jn}(x_{11},x_{21})|^{2+\delta}$$

$$+ |U_{jn}(x_{12},x_{22})|^{2+\delta} + |V_{1jn}(x_{12},x_{22})|^{2+\delta}$$

$$+ |V_{2jn}(x_{12},x_{22})|^{2+\delta}]$$

$$= O\left(\frac{1}{a_n^{\delta}}\right) + O\left(\frac{1}{a_n^{2+\delta-\frac{2\varepsilon}{1+\varepsilon}}}\right), \text{ by Lemma 4.9}$$

$$= O\left(\frac{1}{a_n^{2+\delta-\frac{2\varepsilon}{1+\varepsilon}}}\right).$$

The proof of part (ii) is now complete.

<u>Lemma 4.11</u>. Under Conditions (A.4.3) for some $\delta > 0$

(i) $\lim_{n\to\infty} \{\sum_{j=1}^{n} \sigma_{jn}^2\} = \underline{d}C\underline{d}' > 0$

(ii) $\lim_{n\to\infty} \{\sum_{j=1}^{n} \rho_{jn}^{2+\delta}\} = 0.$

Proof.

(i) By part (i) of Lemma 4.10 we have

$$\sum_{j=1}^{n} \sigma_{jn}^2 = \underset{\sim}{d} C \underset{\sim}{d}' + 0(a_n).$$

To complete the proof we have to show that the submatrices D_1 and D_2 which appear in matrix C (see relation (4.11)) are each positive definite. Due to the symmetry in the definition it suffices to prove that the matrix D_1 is positive definite. This result follows from the fact that for any nonnull vector $\underset{\sim}{\ell} = (\ell_o, \ell_1, \ell_2)$ we have

$$\underset{\sim}{\ell} D_1 \underset{\sim}{\ell}' = g(x_{11}, x_{21}) E[\{\ell_o + \ell_1 Y_1 + \ell_2 Y_2\}^2 | X_1 = x_{11}, X_2 = x_{21}] > 0.$$

(ii) Using the results of part (ii) of Lemma 4.10 we have

$$\{\sum_{j=1}^{n} \rho_{jn}^{2+\delta}\} = 0\left(\frac{1}{n^{\frac{\delta}{2}} a_n^{2+\delta-\frac{2\varepsilon}{1+\varepsilon}}}\right)$$

$$= 0\left(\frac{1}{na_n^{\frac{4}{\delta}+2-\frac{4\varepsilon}{(1+\varepsilon)\delta}}}\right)^{\delta/2}$$

where ε is a positive number for which

$$[E|Y_i|^{(2+\delta)(1+\varepsilon)}] < \infty, \quad i = 1,2.$$

We now let $\varepsilon = \alpha\delta$ $(\alpha > 0)$. Then we have

$$\{\sum_{j=1}^{n} \rho_{jn}^{2+\delta}\} = 0\left(\frac{1}{na_n^{\frac{4}{\delta}+2-\frac{4\alpha}{1+\alpha\delta}}}\right)^{\delta/2}$$

provided $E[|Y_i|^{(2+\delta)(1+\alpha\delta)}] < \infty$, $i = 1,2$. By Conditions 4.3.2

and 4.3.7 the conclusion will follow if we can show that for arbitrarily small positive η there exist $\alpha > 0$ and $\delta > 0$ such that $(2+\delta)(1+\alpha\delta) < 3 + \eta$ and $\frac{4}{\delta} + 2 - \frac{4\alpha}{1+\alpha\delta} < 6$. Now, $\frac{4}{\delta} + 2 - \frac{4\alpha}{1+\alpha\delta} < 6$ if and only if $\alpha\delta^2 + \delta - 1 > 0$. We consider the equation

$$h(\delta) = \alpha\delta^2 + \delta - 1 = 0.$$

The roots of this equation are $\frac{-1 \pm \sqrt{1+4\alpha}}{2\alpha}$. It is easy to verify that the function $h(\delta)$ is strictly increasing and positive in the interval $\left(\frac{-1+\sqrt{1+4\alpha}}{2\alpha}, \infty\right)$, We further note that $\lim_{\alpha \to 0} \left\{\frac{-1+\sqrt{1+4\alpha}}{2\alpha}\right\} = 1$. Hence, for arbitrarily small positive ε_1 there exists $\eta_1 > 0$ such that $\left|\frac{-1+\sqrt{1+4\alpha}}{2\alpha} - 1\right| < \varepsilon_1$ whenever $0 < \alpha < \eta_1$. Hence, if $0 < \alpha < \min(\varepsilon_1, \eta_1)$ and $1 + \varepsilon_1 < \delta < 1 + 2\varepsilon_1$, then

$$(2+\delta)(1+\alpha\delta) < (3+2\varepsilon_1)\{1 + \varepsilon_1(1+2\varepsilon_1)\}$$

$$= 3 + 5\varepsilon_1 + 8\varepsilon_1^2 + 4\varepsilon_1^3,$$

and

$$h(\delta) = \alpha\delta^2 + \delta - 1 > 0.$$

Since ε_1 is arbitrary the proof is complete.

Lemma 4.12. Under Conditions (A.4.3), $\underline{Z}_n$ converges in distribution to $\underline{Z}$ where $\underline{Z}$ is a six dimensional normal random variable with mean vector 0 and covariance matrix C.

Proof. It suffices to prove that the random variable $\underline{d}\underline{Z}_n$ converges in distribution to $\underline{d}\underline{Z}$ as n tends to infinity (see Theorem (Xi) of Rao [30], p.103). We note that

$\underline{d}\underline{Z}_n = \sum_{j=1}^{n} \{\frac{\underline{d}\underline{W}_{jn}}{n^{\frac{1}{2}}}\}$. We also note that the results in Lemma 4.11 enable us to apply Theorem B(ii) in Loeve [21], p.275, and to conclude that $\frac{\underline{d}\underline{Z}_n}{\sqrt{\mathrm{Var}(\underline{d}\underline{Z}_n)}}$ converges in distribution to a standard normal random variable. Using this result and Lemma 4.11 we have

$$\frac{\underline{d}\underline{Z}_n}{\sqrt{\underline{d}C\underline{d}'}} = \frac{\underline{d}\underline{Z}_n}{\sqrt{\mathrm{Var}(\underline{d}\underline{Z}_n)}} \frac{\sqrt{\mathrm{Var}(\underline{d}\underline{Z}_n)}}{\sqrt{\underline{d}C\underline{d}'}}$$

$$= \frac{\underline{d}\underline{Z}_n}{\sqrt{\mathrm{Var}(\underline{d}\underline{Z}_n)}} \{1 + o(1)\}$$

$$= \frac{\underline{d}\underline{Z}_n}{\sqrt{\mathrm{Var}(\underline{d}\underline{Z}_n)}} + 0_p(1)\cdot o(1).$$

The proof follows from this observation.

Lemma 4.13. Under Conditions (A.4.3), $\underline{Z}_n^*$ converges in distribution to $\underline{Z}$ where $\underline{Z}$ is a six dimensional normal random variable with mean vector 0 and covariance matrix C.

Proof. We have by Lemma 4.7

$$\underline{Z}_n^* = \underline{Z}_n - \underline{F}_n$$

$$= \underline{Z}_n + \underline{0}_n,$$

where $\underline{0}_n$ converges to the zero vector as n tends to infinity. The proof now follows from Lemma 4.12.

We now prove the following theorem.

<u>Theorem 4.2</u>. Under Conditions (A.4.3)

$$(na_n^2)^{\frac{1}{2}}(m_{1n}(x_{11},x_{21}) - m_1(x_{11},x_{21}),\ m_{2n}(x_{11},x_{21})$$
$$- m_2(x_{11},x_{21}),\ m_{1n}(x_{12},x_{22}) - m_1(x_{12},x_{22}),$$
$$m_{2n}(x_{12},x_{22}) - m_2(x_{12},x_{22}))'$$

converges in distribution to $\underline{Z}^*$ where $\underline{Z}^*$ is a four dimensional normal random variable with mean vector 0 and covariance matrix Λ.

<u>Proof</u>. We define

$$\underline{T}_n = (T_{1n},T_{2n},T_{3n},T_{4n},T_{5n},T_{6n})$$
$$\underline{\theta} = (\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)$$

where

$$T_{1n} = \frac{1}{n}\sum_{j=1}^{n} U_{jn}^*(x_{11},x_{21}), \quad T_{2n} = \frac{1}{n}\sum_{j=1}^{n} V_{1jn}^*(x_{11},x_{21})$$

$$T_{3n} = \frac{1}{n}\sum_{j=1}^{n} V_{2jn}^*(x_{11},x_{21}), \quad T_{4n} = \frac{1}{n}\sum_{j=1}^{n} U_{jn}^*(x_{12},x_{22})$$

$$T_{5n} = \frac{1}{n}\sum_{j=1}^{n} V_{1jn}^*(x_{12},x_{22}), \quad T_{6n} = \frac{1}{n}\sum_{j=1}^{n} V_{2jn}^*(x_{12},x_{22})$$

$$\theta_1 = g(x_{11},x_{21}), \quad \theta_2 = w_1(x_{11},x_{21}), \quad \theta_3 = w_2(x_{11},x_{21})$$

$$\theta_4 = g(x_{12},x_{22}), \quad \theta_5 = w_1(x_{12},x_{22}), \quad \theta_6 = w_2(x_{12},x_{22}).$$

We also define a function H on E_6 into E_4 by

$$H(\underline{y}) = (H_1(\underline{y}),\ H_2(\underline{y}),\ H_3(\underline{y}),\ H_4(\underline{y}))$$

where

$$\underline{y} = (y_1,y_2,y_3,y_4,y_5,y_6)$$

$$H_1(\underline{y}) = \frac{y_2}{y_1}, \quad H_2(\underline{y}) = \frac{y_3}{y_1}, \quad H_3(\underline{y}) = \frac{y_5}{y_4}, \quad H_4(\underline{y}) = \frac{y_6}{y_4}.$$

We have

$$\underline{Z}_n^* = (na_n^2)^{\frac{1}{2}}(\underline{T}_n - \underline{\theta}).$$

We conclude using Lemma 4.13 in conjunction with Theorem (iii) of Rao [30], p.322 (replacing $n^{\frac{1}{2}}$ by $(na_n^2)^{\frac{1}{2}}$) that

$$(na_n^2)^{\frac{1}{2}}(H(\underline{T}_n) - H(\underline{\theta}))$$

$$= (na_n^2)^{\frac{1}{2}}(m_{1n}(x_{11},x_{21}) - m_1(x_{11},x_{21}), m_{2n}(x_{11},x_{21})$$

$$- m_2(x_{11},x_{21}), m_{1n}(x_{12},x_{22}) - m_1(x_{12},x_{22}),$$

$$m_{2n}(x_{12},x_{22}) - m_2(x_{12},x_{22}))'$$

converges in distribution to $\underline{Z}^*$ where $\underline{Z}^*$ is a four dimensional normal random variable with mean vector 0 and covariance matrix HCH' where $H = \left(\left(\frac{\partial H_i(\underline{\theta})}{\partial \theta_j}\right)\right)_{4\times 6}$ is the matrix of partial derivatives of the functions $H_i(\underline{y})$, $i = 1,2,3,4$ with respect to their arguments evaluated at $\underline{\theta}$. It is easy to verify that $HCH' = \Lambda$.

The proof of the theorem is now complete.

Remark 4.3.1. Theorem 4.2 can be generalized to the situation when $\underline{X} = (X_1,X_2,\ldots,X_{p_1})'$ and $\underline{Y} = (Y_1,Y_2,\ldots,Y_{p_2})'$ and the regression function $E[\underline{Y}|\underline{X} = \underline{x}]$ is computed at q distinct points $\underline{x}_\ell = (x_{1\ell}, x_{2\ell},\ldots,x_{p_1\ell})'$, $\ell = 1,2,\ldots,q$. It can be shown that under certain regularity conditions the limiting distribution of the estimated regression function at these points is $p_2\times q$-variate normal. Except for a multiplier the covariance matrix of this limiting distribution is given by

$$\Lambda = \begin{bmatrix} \Lambda_1 & & & 0 \\ & \Lambda_2 & & \\ & & \ddots & \\ 0 & & & \Lambda_q \end{bmatrix}$$

where

$$\Lambda_\ell = ((\lambda_{\ell ij}))_{p_2 \times p_2}$$

$$\lambda_{\ell ij} = \mathrm{cov}(Y_i, Y_j \mid \underline{X} = \underline{x}_\ell)/g(\underline{x}_\ell), \quad \ell = 1,2,\ldots,q;$$
$$i,j = 1,2,\ldots,p_2.$$

To accomplish the proof we need to replace Conditions 4.3.2 and 4.3.7 respectively by

Condition 4.3.8: For arbitrarily small positive η,

$$E\left[|Y_i|^{2+\frac{p_1}{2}+\eta}\right] < \infty, \quad i = 1,2,\ldots,p_2.$$

Condition 4.3.9: $\lim_{n\to\infty} n a_n^{\delta} = \infty$ for $\delta < p_1 + 4$ and

$$\lim_{n\to\infty} n a_n^{p_1+4} = 0.$$

REFERENCES

[1] Ahmad, I.A. and Lin, Pi-erh (1976), "Nonparametric Sequential Estimation of a Multiple Regression Function". Bulletin of Mathematical Statistics, 17, No. 1-2, 63-75.

[2] Ahmad, I.A. and Lin, Pi-erh (1977), "Nonparametric Estimation of a Vector-Valued Bivariate Failure Rate". Annals of Statistics, 5, 1027-1038.

[3] Aitken, C. and MacDonald, D.G. (1979), "An Application of Discrete Kernel Methods to Forensic Odontology". Applied Statistics, 28, 55-61.

[4] Beran, R. (1974), "Asymptotically Efficient Adaptive Rank Estimates in Location Models". Annals of Statistics, 2, 63-74.

[5] Bhattacharya, P.K. (1967), "Estimation of a Probability Density Function and Its Derivatives". Sankhya Series A, 29, 373-382.

[6] Bhattacharya, P.K. (1967), "Efficient Estimation of a Shift Parameter from Grouped Data". Annals of Mathematical Statistics, 38, 1770-1787.

[7] Cacoullos, T. (1966), "Estimation of a Multivariate Density". Annals of the Institute of Statistical Mathematics, 18, 178-189.

[8] Caroll, R.J. (1976), "On Sequential Density Estimation". Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 36, 137-151.

[9] Chernoff, H. (1956), "Large Sample Theory, Parametric Case". Annals of Mathematical Statistics, 27, 1-22.

[10] Copas, J.B. and Fryer, M.J. (1980), "Density Estimation and Suicide Risks in Psychiatric Treatment". Journal of the Royal Statistical Society Series A, 143, 167-176.

[11] Cramer, H. (1946), Mathematical Methods of Statistics. Princeton University Press.

[12] Davies, H.I. (1973), "Strong Consistency of a Sequential Estimator of a Probability Density Function". Bulletin of Mathematical Statistics, 16, No. 3-4, 49-54.

[13] Davies, H.I. and Wegman, E.J. (1975), "Sequential Nonparametric Density Estimation". IEEE Transactions on Information Theory, IT-21, 619-628.

[14] Epanechnikov, V.A. (1969), "Nonparametric Estimates of a Multivariate Probability Density". Theory of Probability and Its Applications, 14, 153-158.

[15] Hodges, J.L. Jr. and Lehmann, E.L. (1963), "Estimation of Location Based on Rank Tests". Annals of Mathematical Statistics, 34, 598-611.

[16] Keifer, J. and Wolfowitz, J. (1958), "On the Deviations of the Empirical Distribution Function of Vector Chance Variables". _Transactions of the American Mathematical Society_, _87_, 173-186.

[17] Korovkin, P.P. (1961), _Inequalities_. Pergamon Press, London.

[18] LeCam, L. (1953), "On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes' Estimates". _University of California Publications in Statistics_, _1_, 277-330.

[19] Lehmann, E.L. (1951), "Consistency and Unbiasedness of Certain Nonparametric Tests". _Annals of Mathematical Statistics_, _22_, 165-179.

[20] Lehmann, E.L. (1963), "Nonparametric Confidence Intervals for a Shift Parameter". _Annals of Mathematical Statistics_, _34_, 1507-1512.

[21] Loeve, M. (1963), _Probability Theory_, 3rd Ed. Van Nostrand, Princeton.

[22] Mann, H.B. and Wald, A. (1942), "On Stochastic Limit and Order Relationships". _Annals of Mathematical Statistics_, _14_, 217-226.

[23] Mann, H.B. and Whitney, D.R. (1947), "On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other". _Annals of Mathematical Statistics_, _18_, 50-60.

[24] Murthy, V.K. (1965), "Estimation of Jumps, Reliability, and Hazard Rate". Annals of Mathematical Statistics, 36, 1032-1040.

[25] Nadaraya, E.A. (1964), "On Estimating Regression". Theory of Probability and Its Applications, 9, 141-142.

[26] Nadaraya, E.A. (1965), "On Nonparametric Estimates of Density Function and Regression Curves". Theory of Probability and Its Applications, 10, 186-190.

[27] Nadaraya, E.A. (1970), "Remarks on Nonparametric Estimates for Density Functions and Regression Curves". Theory of Probability and Its Applications, 15, 134-136.

[28] Nadaraya, E.A. (1974), "On the Integral Mean Square Error of Some Nonparametric Estimates for the Density Function". Theory of Probability and Its Applications, 19, 133-141.

[29] Parzen, E. (1962), "On the Estimation of Probability Density Function and Mode". Annals of Mathematical Statistics, 33, 1065-1076.

[30] Rao, C.R. (1965), Linear Statistical Inference and Its Applications, Wiley, New York.

[31] Renyi, A. (1957), "On the Asymptotic Distribution of the Sum of a Random Number of Independent Random Variables". Acta Mathematica Scientiarum Hungaricae, 8, 193-199.

[32] Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function". Annals of Mathematical Statistics, 27, 832-837.

[33] Rosenblatt, M. (1969), "Conditional Probability Density and Regression Estimates". Multivariate Analysis II (Krishnaiah, P.R., Editor), 25-31.

[34] Rosenblatt, M. (1971), "Curve Estimates". Annals of Mathematical Statistics, 42, 1815-1842.

[35] Rosenblatt, M. (1975), "A Quadratic Measure of Deviation of Two Dimensional Density Estimates and a Test of Independence". Annals of Statistics, 3, 1-14.

[36] Sacks, J. (1975), "An Asymptotically Efficient Sequence of Estimators of a Location Parameter". Annals of Statistics, 3, 285-298.

[37] Samanta, M. (1973), "Nonparametric Estimation of the Mode of a Multivariate Density". South African Statistical Journal, 7, 109-117.

[38] Samanta, M. (1974), "Efficient Nonparametric Estimation of a Shift Parameter". Sankhya Series A, 36, 272-292.

[39] Samanta, M. (1974), "A Note on Uniform Strong Convergence of Bivariate Density Estimates". Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 28, 85-88.

[40] Schucany, W.R. and Sommers, J.P. (1977), "Improvement of Kernel Type Density Estimators". Journal of the American Statistical Association, 72, 420-423.

[41] Schuster, E.F. (1969), "Estimation of a Probability Density Function and its Derivatives". Annals of Mathematical Statistics, 40, 1187-1195.

[42] Schuster, E.F. (1972), "Joint Asymptotic Distribution of the Estimated Regression at a Finite Number of Distince Points". Annals of Mathematical Statistics, 43, 84-88.

[43] Schuster, E.F. and Yakowitz, S. (1979), "Contribution to the Theory of Nonparametric Regression, with Applications to System Identification". Annals of Statistics, 7, 139-145.

[44] Singh, R.S. and Tracy, D.S. (1977), "Strongly Consistent Estimators of k th Order Regression Curves and Rates of Convergence". Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 40, 339-348.

[45] Singh, R.S. (1977), "Improvement on Some Known Nonparametric Uniformly Consistent Estimators of Derivatives of a Density". Annals of Statistics, 5, 394-399.

[46] Singh, R.S. (1977), "Applications of Estimators of Density and its Derivatives to Certain Statistical Problems". Journal of the Royal Statistical

Society Series B, *39*, 359-363.

[47] Singh, R.S. (1979), "On Necessary ond Sufficient Conditions for Uniform Strong Consistency of Estimators of a Density and its Derivatives". *Journal of Multivariate Analysis*, *9*, 157-164.

[48] Srivastava, R.C. (1973), "Estimation of Probability Density Function Based on Random Number of Observations with Applications". *International Statistical Review*, *41*,,77-86.

[49] Stein, C. (1956), "Efficient Nonparametric Testing and Estimation". *Proceedings of the Third Berkely Symposium on Mathematical Statistics and Probability*, *1*, 187-195.

[50] Stone, C.J. (1975), "Adaptive Maximum Likelihood Estimators of a Location Parameter". *Annals of Statistics*, *3*, 267-284.

[51] Titterington, D.M. (1980), "A Comparative Study of Kernel-Based Density Estimates for Categorical Data". *Technometrics*, *22*, 259-268.

[52] van Eeden, C. (1970), "Efficiency-Robust Estimation of Location". *Annals of Mathematical Statistics*, *41*, 172-181.

[53] Van Ryzin, J. (1969), "On Strong Consistency of Density Estimates". *Annals of Mathematical Statistics*, *40*, 1765-1772.

[54] Wald, A. (1949), "Note on the Consistency of Maximum Likelihood Estimate". _Annals of Mathematical Statistics_, _20_, 595-601.

[55] Watson, G.S. (1964), "Smooth Regression Analysis". _Sankhya Series A_, _26_, 359-372.

[56] Weiss, L. and Wolfowitz, J. (1970), "Asymptotically Efficient Nonparametric Estimators of Location and Scale Parameters". _Z. Wahrscheinlichkeitstheorie Verw. Gebiete_, _16_, 134-150.

[57] Wilcoxon, F. (1945), "Individual Comparisons by Ranking Methods". _Biometrics_, _1_, 80-83.

[58] Wolfowitz, J. (1974), "Asymptotically Efficient Nonparametric Estimators of Location and Scale Parameters II". _Z. Wahrscheinlichkeitstheorie Verw. Gebiete_, _30_, 117-128.

[59] Yamato, H. (1971), "Sequential Estimation of a Continuous Probability Density Function and Mode". _Bulletin of Mathematical Statistics_, _14_, No. 3-4, 1-12.

BIOGRAPHICAL DATA

Rujagaata Xavier Mugisha was born in Rujumbura, Kigezi, Uganda, on August 12, 1950. He spent his first fourteen years of education at Kabura, Kahoko, Nyakibale, Butobere, and Ntare Schools all in Western Uganda. In July, 1971, he joined the University of Dar-es-Salaam, Tanzania, from where he graduated with a B.A. degree in statistics in 1974. From April, 1974, to August, 1975, he worked as a statistician for the East African Community. In September, 1975, he joined the Institute of Statistics and Applied Economics at Makerere University, Kampala, where he worked as a special assistant until September, 1976, when he came to Canada. He joined the University of Manitoba in September, 1976, and graduated from there with an M.Sc. degree in statistics in 1978. Since then he has been working for his Ph.D. degree in statistics at the University of Manitoba.