THE ESTIMATION OF
PROBABILITY DENSITY FUNCTIONS

by

Alfred George Waterman

B.Sc., University of British Columbia, 1963

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics

SIMON FRASER UNIVERSITY

July, 1970

APPROVAL

Name: Alfred George Waterman

Degree: Master of Science

Title of Thesis: The Estimation of Probability Density Functions

Examining Committee:

(R. R. Rennie)
Senior Supervisor

(A. Das)
Examining Committee

(D. J. Mallory)
Examining Committee

(R. A. Holmes)
Examining Committee

Date Approved: July 29, 1970

ABSTRACT

This paper is an exposition on the following non-parametric estimation problem: given a random sample $\{X_j\}_{j=1}^{\infty}$ whose corresponding distribution function is absolutely continuous, how can one estimate the density function $f(x)$? In particular, the techniques suggested by E. Parzen, Loftsgaarden and Quesenberry, and Schwartz are discussed.

Parzen (AMS 33) considered estimates of the form $\hat{f}n(x) = \frac{1}{nh(n)} \sum_{j=1}^{n} K(\frac{x-X_j}{h(n)})$ where K is essentially a probability density function and $h(n)$ is a sequence of positive numbers converging to zero. When x is a point of continuity of $f(x)$, Parzen has established that the sequence of estimates $\{\hat{f}n(x)\}_{n=1}^{\infty}$ is an asymptotically unbiased and consistent (in quadratic mean) estimate of $f(x)$. Consistency in mean integrated square error, rates of convergence, and the construction of estimators with optimal convergence properties are also discussed. (Watson and Leadbetter, AMS 34.)

Loftsgaarden and Quesenberry (AMS 36) have introduced an estimator of the form $\hat{f}n(x) = \{\frac{k(n)-1}{n}\}\{\frac{1}{2r_{k(n)}(x)}\}$ where $k(n)$ is a non-decreasing sequence of positive integers such that $k(n) \to \infty$ as $n \to \infty$, $k(n) = o(n)$ and $r_{k(n)}(x)$ is the distance from x to the $k(n)^{th}$ closest observation among $\{X_j\}_{j=1}^{n}$. At points x where $f(x)$ is positive and continuous, they have shown that $\hat{f}n(x)$ is a consistent estimate of $f(x)$ in the sense that $\hat{f}n(x) \to f(x)$ in probability.

For a density function which is square integrable over the reals, Schwartz (AMS 38) has discussed estimators of the form

$\hat{f}n(x) = \sum_{j=0}^{g(n)} \hat{a}_{jn}\phi_j(x)$ where $\phi_j(x)$ is the j^{th} Hermite function,

$\hat{a}_{jn} = \frac{1}{n} \sum_{i=1}^{n} \phi_j(X_i)$ and $g(n)$ is an integer such that $g(n) = o(n)$.

Conditions on $g(n)$ and $f(x)$ are given such that $\hat{f}n(x)$ is a consistent estimate of $f(x)$ in the mean integrated square error sense and the quadratic mean sense.

This paper also compares the three methods, indicates applications and discusses their generalizations.

TABLE OF CONTENTS

58

Page

ACKNOWLEDGEMENTS

The author is pleased to extend his gratitude to Dr. R. R. Rennie, Department of Mathematics, Simon Fraser University, for suggesting the topic, for his encouragement and for his many helpful suggestions. The author also wishes to thank his wife and children for their encouragement and understanding. Finally, thanks to Miss Sue Smithe, Department of Mathematics, for her superb typing of the manuscript.

GLOSSARY OF SYMBOLS

c.d.f.	cumulative distribution function
d.f.	density function
i.i.d.	independent and identically distributed
i.i.d. r.v.'s	independent and identically distributed random variables
LDCT	Lebesque Dominated Convergence Theorem
MISE	mean integrated square error
MSE	mean square error
r.v.	random variable
R	the set of all real numbers
$L_1(R)$	the set of all Lebesque measurable real valued functions such that $\int f(x)dx$ exists and is finite
$L_2(R)$	the set of all Lebesque measurable real valued functions such that $\int f^2(x)dx$ exists and is finite
[An]	item n in the Appendix

$\int f(x)dx \equiv \int_{-\infty}^{\infty} f(x)dx$

CHAPTER 0 - THE INTRODUCTION

The non-parametric estimation of density functions has received an increased amount of attention during the last few years including that of several outstanding statisticians such as M. S. Bartlett [2], E. Parzen [21], M. Rosenblatt [23], and G. S. Watson [32]. Although there are several approaches to this particular estimation problem, in this paper we shall discuss some of those results which have been obtained by Loftsgaarden and Quesenberry [17], Parzen [21] and Schwartz [26].

In the subsequent sections of the introduction we shall define the problem, trace the historical development of techniques of density estimation and present the form of the estimators which we shall consider.

0.1 Definition of the problem and basic assumptions

Let X be a r.v. whose distribution function $F(x)$ is absolutely continuous over the reals. That is, there is a $f(t) \in L_1(R)$ such that

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad \text{for all} \quad x \in R \tag{0.1.1}$$

$f(t)$ is called the density function of the r.v. X.

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent r.v.'s identically distributed as the r.v. X. We shall consider estimates

$\hat{f}n(x)$ of $f(x)$ of the general form

$$\hat{f}n(x) = \hat{f}n(X_1, X_2, \ldots, X_n ; x) \tag{0.1.2}$$

and assume throughout that we have an absolutely continuous distribution function.

0.2 Historical development

A well known estimate of the density function is the common histogram. However, it is dependent on the arbitrary choice of class intervals and provides only a step function approximation.

Continuous approximations of $F(x)$ and $f(x)$ were considered by Gram-Charlier and Edgeworth, ([7], pp.221-231). These estimates are expressed in terms of the normal distribution function and its derivatives as well as the central moments of the random variable. However, in general, the estimates are unsatisfactory since we must assume that the central moments of all orders are finite. In addition, for pointwise convergence of the estimates we must assume that $\int \exp\left(\frac{x^4}{4}\right) dF(x) < \infty$, $f(x)$ is of bounded variation on R and x is a point of continuity of $f(x)$.

For density functions not satisfying the assumptions required for the Edgeworth series approximation, the first reasonable estimates were expressed in terms of estimates of the distribution function. For this reason, we shall now consider estimates of the c.d.f. Given a random sample $X_1, X_2, \ldots, X_n$, the following function is a well known estimate of $F(x)$, ([6], p.123):

$$\text{for all } x \in R, \quad Fn(x) = \frac{1}{n}\{\text{number of } X_i \leq x \text{ where } i = 1, \ldots, n\} \tag{0.2.1}$$

Fn(x) enjoys the following properties:

Letting $E\{\cdot\}$ denote Mathematical Expectation, we have

$$E\{Fn(x)\} = F(x) \quad \text{for all} \quad x \in R \quad \text{and} \quad n = 1,2,... \qquad (0.2.2)$$

That is, Fn(x) is an unbiased estimate of F(x). ([6], p.123).

In addition to this desirable statistical property, Glivenko-Cantelli have established the convergence property

$$\sup_{x \in R}|Fn(x)-F(x)| \to 0(\text{a.e.}) \quad \text{as} \quad n \to \infty. \quad ([6], \text{p.124}).(0.2.3)$$

From (0.1.1), we have that $F'(x)=f(x)$ (a.e.) and with (0.2.2) and (0.2.3) it seems reasonable to consider an estimate of f(x) of the form

$$\hat{f}n(x) = \frac{Fn(x+h)-Fn(x-h)}{2h} \qquad (0.2.4)$$

In fact, (0.2.4) is the kind of estimator which Rosenblatt, [23], first considered. He constructed a more general estimator by proceeding as follows:

Let $\{K_n(u)\}_{n=1}^{\infty}$ be a sequence of functions with the properties

(a) $K_n(u) \geq 0$ for all $u \in R$ and $n = 1,2,...$

(b) $\int K_n(u)du = 1$ for all n.

(c) for all $\varepsilon > 0$, $\lim_{n\to\infty} \int_{|u|<\varepsilon} K_n(u)du = 1$

Corresponding to each sequence $\{K_n(u)\}_{n=1}^{\infty}$ satisfying (a), (b) and (c) an estimator of the form

$$\hat{f}n(x) = \frac{1}{n}\sum_{j=1}^{n} K_n(x-X_j) \quad (0.2.5)$$

was proposed.

We shall now show that Rosenblatt's estimator (0.2.5) is a generalization of the sample c.d.f. estimator (0.2.4). Suppose that

$$K_n(u) = \frac{1}{h}K(\frac{u}{h}) \quad (0.2.6)$$

where $h = h(n) \to 0$ as $n \to \infty$ and $\int K(u)du = 1$.

If $K(u)$ is defined by

$$K(u) = \begin{cases} \frac{1}{2} & \text{if } |u|<1 \\ 0 & \text{otherwise} \end{cases} \quad (0.2.7)$$

Then

$$\hat{f}n(x) = \frac{1}{n}\sum_{j=1}^{n} K_n(x-X_j) \quad (0.2.8)$$

$$= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h}K(\frac{x-X_j}{h}) \quad \text{by (0.2.6)}$$

$$= \frac{1}{nh}\sum_{j=1}^{n}K(\frac{x-X_j}{h})$$

However, notice that by (0.2.7)

$$K(\frac{x-X_j}{h}) = \frac{1}{2} \text{ if and only if } |\frac{x-X_j}{h}| < 1$$

$$\text{if and only if } x-h < X_j < x + h$$

Hence,

$$\hat{f}n(x) = \frac{1}{nh}\{\frac{\text{no.of } X_j \in (x-h,x+h]}{2}\} \quad \text{by (0.2.7)}$$

$$= \frac{Fn(x+h)-Fn(x-h)}{2h} \quad \text{by } (0.2.1)$$

From (0.2.5), estimators with desired properties may be constructed by choosing suitable weighting functions $K_n(u)$. For example, if $K_n(u)$ is continuous, then $\hat{f}n(x)$ is a continuous estimate of $f(x)$.

With regard to the statistical properties of (0.2.5),

$$E\{\hat{f}n(y)\} = E\{\frac{1}{n}\sum_{j=1}^{n} K_n(y-X_j)\}$$

$$= E\{K_n(y-X)\}$$

$$= \int K_n(y-x)\ f(x)dx$$

If $K_n(\cdot)$ is of the form (0.2.6), we have

$$E\{\hat{f}n(y)\} = \int K(u)f(y+hu)du$$

If f is also continuous on the reals, it follows that f is bounded and by the Lebesque Dominated Convergence Theorem we can show that

$$\lim_{n\to\infty} E\{\hat{f}n(y)\} = f(y).$$

Hence the sequence of estimates $\{\hat{f}n(y)\}_{n=1}^{\infty}$ is asymptotically unbiased. We would like to have

$$E\{\hat{f}n(x)\} = f(x) \quad \text{for every} \quad n = 1,2,...$$

However, Rosenblatt has established that if a density function satisfies relatively mild regularity conditions such as continuity or differentiability then

"There exists no unbiased estimates of the density function." [A1]

Although Rosenblatt was the first to publish estimates of the form (0.2.5), it was Parzen, [21], who presented the most referred to discussion. It is difficult to ascertain who originally conceived this method of density estimation since the method is an application of a technique which has been used in Time Series analysis for the purpose of estimating the spectral density function.

We shall conclude the introduction with a brief description of the three most popular techniques of density estimation.

0.3 Techniques for constructing an estimator

In 1962, Parzen [21], considered estimators of $f(x)$ of the form $\hat{f}n(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)$ which are a special case of the more general classes of estimators defined by Rosenblatt. (0.2.8).

In 1965, Loftsgaarden and Quesenberry [17], introduced an estimator of the form

$$\hat{f}n(x) = \left\{\frac{k(n)-1}{n}\right\}\left\{\frac{1}{2r_{k(n)}(x)}\right\}$$ where $k(n)$ is a non-decreasing sequence of positive integers such that $k(n) \to \infty$ as $n \to \infty$, $k(n) = O(n)$ and $r_{k(n)}(x)$ is a random distance function.

In 1967, Schwartz [26], considered the case of a density function which is square integrable over the reals. Assuming $f(x) \in L_2(R)$, we have the representation

$$f(x) = \sum_{j=0}^{\infty} a_j \phi_j(x)$$ where $\{\phi_j(x)\}_{j=0}^{\infty}$ is an orthonormal subset of $L_2(R)$

and $a_j = \int \phi_j(x) f(x) dx$. We shall examine an estimator of the form

$$\hat{f}n(x) = \sum_{j=0}^{q(n)} \hat{a}_{jn} \phi_j(x) \quad \text{where} \quad \hat{a}_{jn} = \frac{1}{n} \sum_{i=1}^{n} \phi_j(X_i)$$

and $q(n)$ is an integer dependent on n such that $q(n) = o(n)$.

CHAPTER 1 - THE KERNEL METHOD

§1.1 Some Asymptotic Properties of Kernel Estimates

In the introduction, we motivated an estimator of the form

$$\hat{f}n(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right) \tag{1.1.1}$$

and remarked that (1.1.1) is a generalization of the sample distribution function estimator. With $\hat{f}n(x)$ expressed as in (1.1.1) we are made aware of a multitude of possible estimates of $f(x)$. For example,

$h(n) = n^{-p}$ for $p>0$ satisfies the conditions

$h(n)$ is a sequence of positive constants and

$$\lim_{n\to\infty} h(n) = 0$$

Also the choice of $K(\cdot)$ is quite arbitrary. With these remarks in mind, how should we choose K and h ? The following discussion will establish that K should be a Borel measurable function.

Recall that $\{X_i\}_{i=1}^{n}$ is a random sample. Therefore, if we assume that K is a Borel function, then $\{K(\frac{x-X_i}{h})\}_{i=1}^{n}$ is a sequence of independent random variables identically distributed as a r.v. $K(\frac{x-X}{h})$.

Regarding $\{\hat{f}n(x)\}_{n=1}^{\infty}$ as a sequence of r.v., we shall now consider some statistical properties which are desirable for $\hat{f}n(x)$ to satisfy.

Estimators are usually deemed "good" if they satisfy some of the following properties:

(a) unbiasedness

(b) minimum variance

(c) consistency

(d) efficiency

(e) have a known and tractable distribution.

In addition to these statistical properties, if we regard $\{\hat{f}n(x)\}$ as a sequence of functions defined on the reals, we can consider modes of convergence of $\hat{f}n(x)$ to $f(x)$. For example, pointwise convergence, convergence in probability and convergence in mean.

If we wish $\hat{f}n(x)$ to satisfy some statistical or convergence properties, this will force K and h to have properties in addition to those stated. However, K and h will still be fairly general.

Conditions on K and h such that $\{\hat{f}n(x)\}_{n=1}^{\infty}$ is asymptotically unbiased are given in Parzen's Theorem (1.1.3) and its corollary as follows:

Theorem (1.1.3)

Suppose $K(y)$ is a Borel function satisfying

$$\sup_{y\in R} |K(y)| < \infty, \int |K(y)|dy < \infty \quad \underline{\text{and}} \quad \lim_{y\to\infty} |yK(y)| = 0.$$

Let $g(y)$ satisfy $\int |g(y)|dy < \infty$.

Let $h(n)$ satisfy (1.1.2). Define $gn(x)$ as

$$gn(x) = \frac{1}{h(n)} \int K\left(\frac{y}{h(n)}\right) g(x-y)\,dy$$

Then at every point x of continuity of $g(\cdot)$,

$$\lim_{n\to\infty} gn(x) = g(x)\int K(y)\,dy$$

Proof

$$gn(x)-g(x)\int K(y)\,dy = \frac{1}{h(n)}\int K\left(\frac{y}{h(n)}\right)g(x-y)\,dy - g(x)\int K(y)\,dy$$

$$= \frac{1}{h(n)}\int K\left(\frac{y}{h(n)}\right)g(x-y)\,dy - g(x)\frac{1}{h(n)}K\left(\frac{y}{h(n)}\right)dy$$

$$= \int\{g(x-y)-g(x)\}\frac{1}{h(n)}K\left(\frac{y}{h(n)}\right)dy$$ since the indicated integrals exist.

Hence

$$\left|gn(x)-g(x)\int K(y)\,dy\right| = \left|\int\{g(x-y)-g(x)\}\frac{1}{h(n)}K\left(\frac{y}{h(n)}\right)dy\right|$$

$$\le\int|g(x-y)-g(x)|\frac{1}{h(n)}\left|K\left(\frac{y}{h(n)}\right)\right|dy$$

Letting $\delta > 0$ and splitting the region of integration into $|y|\le\delta$ and $|y|>\delta$, we have

$$\int|g(x-y)-g(x)|\frac{1}{h(n)}\left|K\left(\frac{y}{h(n)}\right)\right|dy = \int_{|y|\le\delta}|g(x-y)-g(x)|\frac{1}{h(n)}\left|K\left(\frac{y}{h(n)}\right)\right|dy$$

$$+\int_{|y|\ge\delta}|g(x-y)-g(x)|\frac{1}{h(n)}\left|K\left(\frac{y}{h(n)}\right)\right|dy$$

$$\le\max_{|y|\le\delta}|g(x-y)-g(x)|\int_{|y|\le\delta}\frac{1}{h(n)}\left|K\left(\frac{y}{h(n)}\right)\right|dy$$

$$+\int_{|y|\ge\delta}\{|g(x-y)|+|g(x)|\}\frac{1}{h(n)}\left|K\left(\frac{y}{h(n)}\right)\right|dy$$

Let $z = \frac{y}{h(n)}$, then $|y|\le\delta$ implies $|z|\le\frac{\delta}{h(n)}$ and $|y|\ge\delta$ implies

$$|z| \geq \frac{\delta}{h(n)}$$

Therefore,

$$\int |g(x-y)-g(x)| \left|\frac{1}{h(n)}\right| \left|K\left(\frac{y}{h(n)}\right)\right| dy \leq \max_{|z| \leq \delta} |g(x-y)-g(x)| \int_{|z| \leq \frac{\delta}{h(n)}} |K(z)| dz$$

$$+ \int_{|z| \geq \frac{\delta}{h(n)}} |g(x-y)| \left|\frac{1}{h(n)}\right| \left|K\left(\frac{y}{h(n)}\right)\right| dy$$

$$+ |g(x)| \int_{|y| \geq \delta} \left|\frac{1}{h(n)}\right| \left|K\left(\frac{y}{h(n)}\right)\right| dy$$

$$\leq \max_{|y| \leq \delta} |g(x-y)-g(x)| \int |K(z)| dz$$

$$+ \int_{|y| \geq \delta} \left|\frac{g(x-y)}{|y|}\right| \frac{|y|}{h(n)} \left|K\left(\frac{y}{h(n)}\right)\right| dy$$

$$+ |g(x)| \int_{|z| \geq \frac{\delta}{h(n)}} |K(z)| dz$$

where the second integral follows since $|y| \geq \delta > 0$.

Thus,

$$\int |g(x-y)-g(x)| \left|\frac{1}{h(n)}\right| \left|K\left(\frac{y}{h(n)}\right)\right| dy \leq \max_{|y| \leq \delta} |g(x-y)-g(x)| \int |K(z)| dz$$

$$+ \frac{1}{\delta} \sup_{z \geq \frac{\delta}{h(n)}} |zK(z)| \int |g(y)| dy$$

$$+ |g(x)| \int_{|z| \geq \frac{\delta}{h(n)}} |K(z)| dz$$

since $|y| \geq \delta > 0$ implies $\frac{1}{|y|} \leq \frac{1}{\delta}$

Upon letting $n \to \infty$ and then $\delta \to 0$ the result follows as we now show

For a fixed $\delta > 0$, letting $n \to \infty$ implies $h(n) \to 0$ and $\frac{\delta}{h(n)} \to \infty$

Since $\lim_{z \to \infty} |zK(z)| = 0$ and $\frac{\delta}{h(n)} \to \infty$ we have $\sup_{z \geq \frac{\delta}{h(n)}} |zK(z)| \to 0$.

Also we have $\int |g(y)| dy < \infty$

Thus for a fixed $\delta > 0$, $\frac{1}{\delta} \sup_{z \geq \frac{\delta}{h(n)}} |zK(z)| \int |g(y)| dy \to 0$ as $n \to \infty$

Again for a fixed $\delta > 0$,

$$|g(x)| \int_{|z| \geq \frac{\delta}{h(n)}} |K(z)| dz \to 0 \quad \text{as} \quad n \to \infty \quad \text{since} \quad \int |K(z)| dz < \infty$$

Finally letting $\delta \to 0$, $\max_{|y| \leq \delta} |g(x-y)-g(x)| \int |K(z)| dz \to 0$

since g is continuous at x and the indicated integral is finite.

As a special case of theorem (1.1.3) if $\int K(y) dy = 1$, then at every point x of continuity of $g(\cdot)$, $\lim_{n \to \infty} gn(x) = g(x)$. Hence,

<u>Corollary (1.1.4)</u>

<u>If</u> $\lim_{n \to \infty} h(n) = 0$ <u>where</u> $h(n)$ <u>is a sequence of positive constants</u> and $K(y)$ satisfies $\sup_{y \in R} |K(y)| < \infty$, $\int |K(y)| dy < \infty$, $\lim_{y \to \infty} |yK(y)| = 0$ <u>and</u> $\int K(y) dy = 1$ <u>and if the probability density function</u> $f(x)$ <u>is continuous at</u> x, <u>then</u>

$$\lim_{n \to \infty} E\{\hat{f}n(x)\} = f(x)$$

Proof

For estimators of the form (1.1.1), we have

$$E\{\hat{f}n(x)\} = E\{\frac{1}{h}K(\frac{x-X}{h})\} \tag{1.1.5}$$

$$= \int_{-\infty}^{\infty} \frac{1}{h(n)} K(\frac{x-y}{h(n)}) f(y)dy$$

For fixed x, the transformation $z = x-y$ gives

$$E\{\hat{f}n(x)\} = \int_{\infty}^{-\infty} \frac{1}{h(n)} K(\frac{z}{h(n)}) f(x-z)d(-z)$$

$$= \int_{-\infty}^{\infty} \frac{1}{h(n)} K(\frac{z}{h(n)}) f(x-z)dz$$

Hence by theorem (1.1.3) and the fact that $\int_{-\infty}^{\infty} K(z)dz = 1$

$$\lim_{n\to\infty} E\{\hat{f}n(x)\} = f(x)$$

Thus if K satisfies the conditions of the corollary, and if $f(x)$ is continuous at x, then estimates of the form (1.1.1) are asymptotically unbiased.

We shall call K a weighting function if K is an even function satisfying the conditions in (1.1.4). That is,

Definition (1.1.6) $K(y)$ is a weighting function if

$$K(y)=K(-y),\ \sup_{y\in R}|K(y)| < \infty,\ \int|K(y)|dy < \infty$$

$$\int K(y)dy = 1 \quad \text{and} \quad \lim_{y\to\infty}|yK(y)| = 0$$

The standard normal density function is a weighting function. Other weighting functions and their Fourier transforms, $k(u)$, are displayed in Table 1.

TABLE 1

$K(y)$	$k(u)=\int e^{iuy}K(y)\,dy$	$\int K^2(y)\,dy=\frac{1}{2\pi}\int k^2(u)\,du$
$\begin{cases}\frac{1}{2} \text{ if } \|y\|\le 1\\ 0 \text{ if } \|y\|>1\end{cases}$	$\sin u/u$	$\frac{1}{2}$
$\begin{cases}1-\|y\| & \text{if } \|y\|\le 1\\ 0 & \text{if } \|y\|>1\end{cases}$	$\left\{\frac{\sin(u/2)}{u/2}\right\}^2$	$\frac{2}{3}$
$\begin{cases}\frac{4}{3}-8y^2+8\|y\|^3 & \text{if } \|y\|<\frac{1}{2}\\ \frac{8}{3}(1-\|y\|)^3 & \text{if } \frac{1}{2}\le\|y\|\le 1\\ 0 & \text{if } y>1\end{cases}$	$\left\{\frac{\sin(u/4)}{u/4}\right\}^4$	.96
$(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}y^2}$	$e^{-\frac{1}{2}u^2}$	$(2\pi^{\frac{1}{2}})^{-1}$
$\frac{1}{2}e^{-\|y\|}$	$(1+u^2)^{-1}$	$\frac{1}{2}$
$\frac{1}{\pi}(1+y^2)^{-1}$	$e^{-\|u\|}$	π^{-1}
$\frac{1}{2\pi}\left(\frac{\sin(y/2)}{y/2}\right)^2$	$\begin{cases}1-\|u\| & \text{if } \|u\|\le 1\\ 0 & \text{if } \|u\|>1\end{cases}$	$(3\pi)^{-1}$

The following remark about weighting functions will be useful in subsequent sections of the chapter.

$$\sup_{y\in R}|K(y)| < \infty \quad \text{and} \quad \int |K(y)|\,dy < \infty \quad \text{implies} \quad \int |K(y)|^{2+\delta}\,dy < \infty \qquad (1.1.7)$$

for all $\delta \geq 0$. [A2].

We shall now investigate conditions for estimators of the form (1.1.1) to be consistent. As a preliminary result we have

Theorem (1.1.8)

If $f(x)$ is continuous at x and $\{h(n)\}_{n=1}^{\infty}$ is a sequence of positive constants such that $\lim_{n\to\infty} h(n) = 0$, then

$$\lim_{n\to\infty} nh \operatorname{Var}(\hat{f}n(x)) = f(x)\int K^2(y)dy$$

Proof

From (1.1.1) $\operatorname{Var}(\hat{f}n(x)) = \frac{1}{n}\operatorname{Var}[\frac{1}{h}K(\frac{x-X}{h})]$

since $\{\frac{1}{h}K(\frac{x-X_j}{h})\}_{j=1}^{n}$ are independent r.v. identically distributed

as a r.v. $\frac{1}{h}K(\frac{x-X}{h})$.

By the definition of $\operatorname{Var}(\cdot)$, we have

$$\operatorname{Var}[\frac{1}{h}K(\frac{x-X}{h})] = E\{[\frac{1}{h}K(\frac{x-X}{h})]^2\} - E^2\{\frac{1}{h}K(\frac{x-X}{h})\}$$

Let $gn^*(x) = hE\{[\frac{1}{h}K(\frac{x-X}{h})]^2\}$

$$= \frac{1}{h}E\{K^2(\frac{x-X}{h})\}$$

$$= \frac{1}{h}\int K^2(\frac{x-y}{h})f(y)dy$$

and notice that gn^* satisfies the conditions of theorem (1.1.3).

Since x is a point of continuity of $f(x)$,

$$\lim_{n\to\infty} gn^*(x) = f(x)\int K^2(y)dy$$

However,

$$nH \operatorname{Var}(fn(x)) = h \operatorname{Var}[\frac{1}{h}K(\frac{x-X}{h})]$$

$$= h\{E\{[\frac{1}{h}K(\frac{x-X}{h})]^2\} - E^2\{\frac{1}{h}K(\frac{x-X}{h})\}\}$$

$$= gn^*(x)-hE^2\{\frac{1}{h}K(\frac{x-X}{h})\} \quad \text{by defn. of } gn^*(x)$$

$$= gn^*(x)-hE^2\{\hat{f}n(x)\} \quad \text{since } E\{\hat{f}n(x)\} = E\{\frac{1}{h}K(\frac{x-X}{h})\}$$

Hence,

$$\lim_{n\to\infty} nh\,\mathrm{Var}(\hat{f}n(x)) = \lim_{n\to\infty} gn^*(x) - \lim_{n\to\infty} hE^2\{\hat{f}n(x)\}$$

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

where the first equality follows since the indicated limits exist and are finite. We have the second equality by (1.1.4) and the fact that $\lim_{n\to\infty} h(n) = 0$.

Corollary (1.1.9)

Let $h(n)$ be a sequence of positive constants satisfying $\lim_{n\to\infty} h(n) = 0$ and $\lim_{n\to\infty} nh(n) = \infty$. If $f(x)$ is continuous at x, then

$$\lim_{n\to\infty} \mathrm{Var}(\hat{f}n(x) = 0$$

Proof

$$\mathrm{Var}(\hat{f}n(x)) = \frac{nh\,\mathrm{Var}(\hat{f}n(x))}{nh} \quad \text{since} \quad n,h > 0$$

Observe that $\lim_{n\to\infty} \mathrm{Var}(\hat{f}n(x)) = \lim_{n\to\infty}\{\frac{nh\,\mathrm{Var}(\hat{f}n(x))}{nh}\}$

The result then follows from the hypothesis and theorem (1.1.8).

As an immediate consequence of corollary (1.1.9) we can state conditions under which the estimate $\hat{f}n(x)$ is consistent in quadratic mean in the sense that

$$\lim_{n\to\infty} E\{|\hat{f}n(x)-f(x)|^2\} = 0 \qquad (1.1.10)$$

We have

$$E\{|\hat{f}n(x)-f(x)|^2 = \mathrm{Var}(\hat{f}n(x)) + b^2[\hat{f}n(x)] \qquad (1.1.11)$$

where $b[\hat{f}n(x)] \equiv E\{\hat{f}n(x)\} - f(x)$ is the bias of $\hat{f}n(x)$.

Corollary (1.1.12)

If $\lim_{n\to\infty} h(n) = 0$, $\lim_{n\to\infty} nh(n) = \infty$, *and* x *is a point of continuity of* $f(x)$, *then* $\hat{f}n(x)$ *is a consistent estimate of* $f(x)$ *in quadratic mean.*

Proof

From (1.1.11), corollary (1.1.9) and corollary (1.1.4)

$$\begin{aligned}\lim_{n\to\infty} E\{|\hat{f}n(x)-f(x)|^2\} &= \lim_{n\to\infty} Var(\hat{f}n(x))\} \\ &+ \lim_{n\to\infty} [E\{\hat{f}n(x)\} - f(x)]^2 \\ &= 0\end{aligned}$$

Recall that $\hat{f}n\to f$ in quadratic mean implies $\hat{f}n\to f$ in probability. Therefore, at points of continuity of the density function, estimates of the form (1.1.1) are consistent estimates of $f(x)$ in the sense that $\hat{f}n\to f$ in probability.

For the following development, we shall assume

(a) $\hat{f}n(x)$ is of the form (1.1.1)

(b) K is a weighting function

(c) h satisfies $\lim_{n\to\infty} h(n) = 0$ and $\lim_{n\to\infty} nh(n) = \infty$.

Theorem (1.1.13)

Under assumptions (a), (b) and (c), if x *is a point of continuity of* $f(x)$, *then the sequence of estimates* $\{\hat{f}n(x)\}_{n=1}^{\infty}$ *is asymptotically normal as well as consistent.*

Proof

First note $\hat{f}n(x) = \frac{1}{n}\sum_{k=1}^{n} V_{nk}$ where $V_{nk} \equiv \frac{1}{h(n)} K(\frac{x-X_k}{h(n)})$ and $\{V_{nk}\}_{k=1}^{n}$ are independent r.v. identically distributed as a r.v.

$V_n = \frac{1}{h(n)} K(\frac{x-X}{h(n)})$. Recall that $\{\hat{f}n(x)\}_{n=1}^{\infty}$ is asymptotically normal if and only if

$$\text{for all } c \in R, \lim_{n\to\infty} P[\frac{\hat{f}n(x)-E\{\hat{f}n(x)\}}{\sigma(\hat{f}n(x))} \le c] = \Phi(c)$$

where Φ is the c.d.f. of the standard normal density function.

From Loève ([16], p.316) a necessary and sufficient condition that $\{\hat{f}n(x)\}_{n=1}^{\infty}$ be asymptotically normal is that

$$\text{for all } \varepsilon > 0, nP\{|\frac{V_n - E\{V_n\}}{\sigma(V_n)}| \ge \varepsilon n^{\frac{1}{2}}\} \to 0 \text{ as } n \to \infty \quad (1.1.14)$$

See [A4].

A sufficient condition for (1.1.14) to hold is that

$$\text{for some } \delta > 0, \frac{E\{|V_n - E\{V_n\}|^{2+\delta}\}}{n^{\delta/2}\sigma^{2+\delta}[V_n]} \to 0 \text{ as } n \to \infty \quad (1.1.15)$$

See [A3].

We shall show that

$$\text{for all } \delta > 0, \lim_{n\to\infty} \frac{E\{|V_n - E\{V_n\}|^{2+\delta}\}}{n^{\delta/2}\sigma^{2+\delta}[V_n]} = 0 \quad (1.1.16)$$

which with (1.1.14) establishes the theorem.

Let $\{s(n)\}_{n=1}^{\infty}$ and $\{t(n)\}_{n=1}^{\infty}$ be two sequences of reals.

Recall that, we say

$$s(n) \sim t(n) \text{ if and only if } \lim_{n\to\infty} \frac{s(n)}{t(n)} = 1$$

For $\delta > 0$, we have

$$E\{|V_n|^{2+\delta}\} = \int |\frac{1}{h} K(\frac{x-y}{h})|^{2+\delta} f(y)dy$$

$$= \frac{1}{h^{1+\delta}}\{\frac{1}{h}\int |K(\frac{x-y}{h})|^{2+\delta} f(y)dy\}$$

However, notice that $|K(\frac{x-y}{h})|^{2+\delta}$ satisfies the hypotheses of theorem (1.1.3).

Therefore at every continuity point x of f,

$$\lim_{n\to\infty} \{\frac{1}{h}\int |K(\frac{x-y}{h})|^{2+\delta} f(y)dy\} = f(x)\int |K(y)|^{2+\delta} dy$$

We claim $\int |\frac{1}{h}K(\frac{x-y}{h})|^{2+\delta} f(y)dy \sim \frac{1}{h^{1+\delta}} f(x)\int |K(y)|^{2+\delta} dy$

because

$$\lim_{n\to\infty} \{\frac{\int |\frac{1}{h}K(\frac{x-y}{h})|^{2+\delta} f(y)dy}{\frac{1}{h^{1+\delta}} f(x)\int |K(y)|^{2+\delta} dy}\}$$

$$= \lim_{n\to\infty} \{\frac{\frac{1}{h}\int |K(\frac{x-y}{h})|^{2+\delta} f(y)dy}{f(x)\int |K(y)|^{2+\delta} dy}\}$$

$$= (f(x)\int |K(y)|^{2+\delta} dy)^{-1} \lim_{n\to\infty} \{\frac{1}{h}\int |K(\frac{x-y}{h})|^{2+\delta} f(y)dy\}$$

$$= 1$$

and we conclude

$$E\{|V_n|^{2+\delta}\} \sim \frac{1}{h^{1+\delta}} f(x)\int |K(y)|^{2+\delta} dy \qquad (1.1.17)$$

We also have

$$Var[V_n] \sim \frac{f(x)}{h}\int K^2(y)dy$$

which follows from

$$Var[V_n] = n\ Var[\hat{f}n(x)]$$

$$= \frac{nh\ Var[\hat{f}n(x)]}{h}$$

and $\left(\frac{nh\ \mathrm{Var}[\hat{f}n(x)]}{h}\right) / \left(\frac{f(x)\int K^2(y)dy}{h}\right) = \frac{nh\ \mathrm{Var}[\hat{f}n(x)]}{f(x)\int K^2(y)dy}$

However,

$$\lim_{n\to\infty} \frac{nh\ \mathrm{Var}[\hat{f}n(x)]}{f(x)\int K^2(y)dy} = \left(f(x)\int K^2(y)dy\right)^{-1} \lim_{n\to\infty} nh\ \mathrm{Var}(\hat{f}n(x))$$

$$= 1 \text{ by theorem (1.1.8)}$$

That is,

$$\mathrm{Var}[V_n] \sim \frac{f(x)}{h}\int K^2(y)dy \qquad (1.1.18)$$

The following properties of "$\sim$" are useful in finishing the proof and in the succeeding material dealing with the Berry-Esseen bound.

(a) $u(n) \neq 0$ for all n and $s(n) \sim t(n)$ implies $u(n)\ s(n) \sim u(n)\ t(n)$

(b) $s(n) \sim t(n)$ implies $(s(n))^r \sim (t(n))^r$ for all real r.

(c) $s_1(n) \sim s_2(n)$, $t_1(n) \sim t_2(n)$ implies $\frac{s_1(n)}{t_1(n)} \sim \frac{s_2(n)}{t_2(n)}$

Since $V_n \geq 0$, observe that

$$\frac{E\{|V_n - E\{V_n\}|^{2+\delta}\}}{n^{\delta/2}\sigma^{2+\delta}[V_n]} \leq \frac{E\{|V_n|^{2+\delta}\}}{n^{\delta/2}\sigma^{2+\delta}[V_n]} \qquad (1.1.19)$$

Now by (1.1.18), (a) and (b),

$$\sigma^{2+\delta}[V_n] \sim \left(\frac{f(x)}{h}\int K^2(y)dy\right)^{1+\frac{\delta}{2}}$$

Thereby, by (a), (c), (1.1.17) and this equivalence

$$\frac{E\{|V_n|^{2+\delta}\}}{n^{\delta/2}\sigma^{2+\delta}[V_n]} \sim \frac{\frac{1}{h^{1+\delta}} f(x)\int |K(y)|^{2+\delta}dy}{n^{\delta/2}\left(\frac{f(x)}{h}\int K^2(y)dy\right)^{1+\delta/2}} \qquad (1.1.20)$$

$$= \frac{f(x)\int |K(y)|^{2+\delta} dy}{(nh)^{\delta/2}(f(x)\int K^2(y)dy)^{1+\delta/2}}$$

With the assumption $nh \to \infty$ as $n \to \infty$, we have

$$(nh)^{\delta/2} \to \infty \quad \text{as} \quad n \to \infty \quad \text{for} \quad \delta > 0.$$

Thus by (1.1.20), for $\delta > 0$

$$\lim_{n\to\infty} \frac{E\{|V_n|^{2+\delta}\}}{n^{\delta/2}\sigma^{2+\delta}[V_n]} = 0$$

and this with (1.1.19) gives (1.1.16) and proves the theorem.

Having established that the sequence of estimates $\{\hat{f}n(x)\}_{n=1}^{\infty}$ is asymptotically normal, it is natural to ask "How quickly does the approximating expression converge to its limiting form?" Some idea of the closeness of the normal approximation can be obtained from the Berry-Esseen bound. Following Loève ([16], p.288) we have for a suitable C in the reals,

$$\sup_{a\in R} \left| P\{\frac{\hat{f}n(x)-E\{\hat{f}n(x)\}}{\sigma[\hat{f}n(x)]} \le a\} - \Phi(a) \right| \le C \frac{E\{|V_n|^3\}}{n^{1/2}\sigma^3[V_n]}$$

since $\hat{f}n(x)$ is a sum of independent r.v. V_{nk} identically distributed as the r.v. V_n and $E\{|V_n|^3\} < \infty$.

Notice that

$$\frac{CE\{|V_n|^3\}}{n^{1/2}\sigma^3[V_n]} \sim \frac{1}{(nhf(x))^{1/2}} \left[\frac{\int |K(y)|^3 dy}{(\int K^2(y)dy)^{3/2}}\right]$$

because if $\delta = 1$ in (1.1.17) we obtain

$$E\{|V_n|^3\} \sim \frac{1}{h^2} f(x)\int |K(y)|^3 dy \tag{1.1.21}$$

$$\sigma^3[v_n] = (\sigma^2[v_n])^{3/2} \sim \left(\frac{f(x)}{h}\int K^2(y)dy\right)^{3/2} \qquad (1.1.22)$$

(1.1.21), (1.1.22) and (b) give the result.

§1.2 Uniform Consistency of Kernel Estimates

We shall now investigate conditions under which the sequence of estimates $\{\hat{f}n(x)\}_{n=1}^{\infty}$ converges uniformly in probability to $f(x)$. In this case, if the mode is unique we are able to obtain consistent estimates of the mode.

Recall that $\{\hat{f}n(x)\}_{n=1}^{\infty}$ is said to <u>converge uniformly in probability to</u> $f(x)$ if

$$\text{for all } \varepsilon > 0, \lim_{n\to\infty} P[\sup_{x\in R} |\hat{f}n(x)-f(x)| < \varepsilon] = 1$$

Let $K(y)$ be a weighting function (1.1.6) and $k(u)$ its Fourier transform.

That is,

$$k(u) = \int e^{-iuy}K(y)dy \qquad (1.2.1)$$

Notice that if $K(y)$ is even, then $k(u)$ is even.

If we assume $k(u)$ is absolutely integrable, then ([6], p.143)

$$K(y) = \frac{1}{2\pi}\int \exp(iuy)k(u)du \qquad (1.2.2)$$

and $K(y)$ is uniformly continuous in y.

Let $\phi n(u) = \int \exp(iux)dFn(x)$

$= \frac{1}{n}\sum_{k=1}^{n} e^{iuX_k}$ be the sample characteristic function.

From these comments, we have

$$\hat{f}n(x) = (nh)^{-1} \sum_{k=1}^{n} K(\frac{x-X_k}{h})$$

$$= (nh)^{-1} \sum_{k=1}^{n} K(\frac{X_k-x}{h}) \quad \text{since } K \text{ is even}$$

$$= (nh)^{-1} \sum_{k=1}^{n} \frac{1}{2\pi}\int \exp[iu(\frac{X_k-x}{h})]k(u)du \quad \text{by (1.2.2)}$$

$$= \frac{1}{2\pi}\int (nh)^{-1} \exp(\frac{-iux}{h}) \sum_{k=1}^{n} \exp(\frac{iuX_k}{h})k(u)du$$

$$= \frac{1}{2\pi}\int h^{-1} \exp(\frac{-iux}{h})\phi n(\frac{u}{h})k(u)du \quad \text{by defn. of } \phi n(\cdot)$$

$$= \frac{1}{2\pi}\int \exp(-iux)k(hu)\phi n(u)du \quad \text{with change of variable}$$

Since $k(hu)\phi n(u)$ is the characteristic function of $fn(x)$ we can write

$$\hat{f}n(x) = \frac{1}{2\pi}\int \exp(-iux)\phi_{\hat{f}n(x)}(u)du \tag{1.2.3}$$

Theorem 1.2.4

If $\{h(n)\}_{n=1}^{\infty}$ satisfies $\lim_{n\to\infty} h(n) = 0$ and $\lim_{n\to\infty} nh^2(n) = \infty$ and if $f(x)$ is uniformly continuous, then for all $\varepsilon > 0$

$$\lim_{n\to\infty} P[\sup_{y\in R}|\hat{f}n(x)-f(x)| < \varepsilon] = 1$$

Proof

Since convergence of $\hat{f}n\to f$ in the mean implies convergence of $\hat{f}n\to f$ in probability, it suffices to show that

$$\lim_{n\to\infty} E^{1/2}\{\sup|\hat{f}n(x)-f(x)|^2\} = 0 \tag{1.2.5}$$

From Corollary (1.1.4) we know that

$$\lim_{n\to\infty} |E\{\hat{f}n(x)\} - f(x)| = 0$$

and since $f(x)$ is uniformly continuous, it follows that

$$\lim_{n\to\infty} \sup_{x\in R} |E\{\hat{f}n(x)\} - f(x)| = 0$$

Therefore, (1.2.5) will follow by the triangle inequality if we can show that

$$\lim_{n\to\infty} E^{1/2}\{\sup_{x\in R} |\hat{f}n(x) - E\{\hat{f}n(x)\}|\} = 0 \qquad (1.2.6)$$

From (1.2.3) and Fubini's Theorem, we have

$$E\{\hat{f}n(x)\} = E\{\frac{1}{2\pi}\int \exp(-iux)k(hu)\phi n(u)du\}$$

$$= \frac{1}{2\pi}\int \exp(-iux)k(hu)E\{\phi n(u)\}du$$

Hence for all $x \in R$,

$$|\hat{f}n(x) - E\{\hat{f}n(x)\}| \le \frac{1}{2\pi}\int |\exp(-iux)||k(hu)||\phi n(u) - E\{\phi n(u)\}|du$$

$$\le \frac{1}{2\pi}\int |k(hu)||\phi n(u) - E\{\phi n(u)\}|du$$

since $|\exp(-iux)| \le 1$

$$\sup_{x\in R} |\hat{f}n(x) - E\{\hat{f}n(x)\}| \le \frac{1}{2\pi}\int |k(hu)||\phi n(u) - E\{\phi n(u)\}|du$$

from which we obtain

$$\sup_{x\in R} |\hat{f}n(x) - E\{\hat{f}n(x)\}|^2 \le (\frac{1}{2\pi}\int |k(hu)||\phi n(u) - E\{\phi n(u)\}|du)^2$$

$$\le \frac{1}{4\pi^2}\int |k(hu)|^2|\phi n(u) - E\{\phi n(u)\}|^2du$$

by the Cauchy-Schwarz Inequality.

Hence,

$$E^{1/2}\{\sup_{x\in R} |\hat{f}n(x) - E\{\hat{f}n(x)\}\} \le \frac{1}{2\pi}\int |k(hu)|E^{1/2}\{|\phi n(u) - E\{\phi n(u)\}|^2\}du$$

by Fubini's Theorem.

Applying Minkowski's inequality, we obtain

$$E^{1/2}\{|\phi n(u)-E\{\phi n(u)\}|^2\} \le E^{1/2}\{|\phi n(u)|^2\} + E^{1/2}\{|E\{\phi n(u)\}|^2\}$$

$$\le 2E^{1/2}\{|\phi n(u)|^2\} \quad \text{since} \quad |E\{\phi n(u)\}|^2$$

$$\le E\{|\phi n(u)|^2\}$$

$$\le \frac{2}{n^{1/2}}$$

Therefore,

$$E^{1/2}\{\sup_{x \in R}|\hat{f}n(x)-E\{\hat{f}n(x)\}|^2\} \le \frac{1}{\pi n^{1/2}}\int |k(hu)|du$$

$$\le \frac{1}{n^{1/2}h}\int |k(u)|du$$

With the assumption $nh^2 \to \infty$ as $n \to \infty$, we have $n^{1/2}h \to \infty$ as $n \to \infty$

Hence,

$$\lim_{n\to\infty} E^{1/2}\{\sup_{x \in R}|\hat{f}n(x)-E\{\hat{f}n(x)\}|^2\} = 0$$

from which (1.2.5) follows and the theorem is proved.

§1.3 Rates of Convergence of Kernel Estimates and Related Results

In sections 1.1 and 1.2 we considered the consistency properties of estimates of the form (1.1.1). By Corollary (1.1.12) if $f(x)$ is continuous at x, then $\hat{f}n(x)$ is a consistent estimate of $f(x)$ in quadratic mean in the sense that

$$\lim_{n\to\infty} E\{|\hat{f}n(x)-f(x)|^2\} = 0$$

A natural question to ask is "How quickly does $\hat{f}n$ converge to f in quadratic mean?"

In order to partially answer this question and related questions, we shall consider the results of Watson and Leadbetter. [32]. We shall estimate $f(x)$ by estimates of the form

$$\hat{f}n(x) = \frac{1}{n}\sum_{j=1}^{n} K_n(x-X_j) \tag{1.3.1}$$

which is the kind of estimator proposed by Rosenblatt (0.2.5). However, we do not assume $\{K_n(u)\}_{n=1}^{\infty}$ is a sequence of weighting functions. Rather we suppose that $K_n(\cdot)$ and $f(x)$ are square integrable over the reals.

Definition (1.3.2) For $K_n(\cdot)$, $f(x) \in L_2(R)$ we define the Mean Integrated Square Error (MISE), J_n, of the estimator $\hat{f}n(x)$ as follows:

$$J_n \equiv E\{\int(\hat{f}n(x)-f(x))^2dx\}$$

Let $\{H(n)\}_{n=1}^{\infty}$ be a sequence s.t. $\lim_{n\to\infty} H(n) = \infty$, then

Definition (1.3.3) The estimator $\hat{f}n(x)$ is integratedly consistent of order $H(n)$ if $\lim_{n\to\infty} H(n)J_n = a \neq 0$ and $a \in R$.

In the succeeding development, various types of estimators will be discussed and their orders of integrated consistency investigated. One class of estimators we shall consider is estimators of the Parzen type. (1.1.1). We shall see that the type of estimator which is appropriate depends largely on the behaviour of the characteristic function $\phi_f(t)$ of the probability density $f(x)$ for large t.

Our criterion of a good estimate will be an estimate which minimizes the MISE. Thus, we shall determine the functions $K_n(\cdot)$ in such a way that we minimize J_n.

For $g(x) \varepsilon L_1(R)$, let

$\phi_g(t) = \int \exp(ixt)g(x)dx$ be the Fourier transform of g.

We shall now present some preliminary results with the following objectives in mind:

(a) to express J_n in terms of $\phi_{\hat{f}n}(t)$ and $\phi_f(t)$,

(b) to express $\phi_{\hat{f}n}(t)$ in terms of $\phi_{K_n}(t)$,

(c) to determine $\phi_{K_n}(t)$, and hence $K_n(x)$, such that we minimize J_n.

From (1.3.2) and Parseval's Theorem [A5], we have

$$J_n = E\{\int(\hat{f}n(x)-f(x))^2 dx\} \tag{1.3.3}$$

$$\frac{1}{2\pi}E\{\int|\phi_{fn}(t)-\phi_f(t)|^2 dt\}$$

which is (a). For (b), recall that (1.2.3)

$$\phi_{\hat{f}n}(t) = \hat{\phi}_n(t)\phi_{K_n}(t) \tag{1.3.4}$$

where $\hat{\phi}_n(t)$ is the sample characteristic function.

From (1.3.3) and (1.3.4)

$$2\pi J_n = E\{\int|\hat{\phi}_n(t)\phi_{K_n}(t)-\phi_f(t)|^2 dt\} \tag{1.3.5}$$

$$= \int[n^{-1}|\phi_{K_n}(t)|^2(1-|\phi_f(t)|^2) + |\phi_f(t)|^2 \ (|1-\phi_{K_n}(t)|^2]dt \quad [A6]$$

and (1.3.5) is equivalent to

$$2\pi J_n = \int (\frac{1}{n}+\frac{n-1}{n}|\phi_f(t)|^2)(|\phi_{K_n}(t) - \frac{n|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2}|^2)dt \quad (1.3.6)$$
$$+ \int \frac{|\phi_f(t)|^2(1-|\phi_f(t)|^2)}{1+(n-1)|\phi_f(t)|^2} dt$$

Since the integrands are real and non-negative $2\pi J_n$ is minimized if we choose $K_n = K_n^*$ in such a way that

$$\phi_{K_n}(t) = \frac{n|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2} \quad (1.3.7)$$

Corresponding to this choice of $\phi_{K_n}(t)$, we have

$$2\pi J_n^* = \int \frac{|\phi_f(t)|^2(1-|\phi_f(t)|^2)}{1+(n-1)|\phi_f(t)|^2} dt \quad (1.3.8)$$

Hence we have accomplished objective (c).

In order to discuss orders of integrated consistency, we show

$$J_n^* = \frac{K_n^*(0)}{n} - \frac{1}{2\pi}\int \frac{|\phi_f(t)|^4}{1+(n-1)|\phi_f(t)|^2} dt \quad (1.3.9)$$

$$= \frac{K_n^*(0)}{n} - o(\frac{1}{n})$$

From (1.3.8),

$$J_n^* = \frac{1}{2\pi}\int \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2} dt - \frac{1}{2\pi}\int \frac{|\phi_f(t)|^4}{1+(n-1)|\phi_f(t)|^2} dt$$

and by the Inversion theorem,

$$K_n^*(x) = \frac{1}{2\pi}\int \exp(-itx)\phi_{K_n^*}(t)dt$$

Consequently,

$$K_n^*(0) = \frac{1}{2\pi}\int \phi_{K_n^*}(t)dt$$

$$= \frac{1}{2\pi}\int \frac{n|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2} dt \qquad \text{by (1.3.7)}$$

and

$$\frac{K_n^*(0)}{n} = \frac{1}{2\pi}\int \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2} dt$$

Since $|\phi_f(t)|^2 \le 1$

$$\frac{1}{2\pi}\int \frac{|\phi_f(t)|^4}{1+(n-1)|\phi_f(t)|^2} dt \le \frac{1}{2\pi(n-1)}\int |\phi_f(t)|^2 dt$$

However, $\int |\phi_f(t)|^2 dt < \infty$, and therefore

$$\frac{1}{2\pi(n-1)}\int |\phi_f(t)|^2 dt = O(\frac{1}{n})$$

Hence, (1.3.9) follows.

We give some examples to illustrate (1.3.7), (1.3.9) and a later result (1.3.12)

<u>Example</u>

Suppose $f(x) = \begin{cases} \exp(-x) & \text{for } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$, then $\phi_f(t) = (1-it)^{-1}$ for $t \in R$

Hence $|\phi_f(t)|^2 = (1+t^2)^{-1}$ and by (1.3.7)

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

$$= \frac{n(1+t^2)^{-1}}{1+(n-1)(1+t^2)^{-1}}$$

Thus by the Inversion theorem

$$K_n^*(x) = \frac{1}{2\pi}\int \exp(-itx)\phi_{K_n^*}(t)dt$$

$$= \frac{2}{2\pi}\int_0^\infty \frac{\cos tx \; n(1+t^2)^{-1}}{1+(n-1)(1+t^2)^{-1}}\, dt$$

$$= \frac{n^{1/2}}{2}\exp(-|x|n^{1/2})$$

Therefore, $K_n^*(0) = \frac{n^{1/2}}{2}$

From (1.3.9),

$$J_n^* = \frac{K_n^*(0)}{n} - o(\frac{1}{n})$$

$$= \frac{1}{2n^{1/2}} - o(\frac{1}{n})$$

and $n^{1/2}J_n^* = \frac{1}{2} - o(n^{-1/2})$

Therefore, $\lim_{n\to\infty} n^{1/2}J_n^* = 1/2$

Hence the estimator $\hat{f}n(x)$, formed by $K_n^*(x)$, is integratedly consistent of order $n^{1/2}$. This result is a special case of (1.3.12) in which $p = 1$.

§1.3A Characteristic Functions Which Decrease Algebraically

We shall begin an investigation of orders of integrated consistency of estimators whose characteristic functions decrease algebraically.

Some preliminary definitions and examples.

Definition (1.3.10) $\phi_f(t)$ decreases algebraically of degree $p > 0$ if

$$\lim_{|t|\to\infty} |t|^p|\phi_f(t)| = C^{1/2} > 0$$

Example

In the previous example, we had $\phi_f(t) = (1-it)^{-1}$ for $t \in R$

Therefore, $|\phi_f(t)| = (1+t^2)^{-1/2}$

Hence $\phi_f(t)$ decreases algebraically of degree 1 since

$$\lim_{|t|\to\infty} |t||\phi_f(t)| = \lim_{|t|\to\infty} |t|(1+t^2)^{-1/2}$$
$$= 1$$

The characteristic function of the Gamma density function decreases algebraically whereas that of the normal density does not.

Definition (1.3.11) $\hat{f}n(x) = \frac{1}{n}\sum_{j=1}^{n} K_n(x-X_j)$ has algebraic form if $\phi_{K_n}(t) = h(A_n t)$ where h is a bounded, even square integrable function and $A_n \to 0$ as $n \to \infty$.

Remark This condition is equivalent to specifying that $K_n(x)$ be of the form

$$K_n(x) = \frac{1}{A_n}k\left(\frac{x}{A_n}\right) \quad \text{where} \quad \phi_k(t) = h(t)$$

because $K_n(x) = \frac{1}{A_n}k\left(\frac{x}{A_n}\right)$ if and only if

$$\phi_{K_n}(t) = \phi_{\frac{1}{A_n}k\left(\frac{x}{A_n}\right)}(t)$$

$$= \phi_k(A_n t)$$

$$= h(A_n t)$$

<u>Example</u> For the exponential density and its optimum weighting function we have,

$$K_n^*(x) = \frac{n^{1/2}}{2} \exp(-|x| n^{1/2})$$

$$= \left(\frac{2}{n^{1/2}}\right)^{-1} \exp\left(-|x| \left(\frac{1}{n^{1/2}}\right)^{-1}\right)$$

Let $A_n = n^{-1/2}$ and $k(x) = \frac{1}{2}\exp(-|x|)$, then

$$K_n^*(x) = \frac{1}{A_n} k\left(\frac{x}{A_n}\right)$$

In this section, our estimates are of the form

$$\hat{f}n(x) = \frac{1}{n} \sum_{j=1}^{n} K_n(x-X_j)$$

Under the assumption, $\hat{f}n(x)$ has algebraic form, $K_n(x-X_j) = \frac{1}{A_n} k\left(\frac{x-X_j}{A_n}\right)$

Hence, $\hat{f}n(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{A_n} k\left(\frac{x-X_j}{A_n}\right)$

and this estimate is of the form considered by Parzen (1.1.1)

We shall now investigate the order of integrated consistency of the optimum estimate.

<u>Theorem (1.3.12)</u>

<u>Let</u> $\phi_f(t)$ <u>decrease algebraically of degree</u> $p > 1/2$, C <u>as in</u> (1.3.10) <u>then</u> J_n^*, <u>the minimum MISE satisfies,</u>

$$\lim_{n\to\infty} n^{1-\frac{1}{2p}} J_n^* = \frac{C^{\frac{1}{2p}}}{2\pi} \int \frac{dt}{1+|t|^{2p}}$$

Proof

From (1.3.9),

$$J_n^* = \frac{1}{2\pi}\int \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2}\,dt - O(n^{-1})$$

We first examine

$$J_n^{**} \equiv \int \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2}\,dt$$

Since $\phi_f(t)$ decreases algebraically of degree $p > 1/2$,

$$\lim_{|t|\to\infty} |t|^p|\phi_f(t)| = C^{1/2} > 0$$

It follows that

$$\lim_{|t|\to\infty} |t|^{-2p}|\phi_f(t)|^{-2} = C^{-1}$$ and we conclude that

for all $\varepsilon > 0$, there is a T such that

$$|t| \geq T \text{ implies } \left||t|^{-2p}|\phi_f(t)|^{-2} - C^{-1}\right| < \varepsilon$$

We have

$$n^{1-\frac{1}{2p}} J_n^{**} = n^{1-\frac{1}{2p}} \int_{-T}^{T} \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2}\,dt$$

$$+ n^{1-\frac{1}{2p}} \int_{|t|>T} \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2}\,dt$$

The second integral in the expression for $n^{1-\frac{1}{2p}} J_n^{**}$ can be written as

$$n^{1-\frac{1}{2p}} \int_{|t|>T} \frac{dt}{(n-1)+|\phi_f(t)|^{-2}}$$ since $|\phi_f(t)|^2 > 0$ for ε

sufficiently small.

Then we can write $n^{1-\frac{1}{2p}} \int_{|t|>T} \frac{dt}{(n-1)+|\phi_f(t)|^{-2}(t)^{-2p}|t|^{2p}}$ since

$|t|>T$ implies $|t|^{2p} > 0$

Therefore,

$$n^{1-\frac{1}{2p}} J_n^{**} = n^{1-\frac{1}{2p}} \int_{-T}^{T} \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2} dt$$

$$+ n^{1-\frac{1}{2p}} \int_{|t|>T} \frac{dt}{(n-1)+|\phi_f(t)^{-2}|t|^{-2p}|t|^{2p}}$$

Upon adding $n^{1-\frac{1}{2p}} \int \frac{dt}{(n-1)+|t|^{2p}C^{-1}}$ and its negative to this expression and then regrouping we obtain

$$n^{1-\frac{1}{2p}} J_n^{**} = n^{1-\frac{1}{2p}} \int_{-T}^{T} \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2} dt \qquad (1.3.13)$$

$$+ n^{1-\frac{1}{2p}} \int \frac{dt}{(n-1)+|t|^{2p}C^{-1}}$$

$$- n^{1-\frac{1}{2p}} \int_{-T}^{T} \frac{dt}{(n-1)+|t|^{2p}C^{-1}}$$

$$+ n^{1-\frac{1}{2p}} \int_{|t|>T} \left[\frac{1}{(n-1)+|\phi_f(t)|^{-2}|t|^{-2p}|t|^{2p}} - \frac{1}{(n-1)+|t|^{2p}C^{-1}}\right] dt$$

Consider the first term of (1.3.13). Since $|\phi_f(t)|^2 \leq 1$,

$$n^{1-\frac{1}{2p}} \int_{-T}^{T} \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2} dt \leq \frac{2Tn^{1-\frac{1}{2p}}}{(n-1)} \to 0 \text{ as } n \to \infty.$$

For the third term of (1.3.13), since $C^{-1}>0$, $\frac{1}{(n-1)+|t|^{2p}C^{-1}} \leq \frac{1}{n-1}$

Hence $n^{1-\frac{1}{2p}} \int_{-T}^{T} \frac{dt}{(n-1)+|t|^{2p}C^{-1}} \le \frac{2Tn^{1-\frac{1}{2p}}}{n-1} \to 0$ as $n \to \infty$.

As for the second term of (1.3.13), we have

$$n^{1-\frac{1}{2p}} \int \frac{dt}{(n-1)+|t|^{2p}C^{-1}} = \frac{n^{1-\frac{1}{2p}}}{(n-1)^{1-\frac{1}{2p}}} \int \frac{(n-1)^{1-\frac{1}{2p}}dt}{(n-1)+|t|^{2p}C^{-1}}$$

$$= \left(\frac{n}{n-1}\right)^{1-\frac{1}{2p}} \int \frac{(n-1)^{-\frac{1}{2p}}dt}{1+(n-1)^{-1}|t|^{2p}C^{-1}} = \left(\frac{n}{n-1}\right)^{1-\frac{1}{2p}} \int \frac{(n-1)^{-\frac{1}{2p}}dt}{1+|(n-1)^{-\frac{1}{2p}}tC^{-\frac{1}{2p}}|^{2p}}$$

letting $s = (n-1)^{-\frac{1}{2p}}C^{-\frac{1}{2p}}t$, we have

$$ds = (n-1)^{-\frac{1}{2p}}C^{-\frac{1}{2p}}dt$$

and the limits of integration are unchanged since $C > 0$ and $n-1 > 0$ for $n > 1$.

Therefore, under the transformation,

$$n^{1-\frac{1}{2p}} \int \frac{dt}{(n-1)+|t|^{2p}C^{-1}} = \left(\frac{n}{n-1}\right)^{1-\frac{1}{2p}} \int \frac{C^{\frac{1}{2p}}ds}{1+|s|^{2p}}$$

Hence,

$$\lim_{n\to\infty} \{n^{1-\frac{1}{2p}} \int \frac{dt}{(n-1)+|t|^{2p}C^{-1}} = \lim_{n\to\infty} \left(\frac{n}{n-1}\right)^{1-\frac{1}{2p}} \int \frac{C^{\frac{1}{2p}}ds}{1+|s|^{2p}}$$

$$= C^{\frac{1}{2p}} \int \frac{ds}{1+|s|^{2p}}$$

Finally for the fourth term of (1.3.13)

$$n^{1-\frac{1}{2p}} \int_{|t|>T} \left[\frac{1}{(n-1)+|\phi_f(t)|^{-2}|t|^{-2p}|t|^{2p}} - \frac{1}{(n-1)+|t|^{2p}C^{-1}}\right]$$

$$\leq n^{1-\frac{1}{2p}} \int_{|t|>T} \frac{\left||t|^{2p}C^{-1} - |t|^{2p}|\phi_f(t)|^{-2}|t|^{-2p}\right|}{((n-1)+|\phi_f(t)|^{-2}|t|^{-2p}|t|^{2p})((n-1)+|t|^{2p}C^{-1})} dt$$

$$= n^{1-\frac{1}{2p}} \int_{|t|>T} \frac{|t|^{2p}\left|C^{-1} - |\phi_f(t)|^{-2}|t|^{-2p}\right|}{(\qquad)(\qquad)} dt$$

$$\leq n^{1-\frac{1}{2p}}\varepsilon \int_{|t|>T} \frac{|t|^{2p}}{(\quad)(\quad)} dt$$ since $|t|\geq T$ implies

$\left||\phi_f(t)|^{-2}|t|^{-2p} - C^{-1}\right| < \varepsilon$

$$= n^{1-\frac{1}{2p}}\varepsilon \int_{|t|>T} \frac{dt}{((n-1)|t|^{-2p}+|\phi_f(t)|^{-2}|t|^{-2p})((n-1)+|t|^{2p}C^{-1})}$$

since $|t|>T$ implies $|t|^{2p}>0$

Now, $|\phi_f(t)|^2$ continuous, $|t|^{2p}$ continuous implies that $|\phi_f(t)|^2|t|^{2p}$ is continuous.

Since $\lim_{|t|\to\infty} |\phi_f(t)|^2|t|^{2p}=C$, then $|\phi_f(t)|^2$ is bounded

That is, there is a $B \in R$ such that

$$|\phi_f(t)|^2|t|^{2p}\leq B \quad \text{for all} \quad t \in R$$

This bound with the fact that the integrand is non-negative gives

$$n^{1-\frac{1}{2p}} \int_{|t|>T} \left[\frac{1}{(n-1)+|\phi_f(t)|^{-2}|t|^{-2p}|t|^{2p}} - \frac{1}{(n-1)+|t|^{2p}C^{-1}}\right] dt$$

$$\leq B\varepsilon n^{1-\frac{1}{2p}} \int \frac{dt}{(n-1)+|t|^{2p}C^{-1}}$$ since $|t|>T$ implies

$(n-1)|t|^{-2p}>0$

However, by the argument for the second term in (1.3.13) we have that

$$n^{1-\frac{1}{2p}}\int \frac{dt}{(n-1)+|t|^{2p}C^{-1}} < \infty$$

and since ε depends on T and is independent of n, we conclude that the fourth term of (1.3.13) is arbitrarily small for sufficient large T.

Hence, we conclude that

$$\lim_{n\to\infty} \{n^{1-\frac{1}{2p}} J_n^{**}\} \equiv C^{\frac{1}{2p}}\int \frac{dt}{1+|t|^{2p}}$$

Recall from (1.3.9) that

$$J_n^{*} = \frac{K_n^{*}(0)}{n} - O(n^{-1})$$

$$= \frac{J_n^{**}}{2\pi} - O(n^{-1})$$

Therefore,

$$n^{1-\frac{1}{2p}} J_n^{*} = \frac{n^{1-\frac{1}{2p}} J_n^{**}}{2\pi} - O(n^{-\frac{1}{2p}})$$

Hence,

$$\lim_{n\to\infty} n^{1-\frac{1}{2p}} J_n^{*} = \lim_{n\to\infty} \frac{1}{2\pi} n^{1-\frac{1}{2p}} J_n^{**}$$

$$= \frac{C^{\frac{1}{2p}}}{2\pi}\int \frac{dt}{1+|t|^{2p}} \quad \text{which is (1.3.12)}$$

Discussion

The theorem shows that when $\phi_f(t)$ decreases algebraically of degree $p > 1/2$, the estimate $\hat{f}n(x)$ formed from $K_n^*(x)$, where K_n^* is defined in (1.3.7), is integratedly consistent of order $n^{1-\frac{1}{2p}}$. Recall that the estimate in terms of K_n^* is optimal in the sense that the corresponding MISE, J_n^*, is a minimum. However, in practice, we would not be able to determine K_n^* for K_n^* is expressed in terms of $\phi_f(t)$ and knowing $\phi_f(t)$ would enable us to determine f directly by the Inversion theorem. Hence, we are forced to look at classes of estimates formed from suitably chosen K_n. Such a class of estimates is the estimates of algebraic form as defined in (1.3.11).

We have considered the consistency properties of the optimum estimate in (1.3.12). Let us now look at the consistency properties of an estimate of algebraic form and compare its properties with those of the optimum estimate.

Consistency Properties of an Estimate of Algebraic Type

Theorem (1.3.13)

Let $\hat{f}n(x)$ be an estimate of algebraic type and $\phi_f(t)$ decrease algebraically of degree $p > 1/2$. If $\int |t|^{-2p}(1-h(t))^2 dt$ exists (where $\phi_{K_n}(t) = h(A_n t)$) and if $A_n = Dn^{-\frac{1}{2p}}$ (where $D > 0$),

then $n^{1-\frac{1}{2p}} J_n \to (2\pi D)^{-1}\int h^2(t)dt + \frac{CD^{2p-1}}{2\pi}\int |t|^{-2p}(1-h(t))^2 dt$ as $n \to \infty$

<u>where</u> J_n <u>is the MISE corresponding to</u> K_n <u>and</u> $\lim_{|t|\to\infty} |\phi_f(t)|^2|t|^{2p} = C.$

<u>Proof</u>

Since $\hat{f}n(x)$ is an estimate of algebraic type, $\phi_{K_n}(t) = h(A_n t)$ where $h(t)$ is a bounded, even square integrable function and $A_n \to 0$ as $n \to \infty$. From (1.3.6)

$$2\pi J_n = \int\{n^{-1}|\phi_{K_n}(t)|^2(1-|\phi_f(t)|^2) + |\phi_f(t)|^2(|1-\phi_{K_n}(t)|^2)\}dt$$

and we have

$$|\phi_{K_n}(t)|^2 = |h(A_n t)|^2$$

$$= h(A_n t)\overline{h(A_n t)}$$

$$= h(A_n t)h(-A_n t)$$

$$= h^2(A_n t) \quad \text{since } h \text{ is even.}$$

Also $|1-\phi_{K_n}(t)|^2 = |1-h(A_n t)|^2$

$$= (1-h(A_n t))^2 \quad \text{since } h \text{ is even.}$$

Hence,

$$2\pi n^{1-\frac{1}{2p}} J_n = n^{-\frac{1}{2p}}\int h^2(A_n t)(1-|\phi_f(t)|^2)dt + n^{1-\frac{1}{2p}}\int|\phi_f(t)|^2\{(1-h(A_n t))^2\}dt \tag{1.3.14}$$

For the first term of (1.3.14)

$$n^{-\frac{1}{2p}}\int h^2(A_n t)(1-|\phi_f(t)|^2)dt = n^{-\frac{1}{2p}}\int h^2(A_n t)dt \qquad (1.3.15)$$

$$- n^{-\frac{1}{2p}}\int h^2(A_n t)|\phi_f(t)|^2 dt$$

Under the transformation $u = A_n t = D_n^{-\frac{1}{2p}} t$ we have

$$n^{-\frac{1}{2p}}\int h^2(A_n t)dt = D^{-1}\int h^2(u)du$$

and noting $h(t)$ bounded implies $|h(t)| \le B^{1/2}$ for some $B^{1/2} \in R$

implies $|h(t)|^2 \le B$

implies $h^2(t) \le B$ since h is even

Therefore,

$$n^{-\frac{1}{2p}}\int h^2(A_n t)|\phi_f(t)|^2 dt \le n^{-\frac{1}{2p}} B \int |\phi_f(t)|^2 dt$$ where the

integral is finite.

Hence,

$$0 \le \lim_{n\to\infty} n^{-\frac{1}{2p}}\int h^2(A_n t)|\phi_f(t)|^2 dt$$

$$\le \lim_{n\to\infty} n^{-\frac{1}{2p}} B \int |\phi_f(t)|^2 dt$$

$= 0$ since $p > 0$.

That is,

$$\lim_{n\to\infty} n^{-\frac{1}{2p}}\int h^2(A_n t)|\phi_f(t)|^2 dt = 0$$

Therefore, from (1.3.15) and the above we conclude that

$$\lim_{n\to\infty} n^{-\frac{1}{2p}}\int h^2(A_n t)(1-|\phi_f(t)|^2)dt = D^{-1}\int h^2(u)du \qquad (1.3.16)$$

For the second term in (1.3.14), let

$$I \equiv n^{1-\frac{1}{2p}}\int |\phi_f(t)|^2\{(1-h(A_n t))^2\}dt$$

We shall show

$$I = CD^{-1}\int |t|^{-2p}(1-h(t))^2 dt + D^{-1}\int |t|^{-2p}(1-h(t))^2\,[\,] \qquad (1.3.17)$$

where $[\,] = [\{n^{1/2}|\phi_f(D^{-1}n^{\frac{1}{2p}}t)||t|^p\}^2 - CD^{2p}]dt$

This follows since

$$I = n^{1-\frac{1}{2p}}\int |\phi_f(t)|^2\{(1-h(A_n t))^2\}dt$$

$$= n^{1-\frac{1}{2p}}\int \frac{|\phi_f(t)|^2|t|^{2p}\{(1-h(A_n t))^2\}}{|t|^{2p}}dt \quad \text{since a Lebesque}$$

Integral

$$= n^{1-\frac{1}{2p}}\int \frac{|\phi_f(D^{-1}n^{\frac{1}{2p}}u)|^2|D^{-1}n^{\frac{1}{2p}}u|^{2p}\{(1-h(u))^2\}}{|D^{-1}n^{\frac{1}{2p}}u|^{2p}} D^{-1}n^{\frac{1}{2p}}du$$

where $u = A_n t$

$$= D_n^{-\frac{1}{2p}}t$$

$$= D^{-1}n\int \frac{|\phi_f(D^{-1}n^{\frac{1}{2p}}u)|^2|D^{-1}u|^{2p}\{(1-h(u))^2\}}{|D^{-1}u|^{2p}}du \quad \text{since } n^{\frac{1}{2p}} \neq 0$$

$$= D^{-1}\int \frac{\{n^{1/2}|\phi_f(D^{-1}n^{\frac{1}{2p}}u)||u|^p\}^2(1-h(u))^2D^{-2p}}{|D^{-1}u|^{2p}}\,du$$

$$= D^{-1}\int \frac{([\{n^{1/2}|\phi_f(D^{-1}n^{\frac{1}{2p}}u)||u|^p\}^2-CD^{2p}](1-h(u))^2D^{-2p}+C(1-h(u))^2)}{|D^{-1}u|^{2p}}\,du$$

$$= CD^{-1}\int \frac{(1-h(u))^2}{|D^{-1}u|^{2p}}\,du + D^{-1}\int \frac{[\{n^{1/2}|\phi_f(D^{-1}n^{\frac{1}{2p}}u)||u|^p\}^2-CD^{2p}](1-h(u))^2D^{-2p}}{|u|^{2p}D^{-2p}}\,du$$

$$= CD^{2p^{-1}}\int \frac{(1-h(u))^2}{|u|^{2p}}\,du + D^{-1}\int \frac{(1-h(u))^2}{|u|^{2p}}[\{n^{1/2}|\phi_f(D^{-1}n^{\frac{1}{2p}}u)||u|^p\}^2-CD^{2p}]\,du$$

which is (1.3.17)

We shall now show the second integral in the expression for I is arbitrarily small. The second integral may be written

$$D^{2p^{-1}}\int \frac{(1-h(u))^2}{|u|^{2p}}[|\phi_f(D^{-1}n^{\frac{1}{2p}}u)|^2|D^{-1}n^{\frac{1}{2p}}u|^{2p}-C]\,du$$

Define

$$gn(u) = \frac{(1-h(u))^2}{|u|^{2p}}\left||\phi_f(D^{-1}n^{\frac{1}{2p}}u)|^2|D^{-1}n^{\frac{1}{2p}}u|^{2p}-C\right| \quad \text{if} \quad u \neq 0$$

$$= 0 \quad \text{if} \quad u = 0$$

Then $|gn(u)| \leq (B+C)\dfrac{(1-h(u))^2}{|u|^{2p}}$ since $|\phi_f(t)|^2|t|^{2p} \leq B$

However, $\dfrac{(1-h(u))^2}{|u|^{2p}}$ integrable implies $(B+C)\dfrac{(1-h(u))^2}{|u|^{2p}}$ is integrable.

Also since $\phi_f(t)$ decreases algebraically of degree $p > 0$ and

$$t = D^{-1} n^{\frac{1}{2p}} u, \quad \text{then}$$

$$\lim_{n\to\infty} gn(u) = 0$$

Thus by the L.D.C.T., the second integral is arbitrarily small. In view of (1.3.14), (1.3.16), (1.3.17) and this fact we have established theorem (1.3.13).

Discussion

This theorem shows that if $\hat{f}n(x)$ is an estimate of algebraic form and $\phi_f(t)$ decreases algebraically of degree $p > 1/2$, then J_n, the MISE corresponding to $\hat{f}n(x)$, is integratedly consistent of order $n^{1-\frac{1}{2p}}$. It may seem surprising that the orders of integrated consistency for the optimum estimate and an estimate of algebraic form are the same. However, this result is a consequence of our choice of $A_n = D_n^{-\frac{1}{2p}}$ in the theorem.

In the applications, theorem (1.3.13) means if we assume $\phi_f(t)$ decreases algebraically of degree $p > 1/2$ and we construct an estimate of algebraic type which satisfies the properties

(a) $\int |t|^{-2p}(1-h(t))^2 dt$ exists where $\phi_{K_n}(t) = h(A_n t)$ and h is a bounded, even square integrable function, and

(b) $A_n = D_n^{-\frac{1}{2p}}$

Then J_n is integratedly consistent of order $n^{1-\frac{1}{2p}}$.

In terms of $K_n(x)$, we choose $h(t)$ such that

$$K_n(x) = \frac{1}{A_n} k\left(\frac{x}{A_n}\right) \quad \text{where} \quad \phi_k(t) = h(t)$$

and $h(t)$ satisfies (a). Including (b) we obtain

$$K_n(x) = \frac{1}{D_n^{\frac{1}{2p}}} k\left(\frac{x}{D_n^{\frac{1}{2p}}}\right) \quad \text{as the form of } K_n\text{'s to choose.}$$

However, notice that h, k and K_n are quite arbitrary. In case we do not know p precisely, if we choose $1/2 < r < p$ we have

<u>Theorem</u> (1.3.18)

Let $\hat{f}n(x)$ be an estimate of algebraic type, and $\phi_f(t)$ decrease algebraically of degree $p > 1/2$. Assume $\int |t|^{-2p}(1-h(t))^2 dt$ exists. If $1/2 < r < p$ and $A_n = D_n^{-\frac{1}{2r}}$, then $\hat{f}n(x)$ is integratedly consistent of order $n^{1-\frac{1}{2r}}$

<u>Proof</u>

As in (1.3.14) and (1.3.15),

$$n^{1-\frac{1}{2r}} J_n = \frac{n^{-\frac{1}{2r}}}{2\pi}\int h^2(A_n t)(1-|\phi_f(t)|^2)dt + \frac{n^{1-\frac{1}{2r}}}{2\pi}\int |\phi_f(t)|^2(1-h(A_n t))^2 dt$$

$$= \frac{D^{-1}}{2\pi}\int h^2(t)dt - \frac{n^{-\frac{1}{2r}}}{2\pi}\int h^2(A_n t)|\phi_f(t)|^2 dt$$

$$+ \frac{n^{1-\frac{1}{2r}}}{2\pi}\int |\phi_f(t)|^2(1-h(A_n t))^2 dt$$

since $n^{1-\frac{1}{2r}} = n^{\frac{r-p}{2pr}} n^{1-\frac{1}{2p}}$

$$n^{1-\frac{1}{2r}} J_n = \frac{D^{-1}}{2\pi}\int h^2(t)dt - \frac{n^{-\frac{1}{2r}}}{2\pi}\int h^2(A_n t)|\phi_f(t)|^2 dt$$

$$+ n^{\frac{r-p}{2pr}}\{\frac{n^{1-\frac{1}{2p}}}{2\pi}\int |\phi_f(t)|^2(1-h(A_n t))^2 dt\}$$

Now $\lim_{n\to\infty} n^{\frac{r-p}{2pr}} = 0$ since $1/2 < r < p$ and

$$\lim_{n\to\infty} \frac{n^{-\frac{1}{2r}}}{2\pi}\int h^2(A_n t)|\phi_f(t)|^2 dt = 0$$ since $r > 0$ and the integral is finite.

Also $$\lim_{n\to\infty} \frac{n^{1-\frac{1}{2p}}}{2\pi}\int |\phi_f(t)|^2(1-h(A_b t))^2 dt = \frac{CD^{2p-1}}{2\pi}\int \frac{(1-h(t))^2}{|t|^{2p}} dt$$ which is finite. Hence,

$$\lim_{n\to\infty} n^{1-\frac{1}{2r}} J_n = \frac{1}{2\pi D}\int h^2(t)dt$$

Hence, if we choose $1/2 < r < p$, $A_n = D_n^{-\frac{1}{2r}}$ and $K_n(x)$ so that $\int |t|^{-2p}(1-h(t))^2 dt$ exists, our estimator $\hat{f}n(x)$ will be integratedly consistent of order $n^{1-\frac{1}{2r}}$.

__Existence of an Estimator of Algebraic Form with the Asymptotic Optimum Property__

If $\phi_f(t)$ decreases algebraically of degree $p > 1/2$, we can construct an estimator of algebraic form with the asymptotic optimum property as follows:

(a) Choose $K_n(x)$ such that $h(t) = (1+|t|^{2p})^{-1}$

(b) Choose A_n such that $A_n = D_n^{-\frac{1}{2p}}$.

Now $h(t)$ is a bounded, even integrable function and recall that h bounded and integrable implies that h is square integrable.

Also notice that

$$\frac{(1-h(t))^2}{|t|^{2p}} = \frac{|t|^{2p}}{1+|t|^{2p}} \left(\frac{1}{1+|t|^{2p}}\right)$$

$$\leq \frac{1}{1+|t|^{2p}} \quad \text{since} \quad |t|^{2p} \geq 0.$$

Hence, $\int |t|^{-2p}(1-h(t))^2 dt$ exists.

By theorem (1.3.13),

$$\lim_{n\to\infty} n^{1-\frac{1}{2p}} J_n = \frac{1}{2\pi D}\int h^2(t)dt + \frac{C}{2\pi} D^{2p-1}\int \frac{(1-h|t|)^2}{t^{2p}} dt$$

$$= \frac{1}{2\pi D}\int \frac{1}{(1+|t|^{2p})^2} dt + \frac{CD^{2p-1}}{2\pi}\int \frac{|t|^{2p}}{(1+|t|^{2p})^2} dt \quad \text{by choice of } h(t).$$

$$= \frac{1}{2\pi D}\int \frac{dt}{(1+|t|^{2p})^2} + \frac{1}{2\pi D}\int \frac{CD^{2p-1}|t|^{2p}}{(1+|t|^{2p})^2} dt$$

$$= \frac{1}{2\pi D}\int \frac{1+CD^{2p-1}|t|^{2p}}{(1+|t|^{2p})^2} dt$$

Choosing $D = C^{-\frac{1}{2p}}$ where $\lim_{|t|\to\infty} |\phi_f(t)|^2|t|^{2p} = C > 0$ we have

$$\lim_{n\to\infty} n^{1-\frac{1}{2p}} J_n = \frac{C^{\frac{1}{2p}}}{2\pi}\int \frac{dt}{1+|t|^{2p}} \quad , \tag{1.3.18}$$

which is (1.3.12).

This construction is extremely important in the applications. For in the practical situation we can not determine the functions $K_n^*(x)$ which give rise to a minimum MISE J_n^*. However, by the construction, if $\phi_f(t)$ decreases algebraically of degree $p \geq 1/2$, our estimator $\hat{f}n(x)$ is integratedly consistent of order $n^{1-\frac{1}{2p}}$ and moreover satisfies the asymptotic optimum property.

Therefore, if the asymptotic efficiency is defined as $\lim_{n\to\infty} [\frac{J_n^*}{J_n}]$, choosing h and A_n as above we have

$$\lim_{n\to\infty} [\frac{J_n^*}{J_n}] = \lim_{n\to\infty} [\frac{n^{1-\frac{1}{2p}} J_n^*}{n^{1-\frac{1}{2p}} J_n}]$$

$$= \frac{\lim_{n\to\infty} n^{1-\frac{1}{2p}} J_n^*}{\lim_{n\to\infty} n^{1-\frac{1}{2p}} J_n} \quad \text{since limits exist and} \quad \lim_{n\to\infty} n^{1-\frac{1}{2p}} J_n \neq 0$$

$$= 1 \quad \text{by (1.3.12) and (1.3.18)}$$

In the following section we shall consider classes of estimates for those densities whose characteristic functions decrease exponentially. For such densities we shall define estimates of exponential form. As in the previous section we shall investigate consistency properties of the following:

(a) the optimum estimate

(b) an estimate of exponential type

Also we shall exhibit the construction of an estimate of exponential type with the asymptotic optimum property.

§1.3B Characteristic Functions Which Decrease Exponentially

Definition 1.3.19 $\phi_f(t)$ is said to decrease exponentially of coefficient $p > 0$

if $|\phi_f(t)| \leq Ae^{-\rho|t|}$ for some constant A and all t. (1.3.19a)

$$\text{and} \quad \lim_{v\to\infty} \int_0^1 (1+e^{2\rho v}|\phi_f(vt)|^2)^{-1} dt = 0 \qquad (1.3.19b)$$

Example

If $f(x) = (\pi(1+x^2))^{-1}$ for x in the reals, then $\phi_f(t)=e^{-|t|}$ for $t \in R$

Hence

$$|\phi_f(t)| = e^{-|t|} \quad \text{and} \quad \rho = A = 1.$$

In addition,

$$|\phi_f(vt)|^2 = e^{-2|vt|}$$

$$= e^{-2vt} \quad \text{for } v,t \geq 0.$$

Therefore,

$$\lim_{v\to\infty} \int_0^1 \frac{1}{1+e^{2\rho v}|\phi_f(t)|^2} dt = \lim_{v\to\infty} \int_0^1 \frac{1}{1+e^{2\rho v}e^{-2vt}} dt$$

$$= \lim_{v\to\infty} \int_0^1 \frac{1}{1+e^{2v}e^{-2vt}}\, dt \quad \text{since} \quad \rho = 1$$

$$= \int_0^1 \lim_{v\to\infty} \frac{1}{1+e^{2v(1-t)}}\, dt \quad \text{by the L.D.C.T. (1 is a dominating function)}$$

$$= 0 \quad \text{since} \quad \lim_{v\to\infty} \frac{1}{1+e^{2v(1-t)}} = 0 \quad \text{except when} \quad t = 1.$$

Thus $\phi_f(t) = e^{-|t|}$ decreases exponentially of coefficient 1.

Of course, the characteristic function of the normal density function decreases exponentially.

<u>Definition</u> 1.3.20 An <u>estimate</u> $\hat{f}n(x)$ formed from $K_n(x)$ is said to have <u>exponential form</u> if $\phi_{K_n}(t) = h(A_n e^{\alpha|t|})$ where $A_n \to 0$ as $n \to \infty$, $\alpha > 0$, and h is bounded and square integrable.

<u>Remark</u> $h^2(A_n e^{\alpha|t|})$ is real valued since

$$|\phi_{K_n}(t)|^2 = \phi_{K_n}(t)\overline{\phi_{K_n}(t)}$$

$$= \phi_{K_n}(t)\phi_{K_n}(-t)$$

$$= h(A_n e^{\alpha|t|})h(A_n e^{\alpha|-t|})$$

$$= h^2(A_n e^{\alpha|t|})$$

Consistency Properties of the Optimum Estimate

Theorem (1.3.21)

Let $\phi_f(t)$ decrease exponentially of coefficient ρ. Then J_n^*, the minimum MJSE, satisfies

$$\lim_{n\to\infty} \frac{n}{\log n} J_n^* = \frac{1}{2\pi\rho}$$

Proof

First we show

$$\lim_{n\to\infty} \left| \int \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2}\, dt - \int \frac{e^{-2\rho|t|}}{1+(n-1)e^{-2\rho|t|}}\, dt \right| = 0 \qquad (1.3.22)$$

because

$$\left| \int \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2}\, dt - \int \frac{e^{-2\rho|t|}}{1+(n-1)e^{-2\rho|t|}}\, dt \right|$$

$$= \left| \int \frac{(|\phi_f(t)|^2 - e^{-2\rho|t|})}{(1+(n-1)|\phi_f(t)|^2)(1+(n-1)e^{-2\rho|t|})}\, dt \right|$$

$$\le 2(A^2+1) \int_0^\infty \frac{e^{-2\rho t} dt}{(1+(n-1)|\phi_f(t)|^2)(1+(n-1)e^{-2\rho t})}\, dt$$

$$= 2(A^2+1) \int_0^{\frac{\log(n-1)}{2\rho}} \frac{e^{-2\rho t} dt}{(\ )(\ )} + \int_{\frac{\log(n-1)}{2\rho}}^{\infty} \frac{e^{-2\rho t} dt}{(\ )(\ )}$$

from (1.3.19a) and the fact that $|\phi_f(t)|^2$ and $e^{-2\rho|t|}$ are even functions.

Denote the first integral by I_1 and the second by I_2.

Then $$I_1 = \int_0^{\frac{\log(n-1)}{2\rho}} \frac{e^{-2\rho t}e^{2\rho t}dt}{(1+(n-1)|\phi_f(t)|^2)(e^{2\rho t}+(n-1))}$$ since $e^{2\rho t} > 0$

$$< \frac{1}{n-1}\int_0^{\frac{\log(n-1)}{2\rho}} \frac{dt}{1+(n-1)|\phi_f(t)|^2}$$

Under the transformation, $t = vs$ where $v = \frac{\log(n-1)}{2\rho}$ we have

$$(n-1)|\phi_f(t)|^2 = (n-1)|\phi_f(vs)|^2$$

$$= e^{\log(n-1)}|\phi_f(vs)|^2$$ since $\Theta > 0$ implies $\Theta = e^{\log\Theta}$

$$= \exp\left(\frac{2\rho\log(n-1)}{2\rho}\right)|\phi_f(vs)|^2$$ since $\rho > 0$

$$= e^{2\rho v}|\phi_f(vs)|^2$$ by definition of v.

Hence,

$$I_1 < \frac{1}{n-1}\int_0^{\frac{\log(n-1)}{2\rho}} \frac{dt}{1+(n-1)|\phi_f(t)|^2}$$

$$= \frac{\log(n-1)}{2\rho(n-1)}\int_0^1 \frac{ds}{1+e^{2\rho v}|\phi_f(vs)|^2}$$

As $n \to \infty$, $v = \frac{\log(n-1)}{2\rho} \to \infty$ and we have

$$0 \le \lim_{n\to\infty} I_1 < \lim_{n\to\infty} \left\{\frac{1}{2\rho}\,\frac{\log(n-1)}{n-1} \int_0^1 \frac{ds}{1+e^{2\rho v}|\phi_f(vs)|^2}\right\}$$

$$= \frac{1}{2\rho} \lim_{n\to\infty} \left\{\frac{\log(n-1)}{n-1}\right\} \lim_{n\to\infty} \left\{\int_0^1 \frac{ds}{1+e^{2\rho v}|\phi_f(vs)|^2}\right\}$$

$= 0$ since the indicated limits are zero (1.3.19b)

Thus $\lim_{n\to\infty} I_1 = 0$

Dealing with I_2, since $|\phi_f(t)|^2 \ge 0$ and $\exp(-2\rho t) > 0$,

$$I_2 < \int_{\frac{\log(n-1)}{2\rho}}^{\infty} \exp(-2\rho t)\,dt$$

$$= \frac{1}{2\rho(n-1)}$$

Hence $\lim_{n\to\infty} I_2 = 0$ and we have (1.3.22)

However, letting $z = 1+(n-1)e^{-2\rho t}$ we have

$$\lim_{n\to\infty} \frac{n}{\log n} \int \frac{e^{-2\rho|t|}}{1+(n-1)e^{-2\rho|t|}}\,dt = \frac{1}{\rho} \qquad (1.3.23)$$

and by (1.3.22) and (1.3.23), it follows that

$$\lim_{n\to\infty} \frac{n}{\log n} \int \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2}\,dt = \frac{1}{\rho} \qquad (1.3.24)$$

But by (1.3.9),

$$J_n{}^* = \frac{K_n{}^*(0)}{n} - o(n^{-1})$$

where $$\frac{K_n^*(0)}{n} = \frac{1}{2\pi}\int \frac{|\phi_f(t)|^2}{1+(n-1)|\phi_f(t)|^2}\,dt$$

So by (1.3.24) and (1.3.9),

$$\frac{n}{\log n} J_n^* \to \frac{1}{2\pi\rho} \quad \text{as} \quad n \to \infty \quad \text{which is (1.3.21).}$$

If $\phi_f(t)$ decreases exponentially of coefficient ρ, then J_n^* is integratedly consistent of order $\frac{n}{\log n}$. However, in the practical situation, we are concerned with the consistency properties of estimates of a particular form. In the following discussion, we shall deal with estimates of exponential type.

<u>Consistency Properties of Estimates of Exponential Type</u>

Let $\hat{f}n(x)$ be an estimate of exponential type. That is, $\phi_{K_n}(t) = h(A_n e^{\alpha|t|})$ where h is bounded and square integrable. In addition, suppose $h(t)$ satisfies

$$|1-h(t)| \le B_1|t| \quad \text{for} \quad |t| \le 1 \tag{1.3.25}$$

<u>Theorem</u> (1.3.26)

<u>Let</u> $\phi_f(t)$ <u>decrease exponentially of coefficient</u> $p > 0$ <u>and let</u> $\hat{f}n(x)$ <u>be an estimate of exponential type such that</u> $h(t)$ <u>satisfies</u> (1.3.25). <u>Let</u> $A_n = Dn^{-b}$ <u>for</u> $b > 1/2$ <u>and</u> $\alpha \le 2\rho b$. <u>Then the</u> J_n <u>corresponding to</u> $\hat{f}n(x)$ <u>satisfies</u>

$$\lim_{n\to\infty} \left\{\frac{n}{\log n} J_n\right\} = \frac{1}{2\pi}\left(\frac{2b}{\alpha}\right)$$

To prove the theorem we establish two lemmas:

Lemma 1

Under the above conditions,

$$\lim_{n\to\infty} \left\{\frac{1}{\log n}\int_{D_n^{-b}}^{\infty} \frac{h^2(t)}{t}\,dt\right\} = b \qquad (1.3.27)$$

Proof

$$\frac{1}{\log n}\int_{D_n^{-b}}^{\infty} \frac{h^2(t)}{t}\,dt = \frac{1}{\log n}\int_{D_n^{-b}}^{1} \frac{h^2(t)}{t}\,dt + \frac{1}{\log n}\int_{1}^{\infty}\frac{h^2(t)}{t}\,dt$$

$$= \frac{1}{\log n}\int_{D_n^{-b}}^{1} \frac{dt}{t} - \frac{1}{\log n}\int_{D_n^{-b}}^{1} \frac{1-h^2(t)}{t}\,dt + \frac{1}{\log n}\int_{1}^{\infty}\frac{h^2(t)}{t}\,dt$$

The first term is $b - \frac{\log D}{\log n} \to b$ as $n \to \infty$.

Hence the lemma will follow if we can show that $\int_1^{\infty} t^{-1}h^2(t)dt < \infty$

The finiteness of the integral follows since

h square integrable implies $\int h^2(t)dt < \infty$ which in turn implies

$\int_1^{\infty} h^2(t)dt < \infty$. But for $t \geq 1$, $t^{-1}h^2(t) \leq h^2(t)$

Consequently, $\int_1^{\infty} t^{-1}h^2(t)dt < \infty$ and we have lemma 1.

Lemma 2

Under the conditions of the theorem,

$$\lim_{n\to\infty} \frac{n}{\log n}\int |\phi_f(t)|^2(1-\phi_{K_n}(t))^2 dt = 0 \qquad (1.3.28)$$

<u>Proof</u>

Since $\hat{f}n(x)$ is an estimate of exponential type, $\phi_{K_n}(t) = h(A_n e^{\alpha|t|})$ where $\alpha > 0$ and $\lim_{n\to\infty} A_n = 0$. Also $A_n = Dn^{-b}$ where $b > 1/2$.

Letting $x = D_n^{-b} e^{\alpha|t|}$

$= D_n^{-b} e^{\alpha t}$ for $t \geq 0$ we have

$$\frac{n}{\log n}\int |\phi_f(t)|^2(1-\phi_{K_n}(t))^2 dt = \frac{2n}{\log n}\int_0^\infty |\phi_f(t)|^2(1-\phi_{K_n}(t))^2 dt$$

$$= \frac{2n}{\log n}\int_{D_n^{-b}}^{\infty} |\phi_f(\tfrac{1}{\alpha}\log(\tfrac{n^b}{D}x))|^2 \frac{(1-h(x))^2}{\alpha x} dx$$

$$\leq \frac{2n}{\log n}\int_{D_n^{-b}}^{1} A^2 \exp[\tfrac{-2\rho}{\alpha}\log(\tfrac{n^b}{D}x)] B_1 \frac{2x}{\alpha} dx$$

$$+ \frac{2n}{\log n}\int_1^\infty A^2 \exp[\tfrac{-2\rho}{\alpha}\log(\tfrac{n^b}{D}x)](1+B)\frac{2dx}{\alpha x}$$

For the following reasons:

(a) by (1.3.19a), $|\phi_f(t)| \leq Ae^{-\rho|t|}$ implies $|\phi_f(t)|^2 \leq A^2 e^{-2\rho|t|}$

$= A^2 e^{-2\rho t}$ since $t \geq 0$.

(b) by (1.3.25), for $|x| \le 1$, $\frac{(1-h(x))^2}{\alpha x} \le \frac{B_1^2 |x|^2}{\alpha x} = B_1^2 \frac{x}{\alpha}$

since $x \ge D_n^{-b} > 0$

(c) $|(1-h(x))^2| = |1-2h(x)+h^2(x)|$

$\le 1+2|h(x)|+|h^2(x)|$

$\le 1+2B+B^2$ since h is bounded

$= (1+B)^2$

The conditions $\alpha \le 2\rho b$ and $b > 1/2$ imply both terms tend to 0 which gives us (1.3.28).

<u>Proof of (1.3.26)</u>

From (1.3.5) and the fact that $\hat{f}n(x)$ is an estimate of exponential type we have

$$2\pi \frac{n}{\log n} J_n = \frac{1}{\log n}\int h^2(D_n^{-b} e^{\alpha|t|})dt - \frac{1}{\log n}\int h^2(D_n^{-b} e^{\alpha|t|})|\phi_f(t)|^2 dt$$

$$+ \frac{n}{\log n}\int |\phi_f(t)|^2 (1-\phi_{K_n}(t))^2 dt$$

but

$$\frac{1}{\log n}\int h^2(D_n^{-b} e^{\alpha|t|})dt = \frac{2}{\log n}\int_0^\infty h^2(D_n^{-b} e^{\alpha t})dt$$ since h is even implies h^2 is even

$$= \frac{2}{\log n}\int_{D_n^{-b}}^\infty h^2(x)\frac{dx}{\alpha x}$$ under the transformations

$$= \frac{2}{\alpha}\frac{1}{\log n}\int_{D_n^{-b}}^{\infty} \frac{h^2(x)}{x}\,dx$$

and by lemma 1, $\lim_{n\to\infty} \{\frac{1}{\log n}\int_{D_n^{-b}}^{\infty} \frac{h^2(x)}{x}\,dx\} = b$

Hence, $\lim_{n\to\infty} \{\frac{1}{\log n}\int h^2(D_n^{-b}e^{\alpha|t|})dt\} = \frac{2b}{\alpha}$

Also, $\lim_{n\to\infty} \{\frac{1}{\log n}\int h^2(D_n^{-b}e^{\alpha|t|})|\phi_f(t)|^2 dt\} = 0$ since h is bounded and $\int |\phi_f(t)|^2 dt$ is finite.

By lemma 2,

$$\lim_{n\to\infty} \{\frac{n}{\log n}\int |\phi_f(t)|^2 (1-\phi_{K_n}(t))^2 dt\} = 0$$

Therefore, $\lim_{n\to\infty} \{\frac{n}{\log n} J_n\} = \frac{1}{2\pi}(\frac{2b}{\alpha})$

Remark

We have that _any_ estimate $\hat{f}n(x)$ of exponential type is integratedly consistent of order $\frac{n}{\log n}$ provided h satisfies (1.3.25). Contrast this with estimators of algebraic type, where we had to choose h satisfying $\int |t|^{-2p}(1-h(t))^2 dt$ is finite. (1.3.13).

For the applications, we need

An Estimator of Exponential Type with the Asymptotic Optimum Property

If $\phi_f(t)$ decreases exponentially of coefficient ρ, let $\hat{f}n(x)$ be an estimate of exponential type such that $h(t)$ satisfies

(1.3.25). Let $A_n = Dn^{-b}$ for $b > 1/2$ and $\alpha \le 2\rho b$, then by (1.3.26)

$$\lim_{n\to\infty} \{\frac{n}{\log n} J_n\} = \frac{1}{2\pi}(\frac{2b}{\alpha}).$$

Hence, if we choose $b > 1/2$ such that b satisfies $\alpha = 2\rho b$, (α and ρ are known), then $\lim_{n\to\infty} \{\frac{n}{\log n} J_n\} = \frac{1}{2\pi\rho}$ which is the optimum property (1.3.21).

CHAPTER 2 - THE L AND Q METHOD

Loftsgaarden and Quesenberry, [17], proposed an alternative method of estimating a density function which is fundamentally different from the Kernel method. We shall present the L and Q method for a one dimensional random variable X, extend the presentation for a p-dimensional random variable and then discuss similarities and differences of the Kernel and L and Q techniques.

§2.1 The Estimates for a One Dimensional Random Variable

We wish to estimate f at a point z where f is positive and continuous. Since f is continuous at z, $F'(z) = f(z)$. That is,

$$\lim_{h \to 0} \frac{F(z+h)-F(z-h)}{2h} = f(z) \tag{2.1.1}$$

Let P be the (unique) probability measure corresponding to the c.d.f. $F(x)$.

Then $F(z+h)-F(z-h) = P\{(z-h,z+h]\}$

$= P\{[z-h,z+h]\}$ since F is continuous.

Define the closed sphere of radius h with centre z as

$$Sh,z = \{x \in R \mid |x-z| \leq h\}$$

$$= [z-h, z+h]$$

Letting λ denote Lebesque measure, we have

$$\lambda(Sh,z) = \lambda([z-h,z+h])$$

$$= 2h$$

Consequently (2.1.1) becomes

$$\lim_{h \to 0} \frac{P\{Sh,z\}}{\lambda(Sh,z)} = f(z) \qquad (2.1.2)$$

The L and Q technique is to estimate $P\{Sh,z\}=F(z+h)-F(z-h)$ as follows:

Let $\{k(n)\}_{n=1}^{\infty}$ be a non-decreasing sequence of positive integers such that

$$\lim_{n \to \infty} k(n) = \infty \quad \text{and} \quad k(n) = o(n) \qquad (2.1.3)$$

Having $k(n)$ and a random sample $X_1, X_2, \ldots, X_n$ we <u>define</u> $h_{k(n)}(z)$ as <u>the distance from</u> z <u>to the</u> $k(n)^{th}$ <u>closest</u> X_i where $i = 1,2,\ldots,n$.

Letting $S_{h_{k(n)}(z),\, z}$ denote the sphere of radius $h_{k(n)}(z)$ about z we shall estimate $f(z)$ by $\hat{f}n(z)$ where

$$\hat{f}n(z) \equiv \left\{\frac{k(n)-1}{n}\right\} \frac{1}{\lambda(S_{h_{k(n)}(z),z})} \qquad (2.1.4)$$

$$= \{\frac{k(n)-1}{n}\} \frac{1}{2h_{k(n)}(z)} ,$$

§2.2 Extension for a p Dimensional Random Variable

The previous discussion extends immediately to the case where we have a p dimensional random variable $X = (X_1, X_2, \ldots, X_p)$ with absolutely continuous c.d.f. $F(x_1, x_2, \ldots, x_p)$ and density function $f(x_1, x_2, \ldots, x_p)$. We shall make the notational conventions

$$x = (x_1, x_2, \ldots, x_p) \quad \text{and} \quad z = (z_1, z_2, \ldots, z_p)$$

and we shall estimate $f(z)$ at a point $z \in R^p$ where $f(z)$ is positive and continuous.

Let $d(x,z) = (\sum_{i=1}^{p} (x_i - z_i)^2)^{1/2}$ and as before define

$$Sh,z = \{x \in R^p | d(x,z) \leq h\}$$

Letting λ denote Lebesque measure in R^p, we have

$$\lambda(Sh,z) = \frac{2\pi^{p/2} h^p}{p\Gamma(p/2)} \quad \text{where } \Gamma \text{ is the gamma function.}$$

With $\{k(n)\}_{n=1}^{\infty}$ and $h_{k(n)}(z)$ as previously defined

$$\hat{f}n(z) = \{\frac{k(n)-1}{n}\} \frac{1}{\lambda(S_{h_{k(n)}(z)}, Z)}$$

$$= \{\frac{k(n)-1}{n}\} \{\frac{p\Gamma(p/2)}{2\pi^{p/2} (h_{k(n)}(z))^p}\}$$

§2.3 Comparison of the Kernel and L and Q Methods

The Kernel technique tacitly assumed $f(x) \geq 0$, for x under consideration, whereas L and Q state $f(x) > 0$. In addition, L and Q assume that f is continuous at a point x when estimating $f(x)$. Similarly, Parzen assumed that f is continuous at a point x for such properties as asymptotic unbiasedness, consistency and asymptotic normality.

We can elucidate further differences by considering the form of the estimators. Recall that a Kernel estimator is of the form

$$\hat{f}n(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h} K\left(\frac{x-X_j}{h}\right) \tag{2.3.1}$$

$$= \int \frac{1}{h} K\left(\frac{x-y}{h}\right) dFn(y)$$

where the distance $h = h(n)$ and h is not a random variable. Contrast this with L and Q estimator (2.1.4) where the distance h is a function of $k(n)$ (hence n), the point z and observe that h is a random variable.

From (2.3.1), we could say Kernel estimators are sample c.d.f. oriented whereas L and Q estimators are distance oriented.

We shall now consider some properties of L and Q estimators. Let $\hat{f}n(z)$ be as in (2.1.4). The L and Q basic result is

Theorem

If f is positive and continuous at z, then $\hat{f}n(z)$ is a consistent estimate of $f(z)$.

As Moore and Henrichon, [18], have observed there is a basic error in the proof of this result. For this reason, we shall

(a) present Moore and Henrichon's results

(b) indicate a valid proof of the L and Q theorem.

<u>Let</u> $r_{k(n)}(z)$ <u>denote the distance from</u> z <u>to the</u> $k(n)^{th}$ <u>observation.</u>

The L and Q estimator is

$$\hat{f}n(z) = \{\frac{k(n)-1}{n}\}\{\frac{1}{2r_{k(n)}(z)}\}$$

Moore and Henrichon have introduced a step function approximation $fn^*(z)$ for $\hat{f}n(z)$ as follows:

Let $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ be the order statistics corresponding to $X_1, X_2, \ldots, X_n$. Then

$$fn^*(z) = \begin{cases} 0 & \text{if } z < X_{1,n} \text{ or } z \geq X_{n,n} \\ \hat{f}n(X_{i,n}) & \text{if } X_{i,n} \leq z < X_{i+1,n} \text{ where} \end{cases}$$

$i = 1,2,\ldots n-1.$

For uniform consistency of the estimates

<u>Theorem</u> (2.3.2)

<u>If</u> $f(z)$ <u>is uniformly continuous and positive on</u> $(-\infty,\infty)$ <u>and</u> $k(n)$ <u>is chosen such that</u> $\log n = o(k(n))$ <u>and</u> $k(n) = o(n)$, <u>then for all</u> $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{\sup_{z\in R}|\hat{f}n(z)-f(z)|>\varepsilon\} = 0 \quad (2.3.3)$$

and

$$\lim_{n\to\infty} P\{\sup_{z\in R}|fn^*(z)-f(z)|>\varepsilon\} = 0 \quad (2.3.4)$$

Subsequently, we shall write

$\hat{f}n \to f$ (UP), uniformly in probability, for (2.3.3) and

$a_n \to a$ (P) for convergence in probability.

Proof

Let $U_{k(n)}(z) = F(z+r_{k(n)}(z))-F(z-r_{k(n)}(z))$

The proof consists of showing that

$$\{\frac{n}{k(n)-1}\}U_{k(n)}(z) \to 1 \text{ (UP)} \tag{2.3.5}$$

from which we shall obtain

$$U_{k(n)}(z) \to 0 \text{ (UP)}$$

and $r_{k(n)}(z) \to 0$ (UP)

and $\dfrac{U_{k(n)}}{2r_{k(n)}} \to f$ (UP)

To obtain (2.3.5), recall that by definition of $r_{k(n)}(z)$ the interval $[z-r_{k(n)}(z), z+r_{k(n)}(z)]$ contains exactly $k(n)$ observations one of which is an endpoint of the interval. Without loss of generality, suppose $X_{q,n}$ is the lower end point. That is, $X_{q,n} = z-r_{k(n)}(z)$

Since F is a c.d.f. and $\{X_{p,n}\}_{p=1}^{n}$ are order statistics, we have $F(X_{1,n}) \leq \ldots \leq F(X_{q,n}) \leq \ldots \leq F(X_{q+j,n}) \leq \ldots \leq F(X_{q+k(n),n}) \leq \ldots \leq F(X_{n,n})$ where $j = 1,\ldots,k(n)$.

However, since $X_{q,n} = z-r_{k(n)}(z)$, we have

$$F(X_{q+k(n)-1,n})-F(X_{q,n}) \le U_{k(n)}(z) \le F(X_{q+k(n),n})-F(X_{q,n})$$

because $F(X_{q+k(n)-1,n}) \le F(z+r_{k(n)}(z)) \le F(X_{q+k(n),n})$

However,

$$F(X_{q+k(n)-1,n})-F(X_{q,n}) = \sum_{j=1}^{k(n)-1} \{F(X_{q+j,n})-F(X_{q+j-1,n})\}$$

and

$$F(X_{q+k(n),n})-F(X_{q,n}) = \sum_{j=1}^{k(n)} \{F(X_{q+j,n})-F(X_{q+j-1,n})\}$$

with the conventions $F(X_{o,n}) = 0$ and $F(X_{n+1,n}) = 1$

Therefore,

$$\sum_{j=1}^{k(n)-1} \{F(X_{q+j,n})-F(X_{q+j-1,n})\} \le U_{k(n)}(z) \le \sum_{j=1}^{k(n)} \{F(X_{q+j,n}) - F(X_{q+j-1,n})\} \qquad (2.3.6)$$

from which we obtain

$$\{\frac{n}{k(n)-1}\} \sum_{j=1}^{k(n)-1} \{F(X_{q+j,n})-F(X_{q+j-1,n})\} \le \{\frac{n}{k(n)-1}\} U_{k(n)}(z) \le \{\frac{n}{k(n)-1}\} \sum_{j=1}^{k(n)} \{F(X_{q+j,n})-F(X_{q+j-1,n})\} \qquad (2.3.7)$$

It is known that the random variables $F(X_{1,n}), F(X_{2,n})-F(X_{1,n}), \ldots, F(X_{n,n})-F(X_{n-1,n}), 1-F(X_{n,n})$ have the same joint distribution as the random variables $\frac{Y_1}{S_{n+1}}, \frac{Y_2}{S_{n+1}}, \ldots, \frac{Y_{n+1}}{S_{n+1}}$ where $Y_1, Y_2, \ldots, Y_{n+1}$ are independent exponential random variables with mean 1 and $S_{n+1} = Y_1+Y_2+\ldots+Y_{n+1}$ ([11], p.78).

Hence, if we can show that

$$\max_{0\le i\le n-k(n)+1} \left|\frac{1}{k(n)} \sum_{j=i+1}^{i+k(n)} \frac{Y_j}{n^{-1}S_{n+1}} -1\right| \to 0(P) \tag{2.3.8}$$

Then the upper and lower bounds for $\{\frac{n}{k(n)-1}\} U_{k(n)}(z)$ in (2.3.7) will converge to 1 (UP). [A7]. This in turn will imply that

$$\{\frac{n}{k(n)-1}\} U_{k(n)}(z) \to 1 \text{ (UP)}.$$

Thus we shall now establish (2.3.8).

Since the Y_j's are independent, identically distributed random variables with mean 1,

$n^{-1}S_{n+1} \to 1(p)$ by the strong law of large numbers.

Therefore (2.3.8) will follow if we show that

$$\left|\frac{1}{k(n)} \sum_{j=i+1}^{i+k(n)} Y_j -1\right| \to 0 \text{ (UP)}$$

Let $\varepsilon > 0$ be arbitrary and define P_n as

$$P_n = P\{\text{for some } i, \left|\sum_{j=i+1}^{i+k(n)} (Y_j-1)\right|>k(n)\varepsilon\}$$

$$= P\{\text{for some } i, \sum_{j=i+1}^{i+k(n)} (Y_j-1)>k(n)\varepsilon\} +$$

$$P\{\text{for some } i, \sum_{j=i+1}^{i+k(n)} (Y_j-1)<-k(n)\varepsilon\}$$

$$\le \sum_{i=1}^{n} P\{\sum_{j=i+1}^{i+k(n)} (Y_j-1)>k(n)\varepsilon\} + \sum_{i=1}^{n} P\{\sum_{j=i+1}^{i+k(n)} (Y_j-1)<-k(n)\varepsilon\}$$

We establish a bound for the first term by using the fact that

if X is any r.v. such that $E\{e^{tX}\} < \infty$ and $t > 0$, Then $P\{X>0\} \le E\{e^{tX}\}$, ([16], p.158).

Letting $X = \sum_{j=i+1}^{i+k(n)} (Y_j-1)-k(n)\varepsilon$ and noting that $E(e^{tX}) < \infty$

since the Y_j's are exponential random variables, we obtain

$$P\{\sum_{j=i+1}^{i+k(n)} (Y_j-1)>k(n)\varepsilon\} \le E\{\exp(t(\Sigma Y_j-k(n)-k(n)\varepsilon))\}$$

$$= \exp(-k(n)t-k(n)\varepsilon t)E\{\exp(t\Sigma Y_j)\}$$

$$= \{e^{-t(1+\varepsilon)}\}^{k(n)}\{E\{e^{tY}\}\}^{k(n)}$$ since Y_j's are i.i.d. r.v.

$$= \{\frac{e^{-t(1+\varepsilon)}}{1-t}\}^{k(n)} \quad \text{for} \quad 0 < t < 1$$

Since $E\{e^{tY}\} = (1-t)^{-1}$ for $0 < t < 1$

Differentiating $(1-t)^{-1} e^{-t(1+\varepsilon)}$, we have

$t \min = 1-(1+\varepsilon)^{-1}$ from which we obtain

$$\frac{e^{-t \min(1+\varepsilon)}}{1-t \min} = (1+\varepsilon)e^{-2}e^{-\varepsilon}$$

$$<(1+\varepsilon)e^{-\varepsilon}$$

and we conclude that

$$P\{\sum_{j=i+1}^{i+k(n)} (Y_j-1)>k(n)\varepsilon\} \le \{(1+\varepsilon)e^{-\varepsilon}\}^{k(n)}$$

$$= \{\frac{e^{\varepsilon}}{1+\varepsilon}\}^{-k(n)}$$

$$= \{a(\varepsilon)\}^{-k(n)}$$

where $a(\varepsilon) \equiv \frac{e^{\varepsilon}}{1+\varepsilon} > 1$ if $\varepsilon > 0$.

A similar bound exists for each term of the second sum in (2.3.9)

Namely,

$$P\{\sum_{j=i+1}^{i+k(n)} (Y_j-1)<-k(n)\varepsilon\} \le \{(1-\varepsilon)e^{\varepsilon}\}^{k(n)}$$

$$= \{(1-\varepsilon)^{-1}e^{-\varepsilon}\}^{-k(n)}$$

$$= \{b(\varepsilon)\}^{-k(n)}$$

where $b(\varepsilon) \equiv (1-\varepsilon)^{-1}e^{-\varepsilon} > 1$ if $\varepsilon > 0$.

Let $c(\varepsilon) \equiv \min(a(\varepsilon),b(\varepsilon))$ for $\varepsilon > 0$, then from (2.3.9) and the above bounds we conclude that

$$P_n \le 2n\{c(\varepsilon)\}^{-k(n)} \quad \text{where} \quad c(\varepsilon) > 1 \quad \text{if} \quad \varepsilon > 0.$$

However,

$$2n\{c(\varepsilon)\}^{-k(n)} = \frac{2n}{\{c(\varepsilon)\}^{k(n)}}$$

$$= \frac{2e^{\log n}}{e^{k(n)\log c(\varepsilon)}}$$

$$= 2\{e^{\log n-k(n)\log c(\varepsilon)}\}$$

Since $c(\varepsilon) > 1$, $\log c(\varepsilon) > 0$ and recalling that $\log n = o(k(n))$ we conclude that

$$\log n-k(n)\log c(\varepsilon) \to -\infty \quad \text{as} \quad n \to \infty.$$

Therefore, $\lim_{n\to\infty} P_n = o$

Thus (2.3.8) follows and hence (2.3.5).

With (2.3.5) we easily obtain

$$U_{k(n)}(z) \to 0 \text{ (UP)} \tag{2.3.5a}$$

and

$$r_{k(n)}(z) \to 0 \text{ (UP)} \tag{2.3.5b}$$

(2.3.5a) is immediate. We shall show (2.3.5b).

That $r_{k(n)}(z) \to 0$ (UP) is almost obvious.

For suppose that $r_{k(n)}(z)$ is bounded away from zero in probability. Then recalling that f is everywhere positive and continuous it would follow that $U_{k(n)}(z)$ is bounded away from zero in probability which contradicts $U_{k(n)}(z) \to 0$ (UP).

For (2.3.3) we need that $\frac{U_{k(n)}}{2r_{k(n)}} \to f$(UP). This result follows since

$$\left|\frac{U_{k(n)}(z)}{2r_{k(n)}(z)}\right| - |f(z)| = \left|\frac{F(z+r_{k(n)}(z))-F(z-r_{k(n)}(z))}{2r_{k(n)}(z)} - f(z)\right|$$

$$= \left|\frac{1}{2r_{k(n)}(z)}\int_{z-r_{k(n)}(z)}^{z+r_{k(n)}(z)} f(t)dt - \frac{1}{2r_{k(n)}(z)}\int_{z-r_{k(n)}(z)}^{z+r_{k(n)}(z)} f(z)dt\right|$$

$$= \left|\frac{1}{2r_{k(n)}(z)}\int_{z-r_{k(n)}(z)}^{z+r_{k(n)}(z)} [f(t)-f(z)]dt\right|$$

$$\leq \max_{t\varepsilon[z-r_{k(n)}(z),z+r_{k(n)}(z)]} |f(t)-f(z)| \left|\frac{1}{2r_{k(n)}(z)}\int_{z-r_{k(n)}(z)}^{z+r_{k(n)}(z)} 1dt\right|$$

$$= \max_{t\varepsilon[z-r_{k(n)}(z),z+r_{k(n)}(z)]} |f(t)-f(z)|$$

If $|t-z| = r_{k(n)}(z)$ is sufficiently small, we have $|f(t)-f(z)|$ arbitrarily small. Note that $r_{k(n)} \to 0(UP)$ implies $|t-z|$ is sufficiently small (UP) and this with the fact that f is uniformly continuous on the reals implies that $|f(t)-f(z)|$ is arbitrarily small.

Hence,

$$\frac{U_{k(n)}}{2r_{k(n)}} \to f(UP).$$

We shall now show (2.3.3). From the definition of $\hat{f}n(z)$, we obtain

$$|\hat{f}n(z)-f(z)| = \left|\{\frac{k(n)-1}{n}\}\{\frac{1}{2r_{k(n)}(z)}\} - f(z)\right|$$

$$= \left| \left(\frac{1}{\{\frac{n}{k(n)-1}\} U_{k(n)}(z)} \right) \left(\frac{U_{k(n)}(z)}{2r_{k(n)}(z)} \right) - f(z) \right|$$

From (2.3.5) we have that $\{\frac{n}{k(n)-1}\} U_{k(n)}(z) \to 1(UP)$ and we have just shown that $\frac{U_{k(n)}(z)}{2r_{k(n)}(z)} \to f(z)(UP)$. Hence we conclude that $\hat{f}n(z) \to f(z)(UP)$.

The argument for (2.3.4) is quite similar to that for (2.3.3) and for this reason it will not be discussed. However, $fn^*(z)$ is important in the estimation problem since fn^* is easy to compute.

We shall conclude Chapter 2 with the following remark:

(a) the argument used to show (2.3.3) is a valid argument for the L and Q theorem.

(b) choosing $k(n)$ near $n^{1/2}$ appears to give "good" estimates of $f(z)$.

CHAPTER 3 - SERIES ESTIMATORS

The series technique of density estimation was conceived by Cencov [5], and independently by Schwartz [26]. A basic assumption we shall make is that $f(x) \in L_2(R)$. For example, bounded density functions are square integrable over the reals.

Recall that with an inner product $(,): L_2(R) \times L_2(R) \to R$ which is defined by

$$(f,g) = \int f(x)g(x)dx \quad \text{for} \quad f,g \in L_2(R)$$

$L_2(R)$ is a Hilbert space. ([12], p.235).

The Hermite functions $\phi_j(x) = (2^j j! \pi^{1/2})^{-1/2} e^{-\frac{x^2}{2}} H_j(x)$ where $H_j(x) = (-1)^j e^{x^2} \frac{dj}{dxj}(e^{-x^2})$ for $j = 0,1,2,\ldots$ is the j^{th} Hermite polynomial constitute a complete orthonormal set in $L_2(R)$. ([12], p.416).

Let $\Phi = \{\phi_j(x) \mid j=0,1,\ldots\}$

Since Φ is an orthonormal subset of $L_2(R)$,

The following properties are equivalent ([12], p.245).

(a) the set Φ is complete

(b) for all $f \in L_2(R)$, $f = \sum_{j=0}^{\infty} a_j \phi_j(x)$ where $a_j = (f,\phi_j)$

(c) for all $f \in L_2(R)$, $||f||^2 = (f,f) = \sum_{j=o}^{\infty} |a_j|^2$

One may wonder, why we shall use the Hermite functions rather than the more familiar trigonometric functions as our orthonormal set. The reason is simple. The trigonometric functions are not square integrable over the reals. However, for function spaces whose elements are defined on compact subsets of R, the trigonometric functions are extremely useful in approximating. Kronmal and Tarter, [13], have discussed the estimation of density functions which have support on compact subsets of R making extensive use of the trigonometric functions. For more arbitrary densities, they have devised a truncation procedure.

Proceeding with Schwartz's results,

Since $f(x) \in L_2(R)$ we have

$$f(x) = \sum_{j=0}^{\infty} a_j \phi_j(x) \quad \text{where} \quad a_j = \int f(x)\phi_j(x)\,dx \tag{3.1.1}$$

As an estimate of $f(x)$, Schwartz proposed

$$\hat{f}n(x) = \sum_{j=0}^{g(n)} \hat{a}_{jn}\phi_j(x) \quad \text{where} \tag{3.1.2}$$

$$\hat{a}_{jn} = \frac{1}{n}\sum_{i=1}^{n}\phi_j(X_i) \tag{3.1.3}$$

and $g(n)$ is a sequence of positive integers such that $g(n)=0(n)$ and $g(n) \to \infty$ as $n \to \infty$.

(3.1.2) and (3.1.3) are obvious estimates because assuming $f(x) \in L_2(R)$ we have (3.1.1) where only a_j is dependent upon f. Therefore our problem is reduced to estimating the Fourier coefficients a_j.

Notice that $\hat{a}_{jn}$ is an unbiased estimate of a_j since

$$E\{\hat{a}_{jn}\} = \frac{1}{n}\sum_{i=1}^{n} E\{\phi_j(X_i)\}$$

$$= E\{\phi_j(X)\}$$

$$= \int \phi_j(x) f(x)\,dx$$

$$= a_j$$

Just as in the case of Kernel estimators our criterion of a good estimate will be a consistent estimate.

To discuss consistency of estimates of the form (3.1.2) in terms of $\quad MSE = \lim_{n\to\infty} E\{(\hat{f}n(x)-f(x))^2\}$

and $\quad MISE = \lim_{n\to\infty} E\{\int(\hat{f}n(x)-f(x))^2 dx\}$ we need

Lemma 1

Let $f(x)$ be continuous, of bounded variation, L_1 and L_2 in $(-\infty,\infty)$ then $\sum_{j=0}^{\infty} a_j\phi_j(x)$ converges uniformly to $f(x)$ in any interval interior to $(-\infty,\infty)$. (See [24], section 4.10)

To discuss rates of convergence for the MSE and MISE we need

Lemma 2

Assume $f'(x)$ exists and that $(xf(x)-f'(x)) \varepsilon L_2(R)$ then $a_j, j=1,2,\ldots$ satisfy

$$|a_j| < \frac{C_3}{(2j)1/2} \quad \text{where } C_3 \text{ is the } L_2 \text{ norm of} \qquad (3.1.4)$$

$$(xf(x)-f'(x))$$

Proof

$$a_j = \int f(x)\phi_j(x)dx$$
$$= \int f(x)\,(2^j j!\pi^{1/2})^{-1/2} e^{-\frac{x^2}{2}} H_j(x)dx$$

The Hermite polynomials satisfy the following recursive relation

$$\frac{d}{dx}(H_{j+1}(x)) = 2(j+1)H_j(x) \quad ([12], \text{p.}244)$$

Therefore, $(2(j+1))^{-1}\, dH_{j+1}(x) = H_j(x)dx$

Hence,

$$a_j = (2(j+1))^{-1}\int f(x)\,(2^j j!\pi^{1/2})^{-1/2} e^{-\frac{x^2}{2}} dH_{j+1}(x)$$

Letting $u = f(x)e^{-\frac{x^2}{2}}$ and $dr = (2(j+1))^{-1}(2^j j!\pi^{1/2})^{-1/2} dH_{j+1}(x)$

We have

$$du = (f'(x) - xf(x))e^{-\frac{x^2}{2}}dx$$

$$\text{and} \quad r = (2(j+1))^{-1}(2^j j!\pi^{1/2})^{-1/2} H_{j+1}(x)$$

Integrating by parts, we obtain

$$a_j = \left[f(x)e^{-\frac{x^2}{2}}(2(j+1))^{-1}(2^j j!\pi^{1/2})^{-1/2} H_{j+1}(x)\right]_{-\infty}^{\infty}$$
$$+ (2j+2)^{-1/2}\int (xf(x) - f'(x))e^{-\frac{x^2}{2}}(2^{j+1}(j+1)!\,\pi^{1/2})^{-1/2} H_{j+1}(x)dx$$

The first term in the expression for a_j is 0 since $H_{j+1}(x)$ is a polynomial of degree $j+1$ and f is bounded.

Hence with the definition of $\phi_{j+1}(x)$, we have that

$$a_j = (2j+2)^{-1/2}\int (xf(x)-f'(x))\phi_{j+1}(x)\,dx$$

Since $(xf(x)-f'(x)) \in L_2(R)$, by the Schwarz inequality

$$|a_j| \le (2j+2)^{-1/2}\left(\int (xf(x)-f'(x))^2 dx\right)^{1/2}\left(\int \phi_{j+1}^2(x)\,dx\right)^{1/2}$$

$$= (2j+2)^{-1/2}\,||(xf(x)-f'(x)||\cdot 1 \quad \text{since } \Phi \text{ is orthonormal}$$

$$< \frac{C_3}{(2j)^{1/2}} \quad \text{where } C_3 \equiv ||(xf(x)-f'(x))||$$

<u>Lemma 3</u>

<u>Assume</u> $e^{\frac{x^2}{2}}\frac{d^r}{dx^r}(e^{-\frac{x^2}{2}}f(x)) = \sum_{i=0}^{r}\binom{r}{i}(-1)^i 2^{-\frac{i}{2}} Hi\left(\frac{x}{2^{1/2}}\right)\frac{d^{r-i}}{dx^{r-i}}(f(x))$

<u>exists and is square integrable over the reals. Then</u> $a_j, j=1,2,\ldots$ <u>satisfy</u>

$|a_j| < \frac{C_3(r)}{(2j)^{r/2}}$, <u>where</u> $C_3(r)$ <u>is the</u> L_2 <u>norm of</u> $e^{\frac{x^2}{2}}\frac{d^r}{dx^r}(e^{-\frac{x^2}{2}}f(x))$.

<u>Proof</u>

By repeated application of the method of Lemma 2.

<u>Remark</u> The standard normal density function satisfies the hypotheses of the lemmas.

Cramér ([9], p.208) has established the following bound for the Hermite functions:

$$|\phi_j(x)| < \frac{C_1}{\pi^{1/2}} = C_2 \quad \text{where } C_1 \text{ (hence } C_2\text{) is independent of } x \text{ and } j. \qquad (3.1.6)$$

Theorem (3.1.7)

Assume $f(x)$ is square integrable and that the sequence of positive integers $q(n)$ is chosen such that $q(n)=o(n)$. Then the sequence of estimates defined by (3.1.2) and (3.1.3) is consistent in the sense of MISE. Furthermore, if $f(x)$ satisfies the hypotheses of lemma 3 with $r \geq 2$, then with $q(n) = O(n^{\frac{1}{r}})$ the MISE satisfies

$$E\{\int(\hat{f}n(x)-f(x))^2dx\} = O(\frac{1}{n^{\frac{r-1}{r}}})$$

Proof

Since Φ is an orthonormal subset of $L_2(R)$, by the Pythagorean theorem

$$||\hat{f}n(x)-f(x))||^2=||fq(n)(x)-f(x)||^2+||\hat{f}n(x)-fq(n)(x)||^2 \quad (3.1.8)$$

where for $x \in R$ $fq(n)(x) \equiv \sum_{j=0}^{q(n)} a_j\phi_j(x)$

but $||\hat{f}n(x)-fq(n)(x)||^2=\sum_{j=0}^{q(n)} (\hat{a}_{jn}-a_j)^2$ by definition of $\hat{f}n(x)$ and $fq(n)(x)$

and $||fq(n)-f(x)||^2=\sum_{j=q(n)+1}^{\infty} a_j^2$ since Φ is orthonormal

Hence from (3.1.8),

$$||\hat{f}n(x)-f(x)||^2=\sum_{j=q(n)+1}^{\infty} a_j^2+\sum_{j=0}^{q(n)} (\hat{a}_{jn}-a_j)^2 \quad (3.1.9)$$

and therefore,

$$E\{||\hat{f}n(x)-f(x)||^2\} = E\{\int(\hat{f}n(x)-f(x))^2dx\} \quad (3.1.10)$$

$$= \sum_{j=q(n)+1}^{\infty} a_j^2+\sum_{j=0}^{q(n)} E\{(\hat{a}_{jn}-a_j)^2\}$$

We may obtain a bound for the second term on the right hand side of (3.1.10) as follows:

Since $E\{\hat{a}_{jn}\} = a_j$ we have that

$$E\{(\hat{a}_{jn}-a_j)^2\} = VAR(\hat{a}_{jn})$$

$$= E\{\hat{a}_{jn}^2\} - a_j^2$$

However,

$$(\hat{a}_{jn})^2 = \frac{1}{n^2}\{\sum_{k=1}^{n}\phi_j^2(X_k)+\sum_{\substack{k_1,k_2=1\\ k_1\neq k_2}}^{n}\phi_j(X_{K_1})\phi_j(X_{K_2})\}$$

Therefore

$$E\{(\hat{a}_{jn}-a_j)^2\}=\frac{1}{n^2}E\{\sum_{k=1}^{n}\phi_j^2(X_K)+\sum_{\substack{k_1,k_2=1\\ k_1\neq k_2}}^{n}\phi_j(X_{K_1})\phi_j(X_{K_2})\}-a_j^2$$

$$\leq\frac{1}{n^2}\{nC_2^2+\sum_{\substack{k_1,k_2=1\\ k_1\neq k_2}}^{n}E(\phi_j(X_{K_1}))E(\phi_j(X_{K_2}))\}-a_j^2 \quad \text{by (3.1.6)}$$

and the independence of the r.v.

$$=\frac{1}{n^2}\{nC_2^2+\sum_{\substack{k_1,k_2=1\\ k_1\neq k_2}}^{n}a_j^2\}-a_j^2 \quad \text{since } E\{\phi_j(X)\}=a_j$$

$$=\frac{1}{n^2}\{nC_2^2+n(n-1)a_j^2\}-a_j^2$$

$$=\frac{C_2^2-a_j^2}{n}$$

Hence,

$$E\{(\hat{a}_{jn}-a_j)^2\} \le \frac{C_2^{\,2}-a_j^{\,2}}{n} \qquad (3.1.11)$$

$$= \frac{C_4}{n} \quad \text{where} \quad C_4 = C_2^{\,2}-a_j^{\,2}$$

From (3.1.10) and (3.1.11)

$$E\{\int(\hat{f}n(x)-f(x)^2dx\} \le \sum_{j=q(n)+1}^{\infty} a_j^{\,2} + \frac{q(n)+1}{n} C_4 \qquad (3.1.12)$$

Since $f(x) \in L_2(R)$ and $q(n)=o(n)$, we have

$$\lim_{n\to\infty} E\{\int(\hat{f}n(x)-f(x))^2dx\} = 0$$

Thus, the sequence of estimates $\{\hat{f}n(x)\}_{n=1}^{\infty}$ is consistent in MISE.

To establish the rate of convergence as in (3.1.7), assume $f(x)$ satisfies the hypotheses of lemma 3 with $r \ge 2$ and choose $q(n)$ such that $q(n) = O(n^{\frac{1}{r}})$.

Then,

$$E\{\ (\hat{f}n(x)-f(x))^2dx\} \le \sum_{j=q(n)+1}^{\infty} a_j^{\,2} + \frac{q(n)+1}{n} C_4 \quad \text{by (3.1.12)}$$

$$\le \sum_{j=q(n)+1}^{\infty} \frac{C_3^{\,2}(r)}{(2j)^r} + \frac{q(n)+1}{n} C_4 \quad \text{by lemma 3,}$$

$$|a_j| < \frac{C_3(r)}{(2j)^{r/2}}$$

$$= C_3^{\,2}(r) \sum_{j=q(n)+1}^{\infty} \frac{1}{(2j)^r} + \frac{q(n)+1}{n} C_4$$

$$\le C_3^{\,2}(r) \int_{q(n)}^{\infty} \frac{dx}{x^r} + \frac{q(n)+1}{n} C_4 \quad \text{since } r \ge 2 \text{ implies}$$

the series converges we can use

the Integral test.

$$= \frac{C_3^2(r)}{2^r(r-1)q^{r-1}} + \frac{q(n)+1}{n} C_4$$

Let $q(n)$ be the largest integer less than or equal to $\left(\frac{nC_3^2(r)}{C_4}\right)^{\frac{1}{r}}$.

Then for each n there is a $\delta(n)$ such that $0 \le \delta(n) < 1$ and

$$q(n)+\delta(n) = \left(\frac{nC_3^2(r)}{C_4}\right)^{\frac{1}{r}}$$

Hence, $q(n) = \left(\frac{nC_3^2(r)}{C_4}\right)^{\frac{1}{r}} - \delta(n)$

Therefore,

$$E\left\{\int(\hat{f}n(x)-f(x))^2dx\right\} \le \frac{C_3^2(r)}{2^r(r-1)q^{r-1}} + (q+1)\frac{C_4}{n}$$

$$= \frac{C_3^2(r)}{2^r(r-1)\left[\left(\frac{nC_3^2(r)}{C_4}\right)^{1/r} - \delta(n)\right]^{r-1}} +$$

$$\frac{C_4}{n}\left\{\left(\frac{nC_3^2(r)}{C_4}\right)^{1/r} - \delta(n)+1\right\}$$

$$= K_1(r)\frac{1}{\left[n^{\frac{1}{r}} - \delta(n)\left(\frac{C_4}{C_3^2(r)}\right)^{1/r}\right]^{r-1}} +$$

$$\left(\frac{1}{n}\right)^{1-\frac{1}{r}} K_2(r) - \frac{C_4\delta(n)}{n} + \frac{C_4}{n}$$

$$\le K_1(r)\frac{1}{\left[n^{\frac{1}{r}} - \left(\frac{C_4}{C_3^2(r)}\right)^{1/r}\right]^{r-1}} +$$

$$\left(\frac{1}{n}\right)^{1-\frac{1}{r}} K_2(r) \frac{C_4\ (n)}{n} + \frac{C_4}{n} \quad \text{since } 0 \le \delta(n) < 1$$

Multiplying the first term by $n^{\frac{r-1}{r}}$ yields

$$K_1(r)\left[\frac{n^{\frac{1}{r}}}{n^{\frac{1}{r}}-\left(\frac{C_4}{C_3^{\,2}(r)}\right)^{1/r}}\right]^{r-1} \to 1 \quad \text{as} \quad n \to \infty.$$

Hence the first term is $O\left(\frac{1}{n^{\frac{r-1}{r}}}\right)$

Clearly the other terms are $O\left(\frac{1}{n^{\frac{r-1}{r}}}\right)$ and we have (3.1.7).

For consistency in Mean Square, more conditions are required on the density function and the sequence $q(n)$. Specifically,

<u>Theorem (3.1.13)</u>

<u>Assume $f(x)$ is continuous, of bounded variation, L_1 and L_2 in $(-\infty,\infty)$. Choose $q(n)$ such that $\frac{q^2(n)}{n} \to 0$ as $n \to \infty$. Then the sequence of estimates defined by (3.1.2) and (3.1.3) converges in Mean Square, uniformly in x. Furthermore, assume $f(x)$ satisfies the hypotheses of lemma 3 with $r \geq 3$. Then with $q(n) = O(n^{\frac{1}{r}})$ the mean square error satisfies</u>

$$E\{(\hat{f}n(x)-f(x))^2\} = O\left(\frac{1}{n^{\frac{r-2}{r}}}\right)$$

<u>Proof</u>

As in the preceding theorem, for $x \in R$ define

$f_{q(n)}(x) = \sum_{j=0}^{q(n)} a_j \phi_j(x)$ and notice that

$$E\{\hat{f}n(x)\} = E\{\sum_{j=0}^{q(n)} \hat{a}_{jn}\phi_j(x)\}$$

$$= \sum_{j=0}^{q(n)} E\{\hat{a}_{jn}\}\phi_j(x)$$

$$= \sum_{j=0}^{q(n)} a_j\phi_j(x) \quad \text{since } \hat{a}_{jn} \text{ is an unbiased estimate of } a_j$$

$$= fq(n)(x)$$

That is, $\hat{f}n(x)$ is an unbiased estimate of $fq(n)(x)$.

Therefore, $E\{(\hat{f}n(x)-f(x))^2\}$ (3.1.14)

$$=E\{[(\hat{f}n(x)-fq(n)(x)+(fq(n)(x)-f(x))]^2\}$$

$$=E\{(\hat{f}n(x)-fq(n)(x))^2+2(\hat{f}n(x)-fq(n)(x))(fq(n)(x))+(fq(n)(x)-f(x))^2\}$$

$$=E\{(\hat{f}n(x)-fq(n)(x))^2\}+2(fq(n)(x)-f(x))E\{(\hat{f}n(x)-fq(n)(x))\}+E\{(fq(n)(x)-f(x))^2\}$$

$$\equiv E\{(\hat{f}n(x)-fq(n)(x))^2\}+(fq(n)(x)-f(x))^2 \quad \text{since } E\{\hat{f}n(x)\} = fq(n)(x)$$

For the first term,

$$E\{(\hat{f}n(x)-fq(n)(x))^2\}$$

$$=E\{(\sum_{j=0}^{q(n)} \hat{a}_{jn}\phi_j(x)-\sum_{j=0}^{q(n)} a_j\phi_j(x))^2\}$$

$$=E\{(\sum_{j=0}^{q(n)}(\hat{a}_{jn}-a_j)\phi_j(x))^2\}$$

$$=E\{\sum_{j=0}^{q(n)} (\hat{a}_{jn}-a_j)^2\phi_j^2(x)+\sum_{\substack{j,k=0\\ j\neq k}}^{q(n)}(\hat{a}_{jn}-a_j)(\hat{a}_{kn}-a_k)\phi_j(x)\phi_k(x)\}$$

$$=\sum_{j=0}^{q(n)} E\{(\hat{a}_{jn}-a_j)^2\}\phi_j^2(x)+\sum_{\substack{j,k=0\\ j\neq k}}^{q(n)} E\{(\hat{a}_{jn}-a_j)(\hat{a}_{kn}-a_k)\}\phi_j(x)\phi_k(x)$$

By the Schwarz inequality,

$$|E\{(\hat{a}_{jn}-a_j)(\hat{a}_{kn}-a_k)\}| \le (E\{(\hat{a}_j n-a_j)^2\})^{1/2} (E\{(\hat{a}_{kn}-a_k)^2\})^{1/2}$$

and by (3.1.11),

$$E\{(\hat{a}_{jn}-a_j)^2\} \le \frac{C_4}{n}$$

while by (3.1.6),

$|\phi_j(x)| < C_2$ where C_2 is independent of x and j.

Therefore,

$$E\{(\hat{f}n(x)-fq(n)(x))^2\} \le (q(n)+1)C_2^2\frac{C_4}{n} + (q(n)+1)q(n)C_2^2\frac{C_4}{n}$$

$$= (q(n)+1)^2 C_2^2\frac{C_4}{n}$$

and from (3.1.14),

$$E\{(\hat{f}n(x)-f(x))^2\} \le (fq(n)(x)-f(x))^2+C_2^2C_4\frac{(q(n)+1)^2}{n} \qquad (3.1.15)$$

Letting $q(n) \to \infty$ in such a way that $\frac{q^2(n)}{n} \to 0$, we have

$\frac{(q(n)+1)^2}{n} \to 0$ because $\frac{(q(n)+1)^2}{n} = \frac{q^2(n)}{n} + 2\frac{q(n)}{n} + \frac{1}{n}$

and $\frac{q^2(n)}{n} \to 0$ implies $\frac{q(n)}{n^{1/2}} \to 0$ which in turn implies that

$\frac{q(n)}{n} \to 0$ since $\frac{q(n)}{n} \le \frac{q(n)}{n^{1/2}}$.

Hence from (3.1.15) and lemma 1

$$\lim_{n\to\infty} E\{(\hat{f}n(x)-f(x))^2\} = 0 \text{ uniformly in } x.$$

For the rate of convergence we assume the conditions of lemma 3 with $r \geq 3$.

Consider the first term in (3.1.15)

$$|f(x)-fq(n)(x)| = \left|\sum_{j=0}^{\infty} a_j\phi_j(x) - \sum_{j=}^{q(n)} a_j\phi_j(x)\right|$$

$$= \left|\sum_{j=q(n)+1}^{\infty} a_j\phi_j(x)\right|$$

$$\leq \sum_{j=q+1}^{\infty} |a_j\phi_j(x)|$$

$$\leq C_2 \sum_{j=q+1}^{\infty} |a_j| \quad \text{since } |\phi_j(x)| < C_2 \text{ for all } x \text{ and all } j=0,1,\ldots$$

$$\leq \frac{C_2C_3(r)}{2^{r/2}} \sum_{j=q+1}^{\infty} \frac{1}{j^{r/2}} \quad \text{by lemma 3, } |a_j| < \frac{C_3(r)}{(2j)^{r/2}}$$

$$\leq \frac{C_2C_3(r)}{2^{r/2}} \int_{x=q}^{\infty} \frac{dx}{x^{r/2}} \quad \text{by the Integral test}$$

$$= \frac{C_2C_3(r)}{2^{r/2}(\frac{r}{2}-1)q^{\frac{r}{2}-1}}$$

Therefore, from (3.1.15)

$$E\{(\hat{f}n(x)-f(x))^2\} \leq \frac{C_2{}^2C_3(r)}{2^r(\frac{r}{2}-1)^2 q^{r-2}} + C_2{}^2C_4 \frac{(q(n)+1)^2}{n} \qquad (3.1.16)$$

Finally, to obtain (3.1.13), let $q(n)$ be the largest integer less than or equal to $\left(\frac{(r-2)C_3^2(r)n}{2^r(\frac{r}{2}-1)^2C_4}\right)^{\frac{1}{r}}$. Then for each n there is a $\delta(n)$ such that

$$0 \le \delta(n) < 1 \quad \text{and} \quad q(n)+\delta(n) = \left(\frac{(r-2)C_3^2(r)n}{2^r(\frac{r}{2}-1)^2C_4}\right)^{\frac{1}{r}} \quad \text{from which}$$

$$q(n) = \left(\frac{(r-2)C_3^2(r)n}{2^r(\frac{r}{2}-1)^2C_4}\right)^{\frac{1}{r}} - \delta(n)$$

With this choice of $q(n)$ in (3.1.16), we obtain

$$E\{(\hat{f}n(x)-f(x))^2\}$$

$$\le \frac{K_1(r)}{\left[\left(\frac{(r-2)C_3^2(r)n}{2^r(\frac{r}{2}-1)^2C_4}\right)^{1/r}-\delta(n)\right]^{r-2}} + \frac{C_2^2C_4\left[\left(\frac{(r-2)C_3^2(r)n}{2^r(\frac{r}{2}-1)^2C_4}\right)^{\frac{1}{r}}-\delta(n)+1\right]^2}{n}$$

$$\le \frac{K_1(r)K_2(r)}{\left[n^{\frac{1}{r}}-\delta(n)\left(\frac{2^r(\frac{r}{2}-1)^2C_4}{(r-2)C_3^2(r)}\right)^{\frac{1}{r}}\right]^{r-2}} + \frac{C_2^2C_4\left[\left(\frac{(r-2)C_3^2(r)n}{2^r(\frac{r}{2}-1)^2C_4}\right)^{1/r}+1\right]^2}{n} \quad \text{since } 0 \le \delta(n) < 1$$

$$\le \frac{K_3(r)}{\left[n^{\frac{1}{r}}-\left(\frac{2^r(\frac{r}{2}-1)^2C_4}{(r-2)C_3^2(r)}\right)^{1/r}\right]^{r-2}} + K_4(r)\left(\frac{1}{n}\right)^{\frac{r-2}{r}} + \frac{K_5(r)n^{1/r}}{n} + \frac{K_6}{n}$$

Each of the four terms is $O\left(\frac{1}{n^{\frac{r-2}{r}}}\right)$ as in the previous theorem.

Hence,

$$E\{(\hat{f}n(x)-f(x))^2\} = O\left(\frac{1}{n^{\frac{r-2}{r}}}\right)$$

CHAPTER 4 - COMPARISONS AND APPLICATIONS

4.1 Comparisons

In 2.3 we have compared the Kernel and L and Q techniques in terms of the form of the estimators and the motivation for the form of the estimators. In this section, we shall compare the Kernel and Series methods.

From a computational point of view, the Series estimator is better than the Kernel estimator since the Series estimator is easier to update. That is, suppose we have a second group of observations and wish to construct an estimator using the new observations as well as our first set of observations. In the Series method, we would have to compute new coefficients $\hat{a}_{jn}$ which can be expressed in terms of the old coefficients $\hat{a}_{jn}$ and the new observations. However, if we had constructed a Kernel estimate of the form $\hat{f}n(x) = \frac{1}{nh}\sum_{j=1}^{n} K(\frac{x-X_j}{h})$ an increase in the sample size would usually change our choice of h and we would have to compute a new estimate using all of the observations since in general a recursive relationship does not exist.

An apparent disadvantage of the Series method is the possibility of negative estimates of the density function over non-degenerate subsets of R . However, Anderson [1], and Kronmal and Tarter, [13], have stated that in practice this situation does not arise. In the Kernel method the possibility does not arise since the weighting functions are chosen to be non-negative.

Kernel estimators generally enjoy better convergence properties as we shall now illustrate. Following Watson and Leadbetter theorem (1.3.12) for $f(x) \in L_2(R)$, if $\phi_f(t)$ decreases algebraically of degree $p > 1/2$, then

$$\lim_{n\to\infty}[n^{\frac{2p-1}{2p}} E\{\int(\hat{f}n(x)-f(x))^2dx\}] = a \quad \text{where} \quad a \in R \quad \text{and} \quad a \neq 0$$

That is, $E\{\int(\hat{f}n(x)-f(x))^2dx\} = O(\frac{1}{n^{\frac{2p-1}{2p}}})$

For the Series method to obtain the same rate of convergence we would need to require that the density satisfies

$e^{\frac{x^2}{2}}\{\frac{d^{2p}}{dx^{2p}}[e^{\frac{x^2}{2}}f(x)]$ exists and is square integrable where

$2p \geq 2$ (Lemma 3, p.76). (4.1.1)

A sufficient condition for (4.1.1) is that the functions

$x^i(\frac{d^{2p-i}}{dx^{2p-i}}[f(x)])$ exist and be integrable for $i = 0,1,\ldots,2p$. (4.1.2)

Thus for the Series method to achieve the same rate of convergence as the Kernel method in MISE we have to assume that the density function satisfies the differentiability and integrability properties in (4.1.2).

We shall now consider consistency in MSE. Suppose the Kernel K satisfies $\int x^i K(x)dx = 0$ for $i = 1,2,\ldots,r-1$ (4.1.3)

and $\int x^r|K(x)|dx < \infty$.

If $f(x)$ satisfies $\int|t^r\phi_f(t)|dt < \infty$ (4.1.4)

and $q(n)$ is chosen such that $q(n) = O(n^{-\frac{1}{2r+1}})$, then

$$E\{(\hat{f}n(x)-f(x))^2\} = O(n^{-\frac{2r}{2r+1}}) \qquad ([21], \text{pp.}1072\text{-}1074).$$

Note that property (4.1.4) implies $f^r(x)$ exists.

For example, suppose $r = 3$ and $f^3(x)$ exists. If we choose K satisfying (4.1.3) and $h(n) = O(n^{-7})$ then $E\{(fn(x)-f(x))^2\} = O(n^{-\frac{6}{7}})$. Whereas, in the Series method, with $r = 3$ and assuming that $f^3(x), xf^2(x), x^2f'(x), x^2f(x) \in L_1(R)$ then by choosing $q(n) = O(n^{-1/3})$ we have $E\{(\hat{f}n(x)-f(x))^2\} = O(n^{-1/3})$. (Theorem 3.1.3).

Thus for $r = 3$, and the appropriate assumptions, the Kernel method estimator achieves a rate of convergence in MSE which is more than twice as fast as the Series estimator.

As Schwartz [26], has observed, it is in the estimation of a Multivariate density function that the Series method may prove to be most advantageous. For in the Series method, the rate of convergence in the sense of MSE or MISE is dependent upon the differentiability properties of the density and is independent of the dimension of the density which we are estimating.

However, for the multivariate Kernel estimator, the rate of convergence is dependent upon the dimension of the density being estimated. As a matter of fact, the rate of convergence decreases with increasing dimension. [4].

4.2 Applications

The estimation of density functions has been applied to the testing of hypotheses [27] and to problems of classification [29]. Schwartz [27] has considered the following problem:

Suppose we have a single observation of a r.v. X and we wish to decide between the following simple hypotheses:

$$H_0: \; X \text{ has d.f. } f_0(x)$$

$$H_1: \; X \text{ has d.f. } f_1(x)$$

with apriori probabilities π_0 and $\pi_1 = 1-\pi_0$ respectively. With a suitable test function $T = T(f_0,f_1)$ and a criterion we can decide whether to accept or reject H_0.

Schwartz considered the case where the density functions are unknown but the apriori probabilities are known. Assuming that the unknown density functions are square integrable over the reals, he estimated the densities by the methods of Chapter 3. A test function $\hat{T} = \hat{T}(\hat{f}_0,\hat{f}_1)$ was then defined in terms of the estimated densities. In order to test the hypotheses, he used the criterion in the case where the densities are known.

Van Ryzin, [29], considered a similar problem. However, his results are more general in that the random sample is of arbitrary size and the apriori probabilities as well as the density functions are estimated. In addition, Van Ryzin is concerned with the consistency properties of the classification procedure rather than with just the consistency properties of the density estimates.

CHAPTER 5 - GENERALIZATIONS AND FURTHER RESULTS

In this section, we shall discuss some results which are generalizations of the Kernel and Series methods. However, our main purpose is to summarize some recent results which are helpful in applying the techniques of density estimation.

Murthy [19], has generalized Parzen's results to the case where the c.d.f. is of the form

$$F(x)=F_1(x)+F_2(x)$$

where $F_1(x)$ is everywhere continuous and $F_2(x)$ is a pure step function with steps of magnitude S_v at x_v, $v = 1,2,\ldots$. We are assuming that $F_3(x)$, the singular component of $F(x)$ is identically zero. ([7], pp.52,53).

Analogously to Parzen, Murthy has shown that the sequence of estimates $\{\hat{f}n(x)\}_{n=1}^{\infty}$ is an asymptotically unbiased estimate of $f(x)$, a consistent estimate of $f(x)$ in MSE and is asymptotically normally distributed provided x is a point of continuity of both $F(x)$ and $f(x)$ and the series $\sum_v \frac{S_v}{|x_v - x|}$ converges.

As a further abstraction, Craswell [8] has considered density estimation in a topological group. He has generalized sequences of weighting functions to so called δ sequences of functions and considered estimates of $f(x)$ of the form

$$\hat{f}n(x) = \frac{1}{n}\sum_{j=1}^{n} K_n(x \cdot X_j^{-1})$$ where $\cdot$ is the group operation

Craswell has established that if s and t are distinct continuity points of f with $f(s) + f(t) \neq 0$, then $\{\hat{f}n(s), \hat{f}n(t)\}_{n=1}^{\infty}$ is jointly asymptotically normal and independent.

Schuster [25] has considered estimating a density function and its derivatives. The form of the estimator is $\hat{f}n^{(r)}(x) = \frac{1}{nh^{r+1}} \sum_{i=1}^{n} K^{(r)}\left(\frac{x-X_i}{h}\right)$ for $r = 0,1,\ldots,s$. He has established the converse of Nadaraya's results [20] which gives

<u>Theorem (5.1.1)</u>

<u>A necessary and sufficient condition for the sequence of estimators $\{fn(x)\}_{n=1}^{\infty}$ to converge uniformly to $f(x)$ with probability one is that $f(x)$ is uniformly continuous.</u>

Picklands [22] has been concerned with the efficiency of density function estimates. He has shown how to construct an estimator sequence which is the most efficient among all estimator sequences of algebraic type. (1.3.11). We shall present some preliminary definitions then exhibit the construction.

For $f(x) \in L_2(R)$, the MISE of the sequence of estimates $\{fn(x)\}_{n=1}^{\infty}$ was defined as

$$J_n = E\{\int (fn(x)-f(x))^2 dx\}$$

$$= \frac{1}{2\pi} E\{\int |\phi_{\hat{f}n}(t) - \phi_f(t)|^2 dt\} \quad (1.3.3).$$

We say that the estimator sequence is consistent in MISE if $\lim_{n\to\infty} J_n = 0$. For any characteristic function $\phi_f(t)$, $\phi_{K_n}(t)$ can be chosen to minimize J_n (1.3.7). Hence we define

$I_n \equiv \min J_n$ where the minimum is taken over all possible functions $\phi_{K_n}(t)$.

This leads to the following definition of efficiency,

$\text{Eff} \equiv \lim_{n\to\infty} \{\frac{I_n}{J_n}\}$ where J_n is the MISE of the given estimator.

Suppose $\phi_f(t)$ decreases algebraically of degree $p > 1/2$ (1.3.10) and $\lim_{t\to\infty} t \frac{\partial}{\partial t} \log |\phi_f(t)|^2 = \alpha$ where $\alpha \in R$ and $\alpha > 1$. If we construct an estimator $\hat{f}n(x)$ of algebraic type (1.3.11), then the most efficient estimator is the one where we chose $K_n(x)$ such that the Fourier transform of $K_n(x)$ is of the form

$$\phi_{K_n}(t) = \frac{1}{1+|\frac{t}{tn}|^\alpha} \quad \text{for} \quad 0 < t < \infty$$

where tn satisfies $|\phi_f(tn)|^2 = \frac{1}{n-1}$

In this case, Eff = 1.

Observe that $\phi_{K_n}(t)$ is of the form considered by Watson and Leadbetter in their construction of an estimator of algebraic form with the asymptotic optimum property. (1.3.18).

In the following, we shall consider some further results in the Series method of density estimation. In the introduction to the Series method of density estimation, we remarked that Cencov, [5], conceived the method. His results and presentation are more general than Schwartz's.

Recall that Schwartz was concerned with density functions defined on the reals such that $f(x) \in L_2(R)$. He proposed estimates of such densities in terms of the orthonormal family of Hermite functions.

On the other hand, Cencov considered the case where we have a weighting function in $v(x)$ defined on S where $S \subseteq R$. By means of the inner product $(f,g) = \int_S f(x)g(x)v(x)dx$ a Hilbert space $L_2(v(x))$ is defined. Letting $\{\phi_j(x)\}_{j=1}^{\infty}$ be an orthonormal basis for $L_2(v(x))$, Cencov proposed estimators of the form considered by Schwartz. (3.1.2) and (3.1.3).

He also obtained a result which gives information about the convergence of a histogram when the number of intervals is approximately equal to the cube root of the sample size. His result is

Theorem (5.1.2)

Suppose the r.v. X satisfies $a \leq X \leq b$, $f'(x)$ is continuous and $f'(x) \not\equiv 0$. Let $\hat{f}n(x)$ be the histogram estimate of $f(x)$ constructed with respect to the random sample $\{X_i\}_{i=1}^{n}$. Then for N intervals $(N \simeq n^{1/3})$ of equal length h, $E\{\int(\hat{f}n(x)-f(x))^2dx\}$ is in probability of order $O(N^{-2/3})$.

Motivated by Cencov and Schwartz, Kronmal and Tarter [13], have proposed continuous approximations of $f(x)$ and $F(x)$ in terms of the classical Fourier series involving the trigonometric functions.

Assume $f(x) \in L_2([a,b])$. Recall that $\{\cos k\pi(\frac{x-a}{b-a})\}_{k=0}^{\infty}$ is an orthogonal subset of $L_2([a,b])$. The sample trigonometric moments $\overline{C_k}$ are defined by

$$\overline{C_k} = \frac{2}{(b-a)n} \sum_{i=1}^{n} \cos k\pi(\frac{X_i-a}{b-a}) I_{[a,b]}(X_i) \quad \text{for } k = 0,1,2,\ldots$$

where $I_{[a,b]}(\cdot)$ is the indicator function of $[a,b]$.

Kronmal and Tarter have proposed an estimate of $f(x)$ of the following form:

$$\hat{f}n(x) = \frac{\overline{C_0}}{2} + \sum_{k=1}^{m} \overline{C_k} \cos k\pi\left(\frac{x-a}{b-a}\right)$$

where m is the optimal number of terms for the construction of the estimator and $m = 0(n^{1/2})$.

Similar estimators using the orthogonal subsets $\{\sin k\pi(\frac{x-a}{b-a})\}_{k=0}^{\infty}$ and $\{\cos k\pi(\frac{x-a}{b-a}), \sin k\pi(\frac{x-a}{b-a})\}_{k=0}^{\infty}$ have also been developed.

These types of estimators are certainly desirable from the computational point of view since they may be easily updated. But of even more importance, they are competitive with other types of estimators. For example, in using MISE as a criterion of goodness the Kronmal and Tarter estimators are competitive with Watson and Leadbetters optimal estimator for the Cauchy density function ([13], p.947).

Apparent disadvantages are how to choose $[a,b]$ and how to choose m, the optimal number of terms. The choice of $[a,b]$ is arbitrary with the suggestion that different $[a,b]$ be tried depending upon how much error can be tolerated.

As a means of estimating m, Kronmal and Tarter have devised a stopping rule which is expressed in terms of $n, (b-a)$, $\overline{C_k}$ and $\overline{S_k}$. ([13], p.949).

BIBLIOGRAPHY

[1] Anderson, G., A comparison of probability density estimates. Presented at I.M.S. annual meeting, August, 1969, New York.

[2] Bartlett, M.S., Statistical estimation of density functions. Sankya, Ser.A, 25 (1963) 245-254.

[3] Bhattacharya, P.K., Estimation of a probability density function and its derivatives. Sankya, Ser.A, 29 (1967) 373-382.

[4] Cacoullos, T., Estimation of a multivariate density. Ann. Inst. Statist. Math., 18 (1966) 179-190.

[5] Cencov, N.N., Evaluation of an unknown distribution density from observations. Soviet.Math., 3 (1962) 1559-1562.

[6] Chung, K.L., A Course in Probability Theory. Harcourt, Brace and World, 1968.

[7] Cramér, H., Mathematical Methods of Statistics. Princeton Univ. Press, 1946.

[8] Craswell, K.J., Density estimation in a topological group. Ann. Math. Statist., 36 (1965) 1047-1048.

[9] Erdelyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F.G., Higher Transcondental Functions. McGraw-Hill, New York, 1953.

[10] Ghurye, S.G., and Olkin, I., Unbiased estimation of some multivariate probability densities. Ann. Math. Statist., 40 (1969) 1261-1271.

[11] Hâjek, J., Theory of Rank Tests. Academic Press, New York, 1967.

[12] Hewitt, E. and Stromberg, K., Real and Abstract Analysis. Springer-Verlag, New York, 1965.

[13] Kronmal, R. and Tarter, M., The estimation of probability densities and cumulatives by Fourier series methods. J. Amer. Statist. Assoc., 63 (1968) 925-952.

[14] Kronmal, R. and Tarter, M., On multivariate density estimates based on orthogonal expansions. Ann. Math. Statist., 41 (1970) 718-722.

[15] Lehmann, E., Notes on the theory of estimation. University of California, Berkeley, 1950.

[16] Loève, M., Probability Theory. 3rd ed., Van Nostrand, Princeton, 1963.

[17] Loftsgaarden, D.O. and Quesenberry, C.P., A non-parametric estimate of a multivariate density function. Ann. Math. Statist., 36 (1965) 1049-1051.

[18] Moore, D.S. and Henrichon, E.G., Uniform consistency of some estimates of a density function. Ann. Math. Statist., 40 (1969) 1499-1502.

[19] Murthy, V.K., Estimation of probability density. Ann. Math. Statist., 36 (1965) 1027-1031.

[20] Nadaraya, E.A., On non-parametric estimates of density functions and regression curves. Theor. Prob. Appl., 10 (1965) 186-190.

[21] Parzen, E., On estimation of a probability density and mode. Ann. Math. Statist., 33 (1962) 1065-1076.

[22] Picklands, J., Efficient estimation of a probability density function. Ann. Math. Statist., 40 (1969) 854-864.

[23] Rosenblatt, M., Remarks on some non-parametric estimates of a density function. Ann. Math. Statist., 27 (1956) 832-837.

[24] Sansone, G., Orthogonal Functions. Interscience, New York, 1959.

[25] Schuster, E.F., Estimation of a probability density function and its derivatives. Ann. Math. Statist., 40 (1969) 1187-1195.

[26] Schwartz, S.C., Estimation of probability density by an orthogonal series. Ann. Math. Statist., 38 (1967) 1261-1265.

[27] Schwartz, S.C., Estimation of density functions by orthogonal series and an application to hypothesis testing. Presented at I.M.S. annual meeting, August, 1969, New York.

[28] Schwartz, S.C., On the estimation of a Gaussian convolution probability density. Siam J. Appl. Math., 17 (1969) 447-453.

[29] Van Ryzin, J., Bayes risk consistency of classification procedures using density estimation. Sankya, Ser.A, 28 (1966) 261-270.

[30] Van Ryzin, J., On strong consistency of density estimates. Ann. Math. Statist., 40 (1969) 1765-1772.

[31] Watson, G.S., Density estimation by orthogonal series. Ann. Math. Statist., 40 (1969) 1496-1498.

[32] Watson, G.S., and Leadbetter, M.R., On the estimation of the probability density, I. Ann. Math. Statist., 34 (1963) 480-491.

[33] Wegman, E.J., A note on estimating a unimodal density. Ann. Math. Statist., 40 (1969) 1661-1667.

[34] Wegman, E.J., Maximum likelihood estimation of a unimodal density. Ann. Math. Statist., 41 (1970) 457-471.

[35] Whittle, P., On the smoothing of probability density functions. J. Roy. Statist. Soc., Ser.B, 20 (1958) 334-343.

[36] Wiener, N., The Fourier Integral and Certain of its Applications. Dover, New York, 1933.

[A1] *Let* $X_1,\ldots,X_n$ *be independent random variables identically distributed as a r.v.* X *with the continuous density function* $f(x)$. *Let* $\hat{f}n(y) = \hat{f}n(X_1,\ldots,X_n,y)$ *be an estimate of* $f(y)$ *and suppose that* $\hat{f}n(y) \geq 0$ *for all* y *and* $n = 1,2,\ldots$. *Then* $\hat{f}n(y)$ *is not an unbiased estimate of* $f(y)$.

Proof

Suppose to the contrary that $E\{\hat{f}n(y)\} = f(y)$ for all y. Then since $f(y)$ is continuous on R, we have $E\{\hat{f}n(y)\} < \infty$. Assume that $\hat{f}n(y)$ is a symmetric function of $X_1,X_2,\ldots,X_n$, since the symmetrized n-tuple is a sufficient statistic for the problem. But then $\int_a^b \hat{f}n(y)dy$ is a symmetric estimate of $\int_a^b f(y)dy = F(b)-F(a)$ and moreover $\int_a^b \hat{f}n(y)dy$ is an unbiased estimate of $F(b)-F(a)$ since $E\{\int_a^b \hat{f}n(y)dy\} = \int_a^b E\{\hat{f}n(y)\}dy$ by Fubini's theorem

$$= \int_a^b f(y)dy$$

$$= F(b)-F(a)$$

However, the only unbiased estimate of $F(b)-F(a)$ symmetric in $X_1,\ldots,X_n$ is $Fn(b)-Fn(a)$ where $Fn(\cdot)$ is the sample distribution function [15]. Hence, $Fn(b)-Fn(a) = \int_a^b \hat{f}n(y)dy$ for all a and b and almost all $X_1,\ldots,X_n$ and this implies $Fn(y)$ is absolutely continuous for all y and almost all $X_1,\ldots,X_n$. Consequently, $Fn(y)$ is continuous for all y and almost all $X_1,\ldots,X_n$ which is impossible.

[A2] $\sup_{y \in R} |K(y)| < \infty$ <u>and</u> $\int |K(y)| dy < \infty$ <u>implies</u> $\int |K(y)|^{2+\delta} dy < \infty$ <u>for all</u> $\delta \geq 0$.

<u>Proof</u>

Let $\delta \geq 0$ be arbitrary. Suppose $\sup_{y \in R} |K(y)| = s_1$. Observe that this implies $\sup_{y \in R} |K(y)|^{\delta} < \infty$. Say $\sup_{y \in R} |K(y)|^{\delta} = s_2$.

Then $|K(y)|^{2+\delta} = |K(y)||K(y)|^{\delta}|K(y)|$

$$\leq \sup_{y \in R} \{|K(y)||K(y)|^{\delta}\} |K(y)|$$

$$\leq \sup_{y \in R} |K(y)| \sup_{y \in R} |K(y)|^{\delta} |K(y)|$$

$$= s_1 s_2 |K(y)|$$

However, $|K(y)| \in L_1(R)$ and $s_1 s_2 \in R$ implies $s_1 s_2 |K(y)| \in L_1(R)$ but $0 \leq |K(y)|^{2+\delta} \leq s_1 s_2 |K(y)|$, hence $|K(y)|^{2+\delta} \in L_1(R)$.

The notation in [A3] and [A4] is defined in theorem (1.1.13) (pp. 17-21)

[A3] Suppose that for some $\delta > 0$, $\lim_{n\to\infty} \frac{E|V_n - E(V_n)|^{2+\delta}}{n^{\delta/2}\sigma^{2+\delta}(V_n)} = 0$, for all $\varepsilon > 0$, $\lim_{n\to\infty} P\{|\frac{V_n - E(V_n)}{\sigma(V_n)}| \geq \varepsilon\, n^{1/2}\} = 0$.

Proof

Suppose $\delta > 0$ satisfies the condition in the hypothesis. Define $\phi(x) = |x|^{2+\delta}$. Then ϕ is strictly increasing on $(0,\infty)$ and $\phi(x) = \phi(-x)$.

$$\text{Now}\quad E\{\phi(\frac{V_n - E(V_n)}{n^{1/2}\sigma(V_n)})\} = E\{|\frac{V_n - E(V_n)}{n^{1/2}\sigma(V_n)}|^{2+\delta}\}$$

$$= \frac{1}{n}\,\frac{E|V_n - E(V_n)|^{2+\delta}}{n^{1/2}\sigma^{2+\delta}(V_n)}$$

Since $\lim_{n\to\infty} \frac{E|V_n - E(V_n)|^{2+\delta}}{n^{\delta/2}\sigma^{2+\delta}(V_n)} = 0$, it follows that for sufficiently large n,

$$\frac{E|V_n - E(V_n)|^{2+\delta}}{n^{\delta/2}\sigma^{2+\delta}(V_n)} < \infty$$

Then by Tchebicheff's theorem and the monotonicity of ϕ we have

$$P\{|\frac{V_n - E(V_n)}{n^{1/2}\sigma(V_n)}| \geq \varepsilon\} \leq \frac{E\{\phi(\frac{V_n - E(V_n)}{n^{1/2}\sigma(V_n)})\}}{\phi(\varepsilon)}$$

$$= \frac{1}{n}\,\frac{E|V_n - E(V_n)|^{2+\delta}}{\varepsilon^{2+\delta} n^{\delta/2}\sigma(V_n)}$$

Hence,

$$0 \leq nP\{|\frac{V_n - E(V_n)}{n^{1/2}\sigma(V_n)}| \geq \varepsilon\} \leq \frac{E|V_n - E(V_n)|^{2+\delta}}{\varepsilon^{2+\delta} n^{\delta/2}\sigma(V_n)} \to 0 \quad \text{as} \quad n \to \infty.$$

[A4] We shall now establish (1.1.14). First, some preliminary definitions and results are given. 'Following Loève ([16], pp. 315, 316) for a $\tau > 0$, let

$$\alpha_{nK}(\tau) = \int_{|x|<\tau} x \, d\, F_{nK}$$

and $$\sigma^2_{nK}(\tau) = \int_{|x|<\tau} x^2 \, d\, F_{nK} - \left(\int_{|x|<\tau} x \, d\, F_{nK}\right)^2$$

where F_{nk} is the c.d.f. of the r.v. X_{nK}. If $F_{nK} \to F$ completely ([16], p. 178) we shall write $\mathcal{L}(X_{nK}) \to \mathcal{L}(X)$. Let $\mathcal{N}(\alpha,\sigma^2)$ be the c.d.f. of a normal random variable with the parameters α and σ^2.

The Normal Convergence criterion ([16], p. 316) states that

Let X_{nK} be independent summands. Then $\mathcal{L}(\Sigma X_{nK}) \to \mathcal{N}(\alpha,\sigma^2)$ and $\max_K P\{|X_{nK}| \geq \varepsilon\} \to 0$ as $n \to \infty$ if and only if for all $\varepsilon > 0$ and a $\tau > 0$ $\sum_{k=1}^{n} P\{|X_{nK}| \geq \varepsilon\} \to 0$ as $n \to \infty$, $\sum_{k=1}^{n} \alpha_{nK}(\tau) \to \alpha$ as $n \to \infty$ and $\sum_{k=1}^{n} \sigma^2_{n_K}(\tau) \to \sigma^2$ as $n \to \infty$.

To obtain (1.1.14) using the above theorem we let

$$X_{nK} = \frac{V_{nK} - E(V_{nK})}{\sqrt{n}\, \sigma(V_{nK})}$$

which results in

$$\frac{\hat{f}n(x) - E(\hat{f}n(x))}{\sigma(\hat{f}n(x))} = \sum_{k=1}^{n} X_{nK}$$

Observe that for any $\varepsilon > 0$

$$\sum_{k=1}^{n} P\left\{\left|\frac{V_{nK} - E(V_{nK})}{\sqrt{n}\, \sigma(V_{nK})}\right| \geq \varepsilon\right\} = n\, P\left\{\left|\frac{V_n - E(V_n)}{\sigma(V_n)}\right| \geq \varepsilon\sqrt{n}\right\}$$ since the r.v.'s

$\{V_{nk}\}_{k=1}^{n}$ are identically distributed as the r.v. V_n. One can show that $\sum_{k=1}^{n}\alpha_{nK}(\tau) \to 0$ as $n \to \infty$ and $\sum_{k=1}^{n}\sigma^2_{nK}(\tau) \to 1$ as $n \to \infty$. Hence (1.1.14) follows by the Normal Convergence criterion.

[A5] $J_n = E\{\int(\hat{f}n(x)-f(x))^2 dx\} = \frac{1}{2\pi}E\{\int|\phi_{fn}(t)-\phi_f(t)|^2 dt\}$

Proof

For $g,h \in L_2(R)$ the inner product is defined as $(g,h) = \int g(x)\overline{h(x)}dx$ where $\overline{h(x)}$ denotes the complex conjugate of $h(x)$. The Parseval transform of g, $\Phi g(t)$ is defined by

$$\Phi g(t) = (2\pi)^{-\frac{1}{2}}\int e^{itx}g(x)dx \quad \text{for } g \quad L_1(R)$$

$$= (2\pi)^{-\frac{1}{2}}\phi_g(t) \quad \text{where } \phi_g(t) \text{ is the Fourier transform of } g.$$

Parseval's theorem ([36], p.70) states that

for $g,h \in L_2(R)$ and $(\ ,\ )$ as defined $(g,h) = (\Phi_g, \Phi_h)$ where Φ_g and Φ_h are the Parseval transforms of g and h respectively.

Thus, $\int g(x)\overline{h(x)}dx = \int \Phi_g(t)\overline{\Phi_h(t)}dt$

$$= \frac{1}{2\pi}\int \phi_g(t)\overline{\phi_h(t)}dt$$

Hence,

$$\int(\hat{f}n(x)-f(x))^2 dx = \int(\hat{f}n(x)-f(x))\overline{(\hat{f}n(x)-f(x))}dx$$

$$= \frac{1}{2\pi}\int \phi_{\hat{f}n-f}(t)\overline{\phi_{\hat{f}n-f}(t)}dt$$

$$= \frac{1}{2\pi}\int |\phi_{\hat{f}n-f}(t)|^2 dt$$

$$= \frac{1}{2\pi}\int |\phi_{\hat{f}n}(t)-\phi_f(t)|^2 dt \quad \text{since } \phi \text{ is linear.}$$

Therefore, we have [A5].

[A6] To show that $2\pi J_n = E\{\int|\hat{\phi}_n(t)\phi_{K_n}(t)-\phi_f(t)|^2 dt\}$

$$= \int[\frac{|\phi_{K_n}(t)|^2}{n}(1-|\phi_f(t)|^2) +$$

$$|\phi_f(t)|^2(|1-\phi_{K_n}(t)|^2)]dt$$

<u>Proof</u>

First note that $2\pi J_n = \int E\{|\phi_{\hat{f}n}(t)-\phi_f(t)|^2\}dt$ by Fubini's theorem.

Now $|\phi_{\hat{f}n}(t)-\phi_f(t)|^2 = |\hat{\phi}_n(t)\phi_{K_n}(t)-\phi_f(t)|^2$

$$= |\phi_{K_n}(t)|^2\phi_n(t)\overline{\phi_n(t)}-\phi_f(t)\overline{\phi_{K_n}(t)}\,\overline{\phi_n(t)}-$$

$$\overline{\phi_f(t)}\phi_{K_n}(t)\phi_n(t)+|\phi_f(t)|^2$$

But $E\{\phi_n(t)\} = E\{\frac{1}{n}\sum_{i=1}^{n}\exp(iXit)\}$

$$= E\{\exp(iXt)\}$$

$$= \phi_f(t)$$

and similarly, $E\{\overline{\phi_n(t)}\} = E\{\exp(-iXt)\}$

$$= \phi_f(-t)$$

$$= \overline{\phi_f(t)}$$

Also notice that

$$\phi_n(t)\overline{\phi_n(t)} = (\frac{1}{n}\sum_{i=1}^{n}\exp(iXit))(\frac{1}{n}\sum_{i=1}^{n}\exp(-iXit))$$

$$= \frac{1}{n^2}\{\sum_{i=1}^{n} e^{iXit}e^{-iXit} + \sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} e^{iXit}e^{-iXjt}\}$$

Hence

$$E\{\phi_n(t)\overline{\phi_n(t)}\} = \frac{1}{n^2}\{n+n(n-1)E\{e^{iXt}\}E\{e^{-iXt}\}\}$$ since we have i.i.d. r.v.

$$= \frac{1}{n} + \frac{n-1}{n}\,\phi_f(t)\overline{\phi_f(t)}$$

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

Therefore,

$$E\{|\phi_{K_n}(t)\phi_n(t)-\phi_f(t)|^2\}$$

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

$$|\phi_f(t)|^2$$

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

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

$$(1-\overline{\phi_{K_n}(t)}-\phi_{K_n}(t)+|\phi_{K_n}(t)|^2)|\phi_f(t)|^2$$

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

$$|\phi_f(t)|^2(1-\phi_{K_n}(t))(1-\overline{\phi_{K_n}(t)})$$

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

The notation in the next result is defined in theorem (2.3.2) (p. 63).

[A7] *If* $\max_{0 \le i \le n-k(n)+1} \left| \frac{1}{k(n)} \sum_{j=i+1}^{i+k(n)'} \frac{Y_j}{n^{-1} S_{n+1}} - 1 \right| \to 0 \ (P)$, *then*

$$\left\{\frac{n}{k(n)-1}\right\} \sum_{j=1}^{k(n)-1} \{F(X_{q+j,n})-F(X_{q+j-1,n})\} \to 1 \ (UP) \qquad (A7,1)$$

and $\left\{\frac{n}{k(n)-1}\right\} \sum_{j=1}^{k(n)} F(X_{q+j,n})-F(X_{q+j-1,n})\} \to 1 \ (UP) \qquad (A7,2)$

where the uniform convergence is taken over $q = 0$ *to* $q = n-k(n)+1$

Proof

From $\max_{0 \le i \le n-k(n)+1} \left| \frac{n}{k(n)} \sum_{j=i+1}^{i+k(n)} \frac{Y_j}{S_{n+1}} - 1 \right| \to 0 \ (P)$ we have that

$\left| \frac{n}{k(n)} \sum_{j=i+1}^{i+k(n)} \frac{Y_j}{S_{n+1}} - 1 \right| \to 0 \ (UP)$ where the uniform

convergence is taken over $i = 0$ to $i = n-k(n)+1$. Since the r.v.'s $\left\{\frac{Y_j}{S_{n+1}}\right\}_{j=1}^{n+1}$ are i.i.d. we have that $\frac{nY_j}{S_{n+1}} \to 1 \ (P)$. This implies that

$n[F(X_{j,n})-F(X_{j-1,n})] \to 1 \ (P)$ and hence that

$\left\{\frac{n}{k(n)-1}\right\} \sum_{j=1}^{k(n)} \{F(X_{q+j,n})-F(X_{q+j-1,n})\} \to 1 \ (UP)$ where the

uniform convergence is taken over $q = 0$ to $q = n-k(n)+1$.

The argument for (A7,1) is similar.