ESTIMATION OF A PROBABILITY DENSITY FUNCTION WITH APPLICATIONS IN STATISTICAL INFERENCE

by

Eugene Francis Schuster

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MATHEMATICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 6 8

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

I hereby recommend that this dissertation prepared under my direction by Eugene Francis Schuster

entitled ESTIMATION OF A PROBABILITY DENSITY FUNCTION WITH APPLICATIONS IN STATISTICAL INFERENCE

be accepted as fulfilling the dissertation requirement of the degree of Doctor of Philosophy

P. K. Bhattacharya	April 5, 1968
Dissertation Director	Date

After inspection of the final copy of the dissertation, the following members of the Final Examination Committee concur in its approval and recommend its acceptance:*

P. K. Bhattacharya	4/5/68
Harvey Cohn	4/5/68
Louise C. Lim	4/15/68
Donald L. Webb	4/15/68
D E Myers	[illegible]

*This approval and acceptance is contingent on the candidate's adequate performance and defense of this dissertation at the final oral examination. The inclusion of this sheet bound into the library copy of the dissertation is evidence of satisfactory performance at the final examination.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Eugene F. Schuster

ACKNOWLEDGEMENTS

The author wishes to take this opportunity to thank the many people who helped him make this paper possible. He is particularly indebted to:

Dr. P. K. Bhattacharya, for his guidance and many suggestions during the research and development of this work;

Dr. L. Lim, for her expression of confidence at a time when this goal seemed unattainable and for her help in the editing of this manuscript;

His family, for their patience, prayers and numerous sacrifices during his many years of study.

This research was supported in part by the National Science Foundation Grant number GP-2141.

TABLE OF CONTENTS

Page

ABSTRACT. v

CHAPTER

1. INTRODUCTION 1

2. SURVEY 5

3. ESTIMATION OF A PROBABILITY DENSITY FUNCTION AND ITS DERIVATIVES. 11

4. UNIFORM CONVERGENCE OF ESTIMATING PROBABILITY DENSITY FUNCTIONS. 19

5. APPLICATIONS 32
   Estimation of a Distribution Function. 32
   Estimation of the Hazard Rate. 36

6. ESTIMATION OF A REGRESSION FUNCTION. 42
   Three Estimates and Their Motivation 43
   Bounds for the Supremum Distance on a Finite Interval. . . . 45
   Estimation of the Point at which a Regression Function attains its Maximum. 59
   Joint Asymptotic Distribution of the Estimated Regression Function at a Finite Number of Distinct Points 61

REFERENCES. 72

ABSTRACT

Let $X_1, X_2, \ldots$ be independent identically distributed random variables having a common probability density function f. After a so-called kernel class of estimates f_n of f based on $(X_1, \ldots, X_n)$ was introduced by Rosenblatt (1956), various convergence properties of these estimates have been studied. The strongest result in this direction was due to Nadaraya (1965) who proved that if f is uniformly continuous then for a large class of kernels the estimates f_n converge uniformly on the real line to f with probability one. For a very general class of kernels, we will show that the above assumptions on f are necessary for this type of convergence. That is, if f_n converges uniformly to a function g with probability one, then g must be uniformly continuous and the distribution F from which we are sampling must be absolutely continuous with $F'(x) = g(x)$ everywhere.

When in addition to the conditions mentioned above, it is assumed that f and its first $r + 1$ derivatives are bounded, we are able to show how to construct estimates f_n such that $f_n^{(s)}$ converges uniformly to $f^{(s)}$ at a given rate with probability one for $s = 0, 1, \ldots, r$.

Several applications of the density estimates are considered, the main ones being the proposed estimates of a regression function

which arise quite naturally from the kernel estimates of a bivariate density. Furthermore, various convergence properties of these regression estimates are studied.

CHAPTER 1

INTRODUCTION

In this paper we shall investigate the asymptotic properties of the so-called kernel class of estimates of a probability density function and apply these estimates to some problems in statistical inference. Before stating this problem precisely we will list a few definitions and indicate some notation to be used.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a random variable defined on Ω.

Definition 1.1. The distribution function F corresponding to the random variable X is defined by $F(x) = P\{X \leq x\} = P\{\omega \varepsilon \Omega | X(\omega) \leq x\}$ for all real x.

Definition 1.2. A distribution function F is said to be absolutely continuous if there exists a Borel measurable function f over $(-\infty, \infty)$ such that

$$F(x) = (L) \int_{-\infty}^{x} f(u)du \tag{1}$$

for all real x. The function f is called a probability density function of F.

In this paper $\int_a^b g(u)du$ will be understood to be a Lebesgue integral and $\int_a^b g(u)dF(u)$ the Lebesgue-Stieltjes integral of g with respect to the probability distribution P determined by the distribution function F. The symbol (L) before an integral (as in (1) above) will be used at times to emphasize that the integral is to be understood in the Lebesgue sense while the symbols (R) or (L-S) before an integral will mean the integral is a Riemann or a Lebesgue-Stieltjes integral respectively. Also whenever the integration extends over $(-\infty, \infty)$ no limits of integration will be given.

Definition 1.3. A random variable is called absolutely continuous if it has an absolutely continuous distribution function. In addition, we say that X has probability density f when the distribution function F of X is given by (1) above.

Definition 1.4. A probability density is a non-negative Borel measurable function g with $\int g(u)du = 1$.

If $(\Omega, \mathcal{A}, P)$ is a probability space and X is a random variable defined on Ω then for any Borel measurable function h and Borel set B, we will use the following notation:

$$E_F[h(X)] = (L\text{-}S) \int h(u)dF(u)$$

and

$$P_F\{h(X)\varepsilon B\} = P\{\omega \ \varepsilon \ \Omega | h[X(\omega)]\varepsilon B\}.$$

If F has density f then $E_f[h(X)]$ and $P_f\{h(X)\varepsilon B\}$ may also be used to denote $\int h(u)dF(u)$ and $P\{\omega \ \varepsilon \ \Omega | h[X(\omega)]\varepsilon B\}$. Whenever it is clear from the context that X has distribution function F or density f the subscripts on E and P may be dropped.

Let $X_1, X_2, \ldots,$ be independent identically distributed random variables having a common probability density function f. Functions of the form

$$f_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} k\left(\frac{x-X_i}{a_n}\right) \tag{2}$$

for all real x, where k is any probability density function and $\{a_n\}$ is a sequence of positive numbers converging to zero, will be called kernel estimates of $f(x)$.

We shall investigate the asymptotic properties of the above kernel estimates and apply these estimates to some problems in non-parametric statistical inference.

The dissertation is divided into four parts. Chapter 2 is devoted to a survey of relevant papers in the literature. In Chapter 3 we examine the convergence of the random variables

$$\sup_x b_n |f_n^{(r)}(x) - f^{(r)}(x)|$$

for appropriate b_n with $b_n \to \infty$, where for any function g, $g^{(0)} = g$

and $g^{(r)}$ denotes the r^{th} derivative of g.

In Chapter 4 a necessary and sufficient condition for the uniform convergence of f_n on the real line with probability one is given.

The final two chapters are concerned with applications of the kernel estimates. Our main application is the estimate of a regression function considered in Chapter 6.

CHAPTER 2

SURVEY

The problem of estimating the density function f of a random variable X has recently begun to receive attention in the literature. Fix and Hodges [5] have considered estimates of the form

$$S_n(x) = \frac{F_n(x+h_n) - F_n(x-h_n)}{2h_n}$$

where F_n is the empirical distribution function based on a random sample $(X_1, \ldots, X_n)$ from f and h_n is a sequence of numbers which approach zero as n tends to infinity. Rosenblatt [15] proved that if $(X_1, \ldots, X_n)$ is a random sample of size n from f, then any non-negative Borel estimate $S(x; X_1, \ldots, X_n)$ of $f(x)$ must be biased, i.e., there exist an x and a continuous density function g such that $E_g S(x; X_1, \ldots, X_n) \neq g(x)$. Rosenblatt also noted that the estimates studied by Fix and Hodges were members of a more general class of estimates of $f(x)$ of the form

$$S_n(x) = \frac{1}{n} \sum_{i=1}^{n} w_n(x-X_i)$$

where w_n is a probability density function for each n. If $w_n(u) = k(u/a_n)a_n$ where a_n is a sequence of positive numbers converging to zero and k is a probability density function, then

$$(1) \qquad S_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} k\left(\frac{x-X_i}{a_n}\right)$$

is a kernel estimate of $f(x)$ as defined in Chapter 1. In the following $f_n(x)$ will denote a kernel estimate of $f(x)$ as given in (1) above.

For the case when $\int u\, k(u)du = 0$, $\int |u|^3 k(u)du$ is finite, and f has continuous derivatives of the first three orders, Rosenblatt showed that for fixed x, the sequence $\{f_n(x)\}$ is an asymptotically unbiased estimate of $f(x)$; that is, $\lim_{n\to\infty} E[f_n(x)] = f(x)$, with $E[f_n(x)] - f(x) = O(a_n^2)$. In addition he showed that the sequence $\{f_n(x)\}$ is consistent in quadratic mean for $f(x)$; in other words, $\lim_{n\to\infty} E[f_n(x) - f(x)]^2 = 0$, with $E[f_n(x) - f(x)]^2$ no smaller than $O(n^{-4/5})$.

Parzen [13] established these last two results under somewhat weaker conditions and also determined conditions under which $f_n(x)$ is asymptotically normal when suitably normalized. In addition he proved a much stronger result, namely, that if $f(x)$ is uniformly continuous and $\lim_{n\to\infty} na_n^2 = \infty$, then for a large class of kernels k, the kernel estimates of the form (1) are uniformly consistent in the sense that for every $\varepsilon > 0$

$$\lim_{n\to\infty} P_f\{\sup_x |f_n(x) - f(x)| < \varepsilon\} = 1.$$

The uniform convergence of f_n to f was also studied by Nadaraya [11] who proved that if k is of bounded variation, $\sum_{n=1}^{\infty} \exp(-cna_n^2)$ is finite for all positive c, and $f(x)$ is uniformly continuous, then for an estimate $f_n(x)$ of $f(x)$ of the form (1), $\sup_x |f_n(x) - f(x)|$ converges to zero with probability one. In her proof $\sup_x |f_n(x) - Ef_n(x)|$ was tacitly assumed to be a measurable function; however if the measure P^∞ corresponding to the distribution of the infinite sequence $(X_1, X_2, \ldots)$ is completed, then her result remains valid whether or not $\sup_x |f_n(x) - f(x)|$ is measurable for each fixed n.

Under somewhat stronger assumptions than Parzen's, Bhattacharya [1] showed for a large class of kernels k, how to choose the sequences $\{a_n\}$ and $\{b_n\}$ such that $\sup_{|x|\le b_n} n^c |f_n^{(r)}(x) - f^{(r)}(x)|$ converges to zero with probability one for appropriate positive c (depending on r).

Cacoullos [2] noted that Parzen's results can be adapted in an obvious manner to provide estimates of a multivariate density. The consistency, asymptotic unbiasedness, and the mean square error of the multivariate estimate f_n follow by using straightforward modifications of Parzen's methods. With respect to asymptotic normality

Cacoullos obtained a stronger result than Parzen's; namely, he proved the joint asymptotic normality of the estimates f_n at continuity points of f.

Several authors have considered slight modifications of the kernel estimates (1) of $f(x)$. Bhattacharya [1] has considered estimates of the form

$$f_n^*(x) = \begin{cases} k(x) & \text{when } \sum_{i=1}^{n} I_n(X_i) = 0 \\ \sum_{i=1}^{n} \left[I_n(X_i) \cdot \frac{1}{a_n} k\left(\frac{x-X_i}{a_n}\right)\right] \Big/ \sum_{i=1}^{n} I_n(X_i) & \text{when } \sum_{i=1}^{n} I_n(X_i) > 0 \end{cases}$$

where a_n and k are as in (1) and I_n is the indicator function of $[-b_n, b_n]$, with b_n tending to infinity. He has given sufficient conditions for the convergence of $\sup_x |f_n^*(x) - f(x)|$ to zero with probability one.

Watson and Leadbetter [16] have considered estimates of the form

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_n} k_n\left(\frac{x-X_i}{a_n}\right)$$

and discussed the properties of such estimates on the basis of their mean integrated square errors $E\left[\int \{\hat{f}_n(x) - f(x)\}^2 dx\right]$ (abbreviated MISE). They have shown that the MISE is a minimum if the Fourier transform of the kernel k_n denoted by ϕ_{k_n} is

$$\Phi_{k_n}(t) = n|\Psi_f(t)|^2 \Big/ \{1+(n-1)|\Psi_f(t)|^2\}$$

where Ψ_f is the characteristic function of f which is assumed to be in L_2. By making suitable assumptions on the asymptotic behavior of Ψ_f, they have been able to calculate the maximum rate of convergence of the MISE. They have shown in any case that the MISE cannot decrease faster than n^{-1}.

The asymptotic behavior of the maximum deviation of the estimated density function $f(x)$ from estimates of the form

$$\tilde{f}_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} g\left(x, \frac{x-X_i}{a_n}\right),$$

where g is a suitably chosen non-negative function defined on R^2, has been studied by Woodroofe [17]. His main result gives sufficient conditions for the convergence of

$$\max_{|x|\leq 1} \left(\frac{na_n}{-2 \log a_n}\right)^{1/2} \frac{|\tilde{f}_n(x)-f(x)|}{||g_x||_2 f(x)}$$

to one in probability where $||g_x||_2^2 = \int g^2(x,y)dy$.

Let $\mathcal{B}$ be the class of Borel sets of a Hausdorff topological group $(G, \mathcal{F})$ and u a left and right Haar measure on the measure space $(G, \mathcal{B})$. Let X be a random variable on $(G, \mathcal{B})$ with probability distribution P which is absolutely continuous with respect to

u. Craswell [3] has noted that Parzen's results can be used to construct extimates of g.

When this work began the strongest results in the direction of this paper were given by Nadaraya's theorem [11] on the convergence of $\sup_x |f_n(x) - f(x)|$ to zero with probability one and Bhattacharya's theorem [1] on the convergence of $\sup_{|x| \leq b_n} n^c |f_n^{(r)}(x) - f^{(r)}(x)|$ to zero with probability one referred to above. Both these results are proved by first obtaining exponential bounds for the probability that the supremum distances are greater than a given positive number ε.

For a very general class of kernels we shall show that the assumptions on f made by Nadaraya in proving that $\sup_x |f_n(x) - f(x)|$ converges to zero with probability one are necessary for this type of convergence. That is, we will show that if $\sup_x |f_n(x) - g(x)|$ converges to zero with probability one for some function g, then g must be uniformly continuous and the distribution function F from which we are sampling must be absolutely continuous with $F'(x) = g(x)$ everywhere (F' being the ordinary derivative of F). Thus combining Nadaraya's theorem with the result obtained here, we have a necessary and sufficient condition for the uniform convergence of f_n with probability one.

With the same assumptions on f and k made by Bhattacharya we shall show how to strengthen his conclusion by proving $\sup_x n^c |f_n^{(r)}(x) - f^{(r)}(x)|$ converges to zero with probability one. The methods employed are different from Bhattacharya's.

CHAPTER 3

ESTIMATION OF A PROBABILITY DENSITY FUNCTION AND ITS DERIVATIVES

Let $X_1, \ldots, X_n$ be independent identically distributed random variables with a common distribution function F. Let F_n be the empirical distribution function based on $(X_1, \ldots, X_n)$; i.e., $nF_n(x)$ is the number of X_i with $X_i \leq x$ where $1 \leq i \leq n$.

Lemma 3.1. There exists a universal constant C such that for any $n>0$, $\varepsilon_n>0$ and distribution function F,

$$(1) \qquad P_F \{\sup_x |F_n(x) - F(x)| > \varepsilon_n\} \leq C \exp(-2n\varepsilon_n^2).$$

Proof: For the case when F is continuous, see Dvoretzky, Kiefer and Wolfowitz [4]. If F is discontinuous at some point then there exists a continuous distribution function $\bar{F}$ for which

$$P_F \{\sup_x |F_n(x) - F(x)| > \varepsilon_n\} \leq P_{\bar{F}} \{\sup_x |F_n(x) - \bar{F}(x)| > \varepsilon_n\}$$

(see [7] and [8]). Thus the lemma is true for all univariate F.

Let $f_n(x)$ be a kernel estimate based on $(X_1, \ldots, X_n)$ from F as defined in Chapter 1 with kernel k chosen such that $\int |u| k(u)du$ is finite, and such that $k^{(s)}$ is a continuous function of bounded variation for $s = 0, 1, \ldots, r$. The density function of the standard normal, for example, satisfies all these conditions. The variation of $k^{(s)}$ on $(-\infty, \infty)$ will be denoted by μ_s. In the following lemma we do not require that the X_i's be absolutely continuous or that $\int |u| k(u)du$ be finite. The continuity assumption on $k^{(r)}$ was made solely to ensure that $\sup_x |f_n^{(r)}(x) - Ef_n^{(r)}(x)|$ is a random variable. With the deletion of this assumption the following lemma remains true when we replace the probability P_F by the outer probability P_F^* of P_F. Our proof remains valid in this case.

Lemma 3.2. There exists a universal constant C such that for any $n>0$, $\varepsilon_n>0$, and distribution function F,

$$P_F \{\sup_x |f_n^{(r)}(x) - Ef_n^{(r)}(x)| > \varepsilon_n\} \leq C \exp(-2n\varepsilon_n^2 a_n^{2r+2}/\mu_r^2)$$

where $\{a_n\}$ is a sequence of positive numbers converging to zero and

$$f_n^{(r)}(x) = \frac{1}{na_n^{r+1}} \sum_{i=1}^{n} k^{(r)}\left(\frac{x-X_i}{a_n}\right).$$

Proof: Since $k^{(r)}$ is of bounded variation on $(-\infty, \infty)$, we know (see [12], page 239) that $k^{(r)}$ is bounded and that $\lim_{x\to\infty} k^{(r)}(x)$ and $\lim_{x\to-\infty} k^{(r)}(x)$ both exist. If $r = 0$ then k is non-negative and $\int k(u)du = 1$, so that since $\lim_{x\to\infty} k(x)$ and $\lim_{x\to-\infty} k(x)$ both exist, these limits must be zero. If $r \geq 1$ then the function $k^{(r-1)}$ has a bounded derivative on $[-a, a]$ for any a, and hence (see [12], page 133) $k^{(r)}$ is Lebesgue integrable on $[-a, a]$. Thus (see [12], page 259)

$$\overset{a}{\underset{-a}{V}} [k^{(r-1)}] = \int_{-a}^{a} |k^{(r)}(u)|du.$$

Now

$$\overset{\infty}{\underset{-\infty}{V}} [k^{(r-1)}] = \lim_{a\to\infty} \overset{a}{\underset{-a}{V}} [k^{(r-1)}]$$

$$= \lim_{a\to\infty} \int_{-a}^{a} |k^{(r)}(u)|du = \int_{-\infty}^{\infty} |k^{(r)}(u)|du$$

so that $\int |k^{(r)}(u)|du$ is finite. This fact together with the existence of $\lim\limits_{x\to\infty} k^{(r)}(x)$ and $\lim\limits_{x\to-\infty} k^{(r)}(x)$ imply that these limits must be zero.

Upon integrating by parts and remembering that $\lim\limits_{x\to\infty} k^{(r)}(x) = \lim\limits_{x\to-\infty} k^{(r)}(x) = 0$, we find that

$$\sup_x |f_n^{(r)}(x) - Ef_n^{(r)}(x)|$$

$$= \sup_x \frac{1}{a_n^{r+1}} \left| \int k^{(r)}\left(\frac{x-u}{a_n}\right) dF_n(u) - \int k^{(r)}\left(\frac{x-u}{a_n}\right) dF(u) \right|$$

$$= \frac{1}{a_n^{r+1}} \sup_x \left| [\{F_n(u) - F(u)\} k^{(r)}\left(\frac{x-u}{a_n}\right)]_{-\infty}^{\infty} - \int \{F_n(u) - F(u)\} dk^{(r)}\left(\frac{x-u}{a_n}\right) \right|$$

$$= \frac{1}{a_n^{r+1}} \sup_x \left| \int \{F_n(u) - F(u)\} dk^{(r)}\left(\frac{x-u}{a_n}\right) \right|$$

$$\leq \frac{1}{a_n^{r+1}} \sup_x |F_n(x) - F(x)| \mu_r .$$

Therefore by an application of Lemma 3.1 we have

$$P_F \{\sup_x |f_n^{(r)}(x) - Ef_n^{(r)}(x)| > \varepsilon_n\}$$

$$\leq P_F \{\sup_x |F_n(x) - F(x)| > \frac{\varepsilon_n a_n^{r+1}}{\mu_r}\} \leq C \exp(-2n\varepsilon_n^2 a_n^{2r+2}/\mu_r^2)$$

and the proof is complete.

Lemma 3.3 below is found in [1]; however, we note here that the symmetry condition imposed on k in [1] is not needed and that in the proof given there the absolute integrability of the $k^{(s)}$, $s = 1, 2, \ldots, r$, has been tacitly assumed. From the proof of

Lemma 3.2, the $k^{(s)}$ tend to zero as $x \to +\infty$ or $-\infty$ and $\int |k^{(s)}(u)|du$ is finite for $s = 0, 1, \ldots, r$, so that the proof of Lemma 3.3 can be completed exactly as in [1].

Lemma 3.3. Let X be an absolutely continuous random variable with probability density function f and let a be any positive real number. If f and its first r+1 derivatives are bounded then there exists a constant C, not depending on a, such that

$$\sup_x \left| E_f\left[\frac{1}{a^{r+1}} k^{(r)}\left(\frac{x-X}{a}\right)\right] - f^{(r)}(x)\right| \leq Ca.$$

Lemma 3.4. If f and its first r+1 derivatives are bounded and if $\{\varepsilon_n\}$ is a sequence of positive numbers such that $a_n = o(\varepsilon_n)$ as $n\to\infty$ where $f_n^{(r)}(x) = \frac{1}{na_n^{r+1}} \sum_{i=1}^{n} k^{(r)}\left(\frac{x-X_i}{a_n}\right)$, then there exist positive constants C_1 and C_2 such that

$$P_f \{\sup_x |f_n^{(r)}(x) - f^{(r)}(x)| > \varepsilon_n\} \leq C_1 \exp(-C_2 n\varepsilon_n^2 a_n^{2r+2})$$

for n sufficiently large.

Proof: We have with the aid of Lemma 3.3

$$\sup_x |f_n^{(r)}(x) - f^{(r)}(x)|$$

$$\leq \sup_x |f_n^{(r)}(x) - Ef_n^{(r)}(x)| + \sup_x |Ef_n^{(r)}(x) - f^{(r)}(x)|$$

$$\leq \sup_x |f_n^{(r)}(x) - Ef_n^{(r)}(x)| + Ca_n.$$

Since $a_n = o(\varepsilon_n)$ it follows that for n sufficiently large

$$P\{\sup_x |f_n^{(r)}(x) - f^{(r)}(x)| > \varepsilon_n\} \leq P\{\sup_x |f_n^{(r)}(x) - Ef_n^{(r)}(x)| > \varepsilon_n/2\}$$

An application of Lemma 3.2 yields the desired result.

The theorem below tells us that for special sequences $\{a_n\}$, $\sup_x |f_n^{(r)}(x) - f^{(r)}(x)|$ converges to zero with probability one. A sequence $\{b_n\}$ with b_n going to infinity is introduced to indicate the rate at which the above convergence takes place.

Theorem 3.5. If f and its first $r+1$ derivatives are bounded and if the sequences $\{a_n\}$ and $\{b_n\}$ are such that $a_n b_n = o(1)$ and $\sum_{n=1}^{\infty} \exp(-cna_n^{2r+2}/b_n^2)$ is finite for all positive c, then

$$\lim_{n\to\infty} \sup_x b_n |f_n^{(r)}(x) - f^{(r)}(x)| = 0$$

with probability one.

Proof: For any $\varepsilon > 0$, we obtain by Lemma 3.4 that

$$P_F\ \{\sup_x |f_n^{(r)}(x) - f^{(r)}(x)| > \frac{\varepsilon}{b_n}\} \leq C_1\ \exp(-C_2\varepsilon^2 n a_n^{2r+2}/b_n^2)$$

for n sufficiently large. Since $\sum_{n=1}^{\infty} \exp(-cna_n^{2r+2}/b_n^2)$ is finite for all positive c, it follows that

$$\sum_{n=1}^{\infty} P\ \{\sup_x |f_n^{(r)}(x) - f^{(r)}(x)| > \frac{\varepsilon}{b_n}\}$$

is finite for all positive ε. Consequently, with the aid of the Borel-Cantelli Lemma we see that $\lim_{n\to\infty} \sup_x b_n |f_n^{(r)}(x) - f^{(r)}(x)| = 0$ with probability one.

It can be seen that the assumption that $f^{(r+1)}$ be bounded could be relaxed somewhat and the conclusion would still hold. The fact that $f^{(r+1)}$ is bounded was used in Lemma 3.3 in [1] to ensure that $\sup_x |Ef_n^{(r)}(x) - f^{(r)}(x)| = O(a_n)$. To establish $\lim_{n\to\infty} \sup_x |f_n^{(r)}(x) - f^{(r)}(x)| = 0$ with probability one following the lines of our argument we would only need $\sup_x |Ef_n^{(r)}(x) - f^{(r)}(x)| = o(1)$ which would be true, for instance, if $f^{(r)}$ were uniformly continuous.

A corollary follows which will indicate the rate of convergence for a particular choice of a_n.

Corollary 3.6. If f and its first $r+1$ derivatives are bounded, $a_n = n^{-1/(2r+4)}$ and $0<c<1/(2r+4)$, then

$$\lim_{n\to\infty} \sup_{x} n^{c}\left|f_n^{(r)}(x) - f^{(r)}(x)\right| = 0$$

with probability one.

CHAPTER 4

UNIFORM CONVERGENCE OF ESTIMATING PROBABILITY DENSITY FUNCTIONS

Let $f_n(x)$ be a kernel estimate based on a random sample $(X_1, \ldots, X_n)$ from F as defined in Chapter 1, i.e.,

$$f_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} k\left(\frac{x - X_i}{a_n}\right).$$

In this chapter we shall assume that the sequence $\{a_n\}$ is such that $\sum_{n=1}^{\infty} \exp(-cna_n^2)$ is finite for all positive c and that k is a probability density function satisfying the following conditions:

(i) k is continuous and of bounded variation on $(-\infty, \infty)$.

(ii) $uk(u) \to 0$ as $u \to +\infty$ or $-\infty$.

(iii) There exists a δ in $(0, 1)$ such that

$$u\left(\overset{-u^\delta}{\underset{-\infty}{V}}(k) + \overset{\infty}{\underset{u^\delta}{V}}(k)\right) \to 0 \quad \text{as} \quad u \to \infty.$$

(iv) $\int |u| dk(u)$, the integral of $|u|$ with respect to the signed measure determined by k, is finite.

For example, the density function of any normal or Cauchy distribution satisfies these conditions. Lemmas 4.1 through 4.10 below hold for any distribution function F.

Lemma 4.1. For any distribution function F,

$$\lim_{n\to\infty} \sup_x |f_n(x) - Ef_n(x)| = 0$$

with probability one.

Proof: By Lemma 3.2

$$P_F \{\sup_x |f_n(x) - Ef_n(x)| > \varepsilon\} \leq C \exp(-\alpha n a_n^2)$$

where $\alpha = 2\varepsilon^2/\mu^2$ and $\mu = \overset{\infty}{\underset{-\infty}{V}} (K)$. Since $\sum_{n=1}^{\infty} c \exp(-\alpha n a_n^2)$ is finite, it follows that $\sum_{n=1}^{\infty} P_F \{\sup_x |f_n(x) - Ef_n(x)| > \varepsilon\}$ is finite, and the proof is complete in view of the Borel-Cantelli Lemma.

We note here that this lemma was proved in [11] for continuous distribution functions F. We have extended this lemma to arbitrary F by using Lemma 3.1 to establish inequality 5 on page 187 of [11].

Lemma 4.2. In order for $\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$ with probability one for some function g, it is necessary and sufficient that $\lim_{n\to\infty} \sup_x |Ef_n(x) - g(x)| = 0$.

Proof: This result follows directly from Lemma 4.1 in conjunction with the following inequalities:

$$\sup_x |f_n(x) - g(x)| \leq \sup_x |f_n(x) - Ef_n(x)| + \sup_x |Ef_n(x) - g(x)|$$

and

$$\sup_x |Ef_n(x) - g(x)| \leq \sup_x |f_n(x) - Ef_n(x)| + \sup_x |f_n(x) - g(x)|.$$

<u>Lemma 4.3</u>. If $\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$ with probability one for some function g, then g is uniformly continuous.

Proof: For any $\varepsilon > 0$ there exists by Lemma 4.2 an $M = M(\varepsilon)$ such that $\sup_x |Ef_n(x) - g(x)| < \varepsilon/4$ for $n \geq M$. Conditions (i) and (ii) on k imply that k is uniformly continuous, so that given $\varepsilon' > 0$ there exists a δ' such that $|k(x) - k(y)| < \varepsilon'$ whenever $|x - y| < \delta'$. With $\varepsilon' = \frac{\varepsilon}{2} a_m$ we define δ to be $\delta' a_m$ so that whenever $|x - y| < \delta$, we shall have

$$|g(x) - g(y)| \leq |g(x) - Ef_M(x)| + |Ef_M(x) - Ef_M(y)| + |Ef_M(y) - g(y)|$$

$$\leq |Ef_M(x) - Ef_M(y)| + 2 \sup_x |Ef_M(x) - g(x)|$$

$$\leq |Ef_M(x) - Ef_M(y)| + \frac{\varepsilon}{2}$$

$$= \left|\int \frac{1}{a_M}\{k(\frac{x-u}{a_M}) - k(\frac{y-u}{a_M})\}dF(u)\right| + \frac{\varepsilon}{2}$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Lemma 4.4. If $\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$ with probability one for some function g, then $\lambda\{x\varepsilon(-\infty,\infty)|F'(x) \neq g(x)\} = 0$ (λ represents the Lebesgue measure on the real line).

Proof: Suppose x is a point where $F'(x)$ exists. Using integration by parts we see that

$$Ef_n(x) = \int \frac{1}{a_n} k(\frac{x-u}{a_n})dF(u)$$

$$= -\int \frac{1}{a_n} F(u)dk(\frac{x-u}{a_n})$$

$$= \int \frac{1}{a_n} F(x-a_n u)dk(u)$$

$$= \int \frac{1}{a_n} F(x-a_n u)dk(u) - \frac{F(x)}{a_n}\int dk(u)$$

$$= \int \frac{F(x-a_n u) - F(x)}{a_n} dk(u), \quad \text{since} \int dk(u) = 0.$$

Let δ be such that (iii) holds. Then we may write

$$Ef_n(x) = \int_{|u|>a_n^{-\delta}} \frac{F(x-a_n u) - F(x)}{a_n} dk(u)$$

$$+ \int I_n(u) \frac{F(x-a_n u) - F(x)}{-a_n u} (-u)dk(u)$$

where I_n is the indicator function of $[-a_n^{-\delta}, a_n^{-\delta}]$. We observe by (iii) that

$$\lim_{n\to\infty} \left|\int_{|u|>a_n^{-\delta}} \frac{F(x-a_n u) - F(x)}{a_n} dk(u)\right|$$

$$\leq \lim_{n\to\infty} \frac{2}{a_n}\left(\overset{-a_n^{-\delta}}{\underset{-\infty}{V}} (k) + \overset{\infty}{\underset{a_n^{-\delta}}{V}} (k)\right) = 0.$$

Also, given $\varepsilon > 0$ there exists an $N = N(\varepsilon, x)$ such that for $n > N$

$$\left|I_n(u) \frac{F(x-a_n u) - F(x)}{-a_n u}\right| \leq F'(x) + \varepsilon.$$

By condition (iv) on k we have that $\int (F'(x) + \varepsilon)|u|dk(u)$ is finite. Thus Lebesgue's Dominated Convergence Theorem for signed measures applies and hence

$$\lim_{n\to\infty} \int I_n(u) \frac{F(x-a_n u) - F(x)}{-a_n u} (-u)dk(u)$$

$$= \int \lim_{n\to\infty} I_n(u) \frac{F(x-a_n u) - F(x)}{-a_n} (-u)dk(u)$$

$$= \int F'(x) \ (-u)dk(u)$$

$$= F'(x)$$

since $\int (-u)dk(u) = 1$. Therefore $\lim_{n\to\infty} Ef_n(x) = F'(x)$ whenever $F'(x)$ exists. By Lemma 4.2, $\lim_{n\to\infty} Ef_n(x) = g(x)$ everywhere and hence $F'(x) = g(x)$ whenever $F'(x)$ exists. Since it is well known that the derivative of a monotone function exists almost everywhere, this completes the proof.

Lemma 4.5. If $\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$ with probability one for some function g, then $\int g(u)du \leq 1$.

Proof: Let $F(x) = F_{AC}(x) + F_S(x) + F_D(x)$ where F_{AC}, F_S and F_D denote the absolutely continuous, the singular, and the discrete part of F respectively. Now $F'(x) = F'_{AC}(x)$ almost everywhere, and $F'(x) = g(x)$ almost everywhere by Lemma 4.4, so that $F'_{AC}(x) = g(x)$ almost everywhere. Thus

$$F_{AC}(x) = \int_{-\infty}^{x} F'_{AC}(u)du = \int_{-\infty}^{x} g(u)du$$

which implies $\int g(u)du \leq 1$ since $\lim_{x\to\infty} F_{AC}(x)$ exists and is less than or equal to one.

Lemma 4.6. If $\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$ with probability one for some function g, then

$$\lim_{n\to\infty} \sup_x \int \frac{1}{a_n} k(\frac{x-u}{a_n}) d(F_S(u) + F_D(u)) = 0.$$

Proof: In the proof of Lemma 4.5 we have shown that $F'_{AC}(x) = g(x)$ almost everywhere. Consequently

$$F_{AC}(x) - F_{AC}(a) = (L) \int_a^x F'_{AC}(u)du$$

$$= (L) \int_a^x g(u)du$$

$$= (R) \int_a^x g(u)du$$

since g is uniformly continuous on $[a, x]$ by Lemma 4.3. So $F'_{AC}(x) = g(x)$ by the Fundamental Theorem of Calculus for Riemann integrals.

Lemma 4.7. If $\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$ with probability one for some function g, then

$$\lim_{n\to\infty} \sup_x \int \frac{1}{a_n} k(\frac{x-u}{a_n}) d(F_S(u) + F_D(u)) = 0.$$

Proof: From $Ef_n(x) = \int \frac{1}{a_n} k(\frac{x-u}{a_n}) dF(u)$ we obtain

$$Ef_n(x) - \int \frac{1}{a_n} k(\frac{x-u}{a_n}) dF_{AC}(u) = \int \frac{1}{a_n} k(\frac{x-u}{a_n}) d(F_S(u) + F_D(u)).$$

So for $\delta > 0$ we have with the aid of Lemma 4.6

$$0 \leq \int \frac{1}{a_n} k(\frac{x-u}{a_n}) d(F_S(u) + F_D(u))$$

$$\leq |Ef_n(x) - g(x)| + |g(x) - \int \frac{1}{a_n} k(\frac{x-u}{a_n}) dF_{AC}(u)|$$

$$= |Ef_n(x) - g(x)| + |g(x) - \int \frac{1}{a_n} k(\frac{x-u}{a_n}) g(u) du|$$

$$= |Ef_n(x) - g(x)| + | \int \{g(x) - g(x-u)\} \frac{1}{a_n} k(\frac{u}{a_n}) du|$$

$$\le |Ef_n(x) - g(x)| + \int_{|u|<\delta} |g(x) - g(x-u)|\frac{1}{a_n} k(\frac{u}{a_n})du$$

$$+ \int_{|u|\ge\delta} |g(x) - g(x-u)|\frac{1}{a_n} k(\frac{u}{a_n})du$$

$$\le |Ef_n(x) - g(x)| + \sup_{|u|<\delta} |g(x) - g(x-u)| + 2 \sup_x g(x) \cdot \int_{|u|\ge\delta/a_n} k(u)du.$$

It follows that

$$\text{(1)} \qquad \sup_x \int \frac{1}{a_n} k(\frac{x-u}{a_n})d(F_S(u) + F_D(u))$$

$$\le \sup_x |Ef_n(x) - g(x)| + \sup_x \sup_{|u|<\delta} |g(x) - g(x-u)|$$

$$+ 2 \sup_x g(x) \cdot \int_{|u|\ge\delta/a_n} k(u)du.$$

In view of Lemmas 4.3, 4.5 and 4.6, g is uniformly continuous and non-negative and $\int g(u)du$ is finite, whence g is bounded.

Let $\varepsilon > 0$ be given. Since g is uniformly continuous we can choose δ so small that the second term on the right side of (1) is less than $\varepsilon/3$. Having so chosen δ we can now choose N so large that if $n \ge N$, then the remaining terms on the right side of (1) will each be less than $\varepsilon/3$, since the first term tends to 0 by Lemma 4.2 and the last term goes to zero for any fixed $\delta > 0$. The desired conclusion now follows.

Lemma 4.8. If $\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$ with probability one for some function g, then $F_D(x) = 0$ for all x.

Proof: Suppose there exists an x_o such that $F_D(x_o) - F_D(x_o-0) > 0$. Then

$$\int \frac{1}{a_n} k\left(\frac{x-u}{a_n}\right) dF_D(u) \geq \frac{1}{a_n} k\left(\frac{x-x_o}{a_n}\right)\{F_D(x_o) - F_D(x_o-0)\}.$$

If c is such that $k(c) > 0$ and $x_n = ca_n + x_o$ then

$$\sup_x \int \frac{1}{a_n} k\left(\frac{x-u}{a_n}\right) dF_D(u) \geq \int_{-\infty}^{\infty} \frac{1}{a_n} k\left(\frac{x_n-u}{a_n}\right) dF_D(u) \geq \frac{k(c)}{a_n}\{F_D(x_o) - F_D(x_o-0)\}$$

which contradicts Lemma 4.7. (Recall that $a_n \to 0^+$.)

Lemma 4.9. If $\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$ with probability one for some function g, then 0 is a derived number of F_S at x_o (as defined on page 207 of [12]) for any x_o in $(-\infty, \infty)$.

Proof: Let a be such that $k(a) > 0$. Since k is continuous there exists a number $b > a$ such that $\inf_{a\leq x\leq b} k(x) \geq \frac{k(a)}{2}$. Now

$$\int \frac{1}{a_n} k\left(\frac{x-u}{a_n}\right) dF_S(u) \geq \int_{x-ba_n}^{x-aa_n} \frac{1}{a_n} k\left(\frac{x-u}{a_n}\right) dF_S(u)$$

$$\geq \inf_{x-ba_n \leq u \leq x-aa_n} k\left(\frac{x-u}{a_n}\right) \cdot \frac{F_S(x-aa_n) - F_S(x-ba_n)}{a_n}$$

$$\geq \frac{(b-a)}{2} k(a) \cdot \frac{F_S(x-aa_n) - F_S(x-ba_n)}{(b-a)a_n} \geq 0.$$

Let x_o be an arbitrary but fixed real number and $x_n = x_o + aa_n$. It then follows that

$$\sup_x \int \frac{1}{a_n} k\left(\frac{x-u}{a_n}\right) dF_S(u) \geq \frac{(b-a)}{2} k(a) \cdot \frac{F_S(x_n-aa_n) - F_S(x_n-ba_n)}{(b-a)a_n}$$

$$= \frac{(b-a)}{2} k(a) \cdot \frac{F_S(x_o) - F_S(x_o-(b-a)a_n)}{(b-a)a_n} .$$

From Lemma 4.7 we can easily deduce that $\lim_{n\to\infty} \frac{F_S(x_o) - F_S(x_o-(b-a)a_n)}{(b-a)a_n} = 0.$

Since x_o was arbitrary the proof is complete.

<u>Lemma 4.10</u>. If $\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$ with probability one for some function g, then $F_S(x) = 0$ for all x.

Proof: Let a and b be real numbers with $a < b$, and put

$h(x) = F_S(x) + x$. Then h is strictly increasing on $[a, b]$ and by Lemma 4.9 it has a derived number equal to one at every point. Thus if we take $E = [a, b]$ in Lemma 2 on page 208 of [12] then we have

$$(2) \qquad 0 \leq \lambda^*(h[a, b]) \leq 1 \cdot \lambda^*([a, b])$$

where $\lambda^*(E)$ denotes the Lebesgue outer measure of E and $h[a, b]$ is the image of $[a, b]$ under h. Since $h[a, b] = [a + F_S(a), b + F_S(b)]$ we can rewrite (2) as

$$0 \leq b + F_S(b) - a - F_S(a) \leq b-a$$

which means $F_S(b) = F_S(a)$. Since a and b were arbitrary, F_S must be constant and hence F_S must be identically zero since $\lim_{x\to\infty} F_S(x) = 0$.

We are now ready to obtain the main theorem of this chapter.

Theorem 4.11. A necessary and sufficient condition for

$$\lim_{n\to\infty} \sup_x |f_n(x) - g(x)| = 0$$

with probability one for a function g is that g be the uniformly continuous derivative of F.

Proof: The sufficiency of this condition has been established by

Nadaraya [11] for a larger class of kernels than that considered here.

Conversely, Lemmas 4.8 and 4.9 show that $F = F_{AC}$. Lemma 4.6 states that $F'_{AC}(x) = g(x)$ everywhere and hence $F'(x) = g(x)$ everywhere. Finally Lemma 4.3 yields the uniform continuity of g and the necessity of the condition is established.

CHAPTER 5

APPLICATIONS

In this chapter we will mention some possible applications of the density estimates considered in the earlier chapters. We will first use the density estimates to construct estimates of the unknown distribution function and then use the density with the distribution estimates to construct estimates of the hazard rate (defined below) at a given point x for an absolutely continuous distribution function F having probability density f. The use of density estimates to construct estimates of a hazard rate was suggested by Parzen [13] and studied by Murthy [10] who showed that the estimates considered below are consistent and asymptotically normal at continuity points of F and f.

Estimation of a Distribution Function

Let $(X_1, X_2, \ldots)$ be independent identically distributed random variables having common distribution function F and let the kernel estimate $f_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} k\left(\frac{x-X_i}{a_n}\right)$ be as in Chapter 1. As an estimate of the value $F(x)$ of the distribution function at a given point x we will take

$$\hat{F}_n(x) = \int_{-\infty}^{x} f_n(u)du.$$

If we let

$$G(x) = \int_{-\infty}^{x} k(u)du$$

then we may write

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} G\left(\frac{x-X_i}{a_n}\right)$$

where G is a (continuous) distribution function.

<u>Lemma 5.1</u>. If F is continuous, then for every $\eta > 0$ and $\varepsilon > 0$ there exists a universal constant C such that

$$P_F \{\sup_x |\hat{F}_n(x) - F(x)| > \varepsilon\} \leq C \exp[-(2-\eta)\varepsilon^2 n]$$

for n sufficiently large.

Proof: If $\eta \geq 2$ the statement is trivially true; so let us assume that $0 < \eta < 2$. Let F_n denote the empirical distribution function based on $(X_1, \ldots, X_n)$ as defined in Chapter 3. Upon integrating by parts and observing that $\int dG(\frac{x-u}{a_n}) = 1$, we find that

$$\sup_x |\hat{F}_n(x) - E\hat{F}_n(x)|$$

$$= \sup_x | \int G(\frac{x-u}{a_n})dF_n(u) - \int G(\frac{x-u}{a_n})dF(u)|$$

$$= \sup_x \left| \int \{F_n(u) - F(u)\} dG(\frac{x-u}{a_n}) - [\{F_n(u) - F(u)\} G(\frac{x-u}{a_n})]_{-\infty}^{\infty} \right|$$

$$= \sup_x \left| \int \{F_n(u) - F(u)\} dG(\frac{x-u}{a_n}) \right|$$

$$\leq \sup_x |F_n(x) - F(x)|.$$

Since

$$\sup_x |E\hat{F}_n(x) - F(x)| = \sup_x \left| \int G(\frac{x-u}{a_n}) dF(u) - F(x) \right|$$

$$= \sup_x \left| \int \{F(x-u) - F(x)\} \frac{1}{a_n} k(\frac{u}{a_n}) du \right|$$

we have for any $\delta > 0$

$$(1) \quad \sup_x |E\hat{F}_n(x) - F(x)| \leq \sup_x \sup_{|u|<\delta} |F(x-u) - F(x)| + 2 \sup_x \int_{|u|\geq\delta} \frac{1}{a_n} k(\frac{u}{a_n}) du.$$

Given $\varepsilon' > 0$, the uniform continuity of F allows us to make the first term on the right side of (1) less than $\frac{\varepsilon'}{2}$ by choosing δ sufficiently small; having so chosen δ we can make the second term less than $\frac{\varepsilon'}{2}$ for n large enough. Thus we have shown $\lim_{n\to\infty} \sup_x |E\hat{F}_n(x) - F(x)| = 0$. By combining the above results it is seen that for n sufficiently large

$$P_F\,\{\sup_x|\hat{F}_n(x) - F(x)| > \varepsilon\} \le P_F\,\{\sup_x|\hat{F}_n(x) - F(x)| > \varepsilon(2-\eta)^{-1/2}\}$$

$$\le P_F\,\{\sup_x|F_n(x) - F(x)| > \varepsilon(2-\eta)^{-1/2}\}.$$

An application of Lemma 3.1 completes the proof. (We note that this proof is almost a verbatim repetition of the argument used in Theorem 1 of [11] if we replace there f_n and f by $\hat{F}_n$ and F, respectively.)

Lemma 5.2. If F is continuous, then

$$\lim_{n\to\infty} \sup_x|\hat{F}_n(x) - F(x)| = 0$$

with probability one.

Proof: By applying Lemma 5.1 and the Borel-Cantelli Lemma.

We remark here that the estimate $\hat{F}_n$ is of dubious importance since the well-known estimate F_n (the empirical distribution function) has many desirable properties as well as the added advantage of being readily computed. However, since $\hat{F}_n$ is continuous in x for each sample point, it is easy to verify that certain functions of the type $\sup_{x\in T} g(F_n(x))$ are random variables (see for example the hazard function discussed below).

The results of Chapter 3, together with Lemmas 5.1 and 5.2, indicate that we may choose $\hat{F}_n$ in such a manner that $\sup_x |\hat{F}_n^{(s)}(x) - F^{(s)}(x)|$ converges to zero with probability one for $s = 0, 1, \ldots, r$, provided that $F^{(s)}$ exists and is bounded for $s = 0, 1, \ldots, r + 1$. As discussed in [1], this suggests the possibility of estimating functionals I of F with

$$I(F) = \int_{-\infty}^{\infty} H[F(u), F^{(1)}(u), \ldots, F^{(r)}(u)]du$$

by $I(\hat{F}_n)$.

Estimation of the Hazard Rate

If the random variable X represents the time to failure of an item, then $F(x)$ is the probability of the event that by time x the item has failed and $R(x) = 1 - F(x)$ is the probability that the item survived time x. If $F'(x) = f(x)$, then

$$Z(x) = \lim_{\Delta x \to 0^+} \frac{P_F\{x < X \leq x + \Delta x | x < X\}}{\Delta x} = \frac{f(x)}{1-F(x)} = \frac{f(x)}{R(x)}$$

is called the hazard rate of X at x. We might say $Z(x)$ gives the probability density of almost immediate failure of an item that has survived time x.

Now let $R_n(x) = 1 - \hat{F}_n(x)$. We will consider estimates of $Z_n(x)$ of $Z(x)$ for which

$$Z_n(x) = \begin{cases} 0 & \text{if } R_n(x) = 0 \\ \dfrac{f_n(x)}{R_n(x)} & \text{if } R_n(x) > 0 \end{cases}$$

and obtain two properties concerning them. In order to do this we will need the following.

Lemma 5.3. Let k be a continuous probability density function satisfying the condition $\lim_{|u|\to\infty} |uk(u)| = 0$ and let $\{a_n\}$ be a sequence of positive constants converging to zero. Suppose a and b are elements in the extended real number system. If there exists an open interval containing $[a, b]$ on which g is uniformly continuous and if $\int |g(u)|du$ is finite, then

$$\lim_{n\to\infty} \sup_{a\leq x\leq b} \left| \int \frac{1}{a_n} k(\frac{u}{a_n}) g(x-u)du - g(x) \right| = 0.$$

Proof: For $\delta > 0$, the inequality

$$\left|\frac{1}{a_n} \int k(\frac{u}{a_n}) g(x-u)du - g(x)\right|$$

$$\leq \sup_{|u|\leq\delta} |g(x-u) - g(x)| + \frac{1}{\delta} \sup_{|u|\geq\delta/a_n} |uk(u)| \int |g(u)|du$$

$$+ |g(x)| \int_{|u|\geq\delta/a_n} k(u)du$$

has been established in the proof of Theorem 1A of [13]. So if $M_1 = \sup_{[a,b]} |g(x)|$ and $M_2 = \int |g(u)|du$ it then follows that

$$\sup_{a \le x \le b} \left|\frac{1}{a_n} \int k(\frac{u}{a_n})g(x-u)du - g(x)\right|$$

$$\le \sup_{a \le x \le b} \sup_{|u| \le \delta} |g(x-u) - g(x)| + \frac{M_2}{\delta} \sup_{|u| \ge \delta/a_n} |uk(u)|$$

$$+ M_1 \int_{|u| \ge \delta/a_n} k(u)du$$

which tends to zero as we first let n tend to ∞ and then let δ tend to 0.

Lemma 5.4. Suppose $F(b) < 1$, $f(x)$ is uniformly continuous on $(-\infty, b + \delta)$ for some $\delta > 0$ and the kernel k is a continuous function of bounded variation on $(-\infty, \infty)$ for which $\lim_{|u| \to \infty} |uk(u)| = 0$. Then there exist constants C_1 and C_2 (C_2 depending on b) such that

$$P_F \{\sup_{x \le b} |Z_n(x) - Z(x)| > \varepsilon\} \le C_1 \exp(-C_2 n a_n^2)$$

for n sufficiently large.

Proof: Let $M = \sup_{x\leq b} Z(x)$. If $R_n(b) > 0$, then for $x \leq b$ we may write

$$|Z_n(x) - Z(x)| = \frac{R(x)}{R_n(x)} \left|Z(x) \frac{\hat{F}_n(x) - F(x)}{R(x)} + \frac{f_n(x) - f(x)}{R(x)}\right|.$$

Since $\frac{R(x)}{R_n(x)} > 2$ implies $\frac{R(x) - R_n(x)}{R(x)} > \frac{1}{2}$, it now follows that

$$P\{\sup_{x\leq b}|Z_n(x) - Z(x)| > \varepsilon\}$$

$$\leq P\{\sup_{x\leq b}|Z_n(x) - Z(x)| > \varepsilon,\ R_n(b) > 0\} + P\{R_n(b) = 0\}$$

$$\leq P\{\varepsilon < 2M \sup_{x\leq b}\left|\frac{\hat{F}_n(x) - F(x)}{R(x)}\right| + 2 \sup_{x\leq b}\left|\frac{f_n(x) - f(x)}{R(x)}\right|\}$$

$$+ P\{\sup_{x\leq b} \frac{R(x)}{R_n(x)} > 2,\ R_n(b) > 0\} + P\{R_n(b) = 0\}$$

$$\leq P\{\varepsilon < 4M \sup_{x\leq b}\left|\frac{\hat{F}_n(x) - F(x)}{R(x)}\right|\} + P\{\varepsilon < 4 \sup_{x\leq b}\left|\frac{f_n(x) - f(x)}{R(x)}\right|\}$$

$$+ P\{\frac{1}{2} < \sup_{x\leq b}\left|\frac{R(x) - R_n(x)}{R(x)}\right|\}$$

$$\leq P\{\varepsilon R(b) < 4M \sup_{x\leq b}\left|\frac{\hat{F}_n(x) - F(x)}{R(x)}\right|\} + P\{\varepsilon R(b) < 4 \sup_{x\leq b} f_n(x) - f(x)|\}$$

$$+ P\{\frac{R(b)}{2} < \sup_{x\leq b}\left|\frac{\hat{F}_n(x) - F(x)}{R(x)}\right|\}.$$

As a consequence of Lemmas 3.2 and 5.3

$$P\{\varepsilon\ R(b) < 4 \sup_{x \leq b} |f_n(x) - f(x)|\} \leq P\{\varepsilon\ R(b) < 8 \sup_{x \leq b} |f_n(x) - Ef_n(x)|\}$$

$$\leq P\{\varepsilon\ R(b) < 8 \sup_{x} |f_n(x) - Ef_n(x)|\} \leq C \exp(-\varepsilon^2 R^2(b) n a_n^2/\mu^2)$$

for n sufficiently large (C is the constant in Lemma 3.1 and μ is the variation of k on $(-\infty, \infty)$.) Therefore by Lemma 5.1 and the above we have for n large

$$P_F\ \{\sup_{x \leq b} |Z_n(x) - Z(x)| > \varepsilon\}$$

$$\leq C[\exp(-\varepsilon^2 R^2(b) n/16M^2) + \exp(-\varepsilon^2 R^2(b) n a_n^2/32\mu^2) + \exp(-R^2(b) n/4)].$$

Since $\exp(-\varepsilon^2 R^2(b) n a_n^2/32\mu^2)$ is dominating the result follows.

<u>Lemma 5.5</u>. If $\sum_{n=1}^{\infty} \exp(-c n a_n^2)$ converges for all positive c, then under the conditions of Lemma 5.1

$$\lim_{n\to\infty} \sup_{x \leq b} |Z_n(x) - Z(x)| = 0$$

with probability one.

Proof: Apply Lemma 5.3 in conjunction with the Borel-Cantelli Lemma.

We also mention that in [11] and [13] the kernel estimates f_n have been used in an obvious manner to construct estimates of the mode of a density. If the estimated mode and the true mode are denoted by θ_n and θ respectively, then sufficient conditions under which $\lim_{n\to\infty} \theta_n = \theta$ with probability one are found in [11]. An exponential bound for $P_F\{|\theta_n - \theta| > \varepsilon\}$ can be deduced from Theorems 1 and 2 there.

Our main applications of the kernel estimates are the proposed estimates of a regression function which will be studied in detail in Chapter 6.

CHAPTER 6

ESTIMATION OF A REGRESSION FUNCTION

Let us suppose that X and Y are real-valued random variables having a joint distribution function F. If $E_F|Y|$ is finite then any version of the conditional expectation of Y given X, $E_F[Y|X]$, will be called a regression function (of Y on X). We will be interested in those distributions F which have a density, that is, those for which there exists a Borel-measurable function f such that

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,dv\,du \quad \text{for all real } x \text{ and } y.$$

In this case the regression function m_f associated with a particular density f of F is defined to be $m_f(x) = \dfrac{\int yf(x,y)dy}{\int f(x,y)dy}$ whenever $\int f(x,y)dy > 0$. In this chapter we will assume that a version of the density f can be chosen which satisfies certain regularity conditions to be specified. The problem considered here is: Given a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from a distribution F which has a density f satisfying the regularity conditions, how can m_f be estimated? In order to motivate properly our search for an estimate of a regression function, we shall try to estimate the regression function m_f corresponding to a "smooth" density function f (if such a density exists). However, the arbitrariness resulting from this will be immaterial as is seen in the following case. If m_n denotes an estimate of m_f and if $\sup_{|x|\leq a} |m_n(x) - m_f(x)|$ converges to zero with probability one, then clearly $\operatorname{ess\,sup}_{|x|\leq a} |m_n(x) - m(x)|$ converges to zero with probability one for any other version m of the regression function.

In the sequel the subscript f on m_f will be dropped and a statement like "the regression function m has property A" is to be interpreted as meaning there exists a choice of the density f such that the regression function m_f, corresponding to f, has property A.

Three Estimates and Their Motivation

Let (X_1, Y_1), (X_2, Y_2), . . . be independent identically distributed bivariate random variables having a common probability density f. The kernel estimates considered in the first four chapters can be adapted to provide estimates of the bivariate probability density function f; i.e., if

(1) $$f_n(x,y) = \frac{1}{na_nb_n} \sum_{i=1}^{n} h\left(\frac{x-X_i}{a_n}, \frac{y-Y_i}{b_n}\right)$$

where $\{a_n\}$ and $\{b_n\}$ are sequences of positive numbers going to zero and h is a bivariate probability density function, then the following convergence properties of f_n to f are obtained directly as in the univariate case (see [2]):

(i) The asymptotic unbiasedness of f_n at every continuity point of f.

(ii) The consistency of f_n in quadratic mean at every continuity point of f.

(iii) The joint asymptotic normality of the estimates f_n at continuity points of f.

(iv) The uniform consistency of f_n whenever f is uniformly continuous.

Moreover, if f is uniformly continuous, the results in [1] may be used to construct estimates f_n of the form (1) for which $\sup_{|x|\leq h_n, |y|\leq k_n} |f_n(x,y) - f(x,y)|$ converges to zero with probability one and an exponential bound on $P\left[\sup_{|x|\leq h_n, |y|\leq k_n} |f_n(x,y) - f(x,y)| > \varepsilon\right]$ exists for appropriate h_n and k_n going to infinity.

If the density f were known then the regression function m could be constructed by ordinary Lebesgue integration. This together with the above discussion suggests we might first estimate f and then construct estimates of m by integration; i.e., if f_n denotes an estimate of f of the form (1), we might estimate

$$m(x) = \frac{\int yf(x,y)dy}{\int f(x,y)dy} \quad \text{by a function of the form} \quad m_n^{(1)}(x) = \frac{\int yf_n(x,y)dy}{\int f_n(x,y)dy}$$

(henceforth we will define the regression estimates to be zero at those points where their explicit formulas are logically undefined).

If the bivariate kernel h is chosen to be a product kernel, say $h(x,y) = k(x)k^*(y)$ where k and k^* are univariate probability densities with $\int yk^*(y)dy = 0$, then the above estimate $m_n^{(1)}$ simplifies to an estimate $m_n^{(2)}$ given by

$$m_n^{(2)}(x) = \sum_{i=1}^{n} Y_i k\left(\frac{x-X_i}{a_n}\right) \Big/ \sum_{i=1}^{n} k\left(\frac{x-X_i}{a_n}\right).$$

Note that $m_n^{(2)}$ is somewhat simpler, computationally and otherwise, than $m_n^{(1)}$. For example $m_n^{(2)}$ is independent of the sequence $\{b_n\}$.

In attempting to find an estimate for which $\lim_{n\to\infty} \sup_{|x|\leq a} |m_n(x) - m(x)| = 0$ with probability one for a fixed real number a, we found that such an estimate could be obtained if we replaced the Y_i in the estimate $m_n^{(2)}$ by a truncated version of the Y_i. Thus we were led to our third estimate

$$m_n^{(3)}(x) = \sum_{i=1}^{n} Y_i I_n(Y_i) k\left(\frac{x-X_i}{a_n}\right) \Big/ \sum_{i=1}^{n} k\left(\frac{x-X_i}{a_n}\right)$$

where I_n is the indicator function of $[-c_n, c_n]$ with c_n going to infinity.

Bounds for the Supremum Distance on a Finite Interval

We will now find bounds for $P\{\sup_{|x|\leq a} |m_n^{(i)}(x) - m(x)| > \varepsilon\}$ $(i=1,2,3)$ for those distribution functions having a density which satisfies certain regularity conditions. From here on, a will denote an arbitrary but fixed positive real number.

Let $X_1, X_2, \ldots$ be independent identically distributed random variables having a common univariate probability density function g. Let $g_n(x)$ be a kernel estimate of g as defined in Chapter 1; i.e.,

$$g_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} k\left(\frac{x-X_i}{a_n}\right) \tag{2}$$

where k is a probability density function and $\{a_n\}$ is a sequence of positive constants converging to zero.

<u>Lemma 6.1</u>. Suppose k has a bounded derivative, $\lim_{|u|\to\infty} |uk(u)| = 0$ and there exists an open interval containing $[-a, a]$ on which g is continuous. Then given any $\varepsilon > 0$ there exist positive constants C_1 and C_2 such that

$$P_g\{\sup_{|x|\leq a} |g_n(x) - g(x)| > \varepsilon\} \leq C_1[\exp(-C_2 n a_n^2)]/a_n^2$$

for n sufficiently large.

Proof: Since g is uniformly continuous for $|x|\leq a$ there exists a $\delta > 0$ such that $|g(x) - g(y)| < \frac{\varepsilon}{4}$ for $|x - y| < \delta$ and x,y belonging to $[-a, a]$. Let us take $\delta(n) = \varepsilon a_n^2/4M_1$ where $M_1 = \sup_x |k'(x)|$. For n sufficiently large $\delta(n) \leq \delta$ so that $|g_n(x) - g_n(y)| < \varepsilon/4$ and $|g(x) - g(y)| < \varepsilon/4$ whenever $|x - y| < \delta(n)$ and x,y belong to $[-a, a]$.

The desired conclusion now follows as in Lemma 2 of [1] using our Lemma 5.3 in place of the corollary to Lemma 1 in [1].

Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be independent bivariate random variables identically distributed as a bivariate random variable (X,Y) whose distribution function F has density f. Let k and $\{a_n\}$ be as in (2) above. For simplicity we write:

$$w_n(x) = \frac{1}{na_n} \sum_{s=1}^{n} Y_s k\left(\frac{x-X_s}{a_n}\right)$$

$$w_n(x) = \int yf(x,y)dy$$

$$\psi(x) = \int \exp(iux)k(u)du \qquad \text{where} \quad i^2 = -1$$

$$\phi_n(u) = \frac{1}{n} \sum_{s=1}^{n} Y_s \exp(iuX_s).$$

Lemma 6.2. Let k be a continuous univariate probability density function satisfying the condition $\lim_{|u|\to\infty} |uk(u)| = 0$. If there exists an open interval containing $[-a, a]$ on which the function w is continuous and if $E_f|Y|$ is finite, then

$$\lim_{n\to\infty} \sup_{|x|\leq a} |Ew_n(x) - w(x)| = 0.$$

Proof: Since $\int |w(u)|du \leq E_f|Y|$ and $Ew_n(x) = \int \frac{1}{a_n} k(\frac{u}{a_n})w(x-u)du$ the proof follows directly from Lemma 5.3.

If we assume that the characteristic function ψ is absolutely integrable, then by the standard inversion formula we may write:

$$w_n(x) = \frac{1}{na_n} \sum_{s=1}^{n} Y_s k\left(\frac{x-X_s}{a_n}\right)$$

$$= \frac{1}{2\pi na_n} \sum_{s=1}^{n} Y_s \int \psi(u)\exp[-iu\left(\frac{x-X_s}{a_n}\right)]du$$

$$= \frac{1}{2\pi} \int e^{-iux} \psi(a_n u)\phi_n(u)du.$$

Lemma 6.3. Suppose k satisfies the conditions of Lemma 6.2 and suppose ψ is absolutely integrable. If there exists an open interval containing $[-a, a]$ on which w is continuous and if $E_f Y^2$ is finite, then for every $\varepsilon > 0$ there exists a constant $C > 0$ such that

$$P_f\{ \sup_{|x|\leq a} |w_n(x) - w(x)| > \varepsilon\} \leq C/na_n^2$$

for n sufficiently large.

Proof: From the discussion preceding the lemma and the fact that $\int |\psi(a_n u)| E_f|Y|du$ is finite we have

$$\sup_{|x|\leq a} |w_n(x) - Ew_n(x)|$$

$$= \sup_{|x|\leq a} |\frac{1}{2\pi} \int e^{-iux} \psi(a_n u)\phi_n(u)du - \frac{1}{2\pi}E_f \int e^{-iux} \psi(a_n u)\phi_n(u)du |$$

$$= \sup_{|x|\le a} \left|\frac{1}{2\pi}\int e^{-iux}\,\psi(a_n u)\{\phi_n(u) - E\phi_n(u)\}du\right|$$

$$\le \frac{1}{2\pi}\int |\psi(a_n u)||\phi_n(u) - E\phi_n(u)|du.$$

Thus $\sup_{|x|\le a} |w_n(x) - Ew_n(x)|^2 \le \frac{1}{4\pi^2}\left[\int |\psi(a_n u)||\phi_n(u) - E\phi_n(u)|du\right]^2$

$$= \frac{1}{4\pi^2}\iint |\psi(a_n u)||\phi_n(u) - E\phi_n(u)||\psi(a_n v)||\phi_n(v) - E\phi_n(v)|dudv$$

so that by the Schwartz inequality we have

$$E\sup_{|x|\le a} |w_n(x) - Ew\ (x)|^2$$

$$\le \frac{1}{4\pi^2}\iint |\psi(a_n u)||\psi(a_n v)|E\{|\phi_n(u) - E\phi_n(u)||\phi_n(v) - E\phi_n(v)|\}dudv$$

$$\le \frac{1}{4\pi^2}\iint |\psi(a_n u)||\psi(a_n v)|E^{1/2}|\phi_n(u) - E\phi_n(u)|^2 E^{1/2}|\phi_n(v) - E\phi_n(v)|^2 dudv$$

$$= \left(\frac{1}{2\pi}\int |\psi(a_n u)|E^{1/2}|\phi_n(u) - E\phi_n(u)|^2 du\right)^2.$$

Now $E|\phi_n(u) - E\phi_n(u)|^2 = \frac{1}{n^2}E\left|\sum_{s=1}^{n}\left(Y_s e^{iuX_s} - EY_s e^{iuX_s}\right)\right|^2$

$$= \frac{1}{n}E|Ye^{iuX} - EYe^{iuX}|^2 = \frac{1}{n}\left(E|Ye^{iuX}|^2 - |EYe^{iuX}|^2\right)$$

$$= \frac{1}{n}\left(EY^2 - |EYe^{iuX}|^2\right) \le \frac{1}{n}EY^2.$$

Thus we have shown

$$E \sup_{|x|\le a} |w_n(x) - Ew_n(x)|^2 \le \frac{1}{4\pi^2} \frac{EY^2}{n}\left(\int |\psi(a_n u)|du\right)^2$$

$$= \frac{1}{4\pi^2} \frac{EY^2}{na_n^2}\left(\int |\psi(u)|du\right)^2.$$

By Lemma 6.2, Tchebychev's inequality and the above inequality, we have for n sufficiently large

$$P\{\sup_{|x|\le a} |w_n(x) - w(x)| > \varepsilon\} \le P\{\sup_{|x|\le a} |w_n(x) - Ew_n(x)| > \varepsilon/2\}$$

$$\le \frac{4}{\varepsilon^2} E \sup_{|x|\le a} |w_n(x) - Ew_n(x)|^2 \le \frac{1}{na_n^2} \frac{4}{\varepsilon^2} EY^2\left(\int |\psi(u)|du\right)^2.$$

If we take $C = \frac{4}{\varepsilon^2}EY^2\left(\int |\psi(u)|du\right)^2$ then the proof is complete.

Let h be a bivariate probability density such that h and $k(x) = \int h(x,y)dy$ satisfy the conditions:

(i) $\int yh(x,y)dy$ is bounded and continuous on $(-\infty, \infty)$.

(ii) $\lim_{|x|\to\infty} |xk(x)| = 0$.

(iii) $\sup_x |k'(x)| < \infty$.

We tabulate here some simplified notation:

$$f_n(x,y) = \frac{1}{na_n b_n} \sum_{i=1}^{n} h\left(\frac{x-X_i}{a_n}, \frac{y-Y_i}{b_n}\right) \quad \text{where} \quad a_n \to 0^+, b_n \to 0^+$$

$$g_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} k\left(\frac{x-X_i}{a_n}\right) \quad \text{where} \quad k(x) = \int h(x,y)dy$$

$$w_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} Y_i k\left(\frac{x-X_i}{a_n}\right)$$

$$m_n^{(1)}(x) = \frac{\int yf_n(x,y)dy}{\int f_n(x,y)dy} = \frac{\int yf_n(x,y)dy}{g_n(x)}$$

$$m_n^{(2)}(x) = \frac{\sum_{i=1}^{n} Y_i k\left(\frac{x-X_i}{a_n}\right)}{\sum_{i=1}^{n} k\left(\frac{x-X_i}{a_n}\right)} = \frac{w_n(x)}{g_n(x)}$$

$$g(x) = \int f(x,y)dy$$

$$w(x) = \int yf(x,y)dy$$

$$m(x) = \frac{\int yf(x,y)dy}{\int f(x,y)dy} = \frac{w(x)}{g(x)}.$$

Theorem 6.4. If $b_n = o(a_n)$, if there exists an open interval containing $[-a, a]$ on which

(i) g is continuous and bounded away from zero

(ii) w is continuous

and if $E_f Y^2$ is finite, then corresponding to each $\varepsilon > 0$ there exist positive constants C_1, C_2 and C_3 such that

$$P\{\sup_{|x|\le a} |m_n^{(i)}(x) - m(x)| > \varepsilon\} \le \frac{C_1}{a_n^2} \exp(-C_2 n a_n^2) + \frac{C_3}{n a_n^2}$$

for n sufficiently large $(i=1,2)$.

Proof: Let Q be the set of rationals in $[-a,a]$ and let $\varepsilon > 0$ be given. Since $\int y f_n(x,y)dy$ and $g_n(x)$ are continuous functions of x for each sample point it follows for $i=1,2$ that

$$\{\sup_{|x|\le a} |m_n^{(i)}(x) - m(x)| > \varepsilon\} = \bigcup_{|x|\le a} \{|m_n^{(i)}(x) - m(x)| > \varepsilon\}$$

$$= \left(\bigcup_{|x|\le a} \{|m_n^{(i)}(x) - m(x)| > \varepsilon, g_n(x) > 0\}\right)$$

$$\bigcup \left(\bigcup_{|x|\le a} \{|m(x)| > \varepsilon, g_n(x) = 0\}\right)$$

$$= \left(\bigcup_{x\varepsilon Q} \{|m_n^{(i)}(x) - m(x)| > \varepsilon, g_n(x) > 0\}\right)$$

$$\bigcup \left(\bigcup_{r=1}^{\infty} \bigcap_{j=1}^{\infty} \bigcup_{x\varepsilon Q} \{|m(x)| > \varepsilon + \frac{1}{r}, g_n(x) < \frac{1}{j}\}\right).$$

Now for fixed x, $m_n^{(i)}(x)$ and $g_n(x)$ are both random variables. Hence $\sup_{|x|\le a} |m_n^{(i)}(x) - m(x)|$ is a random variable $(i=1,2)$.

Let $M_1 = \sup_x \left|\int yh(x,y)dy\right|$, $M_2 = \sup_{|x|\le a} |m(x)|$ and $b = \inf_{|x|\le a} g(x)$. For $g_n(x) > 0$ it is seen that

$$|m_n^{(1)}(x) - m(x)| = \left|m_n^{(2)}(x) - m(x) + b_n \sum_{i=1}^{n} \int yh\left(\frac{x-X_i}{a_n}, y\right) dy \Big/ n a_n g_n(x)\right|$$

$$\leq |m_n^{(2)}(x) - m(x)| + \frac{b_n M_1}{a_n g_n(x)}$$

$$= \frac{g(x)}{g_n(x)}\left[\left|m(x)\left(1 - \frac{g_n(x)}{g(x)}\right) + \frac{w_n(x)-w(x)}{g(x)}\right| + \frac{b_n M_1}{a_n g(x)}\right]$$

$$\leq \frac{g(x)}{g_n(x)}\left[\left|m(x)\left(1 - \frac{g_n(x)}{g(x)}\right) + \frac{w_n(x)-w(x)}{g(x)}\right| + \frac{b_n M_1}{a_n b}\right]$$

so that

$$P\{\sup_{|x|\leq a} |m^{(i)}(x) - m(x)| > \varepsilon\}$$

$$\leq P\{\sup_{|x|\leq a} |m^{(i)}(x) - m(x)| > \varepsilon,\ g_n(x) > 0 \quad \text{for} \quad |x|\leq a\}$$

$$+ P\{g_n(x) = 0 \quad \text{for some} \quad x \quad \text{with} \quad |x|\leq a\}$$

$$\leq P\{\sup_{|x|\leq a} \frac{g(x)}{g_n(x)}\left[\left|m(x)\left(1 - \frac{g_n(x)}{g(x)}\right) + \frac{w_n(x)-w(x)}{g(x)}\right| + \frac{b_n M_1}{a_n b}\right] > \varepsilon,$$
$$g_n(x) > 0 \quad \text{for} \quad |x|\leq a\}$$

$$+ P\{g_n(x) = 0 \quad \text{for some} \quad x \quad \text{with} \quad |x|\leq a\}$$

$$\leq P\{\sup_{|x|\leq a} 2\left[\left|m(x)\left(1 - \frac{g_n(x)}{g(x)}\right) + \frac{w_n(x)-w(x)}{g(x)}\right| + \frac{b_n M_1}{a_n b}\right] > \varepsilon\}$$

$$+ P\{\sup_{|x|\leq a} \frac{g(x)}{g_n(x)} > 2,\ g_n(x) > 0 \quad \text{for} \quad |x|\leq a\}$$

$$+ P\{g_n(x) = 0 \quad \text{for some} \quad x \quad \text{with} \quad |x|\leq a\}.$$

Now $\frac{g(x)}{g_n(x)} > 2$ implies $\frac{g(x)-g_n(x)}{g(x)} > \frac{1}{2}$ and $b_n = o(a_n)$ so that for n large the last member in the string of inequalities above is less than or equal to

$$P\{\sup_{|x|\leq a}\left|m(x)\left(1 - \frac{g_n(x)}{g(x)}\right) + \frac{w_n(x)-w(x)}{g(x)}\right| > \frac{\varepsilon}{4}\} + P\{\sup_{|x|\leq a}\left|\frac{g(x)-g_n(x)}{g(x)}\right| > \frac{1}{2}\}.$$

So

$$P\{\sup_{|x|\leq a}|m^{(1)}(x) - m(x)| > \varepsilon\}$$

$$\leq P\{\sup_{|x|\leq a}\left|m(x)\left(1 - \frac{g_n(x)}{g(x)}\right) + \frac{w_n(x)-w(x)}{g(x)}\right| > \frac{\varepsilon}{4}\} + P\{\sup_{|x|\leq a}\left|\frac{g_n(x)-g(x)}{g(x)}\right| > \frac{1}{2}\}$$

$$\leq P\{\sup_{|x|\leq a} M_2\left|\frac{g_n(x)-g(x)}{g(x)}\right| > \frac{\varepsilon}{8}\} + P\{\sup_{|x|\leq a}\left|\frac{w_n(x)-w(x)}{g(x)}\right| > \frac{\varepsilon}{8}\}$$

$$+ P\{\sup_{|x|\leq a}\left|\frac{g_n(x)-g(x)}{g(x)}\right| > \frac{1}{2}\}$$

$$\leq P\{\sup_{|x|\leq a}|g_n(x) - g(x)| > \frac{\varepsilon b}{8M_2}\} + P\{\sup_{|x|\leq a}|w_n(x) - w(x)| > \frac{\varepsilon b}{8}\}$$

$$+ P\{\sup_{|x|\leq a}|g_n(x)-g(x)| > \frac{b}{2}\}.$$

Hence the theorem follows by Lemmas 6.1 and 6.3.

Theorem 6.5. If $\frac{1}{a_n^2}\exp(-Cna_n^2)$ converges to zero for every $C > 0$ then, under the conditions of Theorem 6.4, for every $\varepsilon > 0$

$$\lim_{n\to\infty} P\{\sup_{|x|\le a} |m_n^{(i)}(x) - m(x)| > \varepsilon\} = 0 \ (i=1,2).$$

Proof: Observe that $\frac{1}{a_n^2}\exp(-Cna_n^2) \to 0$ implies that $na_n^2 \to \infty$ and then apply Theorem 6.4.

If we now define

$$v_n(x) = \frac{1}{na_n}\sum_{i=1}^{n} Y_i I_n(Y_i) k\left(\frac{x-X_i}{a_n}\right)$$

where I_n is the indicator function of $[-c_n, c_n]$ with $c_n \to \infty$, we may write

$$m_n^{(3)}(x) = \frac{\sum_{i=1}^{n} Y_i I_n(Y_i) k\left(\frac{x-X_i}{a_n}\right)}{\sum_{i=1}^{n} k\left(\frac{x-X_i}{a_n}\right)} = \frac{v_n(x)}{g_n(x)}$$

(k satisfying conditions (ii) and (iii) on page 50).

Lemma 6.6. If there exists an open interval containing $[-a, a]$ on which

(i) w is continuous

(ii) $\int |y| f(x,y)dy$ is uniformly integrable

and if $E_f|y|$ is finite, then $\lim_{n\to\infty} \sup_{|x|\leq a} |Ev_n(x)-w(x)| = 0$.

Proof: For any $\delta > 0$

$$\sup_{|x|\leq a} \left| \int_{|v|\geq c_n} \int \frac{v}{a_n} k\left(\frac{x-u}{a_n}\right) f(u,v)dudv \right|$$

$$= \sup_{|x|\leq a} \left| \int_{|v|\geq c_n} \int \frac{v}{a_n} k\left(\frac{u}{a_n}\right) f(x-u,v)dudv \right|$$

$$\leq \sup_{|x|\leq a} \left| \int_{|u|\leq\delta} \int_{|v|\geq c_n} \frac{v}{a_n} k\left(\frac{u}{a_n}\right) f(x-u,v)dvdu \right|$$

$$+ \sup_{|x|\leq a} \left| \int_{|u|\geq\delta} \int_{|v|\geq c_n} \frac{v}{a_n} k\left(\frac{u}{a_n}\right) f(x-u,v)dvdu \right|$$

$$\leq \sup_{|x|\leq a+\delta} \left| \int_{|v|\geq c_n} \frac{v}{a_n} f(x,v)dv \right| + \frac{1}{\delta} \sup_{|u|\geq\delta/a_n} |uk(u)| \cdot E_f|y|$$

which approaches zero if we choose δ sufficiently small and then let n tend to infinity. Since

$$\sup_{|x|\le a} |Ev_n(x) - w(x)| \le \sup_{|x|\le a} |Ew_n(x) - w(x)|$$

$$+ \sup_{|x|\le a} \left| \int_{|v|\ge c_n} \int \frac{v}{a_n} k\left(\frac{x-u}{a_n}\right) f(u,v)\,du\,dv \right|$$

the desired result now follows from Lemma 6.2.

Lemma 6.7. Under the condition of Lemma 6.6, for any $\varepsilon > 0$ there exist constants C_1 and C_2 such that

$$P_f\{\sup_{|x|\le a} |v_n(x) - w(x)| > \varepsilon\} \le \frac{C_1 c_n}{a_n^2} \exp(-C_2 n a_n^2 / c_n^2).$$

Proof: Let $M_1 = \sup_x |k'(x)|$ and $M_2 = \sup_x |k(x)|$ and define $\delta(n) = \frac{a_n^2}{n M_1} \cdot \frac{\varepsilon}{4}$ (note $\delta(n) \to 0$). For simplicity assume $\frac{2a}{\delta(n)}$ is an integer B_n and partition $[-a, a]$ into B_n intervals of length $\delta(n)$. Denote these intervals by $J_{n1}\ J_{n2}, \ldots, J_{nB_n}$ and select an x_{nj} from J_{nj} for $i=1,2,\ldots B_n$. Then

$$\sup_{|x|\le a} |v_n(x) - w(x)| = \max_{j=1,2,..B_n} \sup_{x\varepsilon J_{nj}} |v_n(x) - w(x)|$$

$$\le \max_{j=1,..,B_n} \sup_{x\varepsilon J_{nj}} |v_n(x) - v_n(x_{nj})| + \max_{j=1,..B_n} |v_n(x_{nj}) - Ev_n(x_{nj})|$$

$$+ \max_{j=1,..B_n} |Ev_n(x_{nj}) - w(x_{nj})| + \max_{j=1,2,..B_n} \sup_{x\varepsilon J_{nj}} |w(x_{nj}) - w(x)|.$$

It follows by the uniform continuity of w on $[-a, a]$

and by Lemma 6.6 that for n sufficiently large,

$$P\{\sup_{|x|\leq a}|v_n(x) - w(x)| > \varepsilon\} \leq P\{\max_{j=1,..B_n}|v_n(x_{nj}) - Ev_n(x_{nj})| > \frac{\varepsilon}{4}\}.$$

The conclusion is now obtained from Theorem 2 of Hoeffding in [7] with

$$C_1 = \frac{16aM_1}{\varepsilon} \quad \text{and} \quad C_2 = \frac{\varepsilon^2}{32M_2^2(1+E_f|Y|^2)}.$$

Theorem 6.8. If there exists an open interval containing $[-a, a]$ on which

(i) $g(x) = \int f(x,y)dy$ is continuous and bounded away from zero

(ii) $w(x)$ is continuous and $\int |y|f(x,y)dy$ is uniformly integrable

and if $E_f|Y|$ is finite, then given $\varepsilon > 0$ there exist positive constants C_1 and C_2 such that

$$P_f\{\sup_{|x|\leq a}|m_n^{(3)}(x) - m(x)| > \varepsilon\} \leq \frac{C_1 c_n}{a_n^2}\exp(-C_2 n a_n^2/k_n^2)$$

for n sufficiently large.

Proof: The fact $\sup_{|x|\leq a}|m_n^{(3)}(x) - m(x)|$ is a random variable is obtained as in the proof of Theorem 6.4. For $g_n(x) > 0$ we have

$$|m_n^{(3)}(x) - m(x)| = \frac{g(x)}{g_n(x)}\left|m(x)\left(1 - \frac{g_n(x)}{g(x)}\right) + \frac{v_n(x) - w(x)}{g(x)}\right|.$$

Proceeding as in Theorem 6.4 yields the theorem, in view of Lemmas 6.1 and 6.7 and the domination of the error term in Lemma 6.7.

Theorem 6.9. If $\sum_{n=1}^{\infty} \frac{c_n}{a_n^2}\exp(-Cna_n^2/c_n^2)$ is finite for every $C > 0$ then, under the conditions of Theorem 6.8,

$$\lim_{n\to\infty} \sup_{|x|\leq a} |m_n^{(3)}(x) - m(x)| = 0$$

with probability one.

Proof: By Theorem 6.3 in conjunction with the Borel-Cantelli lemma.

Remark: Inspection of the estimates $m_n^{(2)}(x)$ and $m_n^{(3)}(x)$ reveals that $\sup_x|m_n^{(i)}(x)| \leq \max_{j\leq n}|Y_j|$ for $i=2,3$. If f were the density of a bivariate normal with non-zero correlation coefficient and $\int yf(x,y)dxdy = \int xf(x,y)dy = 0$, then the regression function m_f would be of the form $m_f(x) = bx$. Consequently, $\sup_x|m_n^{(i)}(x) - m_f(x)| = \sup_x|m_n^{(i)}(x) - bx| = \infty$ for all n. This shows the impossibility of proving the result in Theorems 6.5 and 6.9 if the supremum were taken over the entire real line.

Estimation of the Point at which a Regression Function attains its Maximum

A problem which frequently arises in practice is to estimate the point θ at which a regression function attains its maximum value. We will now show our regression estimates can be used to construct estimates of θ.

Let θ satisfy $\sup_{|x|\leq a} m(x) = m(\theta)$ and let us choose a kernel h such that for each sample point, $m_n^{(i)}(x)$ is a continuous function of x for $i=1,2,3$. Then there exist random variables $\theta_n^{(i)}$, $i=1,2,3$, mapping the sample space into $[-a, a]$ such that

$$m_n^{(i)}(\theta_n^{(i)}) = \sup_{|x|\leq a} m_n^{(i)}(x).$$

In the following theorem h and k satisfy conditions (i), (ii) and (iii) on page 50 and $k(x) > 0$ for all real x. This added condition on k guarantees the continuity requirement on $m_n^{(i)}$ $(i=1,2,3)$ above.

<u>Theorem 6.10</u>. Let m have a unique maximum on $[-a, a]$ at θ. Then:

(i) Under the conditions of Theorem 6.4, given $\varepsilon > 0$ there exist positive constants C_1 and C_2 such that for n large

$$P\{|\theta_n^{(i)} - \theta| > \varepsilon\} \leq \frac{C_1}{a_n^2}\exp(-C_2 n a_n^2) + \frac{C_3}{n a_n^2} \quad (i=1,2).$$

(ii) Under the conditions of Theorem 6.8, given $\varepsilon > 0$ there exist positive constants C_1 and C_2 such that

$$P\{|\theta_n^{(3)} - \theta| > \varepsilon\} \leq \frac{C_1 c_n}{a_n^2}\exp(-C_2 n a_n^2/c_n^2).$$

(iii) Under the condition of Theorem 6.5

$$\lim_{n\to\infty} \theta_n^{(i)} = \theta \text{ in probability } (i=1,2).$$

(iv) Under the conditions of Theorem 6.9,

$$\lim_{n\to\infty} \theta_n^{(3)} = \theta \text{ with probability one.}$$

Proof: If $b = \inf_{\substack{t\varepsilon[-a,a] \\ |\theta-t|>\varepsilon}} \{m(\theta) - m(t)\}$, then the continuity of m on $[-a, a]$ and the uniqueness of θ imply that $b > 0$ (for ε sufficiently small).

Now $|\theta_n^{(i)} - \theta| > \varepsilon$ implies that t exists in $[-a, a]$ with $|\theta - t| > \varepsilon$ and $m_n^{(i)}(t) - m_n(\theta) \geq 0$. Therefore, $m_n^{(i)}(t) - m(t) + m(t) - m(\theta) + m(\theta) - m_n^{(i)}(\theta) \geq 0$ so that $m_n^{(i)}(t) - m(t) + m(\theta) - m_n^{(i)}(\theta) \geq m(\theta) - m(t) \geq b > 0$. Thus we see that

$$P\{|\theta_n^{(i)} - \theta| > \varepsilon\} \leq P\{\sup_{|x|\leq a} |m_n^{(i)}(x) - m(x)| \geq \frac{b}{2}\}.$$

The theorem now follows from Theorems 6.4, 6.5, 6.8 and 6.9.

Joint Asymptotic Distribution of the Estimated Regression Function at a Finite Number of Distinct Points

In this section we shall study the estimated regression function $m_n^{(2)}$ at two distinct points. The results obtained here remain valid for any finite number of distinct points.

Assume that (X_1, Y_1) (X_2, Y_2), . . . are independent bivariate random variables identically distributed as a bivariate random variable (X,Y) whose distribution function F has density f, that $\{a_n\}$ is a sequence of positive numbers converging to zero and that k is an univariate probability density function. Let x_1 and x_2 be distinct real numbers and let $c = (c_1,d_1,c_2,d_2)$ be an arbitrary point in R^4. We will use the superscript t to denote the transpose. For brevity we define for $i=1,2,\ldots,n$ and $s = 1,2$:

$$U^*_{ni}(x_s) = \frac{1}{a_n}k\left(\frac{x_s - X_i}{a_n}\right)$$

$$U_{ni}(x_s) = \sqrt{a_n}[U^*_{ni}(x_s) - EU^*_{ni}(x_s)]$$

$$V^*_{ni}(x_s) = Y_i U^*_{ni}(x_s)$$

$$V_{ni}(x_s) = \sqrt{a_n}[V^*_{ni}(x_s) - EV^*_{ni}(x_s)]$$

$$U_n(x_s) = \sum_{i=1}^{n} U_{ni}(x_s)$$

$$V_n(x_s) = \sum_{i=1}^{n} V_{ni}(x_s)$$

$$\sqrt{n}Z_n = (U_n(x_1), V_n(x_1), U_n(x_2), V_n(x_2))^t$$

$$W_{ni} = (U_{ni}(x_1), V_{ni}(x_1), U_{ni}(x_2), V_{ni}(x_2))$$

$$g(x) = \int f(x,y)dy$$

$$w(x) = \int yf(x,y)dy$$

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

$$A = ||k||_2^2 \begin{bmatrix} g(x_1) & w(x_1) & 0 & 0 \\ w(x_1) & v(x_1) & 0 & 0 \\ 0 & 0 & g(x_2) & w(x_2) \\ 0 & 0 & w(x_2) & v(x_2) \end{bmatrix} \quad \text{where} \quad ||k||_2^2 = \int k^2(u)du.$$

Let Z be fourvariate normal with mean vector 0 and covariance matrix A.

Lemma 6.11. If x is an arbitrary but fixed point for which $g(x) > 0$, then $g(x)v(x) \neq w^2(x)$.

Proof: Assume that $g(x)v(x) - w^2(x) = 0$. Then

$$0 = \frac{v(x)}{g(x)} - \frac{w^2(x)}{g^2(x)} = V[Y|X = x] \quad \text{(the variance of } Y \text{ given } X = x\text{)},$$

which means $f(x,y) = 0$ almost everywhere (y) and hence

$$g(x) = \int f(x,y)dy = 0.$$

Lemma 6.12. Suppose the density k satisfies the conditions

(i) $k(u)$ and $|uk(u)|$ are bounded

(ii) $\int |u|k(u)du < \infty$

and suppose $na_n^3 \to \infty$. Let $E_f|Y|^3$ be finite and let g', w' and v' exist and be bounded. If $x_1 \neq x_2$ and $g(x_i) > 0$ for $i=1,2$, then $c \cdot Z_n^t$ converges in distribution to $c \cdot Z^t$ for any $c = (c_1, d_1, c_2, d_2)$ in R^4.

Proof: The following hold for $s = 1,2$ and $r = 1,2$ under the assumption that $s \neq r$ whenever s and r appear in the same expression:

(1) $EU_{ni}^2(x_s) = g(x_s)||k||_2^2 + O(a_n)$.

(2) $EV_{ni}^2(x_s) = v(x_s)||k||_2^2 + O(a_n)$.

(3) $EU_{ni}(x_s)V_{ni}(x_s) = w(x_s)||k||_2^2 + O(a_n)$.

(4) $EU_{ni}(x_s)U_{ni}(x_r) = O(a_n)$.

(5) $EV_{ni}(x_s)V_{ni}(x_r) = O(a_n)$.

(6) $EU_{ni}(x_s)V_{ni}(x_r) = O(a_n)$.

We will sketch the proofs of (1) and (4) to illustrate the method. To obtain (1), we see

$$EU_{ni}^2(x_s) = a_n\left[\int \frac{1}{a_n^2}k^2\left(\frac{x_s-u}{a_n}\right)g(u)du - \left(\int \frac{1}{a_n}k\left(\frac{x_s-u}{a_n}\right)g(u)du\right)^2\right]$$

$$= a_n\left[\frac{1}{a_n}\int k^2(u)g(x - a_nu)du - \left(\int k(u)g(x - a_nu)du\right)^2\right].$$

Since g' and $|yk(y)|$ are bounded and $\int |u|k(u)$ is finite, it follows that

$$\left|\int k(u)\{g(x - a_n u) - g(x)\}du\right| \leq \sup_x |g'(x)| a_n \int |u|k(u)du = O(a_n^2)$$

and

$$\left|\int k^2(u)\{g(x - a_n u) - g(x)\}du\right| \leq \sup_x |g'(x)| a_n \int |u|k^2(u)du = O(a_n).$$

Thus we have

$$EU_{ni}^2(x_s) = g(x_s)||k||_2^2 + O(a_n).$$

As for (4), suppose $x_2 > x_1$ and let $\delta = x_2 - x_1$ and $\delta_n = \delta/a_n$. Then

$$EU_{ni}(x_1)U_{ni}(x_2) = a_n \int \frac{1}{a_n^2} k\left(\frac{x_1 - u}{a_n}\right) k\left(\frac{x_2 - u}{a_n}\right) g(u)du + O(a_n)$$

$$= \int k(u)k(\delta_n + u)g(x_1 - a_n u)du + O(a_n)$$

$$= \int_{|u|<\delta_n/2} k(u)k(\delta_n + u)g(x_1 - a_n u)du + O(a_n)$$

$$+ \int_{|u|\geq\delta_n/2} k(u)k(\delta_n + u)g(x_1 - a_n u)du + O(a_n)$$

$$\leq \sup_{|u|\leq\delta_n/2} k(\delta_n + u) \cdot \int k(z)g(x_1 - a_n z)dz$$

$$+ \sup_{|u|\geq\delta_n/2} k(u) \cdot \int k(\delta_n + z)g(x_1 - a_n z)dz + O(a_n)$$

$$\leq \sup_{|u|\geq\delta_n/2} k(u) \cdot O(1) + \sup_{|u|\geq\delta_n/2} k(u) \cdot \int k(z)g(x_2 - a_n z)dz + O(a_n)$$

$$= 2 \sup_{|u|\geq\delta_n/2} k(u) \cdot O(1) + O(a_n)$$

$$\leq \frac{4}{\delta_n} \sup_{|u|\geq\delta_n/2} |uk(u)| \cdot O(1) + O(a_n)$$

$$= \frac{4a_n}{\delta} \sup_{|u|\geq\delta_n/2} |uk(u)| \cdot O(1) + O(a_n)$$

$$= O(a_n) + O(a_n) = O(a_n)$$

which was to be shown.

Now let $\sigma_n^2 = \mathrm{Var}(c \cdot Z_n^t)$ so that by (1)-(6) above, we have

$$\sigma_n^2 = ||k||_2^2 \sum_{s=1}^{n} [c_s^2 g(x_s) + d_s^2 v(x_s) + 2c_s d_s w(x_s)] + O(a_n).$$

Put $\rho_{ni}^3 = E\left|\frac{c \cdot W_{ni}}{\sqrt{n}}\right|^3$ and $\rho_n^3 = \sum_{i=1}^{n} \rho_{ni}^3$ so that

$$\rho_n^3 = n^{-1/2} E|c \cdot W_{n1}|^3$$

$$\leq n^{-1/2}|c|^3 E|W_{n1}|^3$$

$$\leq 8n^{-1/2}|c|^3 \max_{s=1,2} \{E|U_{n1}(x_s)|^3,\ E|V_{n1}(x_s)|^3\}.$$

Since g', w', v' and k are bounded and $E_f|Y|^3$ is finite it follows by arguments similar to those above that $E|U_{ni}(x_s)|^3 = O(a_n^{-1/2})$ and $E|V_{ni}(x_s)|^3 = O(a_n^{-3/2})$ $(s=1,2)$ so that $\rho_n^3 = O(a_n^{-3/2} n^{-1/2})$.

For $c \neq 0$ we can deduce from Lemma 6.8 that A is positive definite whenever $g(x_1) > 0$ and $g(x_2) > 0$. Thus for $c \neq 0$

$$\lim_{n\to\infty} \frac{\sigma_n^2}{||k_2||^2} = cAc^t > 0$$

since cAc^t is a quadratic form associated with the positive definite matrix A. Hence it follows that $\lim_{n\to\infty} \frac{\rho_n}{\sigma_n} = 0$ whenever $c \neq 0$.

An application of the Berry-Esseen Theorem on page 298 of [9] now completes the proof.

<u>Lemma 6.13</u>. Under the conditions of Lemma 6.12 Z_n converges in distribution to Z (recall that Z is multivariate normal with mean vector zero and covariance matrix A).

Proof: Apply Lemma 6.12 together with Theorem (xi) on page 103 of [14].

Let us write

$$Z_n^* = a_n^{1/2} n^{-1/2} \begin{bmatrix} \sum_{i=1}^{n} [U_{ni}^*(x_1) - g(x_1)] \\ \sum_{i=1}^{n} [V_{ni}^*(x_1) - w(x_1)] \\ \sum_{i=1}^{n} [U_{ni}^*(x_2) - g(x_2)] \\ \sum_{i=1}^{n} [V_{ni}^*(x_2) - w(x_2)] \end{bmatrix} .$$

<u>Lemma 6.14</u>. Suppose $\int uk(u)du = 0$, $\int u^2k(u)du$ is finite and $na_n^5 \to 0$. If $g^{(2)}$ and $w^{(2)}$ exist and are bounded then, under the conditions of Lemma 6.13, Z_n^* converges in distribution to Z.

Proof: Let $B_n = \left(g(x_1) - EU_{n1}^*(x_1),\ w(x_1) - EV_{n1}^*(x_1),\ g(x_2) - EU_{n1}^*(x_2),\ w(x_2) - EV_{n1}^*(x_2)\right)^t$. Since $\int uk(u)du = 0$, $\int u^2k(u)$ is finite and $g^{(2)}$ is bounded, it follows that

$$|EU_{n1}^*(x_i) - g(x_i)| = \left|\int k(u)\{g(x_i - a_n u) - g(x_i)\}du\right|$$

$$\leq \sup_x |g^{(2)}(x)| \frac{a_n^2}{2} \int u^2k(u)du = O(a_n^2) \quad (i=1,2).$$

Similarly $|EV_{ni}(x_i) - w(x_i)| = O(a_n^2)$ so that $B_n = O(a_n^2)$. Then $Z_n - Z_n^* = \sqrt{na_n} B_n = O\left(\sqrt{na_n^5}\right) = o(1)$ since $na_n^5 \to 0$. The desired result now follows from standard large sample theory and Lemma 6.13.

Theorem 6.15. Under the conditions of Lemma 6.14

$$\sqrt{na_n}\begin{bmatrix} m_n^{(2)}(x_1) - m(x_1) \\ \\ m_n^{(2)}(x_2) - m(x_2) \end{bmatrix}$$

converges in distribution to Z^* where Z^* is bivariate normal with mean vector zero and covariance matrix C given by

$$C = ||k||_2^2 \begin{bmatrix} \frac{1}{g(x_1)} V[Y|X = x_1] & 0 \\ \\ 0 & \frac{1}{g(x_2)} V[Y|X = x_2] \end{bmatrix}.$$

Proof: Let the function h from R^4 to R^2 be defined so

$$h(y_1,y_2,y_3,y_4) = (h_1(y_1,y_2,y_3,y_4),\ h_2(y_1,y_2,y_3,y_4))^t,$$

where $h_1(y_1,y_2,y_3,y_4) = y_2/y_1$, and $h_2(y_1,y_2,y_3,y_4) = y_4/y_3$ and let $\theta = (g(x_1),\ w(x_1),\ g(x_2),\ w(x_2))$. Then

$$D = \begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} & \frac{\partial h_1}{\partial y_3} & \frac{\partial h_1}{\partial y_4} \\ \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} & \frac{\partial h_2}{\partial y_3} & \frac{\partial h_2}{\partial y_4} \end{bmatrix}_{(y_1,y_2,y_3,y_4) = \theta}$$

$$= \begin{bmatrix} \frac{-w(x_1)}{g^2(x_1)} & \frac{1}{g(x_1)} & 0 & 0 \\ 0 & 0 & \frac{-w(x_2)}{g^2(x_2)} & \frac{1}{g(x_2)} \end{bmatrix}$$

and hence $DAD^t = C$. Let $T_n = (T_{n1}, T_{n2}, T_{n3}, T_{n4})$ where

$$T_{n1} = \frac{1}{n} \sum_{i=1}^{n} U^*_{ni}(x_1)$$

$$T_{n2} = \frac{1}{n} \sum_{i=1}^{n} V^*_{ni}(x_1)$$

$$T_{n3} = \frac{1}{n} \sum_{i=1}^{n} V^*_{ni}(x_2)$$

$$T_{n4} = \frac{1}{n} \sum_{i=1}^{n} V^*_{ni}(x_2).$$

Let us now write

$$Z^*_n = \sqrt{n}a_n \begin{bmatrix} T_{n1} - g(x_1) \\ T_{n2} - w(x_1) \\ T_{n3} - g(x_2) \\ T_{n4} - w(x_2) \end{bmatrix}.$$

We note that the proof of Theorem (ii) on page 321 of [14] remains valid if $\sqrt{n}$ is replace by r_n provided r_n tends to infinity.

Thus it now follows from the above-mentioned theorem and Lemma 6.14 that $\sqrt{na_n}(h(T_n) - h(\theta))$ converges in distribution to Z^* where Z^* is $N(0,DAD^t)$. Since

$$h(T_n) - h(\theta) = \begin{bmatrix} \frac{T_{n2}}{T_{n1}} - \frac{w(x_1)}{g(x_1)} \\ \\ \frac{T_{n4}}{T_{n3}} - \frac{w(x_2)}{g(x_2)} \end{bmatrix} = \begin{bmatrix} m_n^{(2)}(x_1) - m(x_1) \\ \\ m_n^{(2)}(x_2) - m(x_2) \end{bmatrix}$$

the proof is complete.

REFERENCES

[1] Bhattacharya, P. K. (1967). "Estimation of a Probability Density Function and its Derivatives". Sankhya, Series A, 29, Part 4 (in press).

[2] Cacoullos, T. (1966). "Estimation of a Multivariate Density". Ann. Inst. Statist. Math. 18, 179-189.

[3] Craswell, K. J. (1965). "Density Estimation in a Topological Group". Ann. Math. Stat. 36, 1047-1048.

[4] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator". Ann. Math. Stat. 27, 642-669.

[5] Fix, E. and Hodges, J. L., Jr. Discriminatory Analysis, Non-parametric Discrimination: Consistency Properties. USAF School of Aviation Medicine, Project No. 21-49-004, Report No. 4,(N. D.).

[6] Hoeffding, W. (1963). "Probability Inequalities for Sums of Bounded Random Variables". J. Amer. Statist. Assoc. 58, 13-30.

[7] Kiefer, J. and Wolfowitz, J. (1958). "On the Deviations of the Empiric Distribution Function of Vector Chance Variables". Trans. Am. Math. Soc. 87, 173-186.

[8] Kolmogorov, A. (1941). "Confidence Limits for an Unknown Distribution Function". Ann. Math. Stat. 12, 461-463.

[9] Loève, M. (1963). Probability Theory (3rd ed.). D. Van Nostrand, Princeton, New Jersey.

[10] Murthy, V. K. (1965). "Estimation of Jumps, Reliability and Hazard Rate". Ann. Math. Stat. 36, 1032-40.

[11] Nadaraya, E. A. (1965). "On Non-parametric Estimates of Density Functions and Regression Curves". Theory Prob. Appl. 10, 186-190.

[12] Natanson, I. P. (1955). Theory of Functions of a Real Variable (Rev. ed.), Vol. I, Ungar, New York.

[13] Parzen, E. (1962). "On Estimation of a Probability Density Function and Mode". Ann. Math. Stat. 33, 1065-1076.

[14] Rao, C. R. (1965). Linear Statistical Inference and Its Applications. John Wiley and Sons, New York.

[15] Rosenblatt, M. (1956). "Remarks on Some Nonparametric Estimates of a Density Function". Ann. Math. Stat. 27, 832-837.

[16] Watson, G. S. and Leadbetter, M. R. (1963). "On the Estimation of the Probability Density, I". Ann. Math. Stat. 34, 480-491.

[17] Woodroofe, M. (1967). "On the Maximum Deviation of the Sample Density". Ann. Math. Stat. 38, 475-481.