THE UNIVERSITY OF CALGARY

GOODNESS-OF-FIT TESTS BASED ON THE EMPIRICAL PROCESS

by

CHITA PADILLA EVARDONE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

JUNE, 1988

ISBN 0-315-46587-5

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "*Goodness-of-fit Tests Based on the Empirical Process*" submitted by Chita Padilla Evardone in partial fulfillment of the requirements for the degree of Master of Science.

M. D. Burke

M.D. Burke, Chairman
Department of Mathematics & Statistics

John R Collins

J.R. Collins
Department of Mathematics & Statistics

P.C. Consul

P.C. Consul
Department of Mathematics & Statistics

E. S. Edgington

E.S. Edgington
Department of Psychology

DATE: June 20, 1988

ABSTRACT

The classical problem of testing whether a set of observations comes from a population with specified distribution function F(x) has received a fair share of attention in the literature. As early as 1933, Kolmogorov introduced a "distribution-free" statistic based on the empirical process α_n and has derived the asymptotic distribution of this test statistic. Since then a lot of work has been done on this subject.

However, some goodness-of-fit problems arising in practice do not usually specify the parameters of F(x). Our main concern then is to test the hypothesis that a random sample was drawn from a parametric family of distribution functions. One way of testing this composite hypothesis is to adapt the empirical process where the parameter θ is approximated in terms of the random observations. As will be shown in Chapter II, the limiting distribution of test statistics based on the estimated empirical process depends on the underlying distribution function F(x).

Recent works of Burke and Gombay (1988) and Durbin (1976 and 1961) proposed distribution-free procedures to test this composite hypothesis. These are the bootstrap method, half-sample and random substitution method.

It is the purpose of this thesis to show that the asymptotic behaviour of the estimated empirical process based upon the above procedures is the same as the asymptotic behaviour of the empirical process under the specified case. The same is true for the Kolmogorov-Smirnov, Anderson-Darling, and Cramér-von Mises type of statistics.

Numerical results from a computer study are tabulated in Chapter IV to examine these results and see how the results for sample sizes of 50, 100, 150 and 200 compare with the asymptotic values.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor Dr. M.D. Burke for his kind help, suggestions and time during the research and writing of this thesis. My special thanks also go to Dr. P.C. Consul for giving me the encouragement and support and to the members of the examining committee for their helpful suggestions.

I am extremely grateful to my friend Tak Fung, of the Academic Computing Services, for helping me explore the intricacies of Multics and to Maria Dourado for her expert typing of my thesis.

Finally, I would like to acknowledge the financial assistance of the University of Calgary for making this study a reality.

To Tony, Joy and Karl

TABLE OF CONTENTS

Page

CHAPTER

CHAPTER I

TESTS BASED ON THE EMPIRICAL PROCESS WHEN PARAMETERS ARE SPECIFIED

1.1 INTRODUCTION

The problem treated in this chapter is that of testing the hypothesis that n independent, identically distributed random variables have a specified continuous distribution function $F(x)$. In statistical language, this goodness-of-fit problem is to test the simple hypothesis

$$H_0: \quad F(x) = F_0(x). \tag{1.1.1}$$

For example, the population may be specified by the hypothesis to be normal with mean 2 and variance 4, where the corresponding cumulative distribution function

$$F(x) = \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(y-2)^2}{8}} \, dy.$$

The most common method to test the validity of the null hypothesis is the χ^2 test which was originally proposed by Karl Pearson in 1900. It provided one of the earliest methods of statistical inference. By this method, the empirical histogram is compared to the hypothetical histogram. In this thesis, we will only consider those tests which compare the empirical distribution function $F_n(x)$ with the hypothetical distribution function $F(x)$.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous cumulative distribution function $F(x)$. We define the empirical distribution function of the sample by

$$F_n(x) = \sum_{j=1}^{n} \frac{I(X_j \le x)}{n}, \quad -\infty < x < \infty, \tag{1.1.2}$$

where $I(A)$ denotes the indicator function of the event A. An equivalent definition of $F_n(x)$ in terms of the ordered statistics

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

of the random sample $X_1, X_2, \ldots, X_n$ is given as

$$F_n(x) = \begin{cases} 0 & \text{if} \quad X_{(1)} > x \\ \frac{k}{n} & \text{if} \quad X_{(k)} \le x < X_{(k+1)}, \quad k = 1, \ldots, n-1 \\ 1 & \text{if} \quad X_{(n)} \le x. \end{cases} \tag{1.1.3}$$

Asymptotic behaviour of test statistics based on the empirical distribution function will be our main concern. These are the famous Kolmogorov-Smirnov (K-S) statistic

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - F_0(x)|,$$

the Cramér-von Mises (C-vM) statistic

$$W_n^2 = n \int_{-\infty}^{\infty} \left[F_n(x) - F_0(x)\right]^2 \, dF_0(x),$$

and the Anderson-Darling (A-D) statistic

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{\left[F_n(x) - F_0(x)\right]^2 \, d\,F_0(x)}{F_0(x)(1-F_0(x))} \, .$$

Under the null hypothesis, these statistics are asymptotically distribution-free. From a goodness-of-fit point of view, this is a very desirable property.

As will be shown in Section 6, the limiting distribution of the above mentioned statistics will be based on the Brownian motion and Brownian Bridge processes. These Gaussian processes are discussed in the next section.

1.2 BROWNIAN MOTION AND OTHER GAUSSIAN PROCESSES

The physical phenomenon of the Brownian motion was discovered by the English botanist Brown in 1827. A mathematical description of this process was first derived from the laws of physics by Einstein in 1905. Since then the subject has made considerable progress.

The simplest model for a one-dimensional Brownian motion or Wiener process can be given in terms of the random walk model. Assume that the particle is moving on the real line and starting from the origin, it can only move one step to the right or to the left. These steps are assumed independent. If X_i represents the outcome of the i^{th} step of the particle with possible values 1 or -1, then $X_1, X_2, \ldots$ are identically and independently distributed random variables (i.i.d.r.v.) with $P(X=1) = P(X=-1) = \frac{1}{2}$. A typical realization of this stochastic process $\{X_t,\ t \in T\}$ would be 1, 1, -1, 1, -1, -1, 1. This is shown in Figure 1.1 where the ordinate for $t=n$ is the value of X_n.

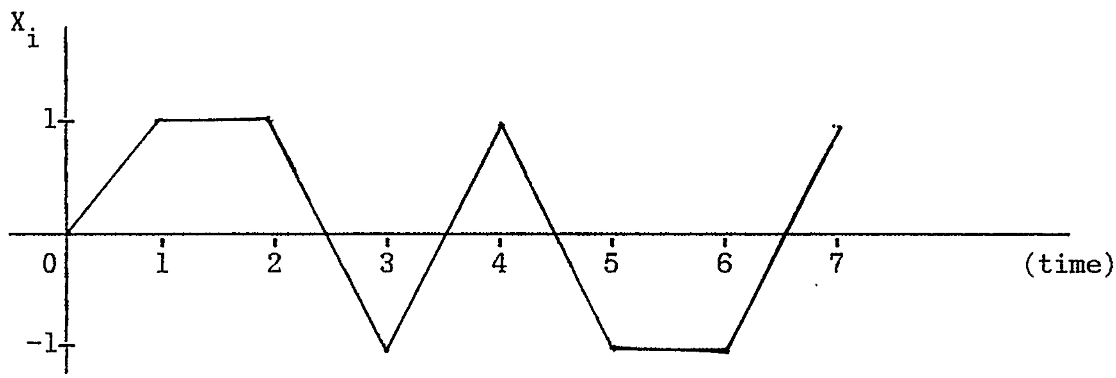

Figure 1.1

A Simple Example of a Stochastic Process

After n steps, this particle will be located at $S_n = X_1 + X_2 + \cdots + X_n$. Thus the created path $S_1, S_2, \ldots$ imitates the Brownian motion quite well if the time unit and steps are short enough. In a more realistic model of a Brownian motion, the particle moves in a continuous time scale and continuous state space.

Definition 1.1

A stochastic process $\{W(t,\omega) = W(t);\ 0 \le t \le 1\}$ where $\omega \in \Omega$, $\{\Omega, A, P\}$ is a probability space, is called a Brownian motion if

(i) $P(W(0) = 0) = 1$,

(ii) $W(t) \sim N(0,t)$, $0 \le t \le 1$,

(iii) $W(t_1) - W(t_0)$, $W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1})$ are independent for $0 \le t_0 \le t_1 \le \cdots \le t_n \le 1$,

(iv) the sample path function $W(t,\omega)$ is continuous in t.

A direct consequence of properties (ii) and (iii) is that $\left[W(t_1), \ldots, W(t_n)\right]$ is multivariate normal with mean 0 and covariance function

$$E\, W(t)\, W(t') = t \wedge t',$$

where $t \wedge t' = \min(t,t')$. The existence of this process on the space $C[0,1]$ of Definition 1.4 was derived by Billingsley (1968, Section 9).

<u>Definition 1.2</u>

A stochastic process $\{B(t);\ 0 \le t \le 1\}$ is called a <u>tied-down</u> <u>Brownian process</u> or <u>Brownian Bridge</u> if

(i) the joint distribution of $B(t_1)$, $B(t_2), \ldots, B(t_n)$ $(0 \le t_1 \le \cdots \le t_n \le 1;\ n = 1, 2, \ldots)$ is Gaussian with $E\, B(t) = 0$,

(ii) the covariance function of $B(t)$ is $E\, B(t)B(t') = t \wedge t' - tt'$,

(iii) the sample path function of $B(t,\omega)$ is continuous in t with probability 1.

(iv) $B(0) = B(1) = 0$ a.s.

The existence of such process is a simple consequence of Lemma 1.4.1 in Csörgő and Révész (1981). Here, the Brownian Bridge can be represented as

$$B(t) = W(t) - t\,W(1), \qquad 0 \le t \le 1,$$

where $\{W(t)\}$ is a Brownian motion. An important property of the Brownian Bridge is that it behaves like a Wiener path W conditioned by the requirement $W(1) = 0$. To show this consider $0 < t_1 < t_2 < 1$. From Definition 1.1 (ii) and (iii), we know that $W(t_i) - W(t_{i-1})$ is normally distributed with mean 0 and variance $t_i - t_{i-1}$ and that $W(t_1)$, $W(t_2) - W(t_1)$, $W(1) - W(t_2)$ are independent. Hence, their joint density is

$$\text{(1.2.1)} \qquad \frac{e^{-\frac{1}{2}\left[\frac{W(t_1)^2}{t_1} + \frac{[W(t_2)-W(t_1)]^2}{t_2-t_1} + \frac{[W(1)-W(t_2)]^2}{1-t_2}\right]}}{(2\pi)^{3/2}\,(t_1(t_2-t_1)(1-t_2))^{1/2}} .$$

Using some simple transformations, we obtain the same density for $W(t_1)$, $W(t_2)$, $W(1)$. We note that the density of $W(1)$ is

$$(2\pi)^{-1/2}\, e^{-\frac{1}{2}W(1)^2} .$$

Thus the conditional joint density of $W(t_1)$ and $W(t_2)$ given $W(1) = 0$ is

$$\frac{e^{-\frac{1}{2}\left[\frac{W(t_1)^2}{t_1} + \frac{[W(t_2)-W(t_1)]^2}{t_2-t_1} + \frac{[W(t_2)]^2}{1-t_2}\right]}}{2\pi\,(t_1(t_2-t_1)(1-t_2))^{1/2}}$$

$$= \frac{e^{-\frac{1}{2}\left[\frac{t_2W(t_1)^2}{t_1(t_2-t_1)} - \frac{-2W(t_1)W(t_2)}{t_2-t_1} + \frac{(1-t_1)W(t_2)^2}{(t_2-t_1)(1-t_2)}\right]}}{2\pi\ (t_1(t_2-t_1)(1-t_2))^{1/2}} .$$

It can easily be verified that this is the joint density of two normal variables with variances $t_1(1-t_1)$ and $t_2(1-t_2)$ and covariance $t_1(1-t_2)$. Similarly, the joint density of any finite set of $W(t)$'s given $W(1) = 0$ is normal with covariance function in Definition 1.2 (ii). Since the distribution of a normal process in C is determined by its finite-dimensional distribution this implies that the distribution of $\{B(t)\}$ is the same as that of $\{W(t)\}$ given $W(1) = 0$.

Definition 1.3

A Kiefer process K, defined on $[0,1] \times (0,\infty)$, is a separable Gaussian process with mean $E\ K(t,n) = 0$ and covariance function

$$E\ K(t_1,n_1)\ K(t_2,n_2) = n_1 \wedge n_2 \left\{(t_1 \wedge t_2) - t_1t_2\right\}.$$

For fixed $n > 0$,

$$n^{-1/2}\ K(t,n) \overset{\mathcal{D}}{=\!=} B(t),$$

where $\overset{\mathcal{D}}{=\!=}$ stands for the equality of all finite-dimensional distributions and $B(t)$ is a Brownian Bridge on $[0,1]$. As the name suggests, this process was first studied by Kiefer (1972).

The above stochastic processes are Gaussian since all its finite dimensional distributions are normal. An important property of the

Gaussian process is that it is determined stochastically by its mean and its covariance function (Durbin, 1973b).

1.3 PRELIMINARIES

In the next two sections, we will be following the line of thinking of Durbin (1973b) and Billingsley (1968) in showing the weak convergence of the empirical process

(1.3.1) $$\alpha_n(x) = \sqrt{n}\,\left[F_n(x) - F_0(x)\right], \qquad x \in R.$$

Along these lines, we set the following definitions:

Definition 1.4

The space C[0,1] is the space of continuous functions in the interval [0,1], where $\mathscr{C}$, the class of Borel sets in C, is generated by the uniform metric

$$c(x,y) = \sup_{0\le t\le 1} |x(t) - y(t)|, \qquad \text{for} \quad x,\ y \in C.$$

Definition 1.5

The space D[0,1] is the space of functions on [0,1] that are right continuous and have left-hand limits, that is,

(i) for $0 \le t < 1$, $x(t+) = \lim_{s\downarrow t} x(s)$ exists and $x(t+) = x(t)$,

(ii) for $0 < t \le 1$, $x(t-) = \lim_{s\uparrow t} x(s)$ exists.

On D, we use the Skorohod metric

$$d(x,y) = \inf_{\lambda\in\Lambda} \left[\sup_{0\le t\le 1} |x(t) - y(\lambda(t))| + \sup_{0\le t\le 1} |t - \lambda(t)| \right]$$

where $x, y \in D$ and Λ is the class of all strictly increasing functions on $[0,1]$ such that $\lambda(0) = 0$ and $\lambda(1) = 1$.

Definition 1.6

Weak convergence in space D is used in the following sense. The stochastic process

$$\{X_n(t)\} \xrightarrow{\mathcal{D}} \{X(t)\}$$

if

(i) the finite - dimensional distribution of $\{X_n(t)\}$ converges weakly to $\{X(t)\}$. That is,

$$\left[X_n(t_1),\ X_n(t_2),\dots,X_n(t_k)\right] \xrightarrow{\mathcal{D}} \left[X(t_1),\dots,X(t_k)\right].$$

(ii) $\{X_n(t)\}$ is tight.

$\left\{X_n(t)\right\}$ is tight if $\left\{P_n\right\}$ is tight, where P_n is the distribution of X_n. Paraphrasing Theorem 15.5 of Billingsley (1968), tightness follows if

(i) for each positive η there exists an a such that

$$P_n\{X\colon |X(0)| > a\} \le \eta, \quad n \ge 1,$$

(ii) for each positive ϵ and η, there exist a δ, $0 < \delta < 1$, and an integer n_0, such that

$$P_n\left\{X:\ w_X(\delta) > \epsilon\right\} \le \eta, \qquad n \ge n_0,$$

where $w_X(\delta)$ is the modulus of continuity defined by

$$w_X(\delta) = \sup_{|s-t|<\delta} |X(s) - X(t)|, \qquad 0 < \delta < 1.$$

In this case, if P is the weak limit of a subsequence $\left\{P'_n\right\}$ of $\left\{P_n\right\}$, then $P(C) = 1$, that is, the sample paths of the limiting process are continuous almost surely.

1.4 THE EMPIRICAL DISTRIBUTION FUNCTION

The empirical distribution function, as defined in (1.1.2) and (1.1.3), is sometimes called the sample distribution function. It is easy to verify that for fixed x, $F_n(x)$ is the relative frequency of successes in a Bernoulli sequence of trials with

$$E\ F_n(x) = F(x) \tag{1.4.1}$$

and

$$\operatorname{var} F_n(x) = \frac{F(x)(1-F(x))}{n}\ . \tag{1.4.2}$$

By the classical strong law of large numbers, for fixed x,

$$F_n(x) \xrightarrow{a.s.} F(x)\ .$$

Hence, $F_n(x)$ is an unbiased and strongly consistent estimator of $F(x)$. As $n \longrightarrow \infty$, $F(x)$ can be uniquely determined with probability one. This idea was embodied in the following

Theorem 1.1 (Cantelli 1917 and Glivenko 1933)

$$\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \xrightarrow{a.s.} 0 .$$

1.5 ASYMPTOTIC DISTRIBUTION OF THE EMPIRICAL DISTRIBUTION FUNCTION

Consider the ordered sample $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ from a continuous distribution $F(x)$. To test the null hypothesis (1.1.1), we let

$$t_{(j)} = F_0(X_{(j)}).$$

Under H_0, $0 \leq t_{(1)} \leq \cdots \leq t_{(n)} \leq 1$ is an ordered sample of n independent observations from a uniform distribution $U(0,1)$. Let

$$F_n'(t) = \sum_{j=1}^{n} \frac{I(t_{(j)} \leq t)}{n} .$$

Viewing $\{F_n'(t),\ 0 \leq t \leq 1\}$ as a stochastic process, we want to show that its distribution is the same as the Poisson process $\{P_n(t)\}$ with the condition $P_n(1) = 1$.

From basic probability we know that the distribution of $t_{(1)}, t_{(2)},\ldots,t_{(n)}$ is

$$\text{(1.5.1)} \qquad dP = n!\, dt_1 dt_2 \cdots dt_n, \quad 0 \leq t_1 \leq t_2 \leq \cdots t_n \leq 1 .$$

This distribution can be realized as the distribution of occurrence times in a Poisson process given that n events occur in $[0,1]$.

Let $\{P_n(t)\}$ be the Poisson process with occurrence rate n and jumps of $\frac{1}{n}$ for $0 \leq t \leq 1$, i.e., $n\left[P_n(t_2) - P_n(t_1)\right]$ has Poisson distribution with mean

(1.5.2) $$n[t_2-t_1] \quad \text{for} \quad 0 \le t_1 \le t_2 \le 1,$$

$P_n(0) = 0$ and increments are independent.

Consider a set of time points $0 < t_1 < \cdots < t_n < 1$ and choose dt_i small enough such that $[t_i, t_i + dt_i)$ are nonoverlapping. The probability of no event in $[0,t_1)$, one event in $[t_1, t_1 + dt_1)$, none in $[t_1 + dt_1, t_2)$, one in $[t_2, t_2 + dt_2)$ $\cdots$ none in $[t_n + dt_n, 1]$ is

$$e^{-nt_1} e^{-ndt_1} ndt_1 \cdot e^{-n(t_2-t_1-dt_1)} \cdots ndt_n \cdot e^{-n(1-t_n-dt_n)} + o\left[\sum_{i=1}^{n} dt_i\right]$$

$$= n^n e^{-n} dt_1 \cdots dt_n + o\left[\sum_{i=1}^{n} dt_i\right].$$

This follows from (1.5.2). The probability of n events in $[0,1]$ is

$$\frac{e^{-n} n^n}{n!} .$$

Thus the conditional probability of an event in $[t_i, t_i + dt_i)$ for $i = 1,\ldots,n$, given n events in $[0,1]$ is

$$dP = \frac{n!\left[e^{-n}n^n dt_1 \cdots dt_n + o\left[\sum_{i=1}^{n} dt_i\right]\right]}{e^{-n} n^n}$$

(1.5.3) $$= n!\, dt_1 \cdots dt_n + o\left[\sum_{i=1}^{n} dt_i\right], \quad 0 < t_1 \cdots < t_n < 1.$$

As $\max(dt_i) \longrightarrow 0$, we see that the densities (1.5.1) and (1.5.3) are the same for $0 < t_1 < \cdots t_n < 1$. Since the events $t_{(i)} = t_{(j)}$, $(i \ne j)$, $t_{(1)} = 0$, $t_{(n)} = 1$ in (1.5.1) have zero probability, the two distributions are the same for $0 \le t_1 \le t_2 \cdots \le t_n \le 1$, i.e., the distribution of the occurrence times of $\{P_n(t)\}$ in $[0,1]$ given $P_n(1) = 1$ is the same as that of the uniform order statistics. Since the mappings from the vector $[t_1,\ldots,t_n]'$ to the space D of functions on $[0,1]$ are the same for both $F_n'(t)$ and $P_n(t)$, it follows that the distribution of the stochastic process $\{F_n'(t)\}$ is the same as that of the process $\{P_n(t)\}$ given $P_n(1) = 1$.

To study the asymptotic behaviour of $F_n'(t)$, we normalize to give the empirical process

$$\alpha_n'(t) = \sqrt{n}\left[F_n'(t)-t\right], \quad 0 \le t \le 1, \tag{1.5.4}$$

where $E\,\alpha_n'(t) = 0$ and the covariance function is

$$E\,\alpha_n'(t)\,\alpha_n'(t') = t \wedge t' - tt'. \tag{1.5.5}$$

These moments follow from (1.4.1) and (1.4.2). The pointwise behaviour of α_n is quite simple. For fixed $x \in R$,

$$\alpha_n(x) \xrightarrow{\mathcal{D}} N(0,F(x)(1-F(x))$$

or for fixed $t \in [0,1]$,

$$\alpha_n'(t) \xrightarrow{\mathcal{D}} N(0,t(1-t)).$$

We take note that the sample paths of $\{\alpha_n'(t)\}$ are elements of space $D[0,1]$ of Definition 1.5. To find the limiting distribution of $\alpha_n'(t)$, we want to find a normal process in the space $D[0,1]$ that coincides with the mean and covariance function of $\alpha_n'(t)$ as shown in (1.5.5). From Definition 1.2, the Brownian Bridge satisfies these conditions. However, since its domain is in space $C[0,1]$, we need to extend its domain to the space $D[0,1]$.

Following Billingsley (1968, Section 16), let $P(W(t) \in A) = \Pr(W(t) \in A \cap C)$, for each $A \in \mathscr{D}$, where the latter probability is being calculated from the distribution of $W(t)$ on $(C, \mathscr{C})$. The same is true for $\{B(t)\}$ because the Brownian Bridge is defined in terms of $W(t)$. Since C is a member of the class $\mathscr{D}$, we have

$$P(W(t) \in C) = \Pr(B(t) \in C) = 1.$$

This means that the stochastic process $\{W(t)\}$ and $\{B(t)\}$ in D have continuous sample paths with probability 1.

Using the multivariate version of the Central Limit Theorem, we have

$$(\alpha_n'(t_1),\ldots,\alpha_n'(t_k)) \xrightarrow{\mathscr{D}} (B(t_1),\ldots,B(t_k)),$$

for any fixed sequence $0 \le t_1 \le \ldots \le t_k \le 1$. This suggests that the finite-dimensional distribution of $\{\alpha_n'(t)\}$ converges weakly to those of $\{B(t)\}$.

1.6 ASYMPTOTIC THEORY OF SOME FUNCTIONALS OF THE EMPIRICAL PROCESS

From Billingsley (1968, Section 5), an important result of the weak convergence theory is that if g is a measurable function in D which is continuous almost everywhere in metric d and with respect to the distribution of $\{B(t)\}$ and if

$$\left\{\alpha_n'(t)\right\} \xrightarrow{\mathscr{D}} \{B(t)\}$$

then

$$g\left[\alpha_n'(t)\right] \xrightarrow{\mathcal{D}} g(B(t)).$$

Due to the results in Section 1.5, it seems natural to replace the $\left\{\alpha_n'(t)\right\}$ process by the $\{B(t)\}$ process as $n \longrightarrow \infty$, and consequently, the function of $\left\{\alpha_n'(t)\right\}$ will converge in distribution to the function of $\{B(t)\}$. The same is true for $\alpha_n(x)$. In particular,

$$\sqrt{n}\, D_n = \sup_{-\infty<x<\infty} |\alpha_n(x)| \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1} |B(t)|,$$

(1.6.1) $$W_n^2 = \int_{-\infty}^{\infty} \alpha_n(x)^2 \, d\, F(x) \xrightarrow{\mathcal{D}} \int_0^1 B(t)^2 \, dt,$$

and $$A_n^2 = \int_{-\infty}^{\infty} \frac{\alpha_n(x)^2 \, dF(x)}{F(x)(1-F(x))} \xrightarrow{\mathcal{D}} \int_0^1 \frac{B(t)^2}{t(1-t)} \, dt.$$

However, from Definition 1.6, the weak convergence of the finite-dimensional distribution of $\left\{\alpha_n'(t)\right\}$ to $\{B(t)\}$ does not imply convergence in D. The "tightness" of $\left\{\alpha_n'(t)\right\}$ has to be shown. To do away with this tedious proof, we quote the result with the best rates for approximating $\left\{\alpha_n'(t)\right\}$ by a sequence of Brownian Bridges. We have

<u>Theorem 1.2</u> (Komlós, Major and Tusnády 1975)

If the underlying probability space is rich enough (that is, an independent sequence of Wiener processes, which is independent of the originally given i.i.d. sequence $\{X_n\}$, can be constructed on the assumed probability space), one can define a Brownian Bridge $\{B_n(t),\ 0 \le t \le 1\}$ for each n and a Kiefer process $\{K(t,y);\ 0 \le t \le 1,\ 0 < y < \infty\}$ such that

$$\sup_{-\infty<x<\infty} |\alpha_n(x) - B_n(F(x))| \overset{a.s.}{=\!=} O\left[n^{-1/2} \log n\right]$$

and

$$\sup_{-\infty<x<\infty} |n^{1/2}\, \alpha_n(x) - K\,(F(x),n)| \overset{a.s.}{=\!=} O\left[\log^2 n\right].$$

Since $B_n(F(x)) =_{\mathcal{D}} B_m(F(x)) =_{\mathcal{D}} B(F(x))$, the above theorem implies that

$$\alpha_n(x) \xrightarrow{\mathcal{D}} B(F(x)).$$

Thus

$$g(\alpha_n(x)) \xrightarrow{\mathcal{D}} g(B(F(x))),$$

where g is a continuous function.

To prove the results in (1.6.1) it is sufficient to show that the functions

$$\text{(i)} \quad g(x(t)) = \sup_{0\le t\le 1} |x(t)|$$

(1.6.2)

$$\text{(ii)} \quad g(x(t)) = \int_0^1 x(t)^2\, dt$$

are continuous in d for all $x(t) \in D$.

Now if $d(x,x') < \epsilon$, there is a $\lambda \in \Lambda$ such that

$$\sup_{0\le t\le 1} |x(t) - x'\{\lambda(t)\}| + \sup_{0\le t\le 1} |t - \lambda(t)| < 2\epsilon$$

so that

$$\sup_{0\le t\le 1} |x(t) - x'\{\lambda(t)\}| < 2\epsilon.$$

We note that

$$\sup_{0\le t\le 1} |x(t)| \le \sup_{0\le t\le 1} |x'\{\lambda(t)\}| + \sup_{0\le t\le 1} |x(t) - x'\{\lambda(t)\}|$$

$$\le \sup_{0\le t\le 1} |x'(t)| + 2\epsilon.$$

Similarly, one can show that

$$\sup_{0\le t\le 1} |x(t)| \ge \sup_{0\le t\le 1} |x'(t)| - 2\epsilon,$$

whence $\sup_{0\le t\le 1} |x(t)|$ is continuous in metric d.

As to the proof of (1.6.2) (ii), we let $z(t) = x(t)^2$ and show that $\int_0^1 z(t)\,dt$ as a function of $z(t)$ is continuous in d. Now for any sequence of functions z_m converging to z in d there exist functions λ_m such that $\lim_{m\to\infty} z_m(\lambda_m(t)) = z(t)$ uniformly in t and $\lim_{m\to\infty} \lambda_m(t) = t$ uniformly in t (Billingsley 1968, page 112). Since every element of D is bounded and has at most a countable number of discontinuities it is Riemann integrable. Take Riemann subdivisions

$0 < t_1 <\cdots< t_p < 1$ for $\int_0^1 z(t)\,dt$ and $0 < \lambda_m(t_1) <\cdots< \lambda_m(t_p) < 1$ for $\int_0^1 z_m(t)\,dt$. As $m \longrightarrow \infty$, the upper and lower sums for the latter integral converge to those for the former. It follows that

$$\int_0^1 z_m(t)\,dt \longrightarrow \int_0^1 z(t)\,dt.$$

Hence $\int_0^1 z(t)\,dt$ is continuous in d.

We take note that this argument does not apply to the function

$$g(x(t)) = \int_0^1 \frac{x(t)^2}{t(1-t)}\,dt$$

since the function $(t(1-t))^{-1}$ is not continuous at $t = 0$ or 1. As suggested by Durbin (1973b, page 31), one could consider the convergence of the statistic obtained by integrating over the range $(\delta, 1-\delta)$ and then let $\delta \longrightarrow 0$.

A significant application of the above results is in testing the null hypothesis in (1.1.1). In practice, if measure $\sqrt{n}\,D_n$ is adapted, the null hypothesis is rejected for those samples for which

$$\sqrt{n}\,D_n = \sup_{-\infty<x<\infty} |\alpha_n(x)| \geq c,$$

where c is calculated from

$$\alpha = \Pr\left(\sup_{0 \le t \le 1} |B(t)| \ge c \right),$$

for a specified α, say .01 or .05.

1.7 NUMERICAL TABULATION OF THE DISTRIBUTION OF THE K-S, C-vM AND A-D TEST STATISTICS

Kolmogorov (1933) introduced the statistic

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|,$$

and showed that it has the following properties which make it useful for judging how "close" $F_n(x)$ is to $F(x)$:

(i) the probability distribution of D_n depends on n but is independent of $F(x)$,

(ii) for large n, the probability distribution of D_n is given by the relationship

$$\lim_{n \to \infty} \mathrm{Prob}\{\sqrt{n}\, D_n < z\} = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 z^2} = L(z). \tag{1.7.1}$$

In his original paper, Kolmogorov derived a system of recursion formulas which make it possible to compute for any finite n the probabilities

$$\mathrm{Prob}\left\{D_n < \frac{c}{n}\right\}, \quad \text{for } c = 1,2,\ldots n.$$

Birnbaum (1952) gave a numerical tabulation of the distribution of K-S statistic for finite sample size. Massey (1950) also tabulated $\mathrm{Prob}\left\{D_n < \frac{c}{n}\right\}$ for selected values of c and n. He obtained a system of recursive formulas, equivalent to Kolmogorov's, as well as a procedure for replacing them by a system of difference equations.

The function $L(z)$ has been tabulated by Smirnov (1948). A new proof of (1.7.1) has been given by Feller (1948) and a heuristic outline of a proof by Doob (1949). Doob's derivation was based on the evaluation of the probability that a sample path of $\{W(t)\}$ crosses one or both of two straight-line boundaries.

Cramér (1928) proposed as a measure of the discrepancy between $F_n(x)$ and $F_0(x)$ the statistic

$$\int_{-\infty}^{\infty} \left[F_n(x) - F_0(x)\right]^2 dx.$$

This was generalized by von Mises (1931) to the form

$$\int_{-\infty}^{\infty} g(x) \left[F_n(x) - F_0(x)\right]^2 dx,$$

where $g(x)$ is a suitably chosen weight function. Smirnov (1936) modified this to

$$n \int_{-\infty}^{\infty} \Psi\left[F_0(x)\right]\left[F_n(x) - F_0(x)\right]^2 d\,F_0(x)$$

so as to give a distribution-free statistic. When $\Psi = 1$, we call this the Cramér-von Mises statistic and write it as

$$W_n^2 = n \int_{-\infty}^{\infty} \left[F_n(x) - F_0(x)\right]^2 d\, F_0(x).$$

When $\Psi\left[F_0(x)\right] = \left[F_0(x)(1-F_0(x))\right]^{-1}$, we have the Anderson-Darling statistic

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{\left[F_n(x)-F_0(x)\right]^2 d\, F_0(x)}{F_0(x)\left[1-F_0(x)\right]} .$$

Little is known about the exact distribution of C-vM statistic. Marshall (1958) has given explicit expressions for the distribution functions of W_1^2, W_2^2 and W_3^2 and Stephens and Maag (1968) have given formulae for the extreme lower-tail probabilities for W_n^2.

We know that for large n, the distribution of the C-vM statistic is approximately the same as the distribution of

$$W^2 = \int_0^1 B(t)^2\, dt.$$

The characteristic function of W^2 is

(1.7.2) $$\phi(\theta) = \prod_{j=1}^{\infty} \left[1 - \frac{2i\theta}{j^2\pi^2}\right]^{-1/2} = \left[\frac{\sqrt{2i\theta}}{\sin\sqrt{2i\theta}}\right]^{1/2}.$$

This was inverted by Smirnov (1936) to give

$$\text{(1.7.3)}\quad \mathrm{Prob}(W^2 \le x) = 1 - \frac{1}{\pi} \sum_{j=1}^{\infty} (-1)^{j-1} \int_{(2j-1)^2\pi^2}^{(2j)^2\pi^2} y^{-1} \left[\frac{-\sqrt{y}}{\sin\sqrt{y}}\right]^{1/2} e^{-\frac{xy}{2}} dy.$$

A different inversion of (1.7.2) was given as a rapidly-converging series by Anderson-Darling (1952) and was used by them to tabulate the distribution function of W^2.

Similarly, the A-D statistic converges in distribution to

$$A^2 = \int_0^1 \frac{B(t)^2}{t(1-t)}\, dt.$$

This has characteristic function

$$\phi_A(\theta) = \left[\frac{-2\pi i\theta}{\cos\left[\pi\sqrt{\frac{1+8i\theta}{2}}\right]}\right]^{1/2}.$$

Anderson and Darling (1954) gave a rapidly converging series for the inverse of $\phi_A(\theta)$ and tabulated a short table of significance points.

Stephens (1970) has provided good approximations to the percentage points of the above statistics in an extremely compact form. In his paper, for each test statistic T, a simple modification T^* is given, and T^* is compared with the given percentage points.

CHAPTER II

WEAK CONVERGENCE OF THE EMPIRICAL PROCESS WHEN PARAMETERS ARE ESTIMATED

2.1 INTRODUCTION

Consider a random sample $X_1, X_2, \ldots, X_n$ from a family of distribution functions

$$\mathcal{F} = \{F(x,\theta):\ x \in R,\ \theta \in S \subset R^P\}.$$

Let $\left\{\hat{\theta}_n\right\}$ be a sequence of estimators of $\theta = \left[\theta_1,\ \theta_2,\ \ldots,\ \theta_p\right]$, a vector of parameters, based on the random sample.

In this chapter, we shall consider the asymptotic behaviour of the estimated empirical process

$$\hat{\alpha}_n(x) = \sqrt{n}\left[F_n(x) - F(x,\hat{\theta}_n)\right], \quad x \in R, \tag{2.1.1}$$

under the null hypothesis (2.1.2) and under a sequence of alternative hypotheses (2.6.2). As will be shown in Section 3, this limiting distribution converges to a Gaussian process that depends not only on F(x) but also θ_0, the true theoretical value of θ. Consequently, procedures based on $\hat{\alpha}_n(x)$ would not be asymptotically distribution-free (Durbin, 1973b).

The results, as well as an extension of the methodology used in this chapter, are employed in Chapter III to obtain some distribution-free procedures for testing the composite hypothesis

$$H_0: \quad F \in \mathcal{F}. \tag{2.1.2}$$

A typical example might be the hypothesis that the data came from a normal distribution with unknown mean and variance.

2.2 PRELIMINARIES

The following exposition is based on the recently developed strong approximation methodology of Kiefer (1972), Csörgő-Révész (1975) and Komlós-Major-Tusnády (1975). The type of estimation of the parameters $\theta \in R^p$ of $F(x,\theta)$ follows from Durbin (1973a).

Under the null hypothesis (2.1.2), we wish to show that the estimated empirical process $\hat{\alpha}_n(x)$ can be approximated asymptotically by the Gaussian process

$$G_n(x) = B_n\left[F(x,\theta_0)\right] - \int \ell(x,\theta_0)\, dB_n\left[F(x,\theta_0)\right][\nabla_\theta F(x,\theta_0)]', \tag{2.2.1}$$

where B_n is a sequence of Brownian Bridges.

The following notation will be used:

(i) The transpose of a vector v will be denoted by v'.

(ii) The norm $\|\cdot\|$ on R^p is defined by $\|y_1,\ldots,y_p\| = \max_{1\le i\le p} |y_i|$.

(iii) $\nabla_\theta F(x,\theta_0)$ denotes the gradient vector of partial derivatives,

$$\left[\frac{\partial F(x,\theta)}{\partial\theta_1}, \frac{\partial F(x,\theta)}{\partial\theta_2}, \ldots, \frac{\partial F(x,\theta)}{\partial\theta_p}\right],$$

evaluated at $\theta = \theta_0 \in R^P$.

(iv) For a vector-valued function $\ell = [\ell_1,\ldots,\ell_p]$, let $\int \ell$ denote the vector $\left[\int \ell_1, \int \ell_2,\ldots,\int \ell_p\right]$.

(v) All integrals are evaluated from $-\infty$ to ∞.

We will assume the following conditions on the estimator sequence.

A1. $\sqrt{n}\left[\hat{\theta}_n - \theta_0\right] = \sum_{j=1}^{n} \frac{\ell(X_j,\theta_0)}{\sqrt{n}} + \epsilon_{1n}$, where θ_0 is the true unknown value of θ; $\ell(\cdot, \theta_0)$ is a measurable p-dimensional vector-valued function and $\epsilon_{1n} \xrightarrow{P} 0$ as $n \longrightarrow \infty$.

A2. $E\ \ell(X_j,\theta_0) = 0$.

A3. $M(\theta_0) = E\ \ell'(X_j,\theta_0)\ \ell(X_j,\theta_0)$ is a finite nonnegative definite matrix.

A4. The vector $\nabla_\theta F(x,\theta)$ is uniformly continuous in x and $\theta \in \Lambda$, where Λ is the closure of the given neighbourhood of θ_0.

A5. Each component of the vector function $\ell(x,\theta_0)$ is of bounded variation on each finite interval.

Lemma 2.1

Suppose that the vector function $\ell(x,\theta_0)$ satisfies conditions A3 and A5. Then, as $n \longrightarrow \infty$,

$$(2.2.2) \qquad L_n = \int \ell(x,\theta_0)\, d\Big[\alpha_n(x) - B_n\big[F(x,\theta_0)\big]\Big] \xrightarrow{P} 0,$$

where $\alpha_n(x)$ is the empirical process and B_n is a sequence of Brownian Bridges.

Proof:

Let

$$(2.2.3) \qquad T(x) = \big[T_1(x),\ldots,T_p(x)\big],$$

where $T_j(x)$ is the total variation of the j^{th} component $\ell_j(\cdot,\theta_0)$ of $\ell(x,\theta_0)$ on the interval $[-x,x]$, $j = 1,\ldots,p$. Choose a sequence of positive numbers u_n tending so slowly to infinity such that

$$(2.2.4) \qquad \|T(u_n)\| \frac{\log n}{\sqrt{n}} \longrightarrow 0 .$$

With this u_n, consider

$$L_n = \int_{|x|>u_n} \ell(x,\theta_0)\, d\,\alpha_n(x) - \int_{|x|>u_n} \ell(x,\theta_0)\, d\,B_n\big[F(x,\theta_0)\big]$$

$$+ \int_{|x|\le u_n} \ell(x,\theta_0)\, d\Big[\alpha_n(x) - B_n\big[F(x,\theta_0)\big]\Big]$$

(2.2.5) $= L_{1n} - L_{2n} + L_{3n}$.

Now, we can write L_{3n} as

$$\int_{|x|\le u_n} \ell(x,\theta_0)\ d\,\alpha_n(x) - \int_{|x|\le u_n} \ell(x,\theta_0)\ d\,B_n\left[F(x,\theta_0)\right].$$

Since $\ell(x,\theta_0)$ is of bounded variation, using integration by parts we have

$$\alpha_n(x)\ \ell(x,\theta_0)\Big|_{x=-u_n}^{u_n} - \int_{|x|\le u_n} \alpha_n(x)\ d\,\ell(x,\theta_0)$$

$$- B_n\left[F(x,\theta_0)\right]\ \ell(x,\theta_0)\Big|_{x=-u_n}^{u_n} + \int_{|x|\le u_n} B_n\left[F(x,\theta_0)\right]\ d\,\ell(x,\theta_0)$$

$$= \left[\left[\alpha_n(x) - B_n\left[F(x,\theta_0)\right]\right]\ \ell(x,\theta_0)\right]_{x=-u_n}^{u_n}$$

$$- \int_{|x|\le u_n} \left[\alpha_n(x) - B_n\left[F(x,\theta_0)\right]\right]\ d\,\ell(x,\theta_0).$$

Therefore,

$$\|L_{3n}\| \le \left\| \int_{|x|\le u_n} \left[\alpha_n(x) - B_n\left[F(x,\theta_0)\right]\right]\ d\,\ell(x,\theta_0)\right\|$$

$$+ \left\|\left[\left[\alpha_n(x) - B_n\left[F(x,\theta_0)\right]\right]\ \ell(x,\theta_0)\right]_{x=-u_n}^{u_n}\right\|.$$

Using Theorem 1.2, (2.2.3) and (2.2.4), we have

(2.2.6) $$\|L_{3n}\| \underset{a.s.}{=} 0\left\{\frac{\log n}{\sqrt{n}}\right\} \|T(u_n)\| \longrightarrow 0.$$

Next, consider $L_{1n}^{(j)}$ and $L_{2n}^{(j)}$, the j^{th} components of L_{1n} and L_{2n} respectively, $j = 1,\ldots,p$. Since

$$E\ \alpha_n(x) = E\ B_n\left[F(x,\theta_0)\right] = 0,$$

we have

(2.2.7) $$E\ L_{1n}^{(j)} = E\ L_{2n}^{(j)} = 0.$$

Thus

$$E\left[L_{1n}^{(j)}\right]^2 = E\left[L_{2n}^{(j)}\right]^2 = \int_{|x|>u_n} \ell_j^2(x,\theta_0)\ d\ F(x,\theta_0)$$

$$- \left[\int_{x<-u_n} \ell_j(x,\theta_0)\ d\ F(x,\theta_0)\right]^2 - \left[\int_{x>u_n} \ell_j(x,\theta_0)\ d\ F(x,\theta_0)\right]^2.$$

But by (2.2.7), we have

$$\int_{|x|>u_n} \ell_j(x,\theta_0)\ d\ F(x,\theta_0) \text{ is bounded.}$$

Hence, by Chebyshev inequality, with $\epsilon > 0$, we have

$$P\left\{\|L_{1n}\| + \|L_{2n}\| > 2\ \epsilon\right\} \leq \frac{2}{\epsilon^2} \sum_{j=1}^{p} \int_{|x|>u_n} \ell_j^2(x,\theta_0)\ d\ F(x,\theta_0).$$

(2.2.8) As $u_n \longrightarrow \infty$, this bound tends to 0 by condition A3. Finally, using the results in (2.2.6) and (2.2.8), we obtain (2.2.2).

2.3 CONVERGENCE OF THE ESTIMATED EMPIRICAL PROCESS

Theorem 2.1 (Burke, Csörgő, M., Csörgő, S., Révész (1979))
Suppose that the sequence $\{\hat{\theta}_n\}$ satisfies conditions A1 to A5 then

(2.3.1) $$\sup_{-\infty<x<\infty} |\hat{\alpha}_n(x) - G_n(x)| \xrightarrow{P} 0,$$

where $\hat{\alpha}_n(x)$ and $G_n(x)$ are defined in (2.1.1) and (2.2.1) respectively.

Proof:

Let us consider

$$\hat{\alpha}_n(x) = \sqrt{n}\left[F_n(x) - F(x,\hat{\theta}_n)\right]$$

$$= \sqrt{n}\left[F_n(x) - F(x,\theta_0)\right] - \sqrt{n}\left[F(x,\hat{\theta}_n) - F(x,\theta_0)\right].$$

Applying the one-term Taylor expansion of F with respect to θ_0, we obtain

(2.3.2) $$\hat{\alpha}_n(x) = \sqrt{n}\left[F_n(x) - F(x,\theta_0)\right] - \sqrt{n}\left[\hat{\theta}_n-\theta_0\right] \cdot \left[\nabla_\theta F(x,\theta_n^*)\right]',$$

where

(2.3.3) $$\|\theta_n^* - \theta_0\| \le \|\hat{\theta}_n - \theta_0\|.$$

The right-hand side of (2.3.2) is equal to

$$\alpha_n(x) - \sqrt{n}\left[\hat{\theta}_n - \theta_0\right]\left[\nabla_\theta F(x,\theta_n^*)\right]'$$

$$= B_n\left[F(x,\theta_0)\right] - \sqrt{n}\left[\hat{\theta}_n - \theta_0\right]\cdot\left[\nabla_\theta F(x,\theta_0)\right]'$$

$$+ \alpha_n(x) - B_n\left[F(x,\theta_0)\right] + \sqrt{n}\left[\hat{\theta}_n - \theta_0\right]\left[\nabla_\theta F(x,\theta_0) - \nabla_\theta F(x,\theta_n^*)\right]'$$

$$= B_n\left[F(x,\theta_0)\right] - \sqrt{n}\left[\hat{\theta}_n - \theta_0\right]\cdot\left[\nabla_\theta F(x,\theta_0)\right]' + \epsilon_{2n}(x) + \epsilon_{3n}(x),$$

where

$$\epsilon_{2n}(x) = \alpha_n(x) - B_n\left[F(x,\theta_0)\right] \quad \text{and}$$

$$\epsilon_{3n}(x) = \sqrt{n}\left[\hat{\theta}_n - \theta_0\right]\left[\nabla_\theta F(x,\theta_0) - \nabla_\theta F(x,\theta_n^*)\right]'.$$

If we can show that

(2.3.4) (i) $$\left\|\sqrt{n}\left[\hat{\theta}_n - \theta_0\right] - \int \ell(x,\theta_0)\, d\, B_n\left[F(x,\theta_0)\right]\right\| \xrightarrow{P} 0,$$

(2.3.5) (ii) $$\sup_{-\infty<x<\infty} |\epsilon_{2n}(x)| \xrightarrow{P} 0,$$

(2.3.6) (iii) $$\sup_{-\infty<x<\infty} |\epsilon_{3n}(x)| \xrightarrow{P} 0,$$

then we get the desired result in (2.3.1). We note that

$$\frac{1}{n}\sum_{j=1}^{n} \ell(x_j,\theta_0) = \int \ell(x,\theta_0)\, d\, F_n(x),$$

where $d\, F_n(x) = \frac{1}{n}$ when $x = X_i$. Using condition A1, we have

$$\sqrt{n}\left[\hat{\theta}_n-\theta_0\right] = \int \ell(x,\theta_0)\ d\ \sqrt{n}\ F_n(x) + \epsilon_{1n}.$$

By condition A2, this is equal to

$$\int \ell(x,\theta_0)\ d\ \sqrt{n}\left[F_n(x) - F(x,\theta_0)\right] + \epsilon_{1n}$$

$$= \int \ell(x,\theta_0)\ d\ \alpha_n(x) + \epsilon_{1n}$$

$$= \int \ell(x,\theta_0)\ dB_n\left[F(x,\theta_0)\right] + \int \ell(x,\theta_0)\ d\left[\alpha_n(x) - B_n\left[F(x,\theta_n)\right]\right] + \epsilon_{1n}$$

$$= \int \ell(x,\theta_0)\ dB_n\left[F(x,\theta_0)\right] + L_n + \epsilon_{1n}.$$

Since $\epsilon_{1n} \xrightarrow{P} 0$ and $L_n \xrightarrow{P} 0$ by Lemma 2.1, (2.3.4) follows.

The result in (2.3.5) is a direct consequence of Theorem 1.2 which says that

$$\sup_{-\infty<x<\infty} \left|\alpha_n(x) - B_n\left[F(x,\theta_0)\right]\right| \xrightarrow{a.s.} 0\left[\frac{\log n}{\sqrt{n}}\right].$$

Lastly, to show (2.3.6), we take note that $\sqrt{n}\left[\hat{\theta}_n-\theta_0\right]$ is asymptotically a normal vector because of conditions A1, A2 and A3. Thus

$$\left\|\hat{\theta}_n-\theta_0\right\| \xrightarrow{P} 0. \tag{2.3.7}$$

From (2.3.3), (2.3.7) and A4, we have

$$\sup_{-\infty<x<\infty} |\epsilon_{3n}(x)| \xrightarrow{P} 0.$$

Q.E.D.

As in Burke et.al. (1979) paper, if the convergence of ϵ_{1n} is almost-sure, then the convergence of Theorem 2.1 is also almost-sure provided conditions A1* - A3* hold. These additional assumptions are as follows:

A1*. The vector $\nabla_\theta F(x,\theta_0)$ is uniformly bounded in x and the vector $\nabla_\theta^2 F(x,\theta)$ is uniformly bounded in x and $\theta \in \Lambda$, where Λ is defined as in A4.

A2*. $\lim_{s\downarrow 0}\left[s \log\log \frac{1}{s}\right]^{1/2} \cdot \left\|\ell\left[F^{-1}(s,\theta_0),\ \theta_0\right]\right\| = 0$

and $\lim_{s\uparrow 1}\left[(1-s)\log\log\frac{1}{1-s}\right]^{1/2} \cdot \left\|\ell\left[F^{-1}(s,\theta_0),\ \theta_0\right]\right\| = 0,$

where $F^{-1}(s,\theta_0) = \inf\left\{x:\ F(x,\theta_0) \geq s\right\}$.

A3*. $s\left\|\frac{\partial \ell(F^{-1}(s,\theta_0),\ \theta_0)}{\partial s}\right\| \leq c, \quad 0 < s < \frac{1}{2},$

and $(1-s)\left\|\frac{\partial \ell(F^{-1}(s,\theta_0),\ \theta_0)}{\partial s}\right\| \leq c, \quad \frac{1}{2} < s < 1,$

for some positive constant c, where the vector of partial derivatives of the components of $\ell\left[F^{-1}(s,\theta_0),\ \theta_0\right]$ with respect to s, $\frac{\partial \ell(F^{-1}(s,\theta_0),\ \theta_0)}{\partial s}$, exists for all $s \in (0,1)$.

If $\epsilon_{1n} =_{a.s.} O\{h(n)\}$, where $h(n) > 0$, $h(n) \longrightarrow 0$, and the above conditions hold, the authors have established that the rate of convergence of Theorem 2.1 is

$$\sup_{-\infty<x<\infty} |\hat{\alpha}_n(x) - G_n(x)| \overset{=}{_{a.s.}} 0\left\{\max(h(n),\ n^{-\epsilon})\right\} \text{ for some } \epsilon > 0.$$

2.4 MOMENTS OF G(x)

In the previous chapter, we know that a sequence of Brownian Bridges converges to a Brownian Bridge, that is,

$$B_n \xrightarrow{\mathcal{D}} B.$$

This implies that

$$G_n \xrightarrow{\mathcal{D}} G,$$

where G is represented as

$$\text{(2.4.1)} \qquad G(x) = B\left[F(x,\theta_0)\right] - \int \ell(x,\theta_0)\ d\ B\left[F(x,\theta_0)\right]\cdot\left[\nabla_\theta\ F(x,\theta_0)\right]'.$$

This Gaussian process $G(x)$ has mean 0 and covariance function

$$\text{(2.4.2)} \qquad E\ G(x)G(y) = F(x,\theta_0)\ \Lambda\ F(y,\theta_0) - F(x,\theta_0)\ F(y,\theta_0)$$

$$- J(x)\left[\nabla_\theta\ F(y,\theta_0)\right]' - J(y)\left[\nabla_\theta\ F(x,\theta_0)\right]'$$

$$+ \left[\nabla_\theta\ F(x,\theta_0)\right]\cdot M(\theta_0)\cdot\left[\nabla_\theta\ F(y,\theta_0)\right]',$$

where $M(\theta_0)$ is defined in A3 and

$$J(x) = \int_{-\infty}^{x} \ell(z,\theta_0)\ d\ F(z,\theta_0).$$

The mean obviously follows from the mean of the Brownian Bridge.

To prove (2.4.2), let us consider

$$G(x) = B\left[F(x,\theta_0)\right] - \int \ell(x,\theta_0)\ d\ B\left[F(x,\theta_0)\right]\cdot\left[\nabla_\theta\ F(x,\theta_0)\right]'$$

$$G(y) = B\left[F(y,\theta_0)\right] - \int \ell(y,\theta_0)\ d\ B\left[F(y,\theta_0)\right]\cdot\left[\nabla_\theta\ F(y,\theta_0)\right]'.$$

Now,

$$\begin{aligned} G(x)G(y) = {} & B\left[F(y,\theta_0)\right]B\left[F(x,\theta_0)\right] - B\left[F(y,\theta_0)\right]\int \ell(z,\theta_0)\ d\ B\left[F(z,\theta_0)\right] \\ & \cdot\left[\nabla_\theta\ F(x,\theta_0)\right]' - B\left[F(x,\theta_0)\right]\int \ell(z,\theta_0)\ d\ B\left[F(z,\theta_0)\right] \\ & \cdot\left[\nabla_\theta\ F(y,\theta_0)\right]' + \int \ell(z,\theta_0)\ d\ B\left[F(z,\theta_0)\right]\cdot\left[\nabla_\theta\ F(x,\theta_0)\right]' \\ & \cdot\int \ell(z,\theta_0)\ d\ B\left[F(z,\theta_0)\right]\cdot\left[\nabla_\theta\ F(y,\theta_0)\right]'. \end{aligned}$$

Let

$$G(x)G(y) = L_1 - L_2 - L_3 + L_4.$$

Thus

$$\text{(2.4.3)} \qquad E\ G(x)G(y) = EL_1 - EL_2 - EL_3 + EL_4.$$

Consider the first term in (2.4.3).

$$EL_1 = E\ B\left[F(y,\theta_0)\right]\ B\left[F(x,\theta_0)\right]$$

$$\text{(2.4.4)} \qquad = F(x,\theta_0)\ \Lambda\ F(y,\theta_0) - F(x,\theta_0)\ F(y,\theta_0).$$

This follows from the covariance function of the Brownian Bridge.

As to the second term of (2.4.3), we have

$$EL_2 = E\, B\left[F(y,\theta_0)\right] \int \ell(z,\theta_0)\, d\, B\left[F(z,\theta_0)\right] \cdot \left[\nabla_\theta\, F(x,\theta_0)\right]'$$

$$= \int \ell(z,\theta_0)\, d\, E\, B\left[F(y,\theta_0)\right] B\left[F(z,\theta_0)\right] \cdot \left[\nabla_\theta\, F(x,\theta_0)\right]'$$

$$= \int \ell(z,\theta_0)\, d\left[F(y,\theta_0) \wedge F(z,\theta_0) - F(y,\theta_0)F(z,\theta_0)\right] \cdot \left[\nabla_\theta F(x,\theta_0)\right]'$$

$$= \int \ell(z,\theta_0)\, d\left[F(y,\theta_0) \wedge F(z,\theta_0)\right] \cdot \left[\nabla_\theta\, F(x,\theta_0)\right]'$$

$$- \int \ell(z,\theta_0)\, d\left[F(y,\theta_0) \cdot F(z,\theta_0)\right] \cdot \left[\nabla_\theta\, F(x,\theta_0)\right]'$$

$$= \int \ell(z,\theta_0)\, d\, F(\min\,(y,z),\theta_0) \cdot \left[\nabla_\theta\, F(x,\theta_0)\right]'$$

$$- F(y,\theta_0) \int \ell(z,\theta_0)\, d\, F(z,\theta_0) \cdot \left[\nabla_\theta\, F(x,\theta_0)\right]'.$$

Since $\int \ell(x,\theta_0)\, d\, F(x,\theta_0) = 0$, we have

$$EL_2 = \int \ell(z,\theta_0)\, d\, F(\min\,(y,z),\theta_0) \cdot \left[\nabla_\theta\, F(x,\theta_0)\right]'$$

$$= \left[\int_{z \le y} \ell(z,\theta_0) dF(\min(y,z),\theta_0) + \int_{z>y} \ell(z,\theta_0) dF(\min(y,z),\theta_0)\right] \cdot \left[\nabla_\theta F(x,\theta_0)\right]'$$

$$= \left[\int_{z \le y} \ell(z,\theta_0) dF(z,\theta_0) + \int_{z>y} \ell(z,\theta_0) dF(y,\theta_0)\right] \cdot \left[\nabla_\theta F(x,\theta_0)\right]'.$$

But $\dfrac{d\, F(y,\theta_0)}{dz} = 0$, hence

$$EL_2 = \int_{-\infty}^{y} \ell(z,\theta_0)dF(z,\theta_0)\cdot\left[\nabla_\theta\ F(x,\theta_0)\right]'$$

(2.4.5) $$= J(y)\ \cdot\ \left[\nabla_\theta\ F(x,\theta_0)\right]'.$$

Using the same argument as in the proof of EL_2, we have

$$EL_3 = \int_{-\infty}^{x} \ell(z,\theta_0)dF(z,\theta_0)\cdot\left[\nabla_\theta\ F(y,\theta_0)\right]'$$

(2.4.6) $$= J(x)\ \cdot\ \left[\nabla_\theta\ F(y,\theta_0)\right]'.$$

Now, let us consider the last term of (2.4.3).

$$EL_4 = E\int\ell(z,\theta_0)dB\left[F(z,\theta_0)\right]\cdot\left[\nabla_\theta F(x,\theta_0)\right]'\cdot\int \ell(z,\theta_0)\ dB\left[F(z,\theta_0)\right]$$
$$\cdot\left[\nabla_\theta F(y,\theta_0)\right]'$$
$$= \left[\nabla_\theta F(x,\theta_0)\right]\ E\int\ell'(z,\theta_0)dB\left[F(z,\theta_0)\right]\cdot\int \ell(z,\theta_0)\ dB\left[F(z,\theta_0)\right]$$
$$\cdot\left[\nabla_\theta F(y,\theta_0)\right]'.$$

Since

$$E\int\ell(z,\theta_0)dB\left[F(z,\theta_0)\right] = \int\ell(z,\theta_0)\ dEB\left[F(z,\theta_0)\right] = 0,$$

this means that

(2.4.7) $$E\int \ell'(z,\theta_0)dB\left[F(z,\theta_0)\right]\cdot\int \ell(z,\theta_0)\ dB\left[F(z,\theta_0)\right]$$

$$= \int\int \ell'(z,\theta_0)\ \ell(z',\theta_0)\ d\ E\ B\left[F(z,\theta_0)\right]\ B\left[F(z',\theta_0)\right]$$

$$= \int\int \ell'(z,\theta_0)\ \ell(z',\theta_0)\ d\ \left[F(z,\theta_0)\ \Lambda\ F(z',\theta_0)\right]$$

$$- \left\{\int\int \ell'(z,\theta_0)\ d\ F(z,\theta_0)\right\} \left\{\int\int \ell(z',\theta_0)\ d\ F(z',\theta_0)\right\}$$

$$(2.4.8) \qquad = \int \ell'(z,\theta_0)\ \ell(z,\theta_0)\ dF(z,\theta_0) - \left\{\int\int \ell'(z,\theta_0)\ dF(z,\theta_0)\right\}$$

$$\cdot \left\{\int\int \ell(z',\theta_0)\ dF(z',\theta_0)\right\}.$$

But $\int \ell(z',\theta_0)\ d\ F(z',\theta_0) = 0$, thus the second term in (2.4.8) goes to 0. Hence (2.4.7) is equal to

$$\int \ell'(z,\theta_0)\ \ell(z,\theta_0)\ dF(z,\theta_0)$$

$$= E\ \ell'(z,\theta_0)\ \ell(z,\theta_0)$$

$$(2.4.9) \qquad = M(\theta_0).$$

This implies that

$$(2.4.10) \qquad EL_4 = \left[\nabla_\theta\ F(x,\theta_0)\right] \cdot M(\theta_0) \cdot \left[\nabla_\theta\ F(y,\theta_0)\right]'.$$

Combining the results in (2.4.4), (2.4.5), (2.4.6) and (2.4.10), we finally get the covariance function in (2.4.2).

2.5 THE MAXIMUM LIKELIHOOD ESTIMATOR CASE

The sequence of maximum likelihood estimator often satisfy A1 with

(2.5.1) $$\ell(x,\theta_0) = \nabla_\theta \log f(x,\theta_0)\ I^{-1}(\theta_0),$$

where f is the density function of F and $I^{-1}(\theta_0)$ is the inverse of the Fisher information matrix

(2.5.2) $$I(\theta_0) = E\left[\nabla_\theta \log f(X_i,\theta_0)\right]' \cdot \left[\nabla_\theta \log f(X_i,\theta_0)\right].$$

Under these sequence of estimators, we can show that the covariance function of $G(x)$ is

$$EG(x)G(y) = F(x,\theta_0) \wedge F(y,\theta_0) - F(x,\theta_0)\ F(y,\theta_0)$$
$$- \left[\nabla_\theta F(x,\theta_0)\right] I^{-1}(\theta_0) \cdot \left[\nabla_\theta F(y,\theta_0)\right]'.$$

To show this, we look at (2.4.2) and obtain

$$J(y)\cdot\left[\nabla_\theta F(x,\theta_0)\right]' = \int_{-\infty}^{y} \ell(z,\theta_0)\ dF(z,\theta_0)\cdot\left[\nabla_\theta F(x,\theta_0)\right]'$$

$$= \int_{-\infty}^{y} \nabla_\theta \log f(z,\theta_0)\ I^{-1}(\theta_0)\ dF(z,\theta_0)\cdot\left[\nabla_\theta F(x,\theta_0)\right]'$$

$$= \int_{-\infty}^{y} \nabla_\theta \log f(z,\theta_0)\ dF(z,\theta_0)\cdot I^{-1}(\theta_0)\cdot\left[\nabla_\theta F(x,\theta_0)\right]'$$

$$= \int_{-\infty}^{y} \frac{\left[\nabla_\theta f(z,\theta_0)\right] f(z,\theta_0)}{f(z,\theta_0)}\ dz\cdot I^{-1}(\theta_0)\cdot\left[\nabla_\theta F(x,\theta_0)\right]'$$

$$= \nabla_\theta \int_{-\infty}^{y} f(z,\theta_0)\, dz \cdot I^{-1}(\theta_0) \cdot \left[\nabla_\theta F(x,\theta_0)\right]'$$

$$= \left[\nabla_\theta F(y,\theta_0)\right] I^{-1}(\theta_0) \left[\nabla_\theta F(x,\theta_0)\right]'$$

$$(2.5.3) \qquad = \left[\nabla_\theta F(x,\theta_0)\right] I^{-1}(\theta_0) \left[\nabla_\theta F(y,\theta_0)\right]'.$$

Similarly,

$$(2.5.4) \quad J(x) \left[\nabla_\theta F(y,\theta_0)\right]' = \left[\nabla_\theta F(x,\theta_0)\right] I^{-1}(\theta_0) \left[\nabla_\theta F(y,\theta_0)\right]',$$

while $\left[\nabla_\theta F(x,\theta_0)\right] \cdot M(\theta_0) \cdot \left[\nabla_\theta F(y,\theta_0)\right]'$ is equal to

$$\left[\nabla_\theta F(x,\theta_0)\right] E\, \ell'(z,\theta_0)\, \ell(z,\theta_0) \left[\nabla_\theta F(y,\theta_0)\right]'$$

$$= \left[\nabla_\theta F(x,\theta_0)\right] E\left\{I^{-1}(\theta_0) \left[\nabla_\theta \log f(z,\theta_0)\right]' \left[\nabla_\theta \log f(z,\theta_0)\right] I^{-1}(\theta_0)\right\}$$

$$\cdot \left[\nabla_\theta F(y,\theta_0)\right]'.$$

By (2.5.2), this is equal to

$$\left[\nabla_\theta F(x,\theta_0)\right] \cdot I^{-1}(\theta_0) \cdot I(\theta_0) \cdot I^{-1}(\theta_0) \cdot \left[\nabla_\theta F(y,\theta_0)\right]'$$

$$(2.5.5) \qquad = \left[\nabla_\theta F(x,\theta_0)\right] \cdot I^{-1}(\theta_0) \cdot \left[\nabla_\theta F(y,\theta_0)\right]'.$$

Combining the results in (2.4.2), (2.5.3), (2.5.4) and (2.5.5) we have

$$EG(x)G(y) = F(x,\theta_0) \wedge F(y,\theta_0) - F(x,\theta_0)F(y,\theta_0)$$

$$- \left[\nabla_\theta F(x,\theta_0)\right] I^{-1}(\theta_0) \left[\nabla_\theta F(y,\theta_0)\right]' - \left[\nabla_\theta F(x,\theta_0)\right] I^{-1}(\theta_0)$$

$$\cdot \left[\nabla_\theta F(y,\theta_0)\right]' + \left[\nabla_\theta F(x,\theta_0)\right] I^{-1}(\theta_0) \left[\nabla_\theta F(y,\theta_0)\right]'$$

$$(2.5.6) \qquad = F(x,\theta_0) \wedge F(y,\theta_0) - F(x,\theta_0)F(y,\theta_0)$$

$$- \left[\nabla_\theta F(x,\theta_0)\right]' I^{-1}(\theta_0) \left[\nabla_\theta F(y,\theta_0)\right]'.$$

2.6 APPROXIMATIONS UNDER A SEQUENCE OF ALTERNATIVES

In this section, the above results are extended to cover the asymptotic approximation of $\hat{\alpha}_n(x)$ under a sequence of alternative hypotheses.

Suppose that the distribution function of the i.i.d.r.v. is $F(x;\beta,\theta)$, where β is a p_1-dimensional vector of parameters which is assumed to be known and θ is a p_2-dimensional vector of unknown parameters which is estimated by $\{\hat{\theta}_n\}$, based on $X_1, X_2, \ldots, X_n$. Consider the null hypothesis

$$(2.6.1) \qquad H_0: \quad (\beta,\theta) = (\beta_0,\theta_0) = \delta_0$$

where θ_0 stands for the theoretical true value of θ. The sequence of alternatives $\{H_n\}$ is defined as follows:

Let $\{\beta_n\}$ be a sequence of p_1-dimensional (nonrandom) vectors satisfying the condition

$$\beta_n = \beta_0 + \gamma \cdot n^{-1/2},$$

where γ is a given constant vector. Let Λ_1 denote the closure of a given neighbourhood of β_0 and let $m = \min\left\{k: \beta_n \in \Lambda_1, \text{ for all } n \geq k > 2\right\}$.

Then consider

(2.6.2) $\qquad H_n: \ (\beta,\theta) = (\beta_n,\theta_0), \quad \text{for} \quad n = m, m+1, \ldots$

Under the sequence of alternatives $\left\{H_n\right\}$ of (2.6.2), we wish to show that the estimated empirical process

$$\hat{\alpha}_n(x) = \sqrt{n}\left[F_n(x) - F(x;\ \beta_0,\hat{\theta}_n)\right], \qquad x \in R,$$

can be estimated by the Gaussian process

(2.6.3) $\qquad Z_n(x) = G_n(x) - \gamma\, A'\left[\nabla_\theta\ F(x;\ \beta_0,\theta_0)\right]' + \gamma\left[\nabla_\beta\ F(x;\ \beta_0,\theta_0)\right]',$

with

$$G_n(x) = B_n\left[F(x;\ \beta_0,\theta_0)\right] - \left\{\int \ell(x;\ \beta_0,\theta_0)\ d\ B_n\left[F(x;\ \beta_0,\theta_0)\right]\right\}\cdot \left[\nabla_\theta\ F(x;\ \beta_0,\theta_0)\right]'$$

and A as defined below in (i).

We can easily verify that the mean of $Z_n(x)$ is

$$E\ Z_n(x) = -\gamma\ A'\left[\nabla_\theta\ F(x;\ \beta_0,\theta_0)\right]' + \gamma\left[\nabla_\beta\ F(x;\ \beta_0,\theta_0)\right]'$$

and its covariance is the same as in (2.4.2), with the obvious changes in notation.

Again we list all the assumptions under H_n.

(2.6.4)

(i) $\sqrt{n}\left[\hat{\theta}_n - \theta_0\right] = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \ell\left[X_j, \beta_0, \theta_0\right] + \gamma A' + \epsilon_{8n}$, where A is a given finite matrix of order $p_2 \times p_1$, ℓ is a measurable p_2-dimensional vector-valued function, and $\epsilon_{8n} \xrightarrow{P} 0$.

(ii) $E\, \ell\left[X_j, \beta_0, \theta_0\right] = 0$ for $n \geq m$.

(iii) $E\, \ell'\left[X_j, \beta_0, \theta_0\right] \ell\left[X_j, \beta_0, \theta_0\right] = M\left[\beta_n, \theta_0\right]$, a finite nonnegative definite matrix for each $n \geq m$ which converges to a finite nonnegative matrix $= M\left[\beta_0, \theta_0\right]$ as $n \longrightarrow \infty$.

(iv) The vector $\nabla_\beta F\left[x; \beta, \theta_0\right]$ is uniformly continuous in x and $\beta \in \Lambda_1$, and the vector $\nabla_\theta F\left[x; \beta_0, \theta\right]$ is uniformly continuous in x and $\theta \in \Lambda_2$, where Λ_2 is the closure of a given neighbourhood of θ_0.

(v) Each component of $\ell\left[x, \beta_0, \theta_0\right]$ is of bounded variation on each finite interval.

<u>REMARK</u>. Additional conditions are set to obtain the almost-sure approximation of $\hat{\alpha}_n(x)$ (Burke, <u>et</u>.<u>al</u>., 1979).

As in Lemma 2.1, we claim that if the vector function $\ell\left[x, \beta_0, \theta_0\right]$ satisfies conditions (2.6.4) (iii) and (v) then as $n \longrightarrow \infty$,

$$\text{(2.6.5)} \qquad \int \ell(x, \beta_0, \theta_0)\, d\left[\alpha_n(x) - B_n\left[F(x;\, \beta_0, \theta_0)\right]\right] \xrightarrow{P} 0.$$

We state the result.

<u>Theorem 2.2</u> (Burke, <u>et</u>.<u>al</u>., 1979)

Suppose that conditions (2.6.4) hold and let

$$\epsilon_{9n} = \sup_{-\infty<x<\infty} |\hat{\alpha}_n(x) - Z_n(x)|.$$

Then under the sequence of alternatives $\{H_n\}$,

$$\epsilon_{9n} \xrightarrow{P} 0.$$

<u>Proof</u>:

By adding and subtracting, we have under H_n

$$\hat{\alpha}_n(x) = \sqrt{n}\left[F_n(x) - F(x;\, \beta_0, \hat{\theta}_n)\right]$$

$$= \sqrt{n}\left[F_n(x) - F(x;\, \beta_n, \theta_0)\right] + \sqrt{n}\left[F(x;\, \beta_n, \theta_0) - F(x;\, \beta_0, \theta_0)\right]$$

$$- \sqrt{n}\left[F(x;\, \beta_0, \hat{\theta}_n) - F(x;\, \beta_0, \theta_0)\right]$$

$$\text{(2.6.6)} \qquad = Q_{1n}(x) + Q_{2n}(x) - Q_{3n}(x).$$

For the first term in (2.6.6) we have

$$Q_{1n}(x) = \sqrt{n}\left[F_n(x) - F(x;\ \beta_n,\theta_0)\right]$$

$$= B_n\ F(x;\ \beta_n,\theta_0)+\left\{\sqrt{n}\left[F_n(x)-F(x;\ \beta_n,\theta_0)\right]-B_n\ F(x;\ \beta_n,\theta_0)\right\}$$

$$= B_n\ F(x;\ \beta_0,\theta_0)+\left\{\sqrt{n}\left[F_n(x)-F(x;\ \beta_n,\theta_0)\right]-B_n\ F(x;\ \beta_n,\theta_0)\right\}$$

$$+\left\{B_n\ F(x;\ \beta_n,\theta_0) - B_n\ F(x;\ \beta_0,\theta_0)\right\}.$$

$$(2.6.7)\qquad = B_n\ F(x;\ \beta_0,\theta_0) + \epsilon_{3n}\left[F(x;\ \beta_n,\theta_0)\right] + \epsilon_{10n}(x).$$

By Theorem 1.2

$$(2.6.8)\qquad \sup_{-\infty<x<\infty} |\epsilon_{3n}\left[F(x;\ \beta_n,\theta_0)\right]| =_{a.s.} 0\left\{n^{-1/2}\log n\right\}.$$

From the modulus of continuity of the Brownian motion (Csörgő and Révész, 1979, Chapter I),

$$(2.6.9)\qquad \sup|\epsilon_{10n}(x)| =_{a.s.} 0\left\{n^{-\delta}\right\},$$

for any δ satisfying $0 < \delta < \frac{1}{4}$.

For the second term, apply the one-term Taylor expansion of F with respect to β_0 to obtain

$$Q_{2n}(x) = \sqrt{n}\left[F(x;\ \beta_n,\theta_0) - F(x;\ \beta_0,\theta_0)\right]$$

$$= \sqrt{n}\left[\beta_n-\beta_0\right]\left[\nabla_\beta\ F(x;\ \beta_n^*,\theta_0)\right]'$$

$$(2.6.10)\qquad = \gamma\left[\nabla_\beta\ F(x;\ \beta_n^*,\theta_0)\right]' +$$

$$\left\{\left[\sqrt{n}\left[\beta_n-\beta_0\right] - \gamma\right]\left[\nabla_\beta\ F(x;\ \beta_n^*,\theta_0)\right]'\right\},$$

where $\|\beta_n^* - \beta_0\| \le \|n^{-1/2}\gamma\|$. Thus the second term in (2.6.10) $\longrightarrow 0$ as $n \longrightarrow \infty$, reducing (2.6.10) to

$$\gamma\left[\nabla_\beta F(x;\ \beta_0,\theta_0)\right]' + \gamma\left[\nabla_\beta F(x;\ \beta_n^*,\theta_0) - \nabla_\beta F(x;\ \beta_0,\theta_0)\right]'$$

$$= \gamma\left[\nabla_\beta F(x;\ \beta_0,\theta_0)\right]' + \epsilon_{11n}(x). \tag{2.6.11}$$

By condition (2.6.4) (iv) and the fact that $\|\beta_n^* - \beta_0\| \longrightarrow 0$,

$$\sup_{-\infty<x<\infty} |\epsilon_{11n}(x)| \longrightarrow 0. \tag{2.6.12}$$

For the third term, we can repeat the proof of Theorem 2.1 to get

$$Q_{3n}(x) = \sqrt{n}\ \left[F(x;\ \beta_0,\hat{\theta}_n) - F(x;\ \beta_0,\theta_0)\right]$$

$$= \sqrt{n}\ \left[\hat{\theta}_n - \theta_0\right]\ \left[\nabla_\theta F(x;\ \beta_0,\theta_n^*)\right] \tag{2.6.13}$$

where $\|\theta_n^* - \theta_0\| \le \|\hat{\theta}_n - \theta_0\|$. As in the proof of (2.3.4), using conditions (2.6.4) (i) and (ii), we have

$$\sqrt{n}\ \left[\hat{\theta}_n - \theta_0\right] = \int \ell(x;\ \beta_0,\theta_0)\ d\ \sqrt{n}\ \left[F_n(x) - F(x;\ \beta_0,\theta_0)\right] + \gamma A' + \epsilon_{8n}.$$

Thus (2.6.13) can be written as

$$\left\{\int \ell(x;\ \beta_0,\theta_0)\ d\ B_n\ F(x;\ \beta_0,\theta_0) + \gamma\ A'\right\}\left[\nabla_\theta\ F(x;\ \beta_0,\theta_0)\right]' \tag{2.6.14}$$

$$+ \left\{\int \ell(x;\ \beta_0,\theta_0)\ d\ \left[\alpha_n(x) - B_n\ F(x;\ \beta_0,\theta_0)\right]\right\}\left[\nabla_\theta\ F(x;\ \beta_0,\theta_0)\right]'$$

$$+ \epsilon_{8n}.$$

By (2.6.5) and the fact that $\epsilon_{8n} \xrightarrow{P} 0$ we have

(2.6.15) $$Q_{3n}(x) = \left\{\int \ell(x;\ \beta_0,\theta_0)\ dB_n F(x;\ \beta_0,\theta_0) + \gamma\ A'\right\}\left[\nabla_\theta\ F(x;\ \beta_0,\theta_0)\right]'.$$

Combining all the results in (2.6.6) to (2.6.15), we finally have

$$\sup_{-\infty<x<\infty} |\hat{\alpha}_n(x) - Z_n(x)| \xrightarrow{P} 0.$$

CHAPTER III

DISTRIBUTION-FREE PROCEDURES

3.1 INTRODUCTION

For testing the composite hypothesis

$$(3.1.1) \qquad H_0: \ F \in \mathcal{F} = \{F(x,\theta): x \in R, \ \theta \in S \subset R^P\},$$

it was shown in the previous chapter that tests based on the estimated empirical process $\hat{\alpha}_n(x)$ are inadequate; they are not distribution-free. It is the objective of this chapter to examine some procedures that will overcome this difficulty. We wish to obtain a version of the empirical process, whose limiting distribution in the composite hypothesis case is the same as that of the usual form of the empirical process in the specified hypothesis

$$H_0: \ F = F_0.$$

Section 2 refers to the bootstrap method proposed by Burke and Gombay (1988). In this method, a bootstrap sample of size n is obtained from the random sample $X_1, X_2, \ldots, X_n$ of $F(x)$. This is used to obtain an estimate of the unknown sequence of parameters $\theta_0 = (\theta_1, \ldots \theta_p)$.

In Sections 6 and 7 we will examine the half-sample and random substitution device suggested by Durbin (1976) and (1961). In the first technique, as initiated by Rao (1972), the unknown parameter vector is estimated from a randomly chosen half-sample of data and the empirical distribution function is constructed as if the estimates were the true value. In the other method the unknown parameter vector is estimated from an external sample with known distribution function.

Using the above mentioned procedures, the estimated empirical process is approximated by a sequence of Gaussian processes. In the maximum likelihood case, it converges to a Brownian Bridge.

A significant implication of this result is that the K-S, C-vM, and A-D type of statistics converge respectively to the following:

$$\sqrt{n}\,\hat{D}_n = \sup_{-\infty<x<\infty} \sqrt{n}\,|F_n(x) - F(x,\hat{\theta}_n)| \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1} |B(t)|$$

(3.1.2)

$$\hat{W}^2_n = n\int_{-\infty}^{\infty} \left[F_n(x) - F(x,\hat{\theta}_n)\right]^2 d\,F(x,\hat{\theta}_n) \xrightarrow{\mathcal{D}} \int_0^1 B(t)^2\,dt$$

$$\hat{A}^2_n = n\int_{-\infty}^{\infty} \frac{\left[F_n(x)- F(x,\hat{\theta}_n)\right]^2}{F(x,\hat{\theta}_n)\left[1-F(x,\hat{\theta}_n)\right]}\,d\,F(x,\hat{\theta}_n) \xrightarrow{\mathcal{D}} \int_0^1 \frac{B(t)^2}{t(1-t)}\,dt,$$

where $\{\hat{\theta}_n\}$, the sequence of estimators of θ, is obtained using the above procedures and B is a Brownian Bridge. These statistics could now be applied to test the hypothesis in (3.1.1).

3.2 THE BOOTSTRAP METHOD

A general method called "bootstrap" was first introduced by Efron (1979) to solve a variety of estimation problems. For example, the estimation of the sampling distribution of the random variable $R(\underset{\sim}{X},F)$ on the basis of the observed data $\underset{\sim}{X}$. This procedure is based on randomization. The bootstrap algorithm is as follows:

(i) Draw a random sample $X_1,X_2,\ldots,X_n$ from a population with distribution function $F(x)$.

(ii) Construct the sample probability distribution $\hat{F}$, putting mass $\frac{1}{n}$ at each point $X_1,X_2,\ldots,X_n$.

(iii) Given the sample, the bootstrap sample $X_1^*,X_2^*,\ldots,X_m^*$ is obtained by sampling with replacement m elements of the set $\{X_1,\ldots,X_n\}$. Hence

$$P\{X_1^* \le x_1,\ X_2^* \le x_2,\ldots,X_m^* \le x_m | X_1,\ldots,X_n\} = \prod_{i=1}^{m} F_n(x_i).$$

(iv) Approximate the sampling distribution of $R(\underset{\sim}{X},F)$ by the bootstrap distribution of $R^* = R(\underset{\sim}{X}^*,F)$.

For our purposes, consider a random sample $X_1,X_2,\ldots,X_n$ from a distribution function $F(x)$. Given this sample, obtain the bootstrap sample $X_1^*,X_2^*,\ldots,X_m^*$. Let $\hat{\theta}_n$ be the maximum likelihood estimator of θ. Let $\tilde{\theta}_m$ be the bootstrapped version of $\hat{\theta}_n$ based on the bootstrapped sample $X_1^*,X_2^*,\ldots,X_m^*$. Our main goal is to show that when $n = m$, the estimated empirical process

$$\tilde{\alpha}_{n,m}(x) = \sqrt{n}\left[F_n(x) - F(x,\tilde{\theta}_m)\right] \tag{3.2.1}$$

converges weakly to a Brownian Bridge, that is, (3.2.2) $$\tilde{\alpha}_{n,n}(x) \xrightarrow{\mathscr{D}} B\left[F(x,\theta_0)\right].$$

To attain this goal, we need to show that $\tilde{\alpha}_{n,m}$ can be approximated asymptotically by the Gaussian process

(3.2.3) $$G_{n,m}(x) = B_n\left[F(x,\theta_0)\right] - \int \ell(x,\theta_0)\ d\ B_n\left[F(x,\theta_0)\right]\cdot\left[\nabla_\theta\ F(x,\theta_0)\right]'$$
$$- \left[nm^{-1}\right]^{1/2} \int \ell(x,\theta_0)\ d\ B_m^*\left[F(x,\theta_0)\right]\cdot\left[\nabla_\theta\ F(x,\theta_0)\right]'.$$

Obviously, the mean function of $G_{n,n}(x)$ is 0. Its covariance function is

(3.2.4) $$E\ G_{n,n}(x)\ G_{n,n}(y) = F(x,\theta_0) \wedge F(y,\theta_0) - F(x,\theta_0)\ F(y,\theta_0).$$

Remarkably, this is the covariance function of a Brownian Bridge. Since a Gaussian process is uniquely determined by its covariance function, we conclude that

$$G_{n,n}(x) \xrightarrow{\mathscr{D}} B\left[F(x,\theta_0)\right].$$

Using this fact, we get the desired result in (3.2.2).

3.3 BOOTSTRAP EMPIRICAL PROCESS

Let $X_1, X_2, \ldots, X_n$ be a random sample of $F(x)$ and obtain the bootstrap sample $X_1^*, X_2^*, \ldots, X_m^*$. We define the bootstrapped empirical distribution function as

$$F_m^*(x) = \sum_{i=1}^{n} \frac{q_i \, I\{X_i \le x\}}{m}, \quad x \in R, \tag{3.3.1}$$

where q_i is the number of times X_i occurs in the bootstrap sample $X_1^*, X_2^*, \ldots, X_m^*$, $i = 1, \ldots, n$. We note that

$$\sum_{i=1}^{n} q_i = m. \tag{3.3.2}$$

We also define the <u>bootstrap empirical process</u> as

$$\sqrt{m}\left[F_m^*(x) - F_n(x)\right]. \tag{3.3.3}$$

<u>Lemma 3.1</u>

Let $U_1, \ldots, U_m$ be i.i.d.r.v. from a uniform distribution $U(0,1)$ which are independent of $X_1, \ldots, X_n$. Define $E_m(u)$ as

$$E_m[u] = \sum_{j=1}^{m} \frac{I(U_j \le u)}{m}, \quad 0 \le u \le 1. \tag{3.3.4}$$

Then

$$\sqrt{m}\left[F_m^*(x) - F_n(x)\right] =_{\mathcal{D}} \sqrt{m}\left[E_m\left[F_n(x)\right] - F_n(x)\right]. \tag{3.3.5}$$

<u>Proof</u>:

For each n,

$$\left\{q_i,\ 1 \le i \le n\right\} =_{\mathcal{D}} \left\{m\left[\sum_{j=1}^{m} \frac{I(U_j \le \frac{i}{n})}{m} - \sum_{j=1}^{m} \frac{I(U_j \le \frac{i-1}{n})}{m}\right]\right\}. \tag{3.3.6}$$

From (3.3.4) we have

$$(3.3.7) \qquad \left\{q_i,\ 1 \le i \le n\right\} =_{\mathcal{D}} \left\{m\left[E_m\left[\frac{i}{n}\right] - E_m\left[\frac{i-1}{n}\right]\right];\quad 1 \le i \le n\right\}$$

and from (3.3.1)

$$(3.3.8) \qquad \left\{F_m^*(x),\ x \in R,\ m = 1,2,\ldots\right\} =_{\mathcal{D}} \left\{\sum_{i=1}^{n}\left[E_m\left[\frac{i}{n}\right] - E_m\left[\frac{i-1}{n}\right]\right] \cdot I\,(X_i \le x);\ x \in R\right\}.$$

Thus

$$(3.3.9) \qquad \left\{\sqrt{m}\,\left[F_m^*(x) - F_n(x)\right],\ x \in R,\ m = 1,2,\ldots\right\}$$

$$=_{\mathcal{D}} \left\{\sqrt{m}\,\left[\sum_{i=1}^{n}\left[E_m\left[\frac{i}{n}\right]-E_m\left[\frac{i-1}{n}\right]\right]\ I(X_i \le x) - \sum_{i=1}^{n}\frac{I(X_i \le x)}{n}\right]\right\}$$

$$=_{\mathcal{D}} \left\{\sqrt{m}\left[\sum_{i=1}^{n}\left[E_m\left[\frac{i}{n}\right]-E_m\left[\frac{i-1}{n}\right]-\frac{1}{n}\right]\ I(X_i \le x)\right],\ x \in R,\ m = 1,2,\ldots\right\}$$

$$(3.3.10) \quad =_{\mathcal{D}} \left\{\sqrt{m}\,\left[\sum_{i=1}^{n}\left[E_m\left[\frac{i}{n}\right]-E_m\left[\frac{i-1}{n}\right]- \frac{1}{n}\right]\ I\,(X_{(i)} \le x)\right],\ x \in R,\ m = 1,2,\ldots\right\},$$

where $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ are the order statistics of the sample $X_1,\ldots,X_n$. (3.3.10) is true since the q_i's depend only on n and m and thus independent of the X_i's. From Csörgo <u>et</u>.<u>al</u>. (1986, Section 17.2), (3.3.10) is

$$=_{\mathcal{D}} \sqrt{m}\,\left[E_m\left[F_n(x)\right] - F_n(x)\right].$$

Hence, we have

$$\left\{\sqrt{m}\left[F_m^*(x) - F_n(x)\right],\ x \in R,\ m = 1,\ldots\right\} \Longrightarrow_{\mathscr{D}} \left\{\sqrt{m}\left[E_m\left[F_n(x)\right] - F_n(x)\right]\right\}$$

as desired.

3.4 PRELIMINARIES

As we go along the proof of (3.2.3), we assume conditions A1–A5 and notations set in the previous chapter. In addition, we state more conditions to be used in the proof of the main result.

We claim the following assumption:

A6. $$\sqrt{m}\left[\tilde{\theta}_m - \hat{\theta}_n\right] = \frac{1}{\sqrt{m}}\left[\sum_{j=1}^{m} \ell(X_j^*, \theta_0) - \frac{m}{n}\sum_{j=1}^{n} \ell(X_j, \theta_0)\right] + \epsilon_{n,m},$$

where $\epsilon_{n,m} \xrightarrow{P} 0$ as $n \wedge m \longrightarrow \infty$.

Asymptotically, this condition implies that

$$\sqrt{m}\left[\tilde{\theta}_m - \hat{\theta}_n\right] = \int \ell(x,\theta_0)\, d\, \sqrt{m}\left[F_m^*(x) - F_n(x)\right]. \tag{3.4.1}$$

To show this, consider

$$\begin{aligned}
\sqrt{m}\left[\tilde{\theta}_m - \hat{\theta}_n\right] &= \sqrt{m}\left[\tilde{\theta}_m - \theta_0\right] - \sqrt{m}\left[\hat{\theta}_n - \theta_0\right] \\
&= \int \ell(x,\theta_0)\, d\, m^{1/2}\left[F_m^*(x) - F(x,\theta_0)\right] \\
&\quad - \int \ell(x,\theta_0)\, d\, m^{1/2}\left[F_n(x) - F(x,\theta_0)\right] \\
&= \int \ell(x,\theta_0)\, d\, m^{1/2}\left[F_m^*(x) - F_n(x)\right].
\end{aligned}$$

This is true because of A1, A2 and A6.

From Csörgo *et.al*. (1986, Theorem 17.8) and Csörgo and Révész (1981, Lemma 4.4.4), we can construct a single probability space with random functions E'_m, F'_n and B^*_m defined on it. Also on this probability space are versions of $\left[\hat{\alpha}_n, G_n\right]$ and $\left[\alpha^*_m, B^*_m\right]$ which are statistically independent. These random functions satisfy the following:

(i) B^*_m is a sequence of Brownian Bridge.

(ii) $\left[E'_m(t), F'_n(x)\right] =_{\mathcal{D}} \left[E_m(t), F_n(x)\right]$, $t \in [0,1]$, $x \in R$.

(iii) F'_n independent of random vector $\left[E'_m, B^*_m\right]$.

(iv) $\sup_{-\infty<x<\infty} \left|\alpha^*_m(x) - B^*_m\left[F(x,\theta_0)\right]\right| =_{a.s.} 0\left\{\frac{(\log n)^{3/4}}{n^{1/4}}\right\}$

as $n \wedge m \longrightarrow \infty$, where

$$\alpha^*_m(x) = \sqrt{m}\left[E'_m\left[F'_n(x)\right] - F'_n(x)\right] \tag{3.4.2}$$

and $0 < \lim_{n,m\to\infty} \inf \frac{m}{n} \leq \lim_{n,m\to\infty} \sup \frac{m}{n} < \infty$.

(v) $\sup_{-\infty<x<\infty} |\hat{\alpha}_n(x) - G_n(x)| \xrightarrow{P} 0$, where $G_n(x)$ is defined in (2.2.1).

3.5 CONVERGENCE OF $\tilde{\alpha}_{n,m}(x)$

Theorem 3.1 (Burke and Gombay 1988)

Under conditions A1-A6 with F continuous, one can define two independent sequences of Brownian Bridges on [0,1], $\{B_n\}$ and $\{B_m^*\}$, such that

$$\sup_{-\infty<x<\infty} \left|\tilde{\alpha}_{n,m}(x) - G_{n,m}(x)\right| \xrightarrow{P} 0, \quad \text{as} \quad n \wedge m \longrightarrow \infty,$$

where $G_{n,m}$ is the Gaussian process in (3.2.3).

Proof:

First we consider

$$\tilde{\alpha}_{n,m}(x) = \sqrt{n}\left[F_n(x) - F(x,\tilde{\theta}_m)\right]$$

$$= \sqrt{n}\left[F_n(x) - F(x,\hat{\theta}_n)\right] - \sqrt{n}\left[F(x,\tilde{\theta}_m) - F(x,\hat{\theta}_n)\right]$$

$$(3.5.1) \qquad = \hat{\alpha}_n(x) - \sqrt{n}\left[F(x,\tilde{\theta}_m) - F(x,\hat{\theta}_n)\right].$$

Apply the one-term Taylor expansion of F with respect to $\hat{\theta}_n$ in (3.5.1). Thus (3.5.1) is

$$(3.5.2) \qquad = \hat{\alpha}_n(x) - \sqrt{n}\left[\tilde{\theta}_m - \hat{\theta}_n\right]\left[\nabla_\theta F(x,\delta_{n,m})\right]', \text{ where}$$

$$(3.5.3) \qquad \left\|\delta_{n,m} - \hat{\theta}_n\right\| \le \left\|\tilde{\theta}_m - \hat{\theta}_n\right\| .$$

From conditions A1-A6,

$$(3.5.4) \qquad \left\|\tilde{\theta}_m - \hat{\theta}_n\right\| \xrightarrow{P} 0 \quad \text{and} \quad \left\|\hat{\theta}_n - \theta_0\right\| \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty.$$

Hence,

$$\left\|\delta_{n,m} - \theta_0\right\| \le \left\|\delta_{n,m} - \hat{\theta}_n\right\| + \left\|\hat{\theta}_n - \theta_0\right\|$$

$$\le \left\|\tilde{\theta}_m - \hat{\theta}_n\right\| + \left\|\hat{\theta}_n - \theta_0\right\|, \text{ from (3.5.3).}$$

Due to (3.5.4) we have

$$\left\|\delta_{n,m} - \theta_0\right\| \xrightarrow{P} 0$$

and by A4,

$$\sup_{-\infty<x<\infty}\left\|\nabla_\theta\, F(x,\delta_{n,m}) - \nabla_\theta\, F(x,\theta_0)\right\| \xrightarrow{P} 0.$$

Therefore

$$(3.5.5) \qquad \tilde{\alpha}_{n,m}(x) = \hat{\alpha}_n(x) - \sqrt{n}\,\left[\tilde{\theta}_m - \hat{\theta}_n\right]\,\left[\nabla_\theta\, F(x,\theta_0)\right]'.$$

Next we consider $\sqrt{n}\,\left[\tilde{\theta}_m - \hat{\theta}_n\right]$. From (3.4.1) we know that

$$\left\|\sqrt{n}\,\left[\tilde{\theta}_m - \hat{\theta}_n\right] - \frac{\sqrt{n}}{\sqrt{m}}\int \ell(x,\theta_0)\; d\,\sqrt{m}\,\left[F_m^*(x) - F_n(x)\right]\right\| \xrightarrow{P} 0.$$

Using Lemma 3.1, we obtain

$$\left\|\sqrt{n}\,\left[\tilde{\theta}_m - \hat{\theta}_n\right]-(nm^{-1})^{1/2}\int \ell(x,\theta_0)\; d\,\sqrt{m}\,\left[E_m\left[F_n(x)\right]-F_n(x)\right]\right\| \xrightarrow{P} 0.$$

$$\left\|\sqrt{n}\,\left[\tilde{\theta}_m - \hat{\theta}_n\right]-(nm^{-1})^{1/2}\int \ell(x,\theta_0)\; d\,\sqrt{m}\,\left[E_m'\left[F_n'(x)\right]-F_n'(x)\right]\right\| \xrightarrow{P} 0,$$

due to (3.4.2) (ii).

$$\left\| \sqrt{n} \left[\tilde{\theta}_m - \hat{\theta}_n\right] - (nm^{-1})^{1/2} \int \ell(x,\theta_0)\ d\ \alpha_m^*(x) \right\| \xrightarrow{P} 0$$

follows from (3.4.2) (iv). We also note that

$$(nm^{-1})^{1/2} \int \ell(x,\theta_0)\ d\ \alpha_m^*(x)$$

$$= (nm^{-1})^{1/2} \int \ell(x,\theta_0)\ d\ B_m^*\left[F(x,\theta_0)\right]$$

$$+ (nm^{-1})^{1/2} \int \ell(x,\theta_0)\ d\left[\alpha_m^*(x) - B_m^*\left[F(x,\theta_0)\right]\right].$$

As in Lemma 2.1 and (3.4.2) (iv) we have

$$(nm^{-1})^{1/2} \int \ell(x,\theta_0)\ d\left[\alpha_m^*(x) - B_m^*\left[F(x,\theta_0)\right]\right] \xrightarrow{P} 0.$$

Thus

$$\text{(3.5.6)} \qquad \left\| \sqrt{n} \left[\tilde{\theta}_m - \hat{\theta}_n\right] - (nm^{-1})^{1/2} \int \ell(x,\theta_0)\ d\ B_m^*\ \left[F(x,\theta_0)\right] \right\| \xrightarrow{P} 0.$$

(3.5.5) could now be written as

$$\tilde{\alpha}_{n,m}(x) = \hat{\alpha}_n(x) - (nm^{-1})^{1/2} \int \ell(x,\theta_0)\ d\ B_m^*\ \left[F(x,\theta_0)\right] \cdot \left[\nabla_\theta\ F(x,\theta_0)\right]'.$$

From (3.4.2) (v), it follows that

$$\sup_{-\infty<x<\infty} \left| \tilde{\alpha}_{n,m}(x) - \left[B_n\left[F(x,\theta_0)\right] - \int \ell(x,\theta_0)\ d\ B_n\ \left[F(x,\theta_0)\right] \cdot \left[\nabla_\theta\ F(x,\theta_0)\right]'\right.\right.$$

$$- (nm^{-1})^{1/2} \int \ell(x,\theta_0) \, d\, B_m^* \left[F(x,\theta_0)\right] \cdot \left[\nabla_\theta \, F(x,\theta_0)\right]' \Big| \xrightarrow{P} 0$$

or

$$\sup \left| \tilde{\alpha}_{n,m}(x) - G_{n,m}(x) \right| \xrightarrow{P} 0.$$

Q.E.D.

As stated in (3.2.4), the covariance function of $G_{n,n}(x)$ is

$$E \, G_{n,n}(x) \, G_{n,n}(y) = F(x,\theta_0) \wedge F(y,\theta_0) - F(x,\theta_0) \, F(y,\theta_0).$$

To prove this consider

$$G_{n,n}(x) = B_n\left[F(x,\theta_0)\right] - \int \ell(x,\theta_0) \, d\, B_n \left[F(x,\theta_0)\right] \cdot \left[\nabla_\theta \, F(x,\theta_0)\right]'$$

$$- \int \ell(x,\theta_0) \, d\, B_n^* \left[F(x,\theta_0)\right] \cdot \left[\nabla_\theta \, F(x,\theta_0)\right]'$$

$$= G_n(x) - \int \ell(x,\theta_0) \, d\, B_n^* \left[F(x,\theta_0)\right] \cdot \left[\nabla_\theta \, F(x,\theta_0)\right]',$$

where B_n^* is independent of B_n and hence of G_n. Also, $\ell(x,\theta_0)$ is defined as in (2.5.1). Now,

$$G_{n,n}(x)G_{n,n}(y) = G_n(x)G_n(y) - G_n(y)\int \ell(x,\theta_0) dB_n^*\left[F(x,\theta_0)\right] \cdot \left[\nabla_\theta \, F(x,\theta_0)\right]'$$

$$- G_n(x)\int \ell(y,\theta_0) dB_n^*\left[F(y,\theta_0)\right] \cdot \left[\nabla_\theta \, F(y,\theta_0)\right]'$$

$$+ \int \ell(z) \, d\, B_n^*\left[F(z)\right] \cdot \left[\nabla_\theta \, F(x,\theta_0)\right]'$$

$$\cdot \int \ell(z)\, d\, B_n^*\left[F(z)\right] \cdot \left[\nabla_\theta F(y,\theta_0)\right]'$$

$$= G_1 - G_2 - G_3 + G_4.$$

$$E\, G_{n,n}(x) G_{n,n}(y) = EG_1 - EG_2 - EG_3 + EG_4.$$

In the maximum likelihood case, applying (2.5.6), we have

$$\text{(3.5.7)} \qquad EG_1 = F(x,\theta_0) \wedge F(y,\theta_0) - F(x,\theta_0)\, F(y,\theta_0) - \left[\nabla_\theta F(x,\theta_0)\right] \cdot I^{-1}(\theta_0)\left[\nabla_\theta F(y,\theta_0)\right]'.$$

Since B_n^* is independent of G_n and the fact that $E\, G_n = E\, B_n = 0$,

$$\text{(3.5.8)} \qquad E\, G_2 = E\, G_3 = 0.$$

From (2.5.5) we have

$$\text{(3.5.9)} \qquad E\, G_4 = \left[\nabla_\theta F(x,\theta_0)\right] \cdot I^{-1}(\theta_0)\left[\nabla_\theta F(y,\theta_0)\right]'.$$

Combining (3.5.7), (3.5.8) and (3.5.9) we get the desired covariance function.

3.6 HALF-SAMPLE METHOD

Suppose we have a sample of independent observations $X_1,\ldots,X_n$ from a continuous distribution function $F(x,\theta_0)$, where $\theta_0 = (\theta_1,\ldots,\theta_p)$. Let $\{\theta_m^*\}$ be a sequence of maximum likelihood estimators of the unknown parameter θ_0, derived without replacement from the half-sample $X_1', X_2', \ldots, X_m'$, where $n = 2m$. Consider the estimated empirical process

(3.6.1) $$\alpha_n^*(x) = \sqrt{n}\left[F_n(x) - F(x,\theta_m^*)\right] \cdot$$

As suggested by Durbin (1976), this procedure converges asymptotically to a Brownian Bridge, that is,

(3.6.2) $$\alpha_n^*(x) \xrightarrow{\mathscr{D}} B\left[F(x,\theta_0)\right].$$

In showing (3.6.2), a slight modification from the proof of Durbin (1976) will be introduced. We will be using the methodology employed in Chapter II. Hence, we assume the same set of conditions for the estimator sequence except for condition A1. In this case, we assume that:

(3.6.3) A1. $$\sqrt{m}\left[\theta_m^* - \theta_0\right] = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \ell(X_j', \theta_0)) + \epsilon_{1n}, \text{ where } \epsilon_{1n} \xrightarrow{P} 0.$$

When $n = 2m$, we have

(3.6.4) $$\sqrt{n}\left[\theta_m^* - \theta_0\right] = \frac{2}{\sqrt{n}} \sum_{j=1}^{m} \ell(X_j', \theta_0) + \epsilon_{1n}.$$

From Chapter 2, we know that this is asymptotically equal to

(3.6.5) $$2 \int \ell(x,\theta_0)\, d\, \sqrt{n}\, F_m(x).$$

Also, we define the function $d_j(x)$ as

(3.6.6) $$d_j(x) = \begin{cases} 1 - F(x,\theta_0) & \text{if} \quad X_j \le x,\ x \in R \\ -F(x,\theta_0) & \text{otherwise.} \end{cases}$$

Then the empirical process $\alpha_n(x)$ can be represented as

(3.6.7) $$\alpha_n(x) = \alpha_n^1(x) + \alpha_n^2(x),\ x \in R,$$

where

(3.6.8) $$\alpha_n^1(x) = \sum_{j=1}^{m} \frac{d_j(x)}{\sqrt{n}}$$

(3.6.9) and $$\alpha_n^2(x) = \sum_{j=m+1}^{n} \frac{d_j(x)}{\sqrt{n}} .$$

Obviously (3.6.8) and (3.6.9) are independent of each other. We note that

$$E\left[d_j(x)\right] = \left[1 - F(x,\theta_0)\right]\left[F(x,\theta_0)\right] + \left[- F(x,\theta_0)\right]\left[1 - F(x,\theta_0)\right]$$

$$= 0$$

and $$E\left[d_j(x)\right]^2 = \left[1 - F(x,\theta_0)\right]^2\left[F(x,\theta_0)\right] + \left[- F(x,\theta_0)\right]^2\left[1 - F(x,\theta_0)\right]$$

$$= F(x,\theta_0)\left[1 - F(x,\theta_0)\right].$$

Thus

$$E\left[\alpha_n^1(x)\right] = E\left[\alpha_n^2(x)\right] = 0$$

and $$E\left[\alpha_n^1(x)\right]^2 = E\left[\alpha_n^2(x)\right]^2 = \frac{m}{n}\left[F(x,\theta_0)\right]\left[1 - F(x,\theta_0)\right].$$

Therefore, $$E\left[\alpha_n^1(x) + \alpha_n^2(x)\right] = 0$$

and $$E\left[\alpha_n^1(x) + \alpha_n^2(x)\right]^2 = \left[F(x,\theta_0)\right]\left[1 - F(x,\theta_0)\right],$$

which are the moments of $\alpha_n(x)$.

From Theorem 1.2, we know that

$$\sup_{-\infty<x<\infty} \left|\alpha_n(x) - B_n\left[F(x,\theta_0)\right]\right| \xrightarrow{P} 0.$$

We claim that there exists two independent Brownian Bridges $B_n^1\left[F(x,\theta_0)\right]$ and $B_n^2\left[F(x,\theta_0)\right]$ such that

(i) $$B_n\left[F(x,\theta_0)\right] = B_n^1\left[F(x,\theta_0)\right] + B_n^2\left[F(x,\theta_0)\right]$$

(3.6.10) (ii) $$\sup_{-\infty<x<\infty} \left|\alpha_n^1(x) - B_n^1\left[F(x,\theta_0)\right]\right| \xrightarrow{P} 0$$

(iii) $$\sup_{-\infty<x<\infty} \left|\alpha_n^2(x) - B_n^2\left[F(x,\theta_0)\right]\right| \xrightarrow{P} 0.$$

Without going into the details, we know from Chapter II that

$$\begin{aligned}\alpha_n^*(x) &= \sqrt{n}\left[F_n(x) - F(x,\theta_m^*)\right] \\ &= \sqrt{n}\left[F_n(x) - F(x,\theta_0)\right] - \sqrt{n}\left[F(x,\theta_m^*) - F(x,\theta_0)\right] \\ &= \alpha_n(x) - \sqrt{n}\left[\theta_m^* - \theta_0\right]\left[\nabla_\theta F(x,\theta_0)\right]' + \epsilon_n, \text{ where } \epsilon_n \xrightarrow{P} 0.\end{aligned}$$

As in the proof of Theorem 2.1 and by (3.6.5), $\alpha_n^*(x)$ is asymptotically equal to

$$G_n^*(x) = B_n\left[F(x,\theta_0)\right] - 2\int \ell(z,\theta_0)\, d\, B_n^1\left[F(z,\theta_0)\right]\left[\nabla_\theta\, F(x,\theta_0)\right]'.$$

Due to (3.6.10) (i) we have

$$G_n^*(x) = B_n^1\left[F(x,\theta_0)\right] + B_n^2\left[F(x,\theta_0)\right] - 2\int \ell(z,\theta_0)\, dB_n^1\left[F(z,\theta_0)\right] \cdot \left[\nabla_\theta\, F(x,\theta_0)\right]'.$$

Consider

$$\begin{aligned} G_n^*(x)G_n^*(y) = {} & B_n\left[F(x,\theta_0)\right]B_n\left[F(y,\theta_0)\right] - 2B_n^1\left[F(y,\theta_0)\right] \\ & \cdot \int \ell(z,\theta_0)\, dB_n^1\left[F(z,\theta_0)\right]\cdot\left[\nabla_\theta\, F(x,\theta_0)\right]' \\ & - 2\, B_n^2\left[F(y,\theta_0)\right]\int \ell(z,\theta_0)\, dB_n^1\left[F(z,\theta_0)\right]\cdot\left[\nabla_\theta\, F(x,\theta_0)\right]' \\ & - 2\, B_n^1\left[F(x,\theta_0)\right]\int \ell(z,\theta_0)\, dB_n^1\left[F(y,\theta_0)\right]\cdot\left[\nabla_\theta\, F(y,\theta_0)\right]' \\ & - 2\, B_n^2\left[F(x,\theta_0)\right]\int \ell(z,\theta_0)\, dB_n^1\left[F(z,\theta_0)\right]\cdot\left[\nabla_\theta\, F(y,\theta_0)\right]' \\ & + 4\int \ell(z,\theta_0)\, dB_n^1\left[F(z,\theta_0)\right]\cdot\left[\nabla_\theta\, F(x,\theta_0)\right]' \\ & \cdot \int \ell(z,\theta_0)\, dB_n^1\left[F(z,\theta_0)\right]\cdot\left[\nabla_\theta\, F(y,\theta_0)\right]' \\ = {} & G_1 - G_2 - G_3 - G_4 - G_5 + G_6 \end{aligned}$$

Since B_n^1 and B_n^2 are independent and $E\left[B_n^i(\cdot)\right] = 0$, $i = 1,2$, we have

(3.6.11) $$EG_3 = EG_5 = 0.$$

From Section 2.5

(3.6.12) $$EG_2 = EG_4 = 2\left[\nabla_\theta\ F(x,\theta_0)\right]\ I^{-1}(\theta_0)\left[\nabla_\theta\ F(y,\theta_0)\right]'$$

(3.6.13) $$EG_6 = 4\ \left[\nabla_\theta\ F(x,\theta_0)\right]\ I^{-1}(\theta_0)\left[\nabla_\theta\ F(y,\theta_0)\right]'.$$

From the covariance of a Brownian Bridge, we have

(3.6.14) $$EG_1 = F(x,\theta_0)\ \Lambda\ F(y,\theta_0) - F(x,\theta_0)\ F(y,\theta_0).$$

Combining results (3.6.11) to (3.6.14), we obtain

$$EG_n^*(x)\ G_n^*(y) = F(x,\theta_0)\ \Lambda\ F(y,\theta_0) - F(x,\theta_0)\ F(y,\theta_0).$$

The mean of $G_n^*(x)$ is obviously 0. These are the moments of a Brownian Bridge. Hence we conclude that

$$\alpha_n^*(x) \xrightarrow{\mathcal{D}} B\left[F(x,\theta_0)\right].$$

3.7 RANDOM SUBSTITUTION METHOD

The object of this technique is to transform the hypothesis (2.6.1) into simple hypothesis. This is done by replacing the maximum likelihood estimator $\hat{\theta}$ of the unknown parameter θ_0 by a corresponding estimator, external to the sample, of a known value of θ.

First we will consider the finite-sample case. Under the composite hypothesis (2.6.1), suppose a sufficient statistic T_1 for θ exists and another statistic T_2 exists with the following property:

(i) T_2 is distributed independently from T_1.

(ii) The distribution of T_2 does not depend on θ. This is true since T_1 is sufficient.

(iii) The transformation

$$\tau: \quad X_1, X_2, \ldots, X_n \longrightarrow T_1, T_2$$

has a unique inverse τ^{-1} such that given the values of T_1 and T_2,

$$\tau^{-1}: \quad T_1, T_2 \longrightarrow X_1, X_2, \ldots, X_n.$$

Suppose T_1 is known to have distribution function $G(T_1, \theta)$. Let θ^* be an arbitrarily selected value of θ and let T_1^* be a random vector from the distribution function $G(T_1^*, \theta^*)$. The independent random variables $X_1^*, \ldots, X_n^*$ are generated by the inverse transformation

$$\tau^{-1}: \quad T_1^*, T_2 \longrightarrow X_1^*, X_2^*, \ldots, X_n^*.$$

The i.i.d.r.v. have a known distribution function $F(x;\ \beta_0, \theta^*)$. Thus the composite hypothesis with unknown θ, based on $X_1, X_2, \ldots, X_n$, may therefore be replaced by the simple hypothesis

$$H_0^*: \quad (\beta, \theta) = (\beta_0, \theta^*) = \delta^*$$

based on $X_1^*, X_2^*, \ldots, X_n^*$. Statistics based on the sample process

$$\alpha_n^*(x) = \sqrt{n}\ \left[F_n(x) - F(x; \beta_0, \theta^*)\right], \qquad x \in R,$$

could now be applied to test this hypothesis.

Next we consider from a heuristic point of view an asymptotic form of this method. Again we use the notations and assumptions for the estimator sequence listed in Chapter II, under the maximum likelihood estimator case. In addition, we have a strong assumption:

A6. $\left[\nabla_\theta F(x,\delta_0)\right] I^{-1}(\delta_0)\left[\nabla_\theta F(y,\delta_0)\right]'$ is independent of θ.

<u>REMARK</u>. The author, Durbin (1976), claims that this assumption holds under general conditions which he failed to formulate in satisfactory form.

Recall from Chapter II that asymptotically

$$\hat{\alpha}_n(x) = \alpha_n(x) - \sqrt{n}\left[\hat{\theta}-\theta_0\right]\left[\nabla_\theta F(x,\delta_0)\right]' + \epsilon_n \tag{3.7.1}$$

$$= B_n\left[F(x,\delta_0)\right] - \int \ell(x,\delta_0)\, dB_n\left[F(x,\delta_0)\right]\cdot\left[\nabla_\theta F(x,\delta_0)\right]' + \epsilon_n'$$

where $\epsilon_n, \epsilon_n' \xrightarrow{P} 0$. This representation shows that $\hat{\alpha}_n(x)$ and $\hat{\theta}$ are uncorrelated and hence independent asymptotically.
We therefore take

$$T_1 = \hat{\theta} \quad \text{and} \quad T_2 = \hat{\alpha}_n(x).$$

The observation $X_1, X_2, \ldots, X_n$ are mathematically equivalent to the sample process $\alpha_n(x)$, $x \in R$. We denote the transformation

$$\tau_n:\ \alpha_n(x) \longrightarrow \hat{\theta},\ \hat{\alpha}_n(x).$$

The inverse transformation is

$$\tau_n^{-1}: \quad \hat{\theta},\ \hat{\alpha}_n(x) \longrightarrow \alpha_n(x).$$

From (3.7.1) this has the form

$$\alpha_n(x) = \hat{\alpha}_n(x) + \sqrt{n}\left[\hat{\theta}-\theta_0\right]\left[\nabla_\theta F(x,\delta_0)\right]' - \epsilon_n.$$

Suppose $\hat{\theta}$ has distribution function $G(\hat{\theta},\theta)$ and let θ^* be an arbitrary selected value of θ. Let $\hat{\theta}^*$ be a random vector, independent of $X_1,\ldots,X_n$, from the distribution function $G(\hat{\theta}^*,\theta^*)$. As in the finite-sample case, $X_1^*,\ldots,X_n^*$ are equivalent to $\alpha_n^*(x)$ where

$$\alpha_n^*(x) = \hat{\alpha}_n(x) + \sqrt{n}\left[\hat{\theta}^*-\theta^*\right]\left[\nabla_\theta F(x,\delta^*)\right]' - \epsilon_n^*,$$

$\epsilon_n^* \xrightarrow{P} 0$ and $\nabla_\theta F(x,\delta^*) = \left[\dfrac{\partial F(x,\beta_0,\theta)}{\partial\theta_1},\ldots,\dfrac{\partial F(x,\beta_0,\theta)}{\partial\theta_{p_2}}\right]$, evaluated at (β_0,θ^*). Recall from Chapter II that under the maximum likelihood case

(i) $\quad E\left[\sqrt{n}\,[\hat{\theta}^*-\theta^*]\right] = 0$

(ii) $\quad nE\left[[\hat{\theta}^*-\theta^*]'[\hat{\theta}^*-\theta^*]\right] = I^{*-1}(\delta^*)$ where

(3.7.2)
$$I^*(\delta^*) = E\left[\nabla_\theta \log f(x,\delta^*)\right]'\left[\nabla_\theta \log f(x,\delta^*)\right]$$

(iii) $\quad nE\left[\left[\nabla_\theta F(x,\delta^*)\right]\cdot[\hat{\theta}^*-\theta^*]'\,[\hat{\theta}^*-\theta^*]\cdot\left[\nabla_\theta F(y,\delta^*)\right]'\right]$

$$= \left[\nabla_\theta F(x,\delta^*)\right]\cdot I^{*-1}(\delta^*)\left[\nabla_\theta F(y,\delta^*)\right]'$$

$$= \left[\nabla_\theta F(x,\delta_0)\right]\cdot I^{-1}(\delta_0)\left[\nabla_\theta F(y,\delta_0)\right]'.$$

The last equality is true since by assumption A7, this is independent of θ.

We note that $E\,\alpha_n^*(x) = 0$ since $E\,\hat{\alpha}_n(x) = 0$ and by (3.7.2)(i). To find the covariance function of $\alpha_n^*(x)$, consider

$$\alpha_n^*(x)\,\alpha_n^*(y) = \hat{\alpha}_n(x)\left[\hat{\alpha}_n(y) + \sqrt{n}\,[\hat{\theta}^*-\theta^*]\left[\nabla_\theta F(y,\delta^*)\right]'\right] + \sqrt{n}\,[\hat{\theta}^*-\theta^*]$$

$$\cdot\left[\nabla_\theta F(x,\delta^*)\right]'\cdot\left[\hat{\alpha}_n(y)+\sqrt{n}\,(\hat{\theta}^*-\theta^*)\cdot\left[\nabla_\theta F(y,\delta^*)\right]'\right]+\epsilon_n,\quad \epsilon_n \xrightarrow{P} 0.$$

$$= \hat{\alpha}_n(x)\,\hat{\alpha}_n(y) + \sqrt{n}\,\hat{\alpha}_n(x)[\hat{\theta}^*-\theta^*]\left[\nabla_\theta F(y,\delta^*)\right]'$$

$$+ \sqrt{n}\,[\hat{\theta}^*-\theta^*]\left[\nabla_\theta F(x,\delta^*)\right]'\,\hat{\alpha}_n(y)$$

$$+ n\,[\hat{\theta}^*-\theta^*]\left[\nabla_\theta F(x,\delta^*)\right]'\,[\hat{\theta}^*-\theta^*]\left[\nabla_\theta F(y,\delta^*)\right]'$$

$$= L_1 + L_2 + L_3 + L_4.$$

From Section 2.5 and Theorem 2.1

$$EL_1 = F(x,\delta_0) \wedge F(y,\delta_0) - F(x,\delta_0)\,F(y,\delta_0)$$

$$- \left[\nabla_\theta F(x,\delta_0)\right]\cdot I^{-1}(\delta_0)\left[\nabla_\theta F(y,\delta_0)\right]'.$$

From the fact that $\hat{\theta}^*$ is independent of $\hat{\theta}$ which is independent of $\hat{\alpha}_n(x)$ and $E\,\hat{\alpha}_n(x) = 0$, we have

$$E\,L_2 = E\,L_3 = 0.$$

From (3.7.2) (iii),

$$E\, L_4 = \left[\nabla_\theta\, F(x,\delta_0)\right]\cdot I^{-1}(\delta_0)\left[\nabla_\theta\, F(y,\delta_0)\right]'.$$

Combining these results we get the covariance function

$$E\, \alpha_n^*(x)\, \alpha_n^*(y) = F(x,\delta_0) \wedge F(y,\delta_0) - F(x,\delta_0)\, F(y,\delta_0),$$

which we know is the covariance of a Brownian Bridge. Hence we say that $\alpha_n^*(x)$ has the same limiting distribution as $\alpha_n(x)$.

For practical reasons, this method is not preferable due to the cumbersome computation to obtain $\hat{\theta}^*$. Moreso, it is not recommended because of its lack of power (Durbin, 1973b).

<u>REMARK</u>. We make note that the above methods are not the only distribution-free procedures. In fact, the well-known χ^2 test is not only used for testing completely specified hypothesis but also for testing composite hypothesis. This is due to the fact that under fairly general assumptions it is known how the probability distribution of χ^2 is approximately affected when parameters are estimated from the sample (loss of one degree of freedom for each parameter estimated). Moreover, it is based on the sample distribution function since for group boundaries $x_1,\ldots,x_{k-1}$ the chi-square statistic can also be represented as

$$\chi^2 = n \sum_{i=1}^{k} \frac{\left[F_n(x_i)-F_n(x_{i-1})-F(x_i)+F(x_{i-1})\right]^2}{F(x_i)-F(x_{i-1})},$$

where $x_0 = -\infty$ and $x_k = +\infty$. As $n \longrightarrow \infty$, it is approximately distribution-free.

CHAPTER IV

NUMERICAL RESULTS

4.1 INTRODUCTION

In this chapter some pertinent numerical and computational results are presented. The main purpose of the sampling experiment is to investigate how fast the following converge:

(4.1.1)

(i) for fixed $x \in R$,

$$\hat{\alpha}_n(x) = \sqrt{n}\left[F_n(x) - F(x,\hat{\theta}_n)\right] \xrightarrow{\mathcal{D}} B\left[F(x,\theta_0)\right]$$

(ii)
$$\sup_{-\infty<x<\infty} |\hat{\alpha}_n(x)| \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1} |B(t)|$$

(iii)
$$\int \hat{\alpha}_n(x)^2 \, d\, F(x,\hat{\theta}_n) \xrightarrow{\mathcal{D}} \int_0^1 B(t)^2 \, dt$$

(iv)
$$\int \frac{\hat{\alpha}_n(x)^2}{F(x,\hat{\theta}_n)(1-F(x,\hat{\theta}_n))} \, d\, F(x,\hat{\theta}_n) \xrightarrow{\mathcal{D}} \int_0^1 \frac{B(t)^2\, dt}{t(1-t)},$$

where $\left\{\hat{\theta}_n\right\}$ is a sequence of maximum likelihood estimators of $\theta_0 = (\theta_1,\ldots,\theta_p)$ derived under the following procedures:

(4.1.2)

(i) Bootstrap Method

(ii) Half-Sample Method.

We also wish to verify that the limiting behaviour of $\hat{\alpha}_n(x)$ is the same as the limiting behaviour of $\alpha_n(x)$. The same is true for its related functionals.

Estimates of large sample percentage points of the statistics of interest are also tabulated for commonly used significance levels. These are used in application to test the null hypotheses

(4.1.3) $$H_0: \quad F \in \mathcal{F} = \left\{F(x,\theta);\ x \in R,\ \theta \in S \subset R^P\right\}.$$

4.2 PROCEDURE

The experiment was conducted by investigating 4 sample sizes $n = 50, 100, 150, 200$ and 2 distribution functions: the normal distribution with mean $\mu = 1.0$ and variance $\sigma^2 = 1.0$ and the gamma distribution with parameter $\alpha = 4.0$ and scale parameter $\beta = 2.0$. Slight modifications in the computer program are necessary should one want to test for various sample sizes, parameters or continuous distribution functions. As stipulated in Chapter III, the maximum likelihood estimators of the parameters must satisfy conditions A1-A5.

As in the normal distribution $N(\mu,\sigma^2)$ with density function

$$f(x) = \frac{e^{-\frac{1}{2}\left[\frac{x-u}{\sigma}\right]^2}}{\sqrt{2\pi}\,\sigma}, \quad x \in R,$$

the maximum likelihood estimators of μ and σ^2 are $\bar{x}$ and s^2 respectively, where

$$\bar{x} = \sum_{i=1}^{n} \frac{X_i}{n}$$

(4.2.1) and

$$s^2 = \sum_{i=1}^{n} \frac{\left[X_i - \bar{x}\right]^2}{n} .$$

These estimators obviously satisfy the given conditions by letting

$$\ell_1(X_j, \theta_0) = X_j - \mu$$

and

$$\ell_2(X_j, \theta_0) = (X_j - \mu)^2 - \sigma^2 .$$

Thus for $p = 2$, the $1 \times p$ dimensional vector

$$\sqrt{n}\left[\hat{\theta}_n - \theta_0\right] = \sqrt{n}\,[\bar{x} - \mu,\; s^2 - \sigma^2]$$

$$= \left[\sum_{i=1}^{n} \frac{\left[X_i - \mu\right]}{\sqrt{n}},\; \sum_{i=1}^{n} \frac{\left[X_i - \mu\right]^2 - \sigma^2}{\sqrt{n}} + \epsilon_n\right].$$

Similarly, the gamma distribution $G(\alpha, \beta)$ with density function

$$f(x) = \frac{e^{-\frac{x}{\beta}}\, x^{\alpha - 1}}{\Gamma(\alpha)\, \beta^{\alpha}}, \quad x > 0,\ \alpha > 0,\ \beta > 0,$$

has maximum likelihood estimators

$$\hat{\alpha} = \frac{\bar{x}^2}{s^2} \quad \text{and} \quad \hat{\beta} = \frac{s^2}{\bar{x}} \quad \text{with}$$

$$\ell_1(X_j, \theta_0) = \frac{2}{\beta}\left[X_j - \alpha\beta\right] - \frac{1}{\beta^2}\left[\left[X_j - \alpha\beta\right]^2 - \alpha\beta^2\right]$$

$$\ell_2(X_j, \theta_0) = \frac{1}{\alpha\beta}\left[X_j - \alpha\beta\right]^2 - \frac{X_j}{\alpha} .$$

Figures 4.1 and 4.2 show the computer program for this experiment. The algorithm of the program runs as follows:

Step 1. Simulate a random sample of size n from a continuous distribution function $F(x,\theta_0)$. Call this the first-stage sample.

Step 2. For each procedure in (4.1.2), derive the second-stage sample. In the bootstrap method, this sample consists of n elements drawn with replacement from the first-stage sample. The half-sample method draws without replacement n/2 elements from the original sample.

Step 3. Using the second-stage sample, calculate the maximum likelihood estimators of θ_0.

Step 4. Calculate all the statistics given in (4.1.1).

For each sample size, procedure and distribution function, the above routine was executed 1000 times, thus generating 1000 values for each of the above mentioned statistics. As a check on the sampling experiment, the first-stage sample was used to compute for $\alpha_n(x)$ and its functions, where θ_0 is specified. In the normal case, $\theta_0 = [1.0, 1.0]$ while in the gamma case $\theta_0 = [4.0, 2.0]$. With this obtained data, the sampling distribution $F_n(x)$, as defined in (1.1.3), was computed for each statistics. The increment of x was arbitrarily chosen. The results are tabulated in Tables 4.1, 4.2, 4.4 and 4.6. The corresponding graph for this data when n = 200 are shown in Figures 4.3, 4.4, 4.5 and 4.6. The $F_i(x)$ column for each table is computed by the formula

$$F_i(x) = \frac{\text{no. of } X_j \le x}{1000}, \quad j = 1,\ldots,n$$
$$i = 1,2,$$

where i represents the method used in the estimation. The bootstrap method is represented by i = 1, and the half-sample method by i = 2 . i = 3 refers to the method whereby $\alpha_n(x)$ is involved under the specified distribution function.

An IGP (Interactive Graphing Package) printout of the graph, as shown in Figures 4.3, 4.4, 4.5 and 4.6, compares the behaviour of the statistics under the two procedures with that of the theoretical values. The horizontal axis corresponds to the x column in the table while the y axis corresponds to the $F_i(x)$ column.

Empirically derived percentage points are also tabulated as shown in Tables 4.3, 4.5 and 4.7. The $F_i^{-1}(p)$, i = 1,2, represents the pth percentile of the distribution of the statistics under method i.

In the tabulation of $\hat{\alpha}_n(x)$, we have fixed x at 2. From the theoretical results in Chapter III, the limiting distribution of this statistics under the hypothesis, say H_0: $F = G(4,2)$, converges to a normal process with mean 0 and variance $F(2,\theta_0)(1 - F(2,\theta_0))$ where

$$F(2,\theta_0) = \int_{-\infty}^{2} \frac{e^{-\frac{x}{2}} x^{4-1}}{\Gamma(4)\cdot 2^4}\, dx.$$

The z column in Table 4.1 represents the corresponding standard normal deviates and the F(z) column is obtained by using the normal table.

The K-S type of statistic was generated by calling the NKS1 package of the well-known IMSL (International Mathematical and Statistical Library). The generation of this statistic was based on Kolmogorov's (1933) recursion formulas. Inasmuch as this package is used to test that a random sample was drawn from a specified distribution, slight changes were introduced in the external subroutine for our purposes. In Table 4.2, the theoretical values in column F(x) are obtained from Smirnov (1948); rounded off to four decimal places.

As to the C-vM and A-D type of statistics, we have made the following versions for computational reasons:

$$\hat{W}_n^2 = \sum_{k=1}^{n} \left[\frac{k}{n} - F(x_{(k)},\hat{\theta}_n)\right]^2$$

$$\hat{A}_n^2 = \sum_{k=1}^{n} \frac{\left[\frac{k}{n} - F\left[x_{(k)},\hat{\theta}_n\right]\right]^2}{F(x_{(k)},\hat{\theta}_n)\left[1-F(x_{(k)},\hat{\theta}_n)\right]},$$

where $\frac{k}{n}$ follows from the definition of $F_n(x)$ in (1.1.3) and $x_{(k)}$ is the kth order statistics of the first-stage sample. Theoretical values of $F(x)$ in the C-vM case can be found in Anderson and Darling (1952). The same authors (1954) gave a short table of some percentage points for the A-D statistic when the levels of significance are .10, .05 and .01. In Figure 4.6, the "+"'s are derived from the values in column $F_3(x)$ of Table 4.6.

4.3 OBSERVATIONS AND CONCLUSIONS

It was found that the experimentally derived distribution function of the above mentioned statistics agreed well with the theoretical distribution function. The agreement was found to be fairly close to 2 decimal places for $\hat{\alpha}_n(x)$, K-S and C-vM type of statistics. As shown in the tables, even for $n = 50$, there seems to be a good agreement to 1 decimal place. Hence we say that the asymptotic value is reached very rapidly. The upper tail of the distribution appears to be the part which comes into agreement most rapidly with n. In application when n is large, it seems reasonable to use the last upper tail - the region that makes for the statistical tests.

From the graphs, there seems to be no significant difference between the 2 procedures, especially at the tails of the distribution. Some minor discrepancies occur between the tails. This is quite a surprising result because theoretically we expect the statistics based on the bootstrap method to converge faster than those under the half-sample method due to the fact that more randomization is introduced and the sample size used for estimating the parameter is larger.

In spite of the approximations made for the C-vM, it appears that its distribution behaves just as "nice" as the K-S. However, the A-D seems a little bit "off" from the theoretical values. It appears, in particular, that the experimental values are always greater than the exact ones. Perhaps this is due to the function $\left[F\left[x,\hat{\theta}_n\right]\left[1-F(x,\hat{\theta}_n)\right]\right]^{-1}$ which gives heavier weight at the tails of the distribution.

As to the effects of the distribution function, this author observed that there is no significant difference between the gamma and the normal case as shown in the graphs of the K-S and C-vM. This is in consonance with our learned theory that these statistics are distribution-free. However, it seems that the distribution of A-D has longer tail in the gamma case compared to the normal. This author cannot explain this phenomenon.

In testing the hypothesis (4.1.3), it is of considerable importance to determine what sample size, test statistics or method should be applied. Basing from these numerical results, it seems safe to use $n > 50$ and the bootstrap method. But if one is concerned about the computer time, then the half-sample method is preferable. This author cannot suggest as to which test statistic is superior than the other the fact that powers of these tests have to be considered. However, for practical purposes, the K-S is recommended since it is reasonably easy to evaluate especially with a computer. Without the machine, the following version can be used:

$$\sup_{1\le k\le n} \left|\frac{k}{n} - F(x_{(k)},\hat{\theta}_n)\right|.$$

So far this thesis has limited its scope of study to the asymptotic behaviour of the estimated empirical process and its related functionals for the univariate case. Of equal importance in hypothesis testing, one might consider the multivariate case and the power of the tests given in this paper.

The enormous manipulation of data in this research was made possible through the aid of some computer packages of the University of Calgary Academic Computing Services. Information and conventions on the routine used are available in the IMSL manual. Experimental data of this study is available in the files of the Mathematics Department.

```
c Computer Program for The Normal Case
c This is to show that the limiting distribution of the estimated
c empirical process, Kolmogorov-Smirnov, Cramer von Mises and
c Anderson-Darling statistics under the 1)Bootstrap Method
c 2) Durbin's Half-Sample Method, converges to the limiting distribution
c of these statistics under the specified hypothesis,
c H:F=Fo, where F is continuous.
      integer i,j,k,nexec,n,kount,isam2(1000),ix(1000)
      real sam1(1000),sam2(1000),ep(1000),ks(1000),cvm(1000),and(1000),
     &pdif(6),sum,var,rmean,x,fn,fnx,f1,f2,cm,ad,y
      double precision dseed
      external pdf
      common rmean,var
      open(27,file='nresearch.data',form ='formatted')
      parameter(aver=1.0,sig=1.0,nexec=1000 )
      do 200 n=50,200,50
      do 100 k=1,3,1
      do 1 j=1,nexec
      kount=0.0
      dseed=9.0+j*k+n
      sum=0.0
      var=0.0
c Take the first-stage sample ~N(1.0,1.0)
      call ggnml(dseed,n,sam1)
      do 2 k2=1,n
      sam1(k2)=sam1(k2)*sig + aver
    2 continue
c Order the sample
      call vsrta(sam1,n)
      dseed=8.0+j*k+n
      if (k .ne. 3) go to 16
      rmean=aver
      var=sig**2
      go to 12
   16 if (k .ne. 1) go to 8
c Bootstrap Method:Take the second-stage sample,it's mean and stdev.
      ng=n
      call ggud(dseed,n,n,isam2)
    9 do 3 k3=1,ng
      i=isam2(k3)
      sam2(k3)=sam1(i)
      sum=sum+sam2(k3)
    3 continue
      rmean=sum/float(ng)
      do 4 k4=1,ng
      var=var + ((sam2(k4)-rmean)**2)/float(ng)
    4 continue
      go to 12
c Durbin's Half-Sample Method
    8 ng=n/2
      iopt=0
      npop=n
      ip=0
      nsamp=ng
      call ggsrs(dseed,iopt,npop,ip,mpop,pop,nsamp,msamp,samp,isam2,ier)
      go to 9
c Calculate ep=sqrt(n)(fnx-f(x*)),where x is fixed
   12 x=2.0
      do 5 k5=1,n
```

Figure 4.1

Fortran Program for the Asymptotic Behaviour of the Estimated Empirical Process and other Related Statistics under a Normal Distribution

```
      if (sam1(k5) .gt. x) go to 6
      kount=kount+1
    5 continue
    6 fnx=float(kount)/float(n)
      x=(x-rmean)/sqrt(var)
      call mdnor(x,f1)
      ep(j)=sqrt(n)*(fnx-f1)
c Calculate ks=sup/fnx-f(x*)/
      call nks1(pdf,sam1,n,pdif,ier)
      ks(j)=pdif(4)
c Calculate cvm=summation(fnx-f(x*))**2
      ad=0.0
      cm=0.0
      do 7 k7=1,n
      fn=float(k7)/float(n)
      y=(sam1(k7)-rmean)/sqrt(var)
      call mdnor(y,f2)
      cm=cm+(fn-f2)**2
c Calculate and=summation(cramer/f(x*)(1-f(x*))
      ad= ad+(fn-f2)**2/(f2*(1.0-f2))
    7 continue
      cvm(j)=cm
      and(j)=ad
    1 continue
      call vsrta(ep,nexec)
      call vsrta(ks,nexec)
      call vsrta(cvm,nexec)
      call vsrta(and,nexec)
      if (k .ne. 1) go to 18
      write(27,10) n
   10 format(20x,23hBOOTSTRAP METHOD FOR N=,i3)
      go to 25
   18 if (k .ne. 2) go to 19
      write (27,20) n
   20 format(20x,34hDURBIN'S HALF-SAMPLE METHOD FOR N=,i3)
      go to 25
   19 write(27,40)n
   40 format(20x,37hUNDER THE SPECIFIED HYPOTHESIS FOR N=,i3)
   25 do 15 k15=1,nexec
      write (27,30)k15,ep(k15),ks(k15),cvm(k15),and(k15)
   30 format(5x,i5,5x,f9.5,5x,f9.5,5x,f9.5,5x,f9.5,5x,f9.5)
   15 continue
  100 continue
  200 continue
      end
      subroutine pdf(x,f)
      real x,f,t
      common rmean,var
      t=(x-rmean)/sqrt(var)
      f=.5*erfc(-.7071068*t)
      return
      end
```

Figure 4.1 (continued)

```
c Computer Program for the Gamma Case
c This is to show that the limiting distribution of the estimated
c empirical process,Kolmogorov-Smirnov, Cramer-von Mises and
c Anderson-Darling statistics under the 1)Bootstrap Method
c 2) Durbin's Half-Sample Method, converges to the limiting distribution
c of these stattistics under the specified hypothesis,
c H:F=Fo, where F is continuous.
      integer i,j,k,nexec,n,kount,isam2(1000)
      real sam1(1000),sam2(1000),ep(1000),ks(1000),cvm(1000),and(1000),
     &pdif(6),sum,var,rmean,x,fn,fnx,f1,f2,cm,ad,y,a,b,alpha,beta,wk(1)
     &,am1(1000)
      double precision dseed
      external pdf
      common beta,alpha
      open(28,file='gresearch.data',form='formatted')
      parameter(a=4.0,b=2.0,nexec=1000)
      do 200 n=50,200,50
      do 100 k=1,3,1
      do 1 j=1,nexec
      kount=0.0
      dseed=9.0+j*k+n
      sum=0.0
      var=0.0
c Take the first-stage sample ~G(4.0,2.0)
      call ggamr(dseed,a,n,wk,sam1)
      do 2 k2=1,n
      sam1(k2)=sam1(k2)*b
    2 continue
c Order the sample
      call vsrta(sam1,n)
      dseed=8.0+j*k+n
      if (k .ne. 3) go to 16
      alpha=a
      beta=b
      go to 12
   16 if (k .ne. 1) go to 8
c Bootstrap Method: Take the second-stage sample, it's mean and stdev.
      ng=n
      call ggud(dseed,n,n,isam2)
    9 do 3 k3=1,ng
      i=isam2(k3)
      sam2(k3)=sam1(i)
      sum=sum+sam2(k3)
    3 continue
      rmean=sum/float(ng)
      do 4 k4=1,ng
      var=var + ((sam2(k4)-rmean)**2)/float (ng)
    4 continue
      alpha=rmean**2/var
      beta=var/rmean
      go to 12
c Durbin's Half-Sample Method
    8 ng=n/2
      iopt=0
      npop=n
      ip=0
      nsamp=ng
      call ggsrs(dseed,iopt,npop,ip,mpop,pop,nsamp,msamp,samp,isam2,ier)
      go to 9
```

Figure 4.2

Fortran Program for the Asymptotic Behaviour of the Estimated Empirical Process and other Related Statistics under a Gamma Distribution

```
c Calculate ep=sqrt(n)(fnx-f(x)*),where x is fixed
   12 x=2.0
      do 5 k5=1,n
      if (sam1(k5) .gt. x) go to 6
      kount=kount +1
    5 continue
    6 fnx=float(kount)/float(n)
      x=x/beta
      call mdgam(x,alpha,f1,ier)
      ep(j)=sqrt(n)*(fnx-f1)
c Calculate ks=sup/fnx-f(x*)/
      do 31 k17=1,n
      am1(k17)=sam1(k17)
   31 continue
      call nks1(pdf,am1,n,pdif,ier)
      ks(j)=pdif(4)
c Calculate cvm=summation(fnx-f(x*))**2
      ad=0.0
      cm=0.0
      do 7 k7=1,n
      fn=float(k7)/float(n)
      y=sam1(k7)/beta
      call mdgam(y,alpha,f2,ier)
      cm=cm+(fn-f2)**2
c Calculate and=summation (cramer)/(f(x*)(1-f(x*))
      ad=ad+(fn-f2)**2/(f2*(1.0-f2))
    7 continue
      cvm(j)=cm
      and(j)=ad
    1 continue
      call vsrta(ep,nexec)
      call vsrta(ks,nexec)
      call vsrta(cvm,nexec)
      call vsrta(and,nexec)
      if (k .ne. 1) go to 18
      write (28,10)n
   10 format(20x,23hBOOTSTRAP METHOD FOR N=,i3)
      go to 25
   18 if (k .ne. 2) go to 19
      write(28,20)n
   20 format (20x,34hDURBIN'S HALF-SAMPLE METHOD FOR N=,i3)
      go to 25
   19 write(28,40)n
   40 format(20x,37hUNDER THE SPECIFIED HYPOTHESIS FOR N=,i3)
   25 do 15 k15=1,nexec
      write(28,30)k15,ep(k15),ks(k15),cvm(k15),and(k15)
   30 format(5x,i5,5x,f9.5,5x,f9.5,5x,f9.5,5x,f9.5)
   15 continue
  100 continue
  200 continue
      end
      subroutine pdf(x,f)
      real x,f
      common beta,alpha
      x=x/beta
      call mdgam(x,alpha,f,ier)
      return
      end
```

Figure 4.2 (continued)

Table 4.1

A Comparison of the Sampling Distribution of the Estimated Empirical Process at x = 2 Using the Methods in (4.1.2)

NORMAL DISTRIBUTION									GAMMA DISTRIBUTION								
		true value	n = 50			n = 100					true value	n = 50			n = 100		
X	Z	F(z)	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	X	Z	F(x)	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$
-1.3	-3.56	.0000	.000	.000	.001	.000	.000	.001	-0.65	-4.65	.0000	.007	.001	.000	.002	.000	.000
-1.1	-3.01	.0013	.002	.000	.002	.000	.001	.001	-0.55	-3.94	.0000	.011	.006	.000	.007	.004	.000
-0.9	-2.46	.0069	.005	.005	.008	.005	.005	.005	-0.45	-3.22	.0006	.025	.014	.000	.015	.013	.000
-0.7	-1.92	.0274	.015	.022	.049	.023	.021	.033	-0.35	-2.51	.0060	.048	.052	.000	.039	.045	.000
-0.5	-1.37	.0853	.070	.075	.084	.059	.090	.088	-0.25	-1.79	.0367	.136	.126	.000	.108	.100	.000
-0.3	-0.82	.2061	.169	.198	.149	.180	.204	.210	-0.15	-1.07	.1423	.277	.261	.000	.242	.251	.163
-0.1	-0.27	.3936	.379	.392	.382	.370	.404	.410	-0.05	-0.36	.3594	.528	.510	.384	.457	.485	.439
0.1	0.27	.6064	.598	.613	.525	.595	.647	.627	0.05	0.36	.6406	.769	.744	.774	.709	.733	.717
0.3	0.82	.7939	.804	.828	.803	.810	.836	.807	0.15	1.07	.8577	.913	.899	.934	.873	.884	.882
0.5	1.37	.9147	.925	.927	.918	.918	.924	.928	0.25	1.79	.9633	.970	.966	.934	.958	.959	.965
0.7	1.92	.9746	.974	.972	.994	.970	.968	.984	0.35	2.51	.9940	.988	.989	.988	.988	.992	.984
0.9	2.46	.9931	.991	.990	.998	.992	.991	.997	0.45	3.22	.9994	.999	.999	.999	.999	1.000	.994
1.1	3.01	.9987	.998	.997	1.000	.997	.997	1.000	0.55	3.94	1.0000	1.000	1.000	.999	1.000		1.000
1.3	3.56	1.0000	1.000	.999		.999	1.000		0.65	4.65				1.000			
1.5	4.11			1.000		1.000			0.75	5.37							

Table 4.1 (continued)

NORMAL DISTRIBUTION								
		true value	n = 150			n = 200		
X	Z	F(z)	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$
-1.3	-3.56	.0000	.000	.000	.000	.000	.001	.000
-1.1	-3.01	.0013	.001	.001	.000	.002	.001	.002
-0.9	-2.46	.0069	.006	.003	.004	.004	.004	.013
-0.7	-1.92	.0274	.014	.017	.022	.017	.020	.038
-0.5	-1.37	.0853	.016	.072	.085	.074	.072	.090
-0.3	-0.82	.2061	.191	.181	.174	.185	.191	.214
-0.1	-0.27	.3936	.377	.368	.342	.395	.394	.333
0.1	0.27	.6064	.605	.603	.604	.623	.616	.556
0.3	0.82	.7939	.806	.801	.772	.796	.805	.788
0.5	1.37	.9147	.929	.918	.918	.931	.922	.931
0.7	1.92	.9746	.974	.975	.973	.975	.974	.981
0.9	2.46	.9931	.992	.992	.997	.991	.994	.995
1.1	3.01	.9987	.998	.999	.999	.998	.998	1.000
1.3	3.56	1.0000	1.000	1.000	1.000	1.000	1.000	
1.5	4.11							

GAMMA DISTRIBUTION								
		true value	n = 150			n = 200		
X	Z	F(x)	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$
-0.65	-4.65	.0000	.001	.002	.000	.001	.000	.000
-0.55	-3.94	.0000	.004	.005	.000	.004	.004	.000
-0.45	-3.22	.0006	.010	.012	.000	.010	.010	.000
-0.35	-2.51	.0060	.032	.042	.000	.033	.036	.000
-0.25	-1.79	.0367	.089	.111	.000	.089	.090	.021
-0.15	-1.07	.1423	.228	.264	.216	.208	.238	.106
-0.05	-0.36	.3594	.436	.495	.417	.401	.454	.478
0.05	0.36	.6406	.681	.730	.694	.672	.705	.669
0.15	1.07	.8577	.857	.894	.840	.849	.873	.819
0.25	1.79	.9633	.951	.962	.942	.956	.964	.959
0.35	2.51	.9940	.990	.993	.992	.989	.988	.986
0.45	3.22	.9994	.999	.997	.997	.998	.998	.996
0.55	3.94	1.0000	1.000	1.000	1.000	.999	1.000	.998
0.65	4.65					1.000		.999
0.75	5.37							1.000

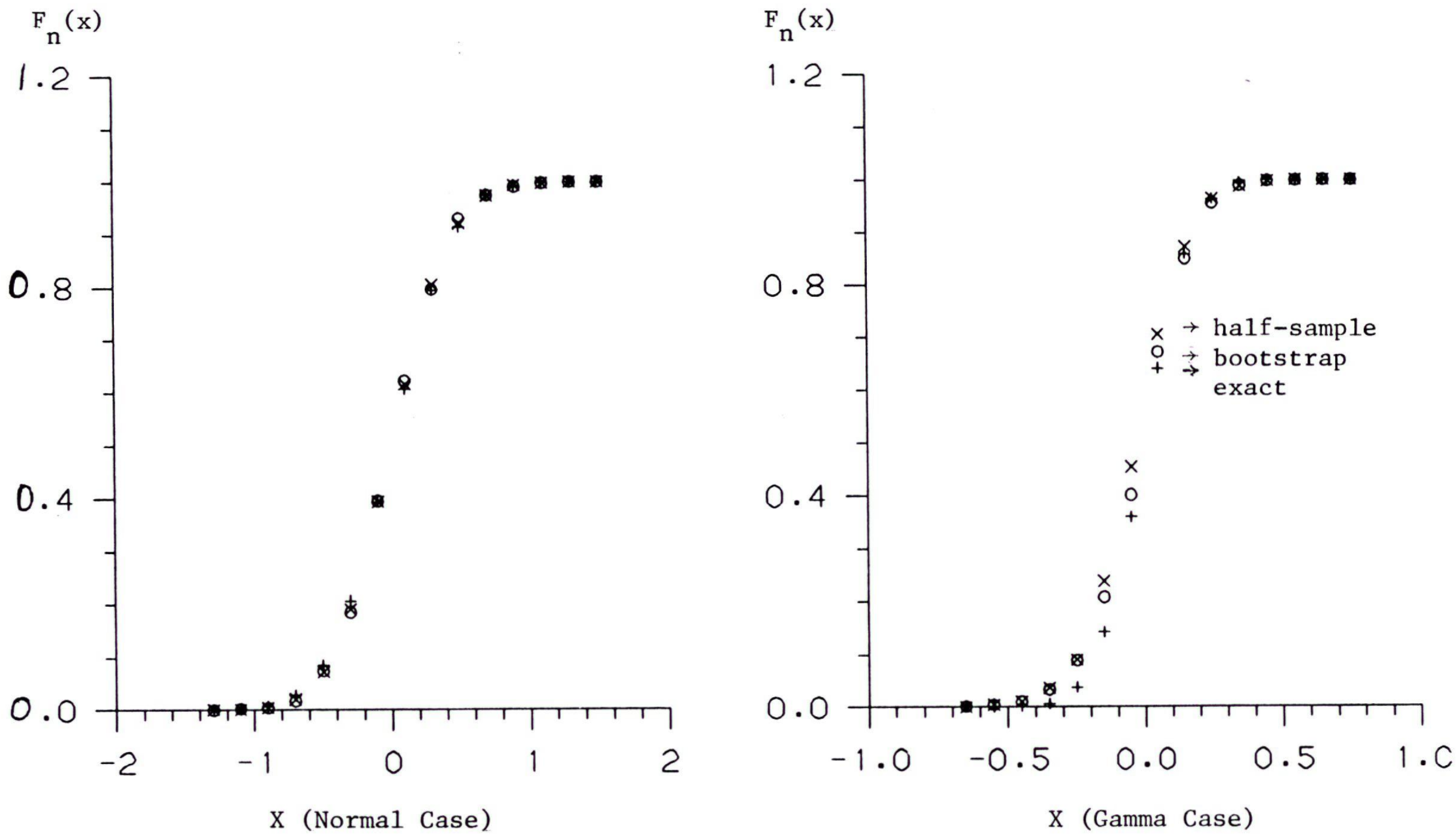

Figure 4.3

An IGP Printout for the Plot of the Data Given in Table 4.1 for the Estimated Empirical Process at x = 2 when n = 200

Table 4.2

A Comparison of the Sampling Distribution of the K-S Type of Statistic Using the Methods in (4.1.2)

		NORMAL DISTRIBUTION						GAMMA DISTRIBUTION					
	true value	n = 50			n = 100			n = 50			n = 100		
X	F(x)	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$
0.35	.0003	.002	.000	.003	.000	.001	.001	.000	.000	.002	.000	.000	.000
0.45	.0126	.022	.029	.014	.015	.018	.010	.014	.018	.024	.014	.013	.013
0.55	.0772	.094	.096	.083	.092	.080	.078	.096	.090	.099	.082	.096	.092
0.65	.2080	.236	.224	.214	.229	.215	.213	.223	.225	.236	.217	.207	.222
0.75	.3728	.381	.392	.385	.402	.380	.401	.389	.397	.403	.375	.356	.377
0.85	.5347	.529	.549	.551	.534	.554	.572	.544	.564	.554	.530	.535	.544
0.95	.6725	.657	.675	.678	.669	.687	.715	.693	.689	.702	.684	.658	.701
1.05	.7798	.744	.779	.791	.784	.778	.813	.794	.796	.789	.794	.772	.795
1.15	.8580	.818	.841	.875	.845	.850	.895	.854	.858	.872	.858	.859	.869
1.25	.9121	.881	.903	.918	.900	.904	.940	.913	.913	.932	.905	.912	.924
1.35	.9478	.920	.944	.947	.941	.939	.962	.945	.947	.961	.942	.954	.961
1.45	.9701	.954	.966	.975	.963	.968	.974	.967	.967	.975	.971	.970	.985
1.55	.9836	.972	.979	.987	.980	.982	.988	.986	.978	.985	.986	.981	.994
1.65	.9914	.988	.985	.994	.991	.993	.991	.995	.991	.991	.989	.994	.996
1.75	.9956	.994	.993	.996	.995	.996	.997	1.000	.998	.996	.994	.998	.997
1.85	.9979	.995	.996	.997	.999	.999	.999		1.000	.998	.998	.999	.998
1.95	.9990	.996	.999	.999	.999	1.000	.999			.999	.999	.999	1.000
2.05	.9996	.997	.999	.999	.999		.999			1.000	1.000	1.000	
2.15	.9998	.999	1.000	1.000	1.000		1.000						

Table 4.2 (continued)

		NORMAL DISTRIBUTION						GAMMA DISTRIBUTION					
	true value	n = 150			n = 200			n = 150			n = 200		
X	F(x)	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$
0.35	.0013	.001	.000	.001	.000	.000	.000	.001	.001	.000	.000	.000	.000
0.45	.0126	.016	.011	.015	.016	.014	.010	.014	.013	.022	.010	.012	.016
0.55	.0772	.088	.076	.085	.094	.077	.085	.082	.081	.100	.071	.083	.077
0.65	.2080	.245	.227	.234	.224	.216	.230	.233	.223	.251	.218	.218	.219
0.75	.3728	.413	.397	.410	.396	.368	.383	.404	.392	.409	.391	.375	.401
0.85	.5347	.571	.531	.577	.529	.545	.572	.561	.544	.567	.535	.532	.569
0.95	.6725	.688	.679	.715	.628	.698	.713	.708	.700	.692	.672	.676	.692
1.05	.7798	.786	.794	.827	.794	.797	.795	.795	.805	.801	.799	.788	.800
1.15	.8580	.871	.867	.900	.864	.866	.873	.867	.868	.881	.873	.871	.877
1.25	.9121	.924	.913	.935	.915	.918	.919	.924	.916	.929	.924	.928	.930
1.35	.9478	.961	.951	.966	.947	.957	.949	.951	.959	.964	.953	.955	.958
1.45	.9701	.983	.973	.981	.974	.976	.971	.966	.978	.983	.969	.981	.982
1.55	.9836	.991	.984	.987	.985	.987	.986	.976	.990	.988	.982	.993	.991
1.65	.9914	.993	.991	.991	.993	.993	.992	.988	.999	.993	.988	.999	.993
1.75	.9956	.996	.997	.998	.995	.994	.996	.995	1.000	.996	.992	1.000	.996
1.85	.9979	.997	.998	.998	.996	.997	.998	.998		.998	.996		.999
1.95	.9990	.999	.999	.999	.998	.998	.999	1.000		.999	.997		1.000
2.05	.9996	.999	1.000	.999	.999	.999	1.000			.999	1.000		
2.15	.9998	1.000		1.000	1.000	1.000				1.000			

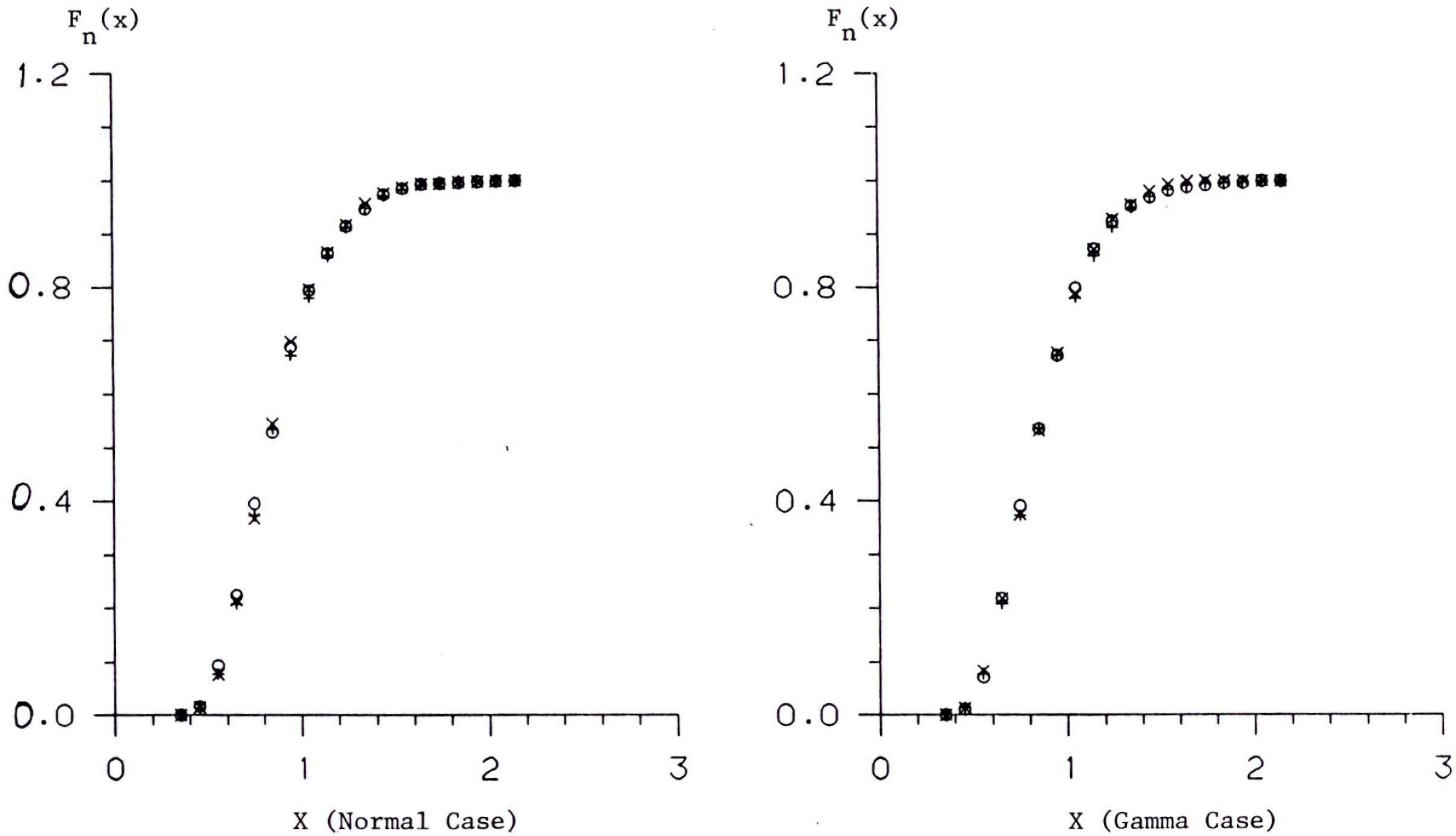

Figure 4.4

An IGP Printout for the Plot of the Data Given in Table 4.2 for the K-S Type of Statistic when n = 200

Table 4.3

Estimated Percentage Points of the K-S Type of Statistic for Commonly Used Significance Level 1-p

		NORMAL DISTRIBUTION				GAMMA DISTRIBUTION			
	true value	n = 50		n = 100		n = 50		n = 100	
1-p	$F^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$
.20	1.073	1.112	1.084	1.083	1.085	1.058	1.055	1.055	1.085
.10	1.224	1.292	1.244	1.251	1.242	1.232	1.220	1.238	1.233
.05	1.358	1.415	1.396	1.384	1.377	1.378	1.362	1.376	1.335
.02	1.517	1.589	1.561	1.550	1.534	1.536	1.555	1.509	1.535
.01	1.628	1.667	1.682	1.642	1.596	1.585	1.638	1.659	1.605
.005	1.731	1.760	1.820	1.737	1.695	1.643	1.676	1.783	1.676

		NORMAL DISTRIBUTION				GAMMA DISTRIBUTION			
	true value	n = 150		n = 200		n = 150		n = 200	
1-p	$F^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$
.20	1.073	1.062	1.058	1.059	1.056	1.057	1.045	1.051	1.068
.10	1.224	1.206	1.202	1.209	1.216	1.206	1.218	1.199	1.197
.05	1.358	1.318	1.341	1.363	1.317	1.341	1.311	1.337	1.331
.02	1.517	1.429	1.503	1.516	1.474	1.573	1.469	1.536	1.433
.01	1.628	1.526	1.600	1.633	1.595	1.675	1.548	1.728	1.525
.005	1.731	1.710	1.726	1.766	1.776	1.760	1.591	1.807	1.595

Table 4.4

A Comparison of the Sampling Distribution of the C-vM Type of Statistic Using the Methods in (4.1.2)

	true value	NORMAL DISTRIBUTION						GAMMA DISTRIBUTION					
		n = 50			n = 100			n = 50			n = 100		
X	F(x)	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$
0.05	.1298	.123	.135	.104	.134	.129	.106	.126	.118	.119	.111	.107	.118
0.15	.6104	.583	.629	.552	.585	.610	.597	.620	.618	.579	.600	.594	.594
0.25	.8116	.796	.822	.755	.804	.813	.820	.805	.805	.799	.814	.799	.804
0.35	.9016	.889	.912	.877	.897	.897	.903	.896	.909	.887	.900	.899	.898
0.45	.9463	.934	.951	.939	.946	.948	.948	.945	.949	.940	.946	.951	.944
0.55	.9702	.969	.968	.962	.973	.974	.972	.971	.970	.961	.971	.980	.976
0.65	.9824	.981	.978	.976	.985	.989	.985	.982	.981	.979	.981	.988	.989
0.75	.9901	.991	.989	.990	.991	.996	.992	.990	.989	.990	.988	.995	.993
0.85	.9923	.994	.996	.994	.996	.998	.997	.994	.993	.994	.993	.998	.995
0.95	.9944	.995	.997	.995	.997	.999	.998	.999	.995	.997	.996	.998	.995
1.05	.9965	.996	.997	.996	.998	.999	.999	1.000	.998	.998	1.000	1.000	.999
1.15	.9986	.997	.999	.998	.999	1.000	1.000		.998	.998			1.000
1.25	1.0000	.998	.999	.999	1.000				1.000	.999			
1.35	1.0000	.999	1.000	.999						.999			
1.45	1.0000	.999		.999						.999			
1.55	1.0000	1.000		.999						1.000			

Table 4.4 (continued)

		NORMAL DISTRIBUTION						GAMMA DISTRIBUTION					
	true value	n = 150			n = 200			n = 150			n = 200		
X	F(x)	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$
0.05	.1298	.132	.123	.114	.125	.116	.129	.142	.109	.137	.107	.112	.116
0.15	.6104	.620	.614	.609	.617	.633	.611	.608	.607	.615	.592	.583	.613
0.25	.8116	.818	.829	.825	.820	.825	.815	.821	.815	.822	.820	.794	.804
0.35	.9016	.911	.918	.918	.907	.913	.904	.908	.911	.905	.908	.910	.905
0.45	.9463	.954	.944	.958	.937	.959	.947	.951	.961	.943	.948	.958	.949
0.55	.9702	.979	.974	.975	.964	.976	.967	.972	.982	.967	.963	.977	.976
0.65	.9824	.986	.988	.984	.979	.989	.979	.977	.990	.982	.978	.992	.986
0.75	.9901	.993	.994	.990	.988	.994	.991	.985	.994	.988	.986	.995	.990
0.85	.9923	.995	.997	.993	.991	.995	.994	.993	.997	.993	.991	.998	.996
0.95	.9944	.997	.998	.993	.995	.996	.998	.998	.998	.995	.995	1.000	.998
1.05	.9965	.999	.998	.995	.996	.998	.999	1.000	.999	.997	.995		.999
1.15	.9986	.999	.999	.999	.997	.999	.999		1.000	.998	.997		1.000
1.25	1.0000	.999	1.000	.999	.998	.999	.999			.999	.998		
1.35	1.0000	.999		1.000	.999	1.000	.999			.999	.999		
1.45	1.0000	.999			.999		.999			1.000	1.000		
1.55	1.0000	1.000			1.000		1.000						

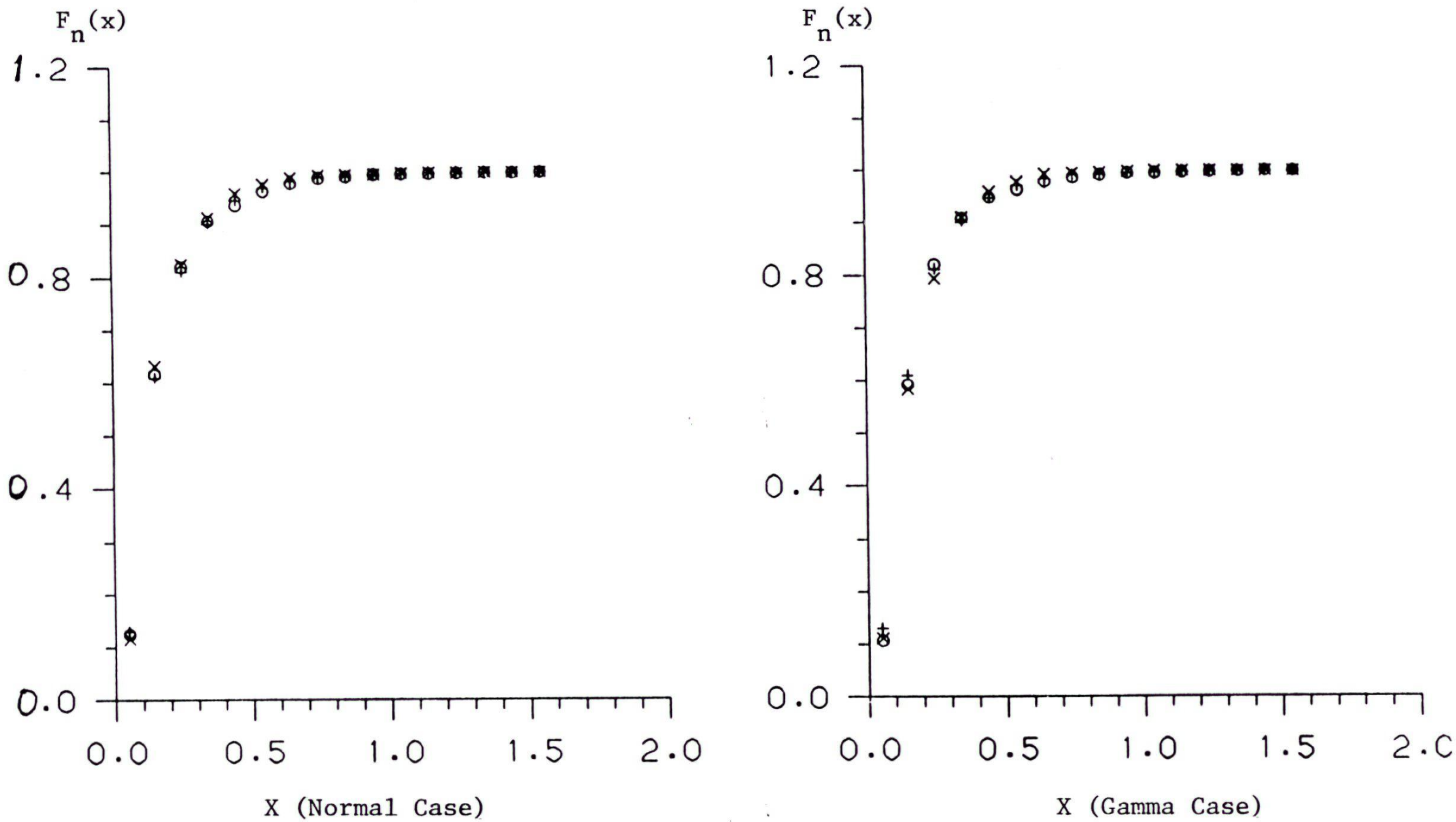

Figure 4.5

An IGP Printout for the Plot of the Data Given in
Table 4.4 for the C -vM Type of Statistic
when n = 200

Table 4.5

Estimated Percentage Points of the C-vM Type of Statistic for Commonly Used Significance Level 1-p

	true value	NORMAL DISTRIBUTION				GAMMA DISTRIBUTION			
		n = 50		n = 100		n = 50		n = 100	
1-p	$F^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$
.20	.241	.252	.232	.246	.242	.249	.245	.241	.251
.10	.347	.374	.333	.360	.354	.353	.338	.350	.353
.05	.461	.480	.443	.468	.457	.463	.464	.471	.447
.02	.620	.619	.669	.601	.586	.633	.626	.639	.554
.01	.743	.723	.759	.735	.667	.736	.764	.760	.673

	true value	NORMAL DISTRIBUTION				GAMMA DISTRIBUTION			
		n = 150		n = 200		n = 150		n = 200	
1-p	$F^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$
.20	.241	.242	.234	.235	.228	.234	.244	.235	.253
.10	.347	.327	.323	.338	.329	.333	.321	.337	.332
.05	.461	.435	.459	.483	.427	.441	.427	.466	.420
.02	.620	.582	.576	.662	.575	.679	.544	.670	.576
.01	.743	.709	.680	.764	.651	.810	.651	.835	.638

Table 4.6

A Comparison of the Sampling Distribution of the A-D Type of Statistic
Using the Methods in (4.1.2)

	NORMAL DISTRIBUTION						GAMMA DISTRIBUTION					
	n = 50			n = 100			n = 50			n = 100		
X	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$
0.3	.035	.071	.052	.053	.062	.048	.012	.070	.059	.053	.061	.048
0.6	.226	.362	.307	.318	.346	.275	.325	.314	.285	.298	.292	.304
0.9	.430	.595	.533	.545	.579	.496	.512	.515	.487	.515	.500	.527
1.2	.551	.723	.676	.687	.705	.654	.644	.654	.639	.654	.635	.670
1.5	.668	.816	.769	.796	.808	.764	.727	.726	.714	.746	.738	.776
2.0	.792	.887	.867	.892	.895	.871	.822	.819	.825	.845	.831	.868
3.5	.947	.972	.970	.987	.987	.971	.936	.922	.949	.959	.952	.976
5.5	.983	.994	.989	1.000	1.000	.996	.974	.963	.979	.991	.987	.994
7.5	.993	.999	.997			.998	.988	.978	.988	.994	.991	.997
9.5	.996	.999	1.000			.999	.993	.988	.993	.994	.994	.998
11.5	.998	1.000				.	.993	.988	.995	.996	.995	.998
13.5	.999					.	.994	.990	.995	.997	.996	.998
15.5	.999					.	.995	.991	.995	.997	.996	.998
17.5	1.000					.	.995	.992	.996	.997	.996	.998
19.5						*	.996	.995	.997	.998	.996	.998
21.5							.996	.995	.997	.998	.997	.998
23.5							.996	.996	.997	.998	.997	1.000
.							.	.	.	.	.	
.							.	.	.	.	.	
.							.	.	.	.	.	
.							*	*	*	*	*	

Table 4.6 (continued)

X	NORMAL DISTRIBUTION						GAMMA DISTRIBUTION					
	n = 150			n = 200			n = 150			n = 200		
	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$
0.3	.048	.062	.041	.043	.058	.060	.069	.063	.061	.050	.064	.064
0.6	.348	.368	.307	.328	.335	.296	.318	.332	.339	.304	.317	.312
0.9	.565	.594	.523	.560	.592	.552	.518	.535	.555	.500	.529	.551
1.2	.714	.739	.682	.726	.835	.803	.668	.689	.712	.659	.672	.696
1.5	.825	.836	.797	.827	.915	.892	.751	.777	.795	.762	.763	.797
2.0	.911	.912	.891	.906	.988	.972	.866	.871	.885	.857	.903	.900
3.5	.979	.985	.979	.977	.998	.997	.965	.978	.973	.966	.975	.978
5.5	.998	.996	.996	.995	.999	.998	.989	.994	.994	.989	.993	.996
7.5	.999	.999	.999	.998	.	.999	.996	.997	.998	.995	.999	.999
9.5	1.000	.999	.999	.999	.	.	.996	.999	.999	.998	1.000	1.000
11.5		.999	.999	.	.	.	.996	.999	1.000	.998		
13.5		.999	.999	.	*	.	.996	.999		.998		
15.5		.999	.999	.		*	.998	1.000		.998		
17.5		.999	.999	*			.998			1.000		
19.5		.999	.999				.998					
21.5		1.000	.999				.998					
23.5			1.000				.999					
.							.					
.							.					
.							.					
.							*					

* Extreme values have been discarded

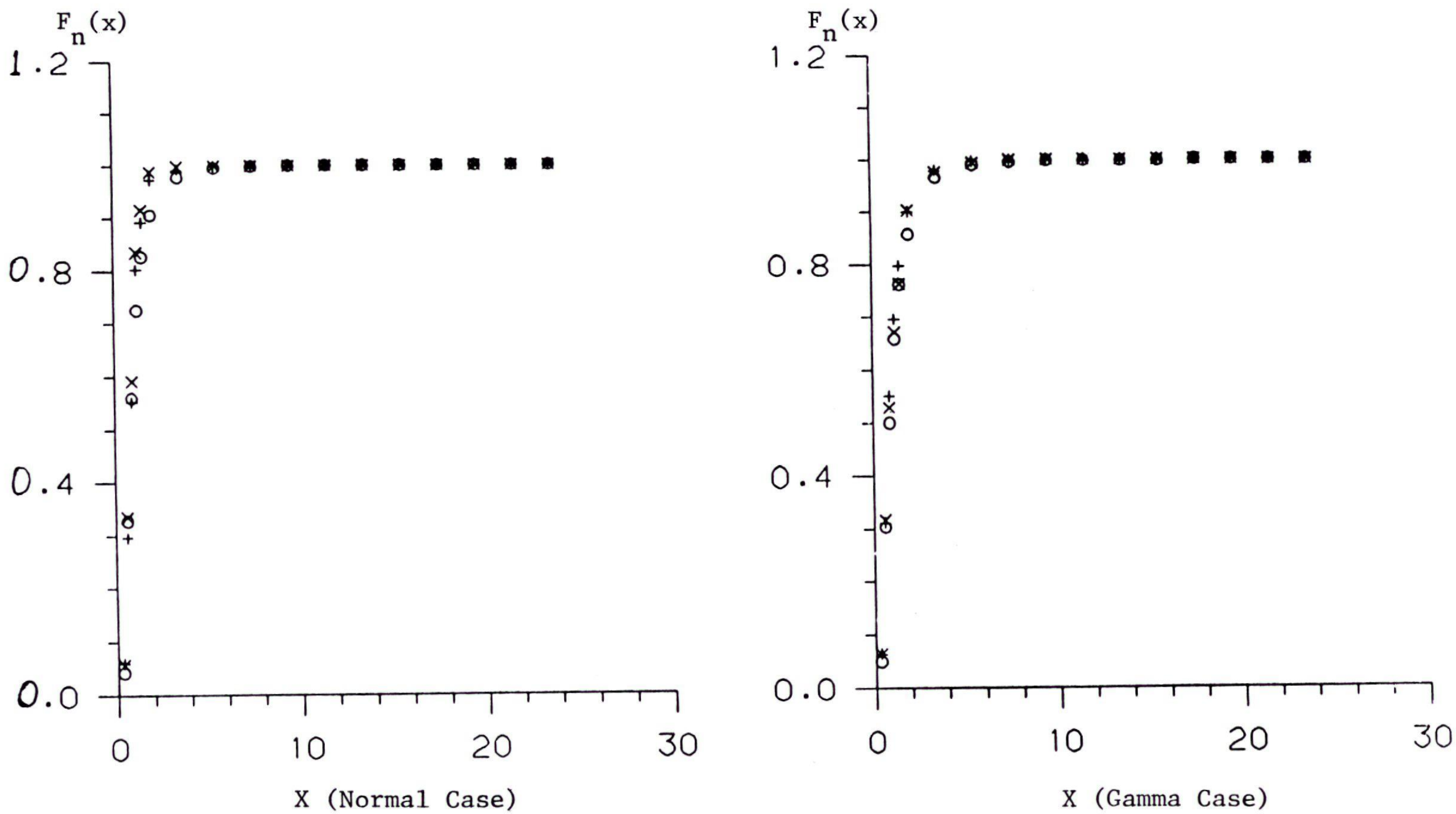

Figure 4.6

An IGP Printout for the Plot of the Data Given in
Table 4.6 for the A-D Type of Statistic
when n = 200

Table 4.7

Estimated Percentage Points of the A-D Type of Statistic
for Commonly Used Significance Level 1-p

		NORMAL DISTRIBUTION				GAMMA DISTRIBUTION			
	true value	n = 50		n = 100		n = 50		n = 100	
1-p	$F^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$
.10	1.933	2.305	2.079	2.048	2.034	2.775	2.841	2.409	2.573
.05	2.492	2.970	2.736	2.737	2.597	3.916	4.495	3.306	3.445
.01	3.857	5.541	4.433	4.066	3.605	7.942	12.963	5.262	7.113

		NORMAL DISTRIBUTION				GAMMA DISTRIBUTION			
	true value	n = 150		n = 200		n = 150		n = 200	
1-p	$F^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$	$F_1^{-1}(p)$	$F_2^{-1}(p)$
.10	1.933	1.904	1.892	1.961	1.869	2.221	2.230	2.329	2.181
.05	2.492	2.404	2.469	2.716	2.413	2.976	2.770	3.134	2.851
.01	3.857	4.399	3.752	4.762	3.693	6.075	4.898	5.792	4.800

BIBLIOGRAPHY

Anderson, T.W., and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann. Math. Statist., 23, 193-212.

Anderson, T.W., and Darling, D.A. (1954). A test of goodness of fit. J. Amer. Statist. Assoc., 49, 765-769.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

Birnbaum, Z.W. (1952). Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. J. Amer. Statist. Assoc., 47, 425-441.

Brown, R. (1828). A brief account of microscopical observations made in the months of June, July and August 1827 on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. Philos. Mag. Ann. of Philos. New Ser. 4, 161-178.

Burke, M.D., and Csörgő, M. (1976). Weak approximations of the empirical process when parameters are estimated. Lecture Notes in Mathematics, 566, 1-16, Springer-Verlag, Berlin.

Burke, M.D., Csörgő, M., Csörgő, S., and Révész, P. (1979). Approximations of the empirical process when parameters are estimated. Ann. Prob., 7, 790-810.

Burke, M.D., and Gombay, E. (1988). On goodness-of-fit and the bootstrap. Statistics and Probability Letters, 6, 287-293.

Cantelli, F.P. (1917). Sulla probabilitá come limita della frequenza. Rend. Accad. Lincei. 26. 39.

Cramér, H. (1928). On the composition of elementary errors. Second paper : Statistical Applications, Skand. Aktuartidskr., 11, 141-180.

Csörgő, M., Komlós, J., Major, P., Révész, P., and Tusnády, G. (1974). On the empirical process when parameters are estimated. Trans. Seventh Prague Conference, Academia, Prague.

Csörgo, M., and Révész, P. (1975). A new method to prove Strassen type laws of invariance principle II. Z. Wahrscheinlichkeitstheorie una verw. Gebiete, 33, 261-269.

Csörgo, M., and Révész, P. (1981). Strong Approximations in Probability and Statistics. Akademiai Kiadó, Budapest and Academic Press, New York.

Csörgo, M., Csörgo, S., and Horváth, L. (1986). An Asymptotic Theory for Empirical Reliability and Concentration Processes. Lecture Notes in Statistics, 33, Springer-Verlag, Berlin.

Doob, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. Ann. Math. Statist., 20, 393-403.

Durbin, J. (1961). Some methods of constructing exact tests. Biometrika, 48, 41-55.

Durbin, J. (1973a). Weak convergence of the sample distribution function when parameters are estimated. Ann. Statist., 1, 279-290.

Durbin, J. (1973b). Distribution Theory for Tests based on the Sample Distribution Function. Regional Conference Series in Applied Math., 9, SIAM, Philadelphia.

Durbin, J. (1976). Kolmogorov-Smirnov tests when parameters are estimated. Lecture Notes in Math., 566, 33-44.

Efron, B. (1979). Bootstrap methods: another look at the jacknife. Ann. Statist., 7, 1-26.

Efron, B. (1982). The Jacknife, the Bootstrap and Other Resampling Plans. Regional Conf. Series on Applied Math., SIAM, Philadelphia.

Einstein, A. (1905). On the movement of small particles suspended in a stationary liquid demended by the molecular-kinetic theory of heat. Ann. Physik, 17, 549-560.

Feller, W. (1948). On the Kolmogorov-Smirnov limit theorems for empirical distributions. Annals of Math. Stat., 19, 177-189.

Glivenko, V. (1933). Sulla determinazione empirica della legge di probabilita. Giorn. Inst. Ital. Attuari., 4, 92-99.

Hogg, R.V., and Craig, A.T. (1978). Introduction to Mathematical Statistics. 4th edition. Macmillan Publishing Co., New York.

Karlin, S. (1966). A First Course in Stochastic Processes. Academic Press, New York.

Kiefer, J. (1972). Skorohod embedding on multivariate r.v.'s and the sample d.f. Z. Wahrscheinlichkeitstheorie verw. Gebiete, 24, 1-35.

Kolmogorov, A. (1933). Sulla determinazione empirica di une legge di distribuzione. Giorn. Inst. Ital. Attuari., 4, 83-91.

Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f. I. Z. Wahrscheinlichkeitstheorie und verw. Gebiete, 32, 111-131.

Marshall, A.W. (1958). The small sample distribution of nw_n^2. Ann. Math. Statist., 29, 307-309.

Massey, F.J. Jr. (1950). A note on the estimation of the distribution function by confidence limits. Ann. Math. Statist., 21, 116-119.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil. Mag., Series 5, 50, 157.

Rao, K.C. (1972). The Kolmogorov, Cramér-von Mises Chisquare statistics for goodness of fit tests in the parametric case. Abstract: Bulletin, Inst. Math. Statist. 1, 87.

Rudin, W. (1953). Principles of Mathematical Analysis. McGraw-Hill Book Company, New York.

Smirnov, N.V. (1936). Sur la distribution de ω^2 (Critérium de M.R. v. Mises). C.R. Acad. Sci., Paris, 202, 449-452.

Smirnov, N.V. (1948). Table for estimating the goodness of fit of empirical distribution. Ann. Math. Statist., 19, 177-189.

Stephens, M.A., and Maag, U.R. (1968). Further percentage points for W_n^2. Biometrika, 55, 428-430.

Stephens, M.A. (1970). Use of the Kolmogorov-Smirnov, Cramér-von Mises and related statistics without extensive tables. J. Roy. Statist. Soc. Ser. B. 32, 115-122.

von Mises, R. (1931). Wahrscheinlichkeitsrechnung, Wein. Leipzig.