#2025

ON SEQUENTIAL ESTIMATION OF THE RENEWAL FUNCTION, OPTIMAL BLOCK REPLACEMENT POLICIES, AND FIXED WIDTH CONFIDENCE BANDS

by

John Crowell

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics.

Chapel Hill

1990

Approved by:

____
Advisor

Gordon Simons
Reader

M. R. Leadbetter
Reader

The Library of the Department of Statistics
North Carolina State University

JOHN CROWELL. On Sequential Estimation of the Renewal Function, Optimal Block Replacement Policies, and Fixed Width Confidence Bands. (Under the direction of PRANAB K. SEN.)

ABSTRACT

Stochastic approximation methodology is applied to the problem of estimating an optimal block replacement policy. Several improvements to minimize the asymptotic mean square error of the estimator are considered. A stopping time is then defined with the objective of a fixed-width confidence interval for the optimal policy.

A nonparametric estimator is used for sequential estimation of the renewal function at a single point and over a bounded interval of real numbers. Several strongly consistent estimates of the asymptotic variance are given. Sequential fixed-width confidence interval methods are shown to be effective for estimating the renewal function at a point. A stopping time is defined for minimum (integrated) risk estimation over an interval. It is shown that bootstrap methods can be used to construct asymptotically correct confidence bands for the renewal function over an interval.

Let R be a statistical function which depends on a distribution F and is defined on an interval of real numbers. Suppose R may be estimated using a random sample from F. Motivated by the problem of a fixed-width confidence band for R, a stopping time is defined that is analogous to those available for fixed-width confidence interval estimation of a real-valued parameter. The proposed stopping time can be used in conjunction with a bootstrap confidence band procedure. Sufficient conditions are given for the sequential confidence band to have the correct asymptotic coverage probability as the desired band width goes to zero. Several properties of the stopping time, such as asymptotic efficiency, are addressed with reference to bootstrap procedures. Stopping times for variable width confidence bands are also considered.

ACKNOWLEDGEMENTS

First and foremost, thanks go to my adviser P. K. Sen for his guidance in the preparation of this dissertation. I hope that if I ever direct a student I will be as helpful as he was.

I would also like to express my appreciation to the members of my committee (I. M. Chakravarti, M. R. Leadbetter, Gordon Simons, and Walter L. Smith) for their assistance, as well as to the entire faculty of the Department of Statistics for their encouragement during my studies.

Thanks also go to the staff, in particular June Maxwell for so many different things and Lee Trimble for the valiant typing of parts of this dissertation.

The constant support of my parents and family are something for which I am always very grateful.

Final thanks go to my fellow students for their support and friendship, especially the one who did not want to be mentioned and my officemate of four years, Prakash Patil.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ARP	age replacement policy
a.s.	almost surely
BRP	block replacement policy
i.i.d.	independent and identically distributed
SA	stochastic approximation
u.c.i.p.	uniformly continuous in probability

LIST OF SYMBOLS

χ, I	the indicator function of a set or proposition
$\xrightarrow{D}$	weak convergence (convergence in distribution)
$\xrightarrow{W}$	weak convergence (convergence in distribution)
$\stackrel{D}{=}$	equal in distribution
$a \wedge b$	minimum of the numbers a and b
$a \vee b$	maximum of the numbers a and b
Δ	difference operator for sequences: $\Delta a_n = a_{n+1} - a_n$
$[\cdot]$	greatest integer function
$\mathbf{E}_{\mathfrak{F}}$	conditional expectation given the σ-field $\mathfrak{F}$

CHAPTER 1

INTRODUCTION

1.1 Introduction

This thesis addresses three separate problems which are introduced and motivated in Section 1.2. A unifying theme is that all three problems are approached from a sequential point of view. An additional link between two of these problems is that both involve the renewal function. Some background and literature review are given in Section 1.3. The final section of this chapter contains a summary of the remaining chapters.

1.2 Three Separate Problems

1.2.1 An optimization problem in reliability theory

An important topic in reliability theory is maintenance policies for systems (mechanical, electrical, etc.) whose components are subject to stochastic failure. Often a primary consideration in maintaining such a system is the cost associated with failure of the system, above and beyond the actual cost of subsequent repair. For instance, if the system is some manufacturing process, then there is the productivity lost between the time of system breakdown and the restoration of production. A partial remedy for this may be to carry out a preventive maintenance plan where aging parts are repaired or replaced according to some schedule. While such a program may diminish the number of system failures or system downtime, it will increase routine maintenance costs since some parts will be replaced or attended to before failure. Ideally, one would like to minimize costs by balancing these two factors.

More specifically, consider a system consisting of a single component or part. Assume that this part can be replaced by a like part at any time and is replaced following failure. Suppose all parts used have independent and identically distributed lifetimes with distribution F. Further assume that the cost C_1 of having to unexpectedly replace a failed part is greater than the cost C_2 of making a replacement at a planned time. If F has increasing failure rate, then it may be less expensive to substitute new parts for aging parts near failure than to just make replacements following failure. Two maintenance plans motivated by this consideration are the block replacement policy (BRP) and the age replacement policy (ARP). Both of these policies are treated in detail by Barlow and Proschan (1965, 1975), where some historical development is also given. We consider the most basic BRP here. The ARP is described in Section 1.3.2.

Suppose a new part is installed at time 0. In the simplest BRP, planned replacements are made at times T, 2T, 3T,... where T is a positive constant; replacements are also made following any failures. It is assumed all replacements are made instantaneously. For this BRP, the expected long-run cost per unit time as a function of T is

$$B(T) = \frac{C_1 H(T) + C_2}{T},$$

where H is the renewal function associated with the lifetime distribution F. (See Section 1.2.2.) Thus H(T) is the expected number of failures for each of the intervals (0, T], (T, 2T], etc.

Suppose there exists a finite positive value ϕ^* which minimizes B(T) in T; ϕ^* is referred to as the optimal BRP since $T = \phi^*$ results in the minimum possible expected cost per unit time. If the lifetime distribution F or associated renewal function H were known, then finding ϕ^* would be a deterministic problem. Here both F and H are taken to be unknown and nonparametric estimation of ϕ^* is undertaken. Frees(1983) applied stochastic approximation methodology to the problem of sequential estimation of an optimal ARP. (See Section 1.3.2.) In Chapters 2 and 3 we consider using similar techniques for sequential estimation of the optimal BRP ϕ^*.

1.2.2 Problem: sequential estimation of the renewal function

Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) nonnegative random variables with distribution function F and positive mean μ_1. Let $S_n = X_1 + X_2 + \ldots + X_n$, the time of the n-th renewal. Denote by N(t) the largest n such that $S_n \leq t$. If $X_1 > t$, then by convention $N(t) = 0$. The renewal function is then $H(t) = \mathbb{E}N(t)$, the expected number of renewals up to and including time t. Noting that $N(t) = \Sigma_{k=1}^{\infty} \chi(S_k \leq t)$, it follows that

$$H(t) = \sum_{k=1}^{\infty} F^{(k)}(t), \tag{1.1}$$

where $F^{(k)}$ is the k-th convolution of F.

The renewal function is important in a number of areas of application. In Section 1.2.1 it was seen that the cost of a BRP depends on the renewal function associated with F. More generally, the renewal function is useful in reliability for characterizing the expected number of parts necessary to keep a system in repair. (See Barlow and Proschan (1965).) Business applications include warranty analysis and inventory policies.

For large values of t the behaviour of the renewal function is well understood. For instance, assuming F is nonarithmetic and has finite second moment $\mu_2 = \mathbf{E} X_1^2$, Smith (1954) showed that

$$(1.2) \qquad H(t) - \frac{t}{\mu_1} \to \frac{\mu_2}{2\mu_1^2} - 1 \quad \text{as } t \to \infty.$$

This result provides a means for estimating H(t) for large values of t from the i.i.d. sample $X_1, X_2, \ldots, X_n$ by first estimating μ_1 and μ_2. (See Section 1.3.3.)

Motivated by (1.1), Frees (1986i, ii) suggested a nonparametric estimator of the renewal function useful for applications such as those previously mentioned where it may be useful to have knowledge of H(t) for small and moderate values of t. This estimator is also applicable when estimating H(t) for large values of t when F does not have a finite second moment. Chapter 4 addresses sequential estimation of H(t) at a single point t using Frees' estimator. A sequential approach to estimating the function H on the interval [0, T], $0 < T < \infty$, is taken in Chapter 5.

1.2.3 The question of sample size for fixed-width confidence bands

In many applications one would like to estimate a statistical function R_F, where R_F depends on a distribution function F. For instance, there might be interest in estimating the renewal function on the interval [0, T] as discussed in Section 1.2.2. Other examples in reliability theory include estimation of the mean and median residual life functions for a lifetime distribution F. The function R_F could be the distribution F or the quantile function associated with F. When estimating such functions, one might want to construct a confidence band of a prescribed width which contains the function of interest with high probability.

Assume that given a random sample $X_1, X_2, \ldots, X_n$ from the distribution F there exists a natural estimator $R_n(\cdot)=R_n(\cdot; X_1, X_2, \ldots, X_n)$ of R_F. Suppose the stochastic process

$$r_n(\cdot)=n^{1/2}\{R_n(\cdot) - R_F(\cdot)\}$$

is such that $r_n \overset{W}{\to} \mathcal{G}_F$, where $\mathcal{G}_F$ is a Gaussian process. A number of authors have exploited weak convergence of r_n to construct confidence bands for a variety of choices of R_F, typically by bootstrapping or simulating the process r_n. Usually it is not clear a priori what sample size n is sufficient to achieve approximately the desired level of precision and confidence. This suggests approaching the problem from a sequential point of view, which is done in Chapter 6.

1.3 Literature Review

1.3.1 Stochastic approximation

Let M(x) be a regression function on the real line. Suppose that for any x we may observe the random variable Y(x), an unbiased estimate of M(x). Robbins and Monro (1951) introduced stochastic approximation (SA) as a technique for finding the (assumed) unique root θ of the equation $M(x)=\alpha$. (Without loss of generality we take $\alpha=0$.) Let $\{a_n\}$ be a decreasing sequence of positive constants such that $\lim_{n\to\infty} a_n = 0$. Given $X_1, X_2, \ldots, X_n$, successive estimates of θ, let Y_n be conditionally unbiased for $M(X_n)$. Robbins and Monro (R-M) proposed the algorithm $X_{n+1} = X_n - a_n Y_n$, where X_{n+1} is the next estimate of θ. Assuming that the a_n are of type $1/n$ and the random variable Y(x) is a.s. bounded for any x, they provided several sets of sufficient conditions on M for weak convergence of X_n to θ.

Kiefer and Wolfowitz (1952) suggested a similar procedure for finding the (assumed) unique θ such that the regression function M(x) is maximized at $x=\theta$. Let $\{a_n\}$ be as before and let $\{c_n\}$ be a second sequence of positive constants decreasing to zero. Suppose Y_{2n} and Y_{2n-1} are conditionally unbiased for $M(X_n+c_n)$ and $M(X_n-c_n)$, respectively. The Kiefer-Wolfowitz (K-W) algorithm is then of the form

$$X_{n+1} = X_n - a_n \frac{(Y_{2n} - Y_{2n-1})}{2c_n} . \tag{1.3}$$

Assuming that M(x) is strictly increasing for $x<\theta$ and strictly decreasing for $x>\theta$, it was shown under additional conditions that X_n is a weakly consistent estimator of θ. We mention some of the important developments in SA since the introduction of the R-M and K-W procedures. Blum (1954i) was the first to use martingale theory to prove a.s. convergence in both settings. In a later paper, Blum (1954ii) proposed several multivariate SA procedures. A useful theorem (Robbins and Siegmund (1971)) for showing a.s. convergence of SA-type procedures is stated and employed in Section 2.2.

Chung (1954) used a "convergence of moments" proof to show asymptotic normality of $a_n^{1/2}(X_n-\theta)$ for the R-M algorithm. The same technique was used by Derman (1956) to establish a limiting normal distribution in the K-W case. Central limit theorems for dependent random variables were the method of Sacks (1958) in weakening the conditions of Chung and Derman. In the same paper is a result on asymptotic normality for multivariate procedures. Fabian (1968) provides a theorem which generalizes the earlier work on asymptotic normality discussed here. This theorem is stated in a univariate form in Section 2.2.

Kersting (1977) demonstrated that the R-M process could be approximated by a weighted sum of i.i.d. random variables. A law of the iterated logarithm and invariance principle followed from this approximation. Ruppert (1982) derived similar properties for a multivariate algorithm which includes both the R-M and K-W cases. A random central limit theorem for a similar process is developed by Frees (1983). A short summary of other works on approximation of SA processes is contained in Ruppert (1982).

Most applications of SA require stopping the algorithm at some point. For the R-M and K-W procedures, a practitioner might want a fixed-width confidence interval of length 2d, for some $d > 0$, that covers θ with some prespecified probability. Sielkin (1973) provides stopping times $N_{d,\gamma}$ and $T_{d,\gamma}$ for the R-M case such that for any $\gamma \in (0, .5)$

$$\lim_{d \to 0} \mathbf{P} \{ \mid X_{T_{d,\gamma}+1} - \theta \mid \le d \} = 1 - 2\gamma$$

and

$$\lim_{d \to 0} \mathbf{P} \{ \mid M(X_{N_{d,\gamma}+1}) - M(\theta) \mid \le d \} = 1 - 2\gamma.$$

McLeish (1976) arrives at a similar stopping time for the R-M procedure using a functional central limit theorem for an appropriately defined process. Frees (1983) considers a stopping rule for a more general process, a rule which is described and then used in Section 3.5 of this work.

1.3.2 Estimating an optimal age replacement policy

Consider the maintenance problem described in Section 1.2.1 with the same assumptions regarding the costs of planned and failure replacements, supposing all replacements are made instantaneously. Under an ARP, if a part has not failed and been replaced prior to reaching some fixed age T, it is preventively replaced at age T. The expected long-run cost per unit time can be shown to be

$$A(T) = \frac{C_1F(T) + C_2S(T)}{\int_0^T S(t)dt},$$

where F is the lifetime distribution of the parts and $S(t) = 1 - F(t)$.

Frees (1983) applies SA to the problem of estimating the optimal age replacement policy when the distribution F is unknown. We summarize some of his results. It is easily checked that

$$\frac{d}{dt}A(t) = \frac{M(t)}{\{\int_0^t S(u)du\}^2}$$

where

$$M(t) = (C_1 - C_2)f(t)\int_0^t S(u)du \; - \; S(t)[C_1F(t) + C_2S(t)].$$

Suppose that ϕ uniquely minimizes A(t) in t, $M(\phi)=0$, and $M(t)(t-\phi)>0$ for $t \neq \phi$. (These are typical assumptions on M for identifiability of ϕ.) It then seems natural to try SA as a method for estimating ϕ.

Let $\{X_{in}\}_{n=1}^{\infty}$, i = 1,2, be sequences of i.i.d. random variables with distribution F such that $\{X_{1n}\}$ and $\{X_{2n}\}$ are independent. Like the K-W method, Frees' procedure requires two sequences of nonnegative real constants, $\{a_n\}$ and $\{c_n\}$. Letting $Z_{in} = \min(X_{in}, \phi_n + c_n)$, $F_{in}(t) = \chi(Z_{in} < t)$, and $S_{in}(t) = 1 - F_{in}(t)$, define an estimate of M(t) by

$$M_n(t) = (C_1 - C_2)f_n(t)\int_0^t S_{2n}(u)du \; - \; S_{1n}(t)[C_1F_{2n}(t) + C_2S_{2n}(t)].$$

Here $f_n(t)$ is an estimate of f(t). Then given ϕ_1, the formula

$$\phi_{n+1} = \phi_n - a_n M_n(\phi_n)$$

defines a sequence $\{\phi_n\}$ of estimates of ϕ. (Frees actually considers an algorithm for finding the zero of $M(g(t))$, where $g: R \to (0, \infty)$ is a known strictly increasing continuous function. This ensures that all estimates of ϕ are positive. Our simplification suffices for this discussion.)

Frees first establishes that if

$$f_n(t) = \frac{\chi(\, t-c_n \le Z_{1n} < t+c_n \,)}{2c_n}$$

then under suitable conditions $\phi_n \to \phi$ a.s. and ϕ_n, properly standardized, has a limiting normal distribution. It is then shown that by using a more refined kernel density estimator, the rate of convergence of the estimate ϕ_n can be improved. The asymptotic mean square error can be minimized by proper choice of the sequences $\{a_n\}$ and $\{c_n\}$. An "adaptive" procedure is defined which estimates the optimal sequences as the algorithm proceeds. With this procedure the minimal asymptotic mean square error is attained. A sequential fixed-width confidence interval scheme is then defined which provides the correct coverage probability asymptotically as the desired confidence interval width goes to zero.

1.3.3 Estimation of the renewal function

Frees (1986i) considers several estimators of $H(t)$ with reference to an application in warranty analysis. Let $X_1,\ldots,X_n$ be i.i.d. random variables with distribution F which has positive mean μ_1 and finite variance σ^2. In practice, observations could be extracted from a realization of the renewal process with underlying distribution F.

The first estimator assumes F is non-arithmetic and depends heavily on the limiting form of the renewal function as described by (1.2):

$$H_{1n}(t) = \frac{t}{\mu_n} + \frac{\sigma_n^2}{2\mu_n^2} - \frac{1}{2}.$$

Here μ_n and σ_n are estimates of μ and σ, respectively. In general, this estimator is not consistent for finite t; however, Frees concludes it may perform well for large t.

Another approach is based on the expression (1.1). Assume $F(x) = F(x; \alpha_1, \ldots, \alpha_p)$ is known up to a finite number of parameters $\alpha_1, \ldots, \alpha_p$. Suppose $\alpha_{1n}, \ldots, \alpha_{pn}$ are consistent estimates of $\alpha_1, \ldots, \alpha_p$. A natural estimate of H(t) is then

$$H_{2n}(t) = \sum_{k=1}^{\infty} \hat{F}_n^{(k)}(t),$$

where $\hat{F}_n^{(k)}$ is the k-fold convolution of $\hat{F}_n(x) = F(x; \alpha_{1n}, \ldots, \alpha_{pn})$. Frees proves that if F is absolutely continuous (discrete) and the density (probability mass function) f is continuous in $\alpha_1, \ldots, \alpha_p$, then $H_{2n}(t) \rightarrow H(t)$ with probability one.

A third estimator is also suggested by (1.1). Let $X_{i1}, \ldots, X_{ik}$ represent any subset of k observations and Σ_C denote the summation over all $\binom{n}{k}$ subsets of k observations. An unbiased estimate of the k-fold convolution of F is the U-statistic

$$F_n^{(k)}(t) = \frac{1}{\binom{n}{k}} \sum_C \chi(X_{i1} + \ldots + X_{ik} \leq t).$$

An estimate of the renewal function is then

$$H_{3n}(t) = \sum_{k=1}^{m} F_n^{(k)}(t),$$

where $m = m(n) \leq n$ and $m \rightarrow \infty$ as $n \rightarrow \infty$. This estimator is strongly consistent assuming some mild conditions on F and m (Frees (1986ii)).

Frees draws some conclusions about these estimators from their theoretical properties and their performance on some data and in a Monte Carlo study. By definition, $H_{1n}(t)$ is appropriate only for t large relative to μ_1. If F is known up to a finite number of parameters, $H_{2n}(t)$ should perform well for both small and large t. However, problems of robustness may arise if F does not fully meet the assumptions made. Finally, $H_{3n}(t)$ would be the preferred estimator for small t when no parametric assumptions can be made.

The statistical properties of $H_{3n}(t)$ are further investigated in Frees(1986ii), reviewed here in Section 4.1. In this paper Frees discusses yet another estimator of H(t). If $\tilde{F}_n^{(1)}(t) = F_n^{(1)}(t)$ is the ordinary sample distribution function, let

$$\tilde{F}_n^{(k)}(t) = \int \tilde{F}_n^{(k-1)}(t-u)\, d\tilde{F}_n^{(1)}(u)$$

be its k-fold convolution. (Each of these is a V-statistic.) Then define $\tilde{H}_n(t)$ by

$$\tilde{H}_n(t) = \sum_{k=1}^{m} \tilde{F}_n^{(k)}(t).$$

Frees asserts that under fairly weak conditions this estimate is consistent and asymptotically normal.

1.4 Summary

1.4.1 Chapters 2 and 3

Chapter 2 introduces a stochastic approximation algorithm for estimating an optimal block replacement policy. Sufficient conditions are given for the procedure to yield a strongly consistent sequence of estimates. A result on the asymptotic normality of the estimator provides information about its rate of convergence. We discuss how the procedure can be carried out concurrently with an actual maintenance program to achieve real savings in cost while estimating the optimal policy. The final section describes a major drawback of the procedure, namely that it is typically impossible to verify a critical assumption.

In Chapter 3 we undertake various improvements of the algorithm defined in Chapter 2. It is indicated how under certain circumstances one can significantly improve the rate of convergence of the estimator. Adaptive stochastic approximation techniques are applied to develop a procedure which minimizes the asymptotic mean square error. Finally, the procedure is made fully sequential by defining a stopping time for fixed-width confidence interval estimation of an optimal BRP.

1.4.2 Chapters 4 and 5

In Chapter 4 concern lies with sequential estimation of the renewal function at a point using the estimator considered by Frees (1986ii). (This is the estimator H_{3n} defined in Section 1.3.3.) We first note how this estimator is applicable under slightly more general conditions than given in Frees (1986ii). A jackknife estimator of the asymptotic variance as well as an estimator considered by Frees for the same quantity are shown to be strongly consistent. These variance estimators are used in defining the usual type of stopping time for a fixed-width confidence interval; with the help of an appropriate invariance principle, developed in Section 4.3, the corresponding fixed-width confidence interval can be shown to have the correct asymptotic coverage probability as the interval width goes to zero.

In Chapter 5 we move on to consider estimating the renewal function simultaneously at all points t in the interval [0, T], again with the estimator considered by Frees. First, a stopping time for minimum risk estimation is considered, where the risk function is composed of an integrated mean square error term and a term which accounts for the cost of the sample. It is seen that such a stopping time can provide guidance in balancing the cost of error with the cost of the sample. Next, the behaviour of the function estimator is more carefully characterized in an invariance principle. Motivated by this result, bootstrap methods are advocated for the construction of both fixed and variable width confidence bands for the renewal function on [0, T]. Stopping times for the construction of these two types of confidence bands are then introduced, directly motivating Chapter 6.

1.4.3 Chapter 6

We consider stopping times for the construction of fixed-width confidence bands to partially answer the problem posed in Section 1.1.3; the objective is to provide some guidance in deciding when an adequate sample size has been attained for the desired levels of confidence and precision. The theoretical development is similar to that available for fixed-width confidence interval estimation of a real-valued parameter. Properties of the stopping time such as asymptotic efficiency and normality (as the band width goes to zero) are explored. We then consider some consequences of using the stopping time in conjunction with a bootstrap procedure for constructing confidence bands. That these same sequential methods can be applied in constructing a confidence region for a statistical function R_F defined on $\mathbf{R}^p$, $p \geq 2$, is shown by means of an example. Stopping times for variable width confidence bands, where the width of the band varies with the argument t, are also considered.

CHAPTER 2

ESTIMATING AN OPTIMAL BLOCK REPLACEMENT POLICY

2.1 Introduction

In this chapter stochastic approximation methodology is applied to the problem of estimating an optimal block replacement policy.

Section 2.2 introduces a basic stochastic approximation algorithm which yields a sequence of estimates of the optimal replacement interval. Sufficient conditions are given for consistency of the procedure; a theorem on the asymptotic normality of the estimate provides information about the rate of convergence. Section 2.3 contains proofs of these results.

Section 2.4 describes how the theoretical procedure of Section 2.2 relates to what might be done in practice. Successive estimates of the optimal time interval could be used to carry out a random BRP, where at the time of each planned replacement the current estimate is used to determine the time of the next planned replacement. We verify that if a consistent estimate of the optimal policy is available, then a random BRP can be implemented such that the incurred cost per unit time converges to the optimum.

Section 2.5 reveals an important drawback of the procedure, namely its dependence on an assumption that may not hold in practice or may be impossible to verify.

2.2 The Basic Algorithm

Recall that with notation previously defined in Section 1.2.1 the cost function for the BRP is

$$B(t) = \frac{C_1 H(t) + C_2}{t}$$

where H is the renewal function associated with F and $C_1 > C_2$. The objective is to minimize B in t. Assume that F is absolutely continuous so that the renewal density h $= H^{(1)}$ exists. Define

$$D(t) = \frac{t C_1 h(t) - C_1 H(t) - C_2}{t^2} ,$$

so that $D(t) = B^{(1)}(t)$. Suppose that B(t) has a unique minimum at $t = \phi^*$ with $D(\phi^*) = 0$. Let $\{a_n\}$ be a sequence of positive constants decreasing to zero. Given an initial estimate ϕ_1^* of ϕ^*, the Kiefer-Wolfowitz algorithm (1.3) suggests defining successive estimates of ϕ^* by the formula

$$\phi_{n+1}^* = \phi_n^* - a_n D_n(\phi_n^*)$$

where $D_n(\phi_n^*)$ is an estimate of $D(\phi_n^*)$. This procedure would have the disadvantage of possibly generating negative estimates when necessarily ϕ^* is greater than zero. If it is given that ϕ^* is greater than some known $\delta > 0$, then one might use the algorithm

$$\phi_{n+1}^* = \max\left(\delta, \phi_n^* - a_n D_n(\phi_n^*)\right).$$

Rather than constraining the estimates in this way, the following approach is taken.

Assume there exists a known δ, $0 < \delta < \phi^*$. Let g: $\mathbf{R} \to (\delta, \infty)$ be a known, continuous, strictly increasing function having s+1 bounded derivatives. Suppose ϕ

minimizes $B\circ g(t)$ in t, so that $\phi^* = g(\phi)$. Define

$$Dg(t) = \frac{g(t)C_1 H\circ g^{(1)}(t) \ - \ g^{(1)}(t)[\ C_1 H\circ g(t) + C_2\ ]}{[\ g(t)\ ]^2}.$$

Since $Dg(t) = B\circ g^{(1)}(t)$ and $Dg(\phi) = 0$, we now consider using a stochastic approximation procedure to estimate the zero of Dg. Given an initial estimate ϕ_1, define successive estimates of ϕ by

$$\phi_{n+1} = \phi_n \ - \ a_n Dg_n(\phi_n), \tag{2.1}$$

where $Dg_n(\phi_n)$ is an estimate of $Dg(\phi_n)$. Given ϕ_n, take $g(\phi_n)$ as an estimate of ϕ^*.

The next step is to define an estimator $Dg_n(t)$ of $Dg(t)$. Let $\{X_{ij}\}_{i,j=1}^{\infty}$ be independent and identically distributed random variables with distribution function F defined on a probability space $(\Omega, \mathcal{F}, P)$. Take $N_n = \{N_n(t), t \geq 0\}$ to be the renewal process associated with the sequence $\{X_{nj}\}_{j=1}^{\infty}$ with renewals occurring at X_{n1}, $X_{n1}+X_{n2}$, $X_{n1}+X_{n2}+X_{n3}$, ... for n = 1, 2, ... Let $\{c_n\}$ be a sequence of positive constants decreasing to zero. Define for $t \in \mathbf{R}$

$$hg_n(t) = \frac{N_n(g(t+c_n)) \ - \ N_n(g(t-c_n))}{2c_n},$$

an estimate of $H\circ g^{(1)}(t)$. Finally, let

$$Dg_n(t) = \frac{g(t)C_1 hg_n(t) \ - \ g^{(1)}(t)[\ C_1 N_n(g(t)) + C_2\ ]}{[\ g(t)\ ]^2}, \tag{2.2}$$

with estimates of ϕ defined recursively by (2.1).

The following assumptions will be employed.

A1: The distribution funtion F has increasing failure rate. F is absolutely continuous with bounded continuous density f, F(0)=0, and $\mu_2=\int_0^\infty t^2 dF(t) < \infty$.

A2: There exists known δ such that $0 < \delta < \phi^*$. For $t \in (\delta, \infty)$, $D(t)(t-\phi^*) > 0$ for $t \neq \phi^*$.

A3: $g: \mathbf{R} \rightarrow (\delta, \infty)$ is a known, continuous, strictly increasing function with s+1 bounded derivatives such that $g^{(2)}$ is continuous.

A4: $\{a_n\}$ and $\{c_n\}$ are sequences of positive constants such that $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = 0$, $\Sigma_{n=1}^\infty a_n = \infty$, $\Sigma_{n=1}^\infty a_n c_n < \infty$, and $\Sigma_{n=1}^\infty a_n^2/c_n < \infty$.

A5: Assume $H\circ g^{(1)}$ is continuous on **R**.

(a) $H\circ g^{(2)}(t)$ exists for each $t \in \mathbf{R}$ and is bounded.

(b) A5 (a) holds and $H\circ g^{(2)}$ is continuous in a neighborhood of ϕ.

(c) $H\circ g^{(2)}(t)$ exists for each $t \in \mathbf{R}$ and for some constant $K_1 > 0$

$$| H\circ g^{(2)}(x_1) - H\circ g^{(2)}(x_2) | \leq K_1 | x_1 - x_2 | \quad \text{for all } x_1, x_2 \in \mathbb{R}.$$

A5′: Assume $H\circ g^{(2)}$ is continuous on **R**.

(a) $H\circ g^{(3)}(t)$ exists for each $t \in \mathbf{R}$ and is bounded.

(b) A5′ (a) holds and $H\circ g^{(3)}$ is continuous at ϕ.

A6: For constants A, $C > 0$, let $\Gamma = A\frac{d}{dt}Dg(t)|_{t=\phi}$, $a_n = An^{-1}$, and $c_n = Cn^{-\gamma}$, where $\gamma \in (0, 1)$ is such that $1-\gamma < 2\Gamma$.

Theorems 2.1 and 2.2 give sufficient conditions for a.s. convergence and asymptotic normality of the estimates generated by the procedure.

Theorem 2.1: Assume A1-A4 hold. Further suppose that any one of the assumptions A5(a), A5(c), or A5′(a) holds. Then $\phi_n \to \phi$ a. s.

Theorem 2.2: Assume A1-A3, and A6 hold. Let $\Sigma = \dfrac{C_1^2 \, H \circ g^{(1)}(\phi)}{2[g(\phi)]^2}$. With Γ defined by A6, if either

(a) A5(b) holds with $\frac{1}{3} \le \gamma < 1$, or

(b) A5(c) or A5′(a) holds with $\frac{1}{5} < \gamma < 1$, then

$$n^{(1-\gamma)/2}(\phi_n - \phi) \xrightarrow{D} N\left(0, \frac{A^2 C^{-1} \Sigma}{2\Gamma - (1-\gamma)}\right).$$

(c) If A5′(b) holds with $\gamma = \frac{1}{5}$, then

$$n^{2/5}(\phi_n - \phi) \xrightarrow{D} N\left(-\frac{AC^2 C_1 H \circ g^{(3)}(\phi)}{6\, g(\phi)[\Gamma - \frac{2}{5}]}, \frac{A^2 C^{-1} \Sigma}{2\Gamma - \frac{4}{5}}\right).$$

2.3 Proofs

This section is devoted to proofs of Theorems 2.1 and 2.2. The following theorem due to Robbins and Siegmund(1971) is the main tool used in establishing Theorem 2.1.

Theorem(Robbins-Siegmund):

Let $\mathcal{F}_n$ be a nondecreasing sequence of sub σ-fields of $\mathcal{F}$. Suppose that X_n, β_n, η_n, and γ_n are nonnegative $\mathcal{F}_n$-measurable random variables such that

$$\mathbf{E}_{\mathcal{F}_n} X_{n+1} \le X_n(1+\beta_n) + \eta_n - \gamma_n \quad \text{for } n = 1, 2, \ldots .$$

Then, $\lim_{n\to\infty} X_n$ exists and $\sum_{n=1}^{\infty} \gamma_n < \infty$ on the set such that $\sum_{n=1}^{\infty} \beta_n < \infty$ and $\sum_{n=1}^{\infty} \eta_n < \infty$.

Before proceeding with the proofs, some simple consequences of the assumptions are noted.

For any sequence of real numbers $\{x_n\}$, let $Y_n = N_n(g(x_n+c_n)) - N_n(g(x_n-c_n))$. Assuming any part of A5 or A5$'$ holds, by an appropriate Taylor expansion, A1, and A4,

$$\begin{aligned} \mathbf{E} Y_n &= H(g(x_n+c_n)) - H(g(x_n-c_n)) \\ &= H \circ g^{(1)}(x_n) 2c_n + o(c_n) . \end{aligned}$$

By A1, for $k \ge 2$,

$$\begin{aligned} (2.3) \quad & \mathbf{P}\{Y_n = k\} \\ &\le \sum_{j=1}^{\infty} [\, F^{(j)}(g(x_n+c_n)) - F^{(j)}(g(x_n-c_n)) \,] \, [\, F(g(x_n+c_n) - g(x_n-c_n)) \,]^{k-1} \\ &= [\, H(g(x_n+c_n)) - H(g(x_n-c_n)) \,] \, [\, F(g(x_n+c_n) - g(x_n-c_n)) \,]^{k-1} \\ &\le K_1 c_n [F(g(x_n+c_n) - g(x_n-c_n))]^{k-1} \end{aligned}$$

Since $c_n \downarrow 0$, ultimately $F(g(x_n+c_n) - g(x_n-c_n)) < 1$, and thus for any integer $p > 0$,

$$\begin{aligned} (2.4) \qquad \mathbf{E}\{\, Y_n^p \, \chi[\, Y_n \ge 2 \,] \,\} &= K_1 \, c_n \sum_{k=2}^{\infty} k^p \, [\, F(g(x_n+c_n) - g(x_n-c_n)) \,]^{k-1} \\ &= o(c_n). \end{aligned}$$

The above implies that

$$\mathbf{P}\{ Y_n = 1 \} = H \circ g^{(1)}(x_n) 2c_n + o(c_n)$$

and that for any real number $p > 0$

(2.5)
$$\mathbf{E} Y_n^p = H \circ g^{(1)}(x_n) 2c_n + o(c_n).$$

We will also need some properties of renewal processes for which the underlying distribution has increasing failure rate. A lifetime distribution F is increasing (decreasing) failure rate (IFR and DFR, respectively) if $\log[1 - F(x)]$ is concave (convex) in x. If F has density f, F being IFR (DFR) is equivalent to the hazard rate

$$r(x) = \frac{f(x)}{1 - F(x)}$$

being non-decreasing (non-increasing) in x. IFR distributions are of particular importance in the study of maintenance policies, since increased probability of failure upon aging is the motivation for preventive replacement of parts.

Barlow, Marshall, and Proschan (1963) and later Barlow and Proschan (1964) explore properties of renewal processes with underlying IFR and DFR distributions. (See also Barlow and Proschan (1965) for some of this work.) Several of these properties will be used here. Suppose F is an IFR (DFR) distribution with mean μ_1 and associated renewal function H. Then

(2.6)
$$\mathbf{Var}\, N(t) \le (\ge)\; H(t) \le (\ge)\; \frac{t}{\mu_1}, \quad \text{for all } t \ge 0;$$

(2.7)
$$\mathbf{E}[N(t)]^k \le e^{-t/\mu} \sum_{j=0}^{\infty} \frac{j^k (t/\mu)^j}{j!}.$$

Define the σ-fields $\mathcal{F}_n = \sigma(X_{i,j},\ i=1,\ldots, n-1,\ j=1, 2,\ldots)$. Use $\mathbf{E}_{\mathcal{F}_n}$ to denote expectation with respect to the σ-field $\mathcal{F}_n$. Since N_{n-1} is $\mathcal{F}_n$-measurable, by (2.1) and (2.2) the estimate ϕ_n is $\mathcal{F}_n$-measurable. Define

$$\Delta_n = \mathbf{E}_{\mathcal{F}_n}[\, Dg_n(\phi_n) - Dg(\phi_n)\,] ,$$

and

$$V_n = c_n^{1/2}[\, Dg_n(\phi_n) - Dg(\phi_n) - \Delta_n\,] . \tag{2.8}$$

In Lemmas 2.1 and 2.2 bounds are derived for Δ_n and $\mathbf{E}_{\mathcal{F}_n}[Dg_n(\phi_n)]^2$, respectively, to be used in the application of the Robbins-Siegmund result. V_n plays an important role in the proof of Theorem 2.2. The asymptotic variance of V_n will be

$$\sum = C_1^2\, H\circ g^{(1)}(\phi)/2[g(\phi)]^2 .$$

<u>Lemma 2.1:</u> Assume A1, A2, A3, and A4 hold.

(a) If A5(a) holds, then $|\Delta_n| \leq K_1 c_n$.

(b) If A5(c) or A5$'$(a) holds, then $|\Delta_n| \leq K_2 c_n^2$.

<u>Proof:</u> By definition, since N_n is independent of $\mathcal{F}_n$ while ϕ_n is $\mathcal{F}_n$-measurable,

$$\Delta_n = C_1\left[\frac{H\circ g(\phi_n+c_n) - H\circ g(\phi_n-c_n)}{2c_n} - H\circ g^{(1)}(\phi_n)\right]\Big/ g(\phi_n) .$$

If A5(a) holds, then

$$H\circ g(\phi_n+c_n) - H\circ g(\phi_n-c_n) = 2c_n H\circ g^{(1)}(\phi_n) + \tfrac{1}{2}c_n^2[H\circ g^{(2)}(\eta_1) - H\circ g^{(2)}(\eta_2)],$$

with $|\eta_i-\phi_n|\leq c_n$, i= 1, 2. Thus, since $g(\phi_n)>\delta$,

$$|\Delta_n| \leq K_1|H\circ g^{(2)}(\eta_1) - H\circ g^{(2)}(\eta_2)|c_n \leq K_2 c_n.$$

If A5(c) holds, then

$$|\Delta_n| \leq K_3|H\circ g^{(2)}(\eta_1) - H\circ g^{(2)}(\eta_2)|c_n$$

$$\leq K_4 c_n^2 .$$

by the Taylor expansion for g. If A5′(a) holds, then

$$H\circ g(\phi_n+c_n) - H\circ g(\phi_n-c_n) = 2c_n H\circ g^{(1)}(\phi_n) + \tfrac{1}{6}c_n^3[H\circ g^{(3)}(\eta_1) + H\circ g^{(3)}(\eta_2)],$$

where again $|\eta_i - \phi_n| \leq c_n$, i= 1, 2. In this case

$$|\Delta_n| \leq \tfrac{1}{12}C_1|H\circ g^{(3)}(\eta_1) + H\circ g^{(3)}(\eta_2)|\, c_n^2 \Big/ g(\phi_n)$$

$$\leq K_5 c_n^2$$

completing the proof of part (b) of the lemma. □

Lemma 2.2: Assume A1-A4 hold. Further suppose that A5(a), A5(c), or A5′(a) holds. Then $\mathbf{E}_{\mathcal{F}_n}[Dg_n(\phi_n)]^2 \leq K_1/c_n$.

Proof: Using the definition of Dg_n and the fact that $(a+b)^2 \leq 4(a^2+b^2)$, one has that

$$Dg_n^2(\phi_n) \leq K_1\left[\frac{N_n(g(\phi_n+c_n)) - N_n(g(\phi_n-c_n))}{2c_n}\right]^2 \Big/ [\, g(\phi_n)\,]^2$$

$$+ K_2[g^{(1)}(\phi_n)]^2[C_1 N_n(g(\phi_n)) + C_2]^2 \Big/ [\, g(\phi_n)\,]^4.$$

By (2.5), A1, and A3,

$$\mathbf{E}_{\mathcal{F}_n}\left[\frac{N_n(g(\phi_n+c_n)) - N_n(g(\phi_n-c_n))}{2c_n}\right]^2 [g(\phi_n)]^{-2}$$

$$\leq K_3 \frac{1}{4c_n^2} [\, H \circ g^{(1)}(\phi_n)2c_n + o(c_n) \,]$$

$$\leq \frac{K_4}{c_n}.$$

Since F has increasing failure rate, it follows from (2.6) and A3 that

$$\mathbf{E}_{\mathcal{F}_n}[g^{(1)}(\phi_n)]^2[C_1N_n(g(\phi_n)) + C_2]^2 \Big/ [\, g(\phi_n) \,]^4$$

$$\leq \left\{K_5 + K_6\mathbf{E}_{\mathcal{F}_n}[N_n(g(\phi_n))]^2 \right\} \Big/ [\, g(\phi_n) \,]^4$$

$$\leq \left\{K_5 + K_7[g(\phi_n)] + K_8[g(\phi_n)]^2 \right\} \Big/ [\, g(\phi_n) \,]^4$$

$$\leq K_9.$$

Noting that $c_n = o(1)$, the lemma follows from the inequality for Dg_n^2 at the beginning of the proof. □

<u>Proof of Theorem 2.1:</u> Letting $U_n = \phi_n - \phi$, we have by (2.1) that

$$U_{n+1}^2 = U_n^2 + a_n^2 Dg_n^2(\phi_n) - 2a_nU_nDg_n(\phi_n).$$

Since ϕ_n is $\mathcal{F}_n$-measurable,

$$\mathbf{E}_{\mathcal{F}_n}U_{n+1}^2 = U_n^2 + a_n^2\mathbf{E}_{\mathcal{F}_n}Dg_n^2(\phi_n) - 2a_nU_n\mathbf{E}_{\mathcal{F}_n}[Dg_n(\phi_n)].$$

Noting that $\mathbf{E}_{\mathcal{F}_n}Dg_n(\phi_n) = Dg(\phi_n) + \Delta_n$, it follows that

$$\mathbf{E}_{\mathcal{F}_n} U_{n+1}^2 \leq U_n^2 + a_n^2 \mathbf{E}_{\mathcal{F}_n} Dg_n^2(\phi_n) + 2a_n|U_n||\Delta_n| - 2a_n(\phi_n - \phi)Dg(\phi_n).$$

If A5(a) holds, then by Lemma 2.1

$$|U_n\Delta_n| \leq K_1|U_n|c_n \leq [\, K_2 + K_3U_n^2 \,]c_n.$$

Similarly, under A5(c) or A5$'$(a),

$$|U_n\Delta_n| \leq K_4|U_n|c_n^2 \leq [\, K_5 + K_6U_n^2 \,]c_n,$$

since $c_n = o(1)$. Thus by Lemmas 2.1 and 2.2,

$$\mathbf{E}_{\mathcal{F}_n} U_{n+1}^2 \leq U_n^2 + a_n^2\, K_1/c_n + 2a_n[\, K_2 + K_3U_n^2 \,]\, c_n - 2a_n(\phi_n - \phi)Dg(\phi_n)$$

$$= U_n^2[\, 1 + 2K_3a_nc_n \,] + [\, K_1a_n^2/c_n + 2K_2a_nc_n \,] - 2a_n(\phi_n - \phi)Dg(\phi_n).$$

By A2 and A4 the assumptions of the Robbins-Siegmund result are met. Thus for some finite random variable ξ one has that $\phi_n \to \xi$ a.s. and $\sum_{n=1}^{\infty} a_n(\phi_n - \phi)Dg(\phi_n) < \infty$. Since $(\phi_n - \phi)Dg(\phi_n) > 0$ for $\phi_n \neq \phi$ by A2 and $\sum_{n=1}^{\infty} a_n = \infty$ by A4, $\phi_n \to \phi$ a.s. □

Before beginning a sequence of lemmas that leads to the proof of Theorem 2.2, we quote a theorem of Fabian(1968) on asymptotic normality of stochastic approximation procedures. The univariate version of this theorem given here is taken from Frees(1983).

Theorem (Fabian):

Suppose $\mathcal{F}_n$ is a nondecreasing sequence of sub σ-fields of $\mathcal{F}$. Suppose U_n, V_n, T_n, Γ_n, and Φ_n are random variables such that Γ_n, Φ_{n-1}, and V_{n-1} are $\mathcal{F}_n$-measurable. Let α, β, T, Σ, Γ, and Φ be real constants with $\Gamma > 0$ such that

$$\Gamma_n \to \Gamma, \quad \Phi_n \to \Phi, \quad T_n \to T \text{ or } \mathbf{E}|T_n - T| \to 0, \quad \mathbf{E}_{\mathfrak{F}_n} V_n = 0,$$

and there exists a constant C such that $C > |\mathbf{E}_{\mathfrak{F}_n} V_n^2 - \Sigma| \to 0$.

Suppose, with $\sigma_{j,r}^2 = \mathbf{E}\, \chi[\, V_j^2 \geq r j^\alpha \,]\, V_j^2$, that

$$\lim_{j \to \infty} \sigma_{j,r}^2 = 0 \text{ for all r}$$

or

$$\alpha = 1 \text{ and } \lim n^{-1} \sum_{j=1}^{n} \sigma_{j,r}^2 = 0 \quad \text{for all r.}$$

Let $\beta_+ = \beta\, \chi[\, \alpha = 1 \,]$ where $0 < \alpha \leq 1$, $\quad 0 \leq \beta$, $\quad \beta_+ < 2\Gamma$, and

$$U_{n+1} = U_n\, [\, 1 - n^{-\alpha}\Gamma_n \,] + n^{-(\alpha+\beta)/2} \Phi_n V_n + n^{-\alpha-\beta/2} T_n .$$

Then,

$$n^{\beta/2} U_n \xrightarrow{D} N\Big(\frac{T}{(\Gamma - \beta_+/2\,)}, \ \frac{\Sigma\Phi^2}{(2\Gamma - \beta_+\,)} \Big).$$

<u>Lemma 2.3:</u> Assume A1-A3 and A6 hold.

(a) If A5(b) holds, then $\Delta_n = o(c_n)$.

(b) If A5(c) or A5$'$(a) holds, then $\Delta_n = O(c_n^2)$.

(c) If A5$'$(b) holds, then $\lim_{n \to \infty} \Delta_n / c_n^2 = \dfrac{C_1 \; H \circ g^{(3)}(\phi)}{6\; g(\phi)}$.

<u>Proof:</u> If A5(b) holds, then from the proof of Lemma 2.1 and since $g(\phi_n) > \delta$,

$$\begin{aligned} |\Delta n| \quad &\leq \tfrac{1}{4} C_1 |H \circ g^{(2)}(\eta_1) - H \circ g^{(2)}(\eta_2)| c_n \Big/ [\, g(\phi_n)\,] \\ &\leq K_1 |H \circ g^{(2)}(\eta_1) - H \circ g^{(2)}(\eta_2)| c_n \end{aligned}$$

with $|\eta_i - \phi_n| \le c_n$, i= 1, 2. Since $|\phi_n - \phi| \to 0$ a.s. by Theorem 2.1 and $H \circ g^{(2)}$ is continuous in a neighborhood of ϕ, part (a) holds.

Part (b) is identical to Lemma 2.1 (b).

Also from the proof of Lemma 2.1, when A5′(b) holds,

$$\Delta_n = \tfrac{1}{12} C_1 [\, H \circ g^{(3)}(\eta_1) + H \circ g^{(3)}(\eta_2) \,]\, c_n^2 \Big/ g(\phi_n)$$

with $|\eta_i - \phi_n| \le c_n$, i= 1, 2. Since $\phi_n \to \phi$ a.s., (c) follows when $H \circ g^{(3)}$ is continuous in a neighborhood of ϕ. □

<u>Lemma 2.4:</u> Assume A1-A3 . If A5(b), A5(c), A5′(a), or A5′(b) holds, then for $t \ge 1$,

$$\mathbf{E}_{\mathcal{F}_n}|V_n|^t \le c_n^{t/2}[\, K_1 + K_2 \mathbf{E}_{\mathcal{F}_n}|hg_n(\phi_n)|^t + K_3 |\, \mathbf{E}_{\mathcal{F}_n} hg(\phi_n)\, |^t \,].$$

where V_n is defined by (2.8).

<u>Proof:</u> By definition

$$V_n = c_n^{1/2} C_1 \left\{ \frac{[\, hg_n(\phi_n) - \mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n)\,]}{g(\phi_n)} - \frac{g^{(1)}(\phi_n)[\, N_n(g(\phi_n)) - H(g(\phi_n))\,]}{[\, g(\phi_n)]^2} \right\}.$$

Since for any real numbers a, b, c, and d, $|a+b+c+d|^t \le 4^t\{|a|^t + |b|^t + |c|^t + |d|^t\}$,

$$|V_n|^t \le c_n^{t/2} K_1 \Big\{ |hg_n(\phi_n)|^t [g(\phi_n)]^{-t} + |\mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n)|^t [g(\phi_n)]^{-t}$$
$$+ |N_n(g(\phi_n))|^t [g(\phi_n)]^{-2t} + |H(g(\phi_n))|^t [g(\phi_n)]^{-2t} \Big\}.$$

Since F has increasing failure rate, by (2.7)

$$E_{\mathcal{F}_n} | N_n(g(\phi_n)) |^t [g(\phi_n)]^{-2t} \le E_{\mathcal{F}_n} | N_n(g(\phi_n)) |^{[t+1]} [g(\phi_n)]^{-2t}$$

$$\le [g(\phi_n)]^{-2t} e^{-[g(\phi_n)/\mu]} \sum_{j=0}^{\infty} \frac{j^{[t+1]}}{j!} \left[\frac{g(\phi_n)}{\mu}\right]^j$$

where [t+1] is the greatest integer less than or equal to $t+1$. Thus

$$E_{\mathcal{F}_n} \frac{| N_n(g(\phi_n)) |^t}{[g(\phi_n)]^{2t}} \le e^{-[g(\phi_n)/\mu]} \sum_{j=1}^{\infty} \frac{j^{[t+1]} [g(\phi_n)]^{j-2t}}{j! \ \mu^j}$$

$$\le e^{-[g(\phi_n)/\mu]} \sum_{j=1}^{[t+1]} \frac{j^{[t+1]} [g(\phi_n)]^{j-2t}}{j! \ \mu^j}$$

$$+ e^{-[g(\phi_n)/\mu]} \sum_{j=[t+1]+1}^{\infty} \frac{j^{[t+1]} [g(\phi_n)]^{j-[t+1]}}{j! \ \mu^j}$$

$$\le K_2$$

since $[t+1] \le 2t$ for $t \ge 1$. By (2.6) and A2,

$$[H(g(\phi_n))]^t [g(\phi_n)]^{-2t} \le [g(\phi_n)/\mu]^t [g(\phi_n)]^{-2t} \le K_3.$$

Hence, again using the fact that $g(\phi_n) > \delta$,

$$E_{\mathcal{F}_n} |V_n|^t \le c_n^{t/2} [K_4 E_{\mathcal{F}_n} |hg_n(\phi_n)|^t + K_4 | E_{\mathcal{F}_n} [hg(\phi_n)] |^t + K_5 + K_6],$$

which establishes the result. □

Lemma 2.5: Assume A1-A3 and A6 hold. Further suppose that one of the assumptions A5(b), A5(c), A5′(a), or A5′(b) holds. Then

(a) $\mathbf{E}_{\mathcal{F}_n} V_n = 0$ and

(b) $\lim_{n\to\infty} \mathbf{E}_{\mathcal{F}_n} V_n^2 = \Sigma = \dfrac{C_1^2\, H\circ g^{(1)}(\phi)}{2\,[g(\phi)]^2}$.

Proof: For part (a) we note that $\mathbf{E}_{\mathcal{F}_n} V_n = 0$ by definition.

Next, let

$$X_n = [\, g(\phi_n)\,]^{-1} \left[\frac{N_n(g(\phi_n+c_n)) - N_n(g(\phi_n-c_n))}{2c_n} \right],$$

$$Y_n = [\, g(\phi_n)\,]^{-1} \left[\frac{H\circ g(\phi_n+c_n) - H\circ g(\phi_n-c_n)}{2c_n} \right], \text{ and}$$

$$Z_n = g^{(1)}(\phi_n)[\, g(\phi_n)\,]^{-2}[\, N_n(g(\phi_n)) - H\circ g(\phi_n)\,],$$

so that $V_n^2 = c_n C_1^2[\, X_n - Y_n - Z_n\,]^2$. To show (b), we consider separately the expectations $\mathbf{E}_{\mathcal{F}_n} X_n^2$, $\mathbf{E}_{\mathcal{F}_n} X_n Y_n$, etc.

(i) By (2.5), Theorem 2.1, and the continuity of $H\circ g^{(1)}$,

$$c_n \mathbf{E}_{\mathcal{F}_n} X_n^2 = c_n\, [g(\phi_n)]^{-2}\, \frac{1}{4c_n^2}\, [\, H\circ g^{(1)}(\phi_n) 2c_n + o(c_n)\,]$$

$$\longrightarrow \frac{H\circ g^{(1)}(\phi)}{2\,[g(\phi)]^2} \text{ as } n \longrightarrow \infty.$$

(ii) Using (2.5) and the fact that $g(\phi_n) > \delta$,

$$c_n \mathbf{E}_{\mathcal{F}_n} X_n Y_n = c_n^{-1}[g(\phi_n)]^{-2}[H\circ g(\phi_n+c_n) - H\circ g(\phi_n-c_n)][H\circ g^{(1)}(\phi_n)\, 2c_n + o(c_n)]/4$$

$$\longrightarrow 0$$

as $n \longrightarrow \infty$ by the continuity of H.

(iii) By the conditional version of the Cauchy-Schwarz inequality,

$$c_n \mathbf{E}_{\mathcal{F}_n}|X_n Z_n|$$

$$= [g(\phi_n)]^{-3} g^{(1)}(\phi_n) \tfrac{1}{2} \mathbf{E}_{\mathcal{F}_n}[\, N_n(g(\phi_n+c_n)) - N_n(g(\phi_n-c_n))\,]\,|N_n(g(\phi_n)) - H\circ g(\phi_n)|$$

$$\leq K_1 \left\{ \mathbf{E}_{\mathcal{F}_n}[\, N_n(g(\phi_n+c_n)) - N_n(g(\phi_n-c_n))\,]^2 \right\}^{1/2}$$

$$\cdot \left\{ \frac{\mathbf{E}_{\mathcal{F}_n}[\, N_n(g(\phi_n)) - H\circ g(\phi_n)\,]^2}{g(\phi_n)} \right\}^{1/2} .$$

Thus by (2.5) and (2.6),

$$c_n \mathbf{E}_{\mathcal{F}_n}|X_n Z_n| \leq K_1\{\, 2H\circ g^{(1)}(\phi_n)c_n + o(c_n)\,\}^{1/2}\left[\frac{g(\phi_n)}{\mu} \cdot \frac{1}{g(\phi_n)}\right]^{1/2}$$

$$\leq K_2\, o(1) \text{ as } n \to \infty.$$

Thus $c_n \mathbf{E}_{\mathcal{F}_n}|X_n Z_n| \longrightarrow 0$ as $n \longrightarrow \infty$.

(iv) Again by (2.5),

$$c_n \mathbf{E}_{\mathcal{F}_n} Y_n^2 = [g(\phi_n)]^{-2} \frac{1}{4c_n} [\, H\circ g(\phi_n+c_n) - H\circ g(\phi_n-c_n)\,]^2$$

$$\leq K_1 \frac{1}{c_n} [\, 2H\circ g^{(1)}(\phi_n)c_n + o(c_n)\,]^2 \to 0 \text{ as } n \to \infty .$$

(v) For all n, $\mathbf{E}_{\mathcal{F}_n} Y_n Z_n = 0$.

(vi) It follows from (2.6) that

$$\mathbf{E}_{\mathcal{F}_n} Z_n^2 = [\, g(\phi_n)\,]^{-4}\, [\, g^{(1)}(\phi_n)\,]^2\, \mathbf{E}_{\mathcal{F}_n}[\, N_n(g(\phi_n)) - H \circ g(\phi_n)\,]^2$$

$$\leq K_1\, [\, g(\phi_n)\,]^{-4} \left[\frac{g(\phi_n)}{\mu}\right]$$

$$\leq K_2.$$

Thus $\lim_{n\to\infty} c_n \mathbf{E}_{\mathcal{F}_n} Z_n^2 = 0$.

The lemma follows immediately from (i)-(vi) since by definition

$$\mathbf{E}_{\mathcal{F}_n} V_n^2 = c_n C_1^2\, \mathbf{E}_{\mathcal{F}_n}[\, X_n - Y_n - Z_n\,]^2. \qquad \square$$

<u>Lemma</u> <u>2.6:</u> Assume A1-A3 and A6 hold. Further suppose that one of the assumptions A5(b), A5(c), A5′(a), or A5′(b) holds. Then

(a) there exists $K > 0$ such that $\mathbf{E}_{\mathcal{F}_n} V_n^2 \leq K$ for all n, and

(b) for any $t \geq 1$, $\lim_{n\to\infty} \mathbf{E}_{\mathcal{F}_n} (c_n^{1/2} V_n)^t = 0$.

<u>Proof:</u> Since the conditions of Lemma 2.4 hold,

$$\mathbf{E}_{\mathcal{F}_n} V_n^2 \leq c_n\, [\, K_1 + K_2 \mathbf{E}_{\mathcal{F}_n} |hg_n(\phi_n)|^2 + K_3 |\, \mathbf{E}_{\mathcal{F}_n}[hg(\phi_n)]\,|^2\,].$$

By A1 and (2.5)

$$c_n \mathbf{E}_{\mathcal{F}_n} |hg_n(\phi_n)|^2 = \frac{1}{4c_n}\, [\, 2H \circ g^{(1)}(\phi_n) c_n + o(c_n)\,] \;\leq\; K_1 .$$

Again by A1 and (2.5),

$$c_n |\, \mathbf{E}_{\mathcal{F}_n}[hg_n(\phi_n)]\,|^2 = \frac{1}{4c_n}\, [\, 2H \circ g^{(1)}(\phi_n) c_n + o(c_n)\,]^2 \leq K_2 .$$

This establishes (a).

By the definition of hg_n and (2.5) one has that

$$c_n^t \, \mathbf{E}_{\mathcal{F}_n}|hg_n(\phi_n)|^t \le \frac{1}{2^t} \, \mathbf{E}_{\mathcal{F}_n}[\, N_n(g(\phi_n+c_n)) - N_n(g(\phi_n-c_n)) \,]^{[t+1]}$$

$$= \frac{1}{2^t} [\, 2H \circ g^{(1)}(\phi_n)c_n + o(c_n) \,].$$

Thus $\lim_{n\to\infty} c_n^t \, \mathbf{E}_{\mathcal{F}_n}|\, hg_n(\phi_n)|^t = 0.$

Also by (2.5),

$$c_n^t \,|\, \mathbf{E}_{\mathcal{F}_n}[hg(\phi_n)] \,|^t = \frac{1}{2^t} [2H \circ g^{(1)}(\phi_n)c_n + o(c_n)]^t \to 0.$$

This establishes (b) using Lemma 2.4. □

Lemma 2.7: Let $\sigma_{n,r}^2 = \mathbf{E}\{ V_n^2 \, \chi[V_n^2 \ge rn] \}$ for $r = 1, 2, \ldots$ and $n = 1, 2, \ldots$. Assume Lemma 2.6 and A6 hold. Then for each r, $\lim_{n\to\infty} \sigma_{n,r}^2 = 0$.

Proof: For the $\gamma \in (0, 1)$ of assumption A6, take $q > 1$ such that $\gamma \le \frac{1}{q}$. Then take p such that $\frac{2}{p} + \frac{1}{q} = 1$. The proof is then identical to that of Lemma 2.7 of Frees(1983), using the fact that by Lemma 2.6, $\mathbf{E}(c_n^{1/2} V_n)^p = o(1)$. □

Proof of Theorem 2.2: We apply the result due to Fabian(1968) in the form quoted earlier.

By the definition of V_n

$$c_n^{-1/2} V_n = Dg_n(\phi_n) - Dg(\phi_n) - \Delta_n.$$

Expanding $Dg(\phi_n)$ about $Dg(\phi) = 0$ one has that

$$Dg(\phi_n) = (\phi_n - \phi)\ \tfrac{d}{dt}Dg(t)|_{t=\eta_n},$$

where $|\eta_n - \phi| \le |\phi_n - \phi|$. As before let $U_n = \phi_n - \phi$. Then

$$U_{n+1} = U_n - a_n[\, c_n^{-1/2} V_n + Dg(\phi_n) + \Delta_n \,]$$

$$= U_n\,[\,1 - a_n \tfrac{d}{dt}Dg(t)|_{t=\eta_n}\,] - An^{-1}C^{-1/2}n^{\gamma/2}V_n - a_n\Delta_n.$$

Set $\Gamma_n = A\,\tfrac{d}{dt}Dg(t)|_{t=\eta_n}$, $\Phi_n = \Phi = -\,AC^{-1/2}$, $T_n = -\,An^{(1-\gamma)/2}\Delta_n$, $\alpha = 1$, and $\beta = 1-\gamma$. Then

$$U_{n+1} = U_n\,[\,1 - n^{-1}\Gamma_n\,] + n^{(\frac{\gamma}{2}-1)}\Phi_n V_n + n^{-1}n^{-(1-\gamma)/2}T_n$$

$$= U_n\,[\,1 - n^{-\alpha}\Gamma_n\,] + n^{-(\alpha+\beta)/2}\Phi_n V_n + n^{-\alpha-(\beta/2)}T_n.$$

Let $\Gamma = A\,\tfrac{d}{dt}Dg(t)|_{t=\phi}$. Since $\tfrac{d}{dt}Dg(t)$ is continuous at ϕ and $\phi_n \to \phi$ a.s. by Theorem 2.1, it follows that $\Gamma_n \to \Gamma$ as $n \to \infty$. With $\Sigma = \dfrac{C_1^2\ H\circ g^{(1)}(\phi)}{2\,[g(\phi)]^2}$, by Lemmas 2.5 and 2.6 there exists a constant K such that $|\mathbf{E}_{\mathcal{F}_n} V_n^2 - \Sigma| < K$ for all n and $\mathbf{E}_{\mathcal{F}_n} V_n^2 \to \Sigma$ as $n \to \infty$. Using Lemma 2.7, $\lim_{j\to\infty} \sigma_{j,r}^2 = \lim_{j\to\infty} \mathbf{E}\{\, V_j^2\ \chi[\, V_j^2 \ge rj^{\alpha}]\,\} = 0$ for all r. By A6, $1-\gamma < 2\Gamma$.

It remains to establish the asymptotic behaviour of T_n. If A5(b) holds and $\gamma \ge \frac{1}{3}$, then by Lemma 2.3

$$|T_n| \le K_1 n^{(1-\gamma)/2} o(n^{-\gamma}) = K_1 n^{(1-3\gamma)/2} o(1) \to 0.$$

Under A5(c) or A5′(a), with $\gamma > \frac{1}{5}$,

$$|T_n| \le K_1 n^{(1-\gamma)/2} O(n^{-2\gamma}) = K_2 n^{(1-5\gamma)/2} \to 0.$$

If $A5'(b)$ holds and $\gamma = \frac{1}{5}$, then

$$T_n = -\frac{A\, c_n^2\, n^{(1-\gamma)/2} \Delta_n}{c_n^2} = -\frac{A\, C^2 \Delta_n}{c_n^2} \to -\frac{AC^2 C_1 H \circ g^{(3)}(\phi)}{6\, g(\phi)}$$

by Lemma 2.3. The limiting normal distributions indicated in Theorem 2.2 then follow from the Fabian result. □

2.4 Random Block Replacement Policies

For the estimation procedure of Section 2.2 it is of course not necessary to observe in entirety the renewal processes $N_n = \{N_n(t),\ t \ge 0\}$, $n \ge 1$; one needs only the randomly truncated processes $\{N_n(t),\ t \in [0, g(\phi_n + c_n)]\}$, $n \ge 1$. Thus in practice one could make planned replacements at time intervals of random length $g(\phi_n + c_n)$ and also make replacements at any intervening failures, possibly reducing costs.

Consider those procedures for estimating ϕ^* which, like the SA algorithm just introduced, can be carried out concurrently with a random BRP. This means that given a planned replacement at time t, the next planned replacement will be made at time $t + T_n$, where T_n is a random variable observable at time t. T_n may be an estimate of ϕ^*. For such a procedure let $N_F(t)$ and $N_P(t)$ denote the number of failure and planned replacements through time t. Define $R(t) = C_1 N_F(t) + C_2 N_P(t)$. Then $R(t)/t$ is the incurred cost per unit time of using the procedure through time t. One desirable property that such a procedure might possess is that $\lim_{t \to \infty} R(t)/t = B(\phi^*)$, the minimum possible expected long-run cost per unit time. The following theorem shows that this is indeed the case when $T_n \to \phi^*$ a.s., F has increasing failure rate, and H is continuous at

ϕ^*. The result and proof are similar to Theorem 1.1 of Frees(1983), which dealt with random age replacement policies.

For the remainder of this chapter let $\{X_{ij}\}_{i,j=1}^{\infty}$ be independent and identically distributed random variables with distribution function F defined on a probability space $(\Omega, \mathfrak{F}, P)$. Take $N_n = \{ N_n(t), t \geq 0 \}$ to be the renewal process associated with the sequence $\{X_{nj}\}_{j=1}^{\infty}$ with renewals occurring at $X_{n1}, X_{n1}+X_{n2}, X_{n1}+X_{n2}+X_{n3}, \ldots$ for $n = 1, 2, \ldots$. Define the σ-fields $\mathfrak{F}_n = \sigma(X_{ij}, i=1, 2, \ldots, n-1, j=1, 2, \ldots)$.

<u>Theorem</u> <u>2.3:</u> Assume F has increasing failure rate and finite mean μ_1. Suppose $\{ T_n \}$ is a sequence of random variables such that T_n is $\mathfrak{F}_n$-measurable and $T_n \to \phi^*$ a.s. Further assume H is continuous at ϕ^*. Define $R_n = C_1 \sum_{j=1}^{n} N_j(T_j) + C_2 n$, the incurred cost through the time of the n-th planned replacement. With N(t) the number of planned replacements through time t,

$$\lim_{t\to\infty} \frac{R_{N(t)}}{t} = B(\phi^*) .$$

<u>Proof:</u> Let $U_n = \sum_{j=1}^{n} [N_j(T_j) - H(T_j)]$, so that $\{ U_n, \mathfrak{F}_{n+1} \}$ is a martingale. By the increasing failure rate assumption and (2.6),

$$\sum_{n=1}^{\infty} E_{\mathfrak{F}_n} [U_{n+1} - U_n]^2 / n^2 \leq \frac{1}{\mu_1} \sum_{n=1}^{\infty} \frac{T_n}{n^2} \leq \infty \text{ a.s.}$$

This implies that $U_n/n \to 0$ a.s. (See Theorem 5 of Chow (1965)). By the continuity of H at ϕ^* and since $T_n \to \phi^*$ a.s.,

$$n^{-1} \sum_{j=1}^{n} H(T_j) \to H(\phi^*) \text{ a.s.}$$

Hence it follows that

$$n^{-1}\sum_{j=1}^{n} N_j(T_j) \to H(\phi^*) \text{ a.s.}$$

and

$$R_n/n \to C_1 H(\phi^*) + C_2 \text{ a.s.}$$

Define $L_n = \sum_{j=1}^{n} T_j$, the time of the n-th planned replacement. Necessarily,

$$L_{N(t)} \le t \le L_{N(t)+1}.$$

Thus the inequality

$$\frac{N(t)}{N(t)+1} \frac{R_{N(t)}/N(t)}{L_{N(t)+1}/[N(t)+1]} \le \frac{R_{N(t)}}{t} \le \frac{R_{N(t)}/N(t)}{L_{N(t)}/N(t)}$$

holds. (This is identical in form to (1.8) of Frees(1983).) Thus with probability one $T_n \to \phi^*$ and $L_n/n \to \phi^*$. The theorem follows by letting $t \to \infty$. □

2.5 A Critical Assumption

This section concerns the interrelationship between the existence of an optimal BRP and assumption A2. At this time we have not located in the literature any general characterization of when an optimal BRP exists. Thus we emphasize by means of examples the dependence of a unique optimal policy on the relative magnitudes of the costs C_1 and C_2. Throughout the remainder of this section we assume that the derivative function D(t) is continuous in t.

Assume A2 holds, so that ϕ^* minimizes B(t) among all finite values of t. A2 also implies that B(t) attains no local maximum and thus $\lim_{t\to\infty} B(t) > B(\phi^*)$. Thus ϕ^* is the optimal BRP and is to be preferred to only making part replacements at failure. Of course A2 is not necessary for the existence of an optimal BRP, as can be seen in an example that will be presented shortly. However, it does play an important role in the proof of the strong convergence of the algorithm.

The important part of A2 can be written as the requirement that

(2.9) $t h(t) - H(t) < C_2/C_1$ for $t < \phi^*$ and $t h(t) - H(t) > C_2/C_1$ for $t > \phi^*$,

which expresses how the existence of an optimal BRP partially depends on the ratio of the costs C_1 and C_2. To see this more explicitly, we will examine when (2.9) holds for three lifetime distributions of the gamma type. The probability density function for these distributions is given by

$$f(x) = \begin{cases} x^{\nu-1} e^{-x}/\Gamma(\nu), & x > 0, \\ 0, & x \leq 0, \end{cases}$$

with $\nu = 2$, 4, and 6.

Figures 2.1(a), 2.1(b), and 2.1(c) show the function $t h(t) - H(t)$ for the gamma distributions with shape parameters 2, 4, and 6, respectively. It is immediately apparent that in all three cases the ratio C_2/C_1 must be sufficiently small for (2.9) and A2 to be satisfied. Also, if C_2/C_1 is too large, then $D(t)$ is negative for all finite values of t and no finite optimal BRP exists. It can be seen from the figures that for certain intermediate values of C_2/C_1, the function $D(t)$ may have at least two zeros. One of these could represent an optimal BRP, but since A2 fails to hold, the proposed stochastic approximation algorithm may also fail.

The examples suggest the following general observations. If the line $y = C_2/C_1$ intersects the function $t h(t) - H(t)$ exactly once, there exists a finite optimal BRP and the critical expression (2.9) holds. Otherwise, an optimal policy may or may not exist, and if it does the stochastic approximation procedure defined in Section 2.2 may well fail to be consistent. Thus we can imagine that for many lifetime distributions failure replacements must be significantly more expensive than planned replacements for the procedure to be useful.

Figure 2.1 (a): gamma(1,2)

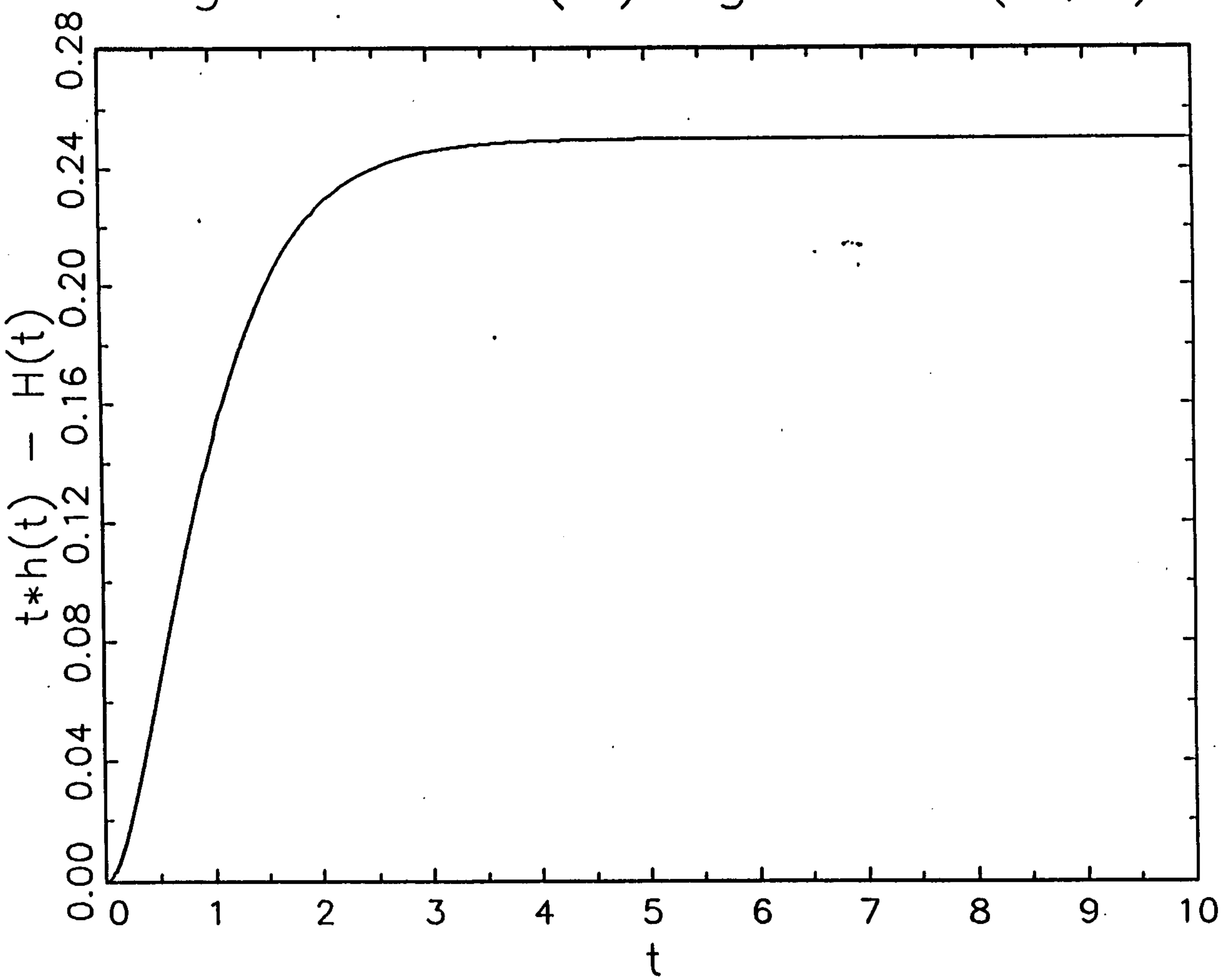


Figure 2.1(b): gamma(1,4)

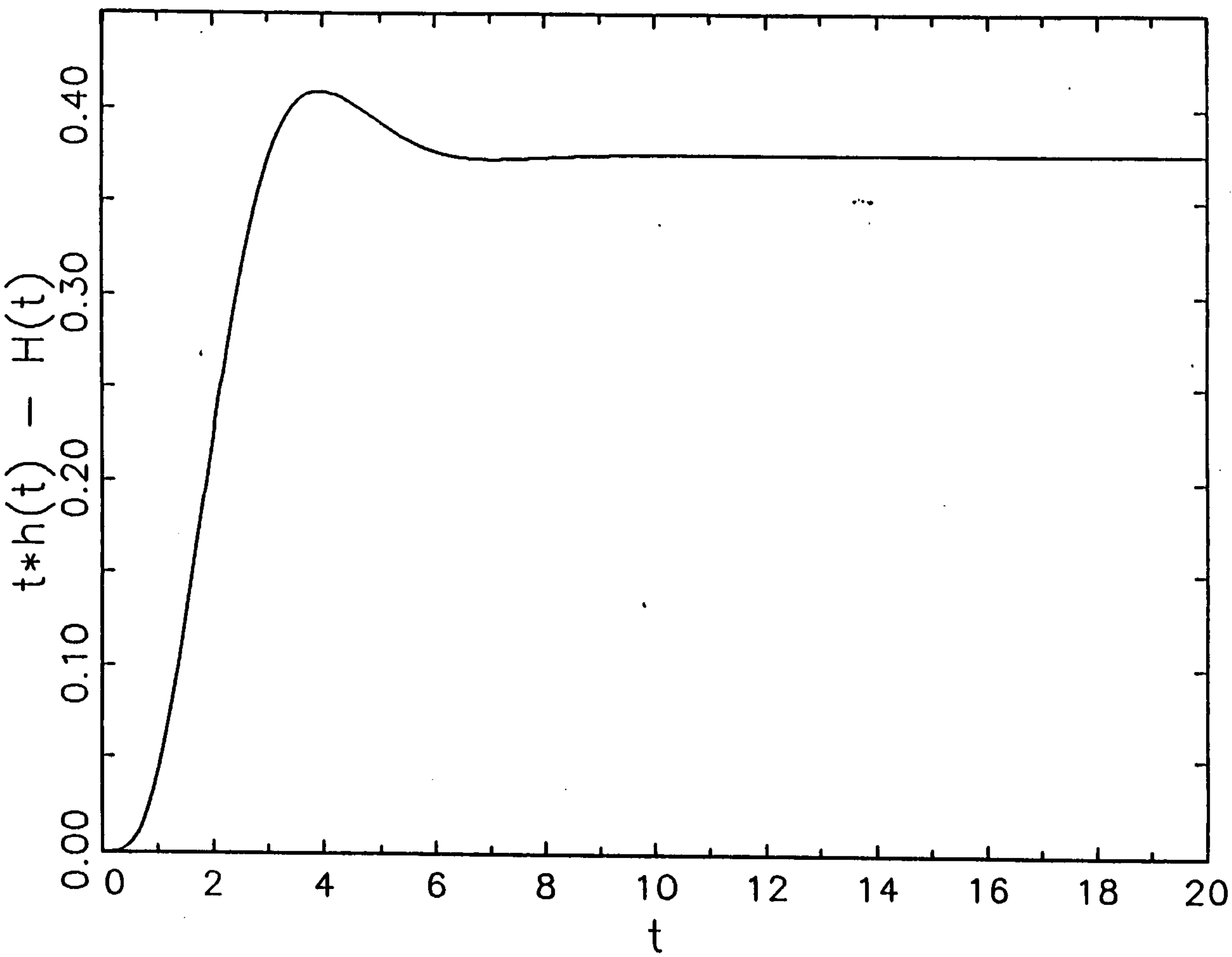
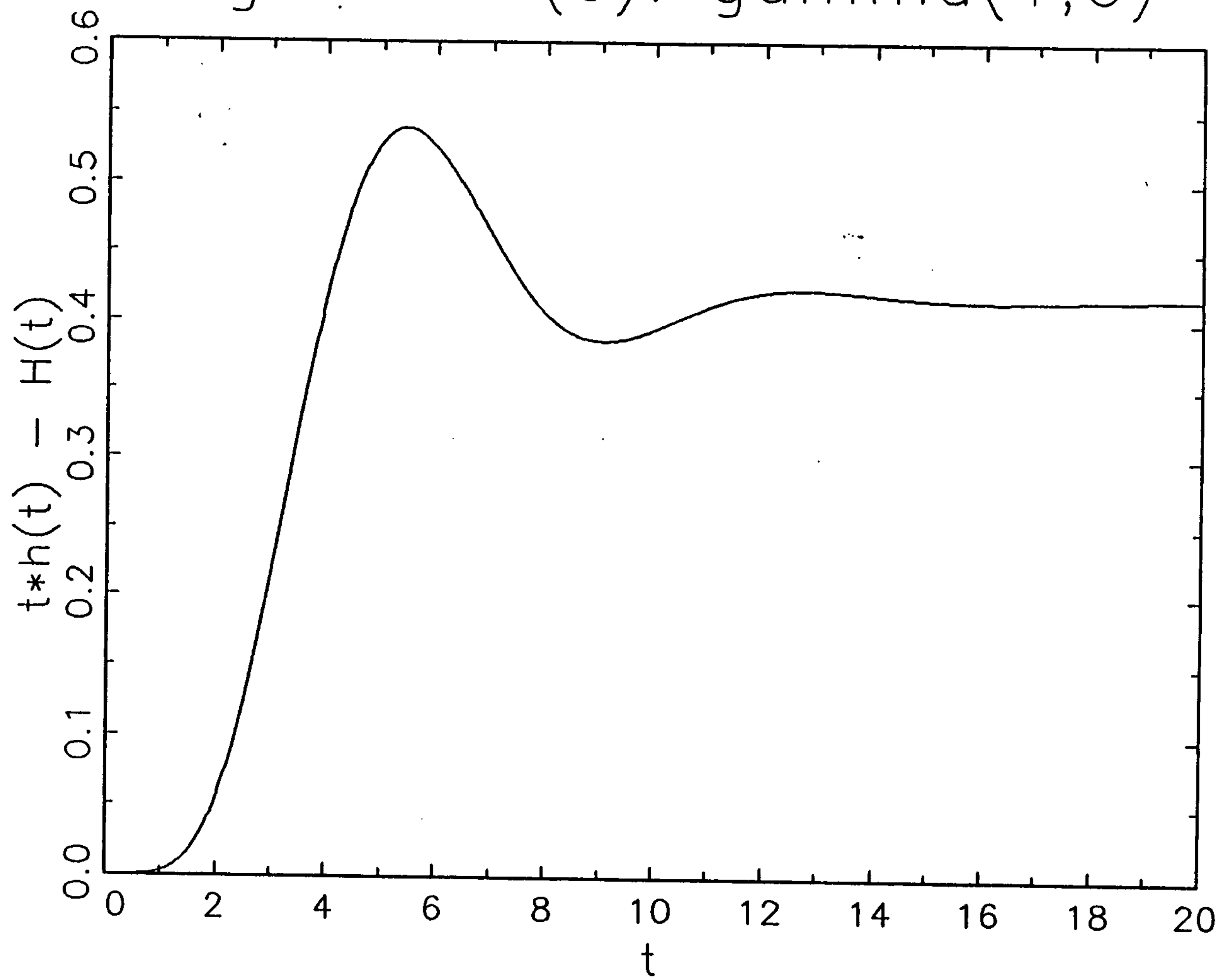


Figure 2.1(c): gamma(1,6)



CHAPTER 3

REFINEMENTS AND A FULLY SEQUENTIAL ESTIMATION PROCEDURE

3.1 Introduction

In this chapter we consider several refinements to the procedure introduced in Section 2.2.

In Section 3.2 the simple renewal density estimator used in (2.2) is replaced by a more general kernel estimator. Under certain conditions this modification reduces the bias at each step of the procedure and consequently results in a faster rate of convergence.

As seen in Theorem 2.2, the asymptotic mean square error of the estimate depends on the tuning constants A and C, which can be chosen by the user. Sections 3.3 and 3.4 address the problem of an adaptive stochastic approximation procedure that estimates the optimal choices of A and C so as to achieve the minimum possible asymptotic mean square error.

Finally, in Section 3.5 construction of a fixed-width confidence interval for the optimal block replacement policy is considered. This leads to the introduction of a stopping time which can provide guidance in deciding when the algorithm has been iterated a sufficient number of times for a desired level of confidence.

3.2 Improving The Rate Of Convergence

This section is principally concerned with achieving a faster rate of convergence for the procedure (2.1) by using a kernel estimator for $Hog^{(1)}$. Kernel estimation of $Hog^{(p+1)}$ for integer p greater than zero is also considered here for applications in Sections 3.4 and 3.5. The basic approach is that taken by Singh (1977) for kernel estimation of derivatives of a distribution function. Frees (1983) employed estimators like those suggested by Singh to speed the rate of convergence of a stochastic approximation procedure. The development undertaken here is like that of Frees except that interest now lies in estimating derivatives of Hog.

Let $0 \le p < r$ where p and r are integers. We will consider the problem of estimating $Hog^{(p+1)}(x)$ assuming that $Hog^{(r+1)}(x)$ exists and is bounded. Let hg_n continue to represent an estimator of $Hog^{(1)}$. For $p \ge 0$, $hg_n^{(p)}$ will denote an estimator of $Hog^{(p+1)}$. It is desirable to have estimators hg_n and $hg_n^{(p)}(x)$ which satisfy the following conditions:

A7: (a) $\sup_x | \mathbf{E} hg_n^{(p)}(x) - Hog^{(p+1)}(x) | = O(n^{-(r-p)/(2r+1)})$.

(b) For $t > -1$, $\sup_x \mathbf{E} | hg_n(x) |^{t+1} = O(n^{+t/(2r+1)})$.

A8: For any integer $t \ge 0$ and sequence $\{x_n\}$ such that $\lim_{n \to \infty} x_n = x_o$, there exist constants Φ_1 and $\Phi_{2,t}$ such that

(a) $\lim_{n \to \infty} n^{(r-p)/(2r+1)} [\mathbf{E} hg_n^{(p)}(x_n) - Hog^{(p+1)}(x_n)] = \Phi_1$ and

(b) $\lim_{n \to \infty} n^{-t/(2r+1)} \mathbf{E}[hg_n(x_n)]^{t+1} = \Phi_{2,t}$.

It will be seen that if an estimator $hg_n(x)$ of $H\circ g^{(1)}(x)$ such that A7 and A8 hold is used in the SA algorithm considered in Section 2.2 and $r>2$, then ϕ_n will converge to ϕ at a faster rate than implied by Theorem 2.2. As a first step a kernel estimator of $H\circ g^{(p+1)}(x)$ is defined which satisfies A7 and A8.

Denote by B the class of all Borel-measurable bounded functions K that vanish outside the interval $(-1, 1)$. Take

$$\mathbf{K}_p = \left\{ K \in B \text{ such that } \frac{1}{j!}\int_{-1}^{+1} u^j K(u)\,du = \begin{cases} 1, & \text{if } j=p, \\ 0, & \text{if } j\neq p,\ j=0, 1, \ldots, r-1 \end{cases} \right\}.$$

As in Section 2.2, let N_n, $n=1, 2, \ldots$ be independent renewal processes, each having underlying distribution F with renewals at times $S_{ni}=X_{n1}+X_{n2}+\cdots+X_{ni}$, $i=1, 2, \ldots$. For $K \in \mathbf{K}_p$ define

$$(3.1) \qquad hg_n^{(p)}(x) = \frac{1}{c_n^{p+1}} \int_0^\infty K\left(\frac{g^{-1}(u)-x}{c_n}\right) dN_n(u) = \frac{1}{c_n^{p+1}} \sum_{i=1}^\infty K\left(\frac{g^{-1}(S_{ni})-x}{c_n}\right)$$

where $\{c_n\}$ is a sequence of positive constants.

The next several results are analogous to those of Singh(1977) and Frees(1983) concerning estimation of the derivatives of a distribution function. The lemmas will provide sufficient conditions for A7 and A8 to hold for the estimator $hg_n^{(p)}(x)$. Part (a) of Lemma 3.1 corresponds to Theorem 3.1 of Singh, while part (b) is similar in spirit to Lemma 3.1 (b) of Frees.

Lemma 3.1: Let $K \in \mathbf{K}_p$ and $hg_n^{(p)}(x)$ be defined by (3.1). Suppose $Hog^{(r+1)}$ is bounded. Assume $c_n = o(1)$. Then

(a) $\sup_x | \mathbf{E} hg_n^{(p)}(x) - Hog^{(p+1)}(x) | = O(c_n^{r-p})$.

(b) Also suppose that $\{x_n\}$ is a sequence such that $\lim_{n\to\infty} x_n = x_o$, where $Hog^{(r+1)}(x)$ is continuous in a neighborhood of x_o. Then

$$\lim_{n\to\infty} c_n^{-(r-p)}[\, \mathbf{E} hg_n^{(p)}(x_n) - Hog^{(p+1)}(x_n)\,] = Hog^{(r+1)}(x_o) \frac{1}{r!} \int_{-1}^{+1} u^r K(u)\, du .$$

Proof: By the substitution $t = \dfrac{g^{-1}(u) - x}{c_n}$ and an application of bounded convergence one has that

$$\begin{aligned} \mathbf{E} hg_n^{(p)}(x) &= \frac{1}{c_n^{p+1}} \mathbf{E}\left\{ \int_0^\infty K\left(\frac{g^{-1}(u) - x}{c_n}\right) dN_n(u) \right\} \\ &= \frac{1}{c_n^{p+1}} \int_0^\infty K\left(\frac{g^{-1}(u) - x}{c_n}\right) h(u)\, du \\ &= \frac{1}{c_n^{p}} \int_{-1}^{+1} K(t)\, hog(x+c_n t)\, g^{(1)}(x+c_n t)\, dt \\ &= \frac{1}{c_n^{p}} \int_{-1}^{+1} K(t)\, Hog^{(1)}(x+c_n t)\, dt. \end{aligned}$$

Expanding $Hog^{(1)}(x+c_n t)$ in a Taylor series about x,

$$Hog^{(1)}(x+c_n t) = \sum_{j=0}^{r-1} \frac{Hog^{(j+1)}(x)}{j!} (c_n t)^j + R_n(t)$$

where $R_n(t) = \dfrac{Hog^{(r+1)}(\eta(t))}{r!} (c_n t)^r$ with $|\eta(t) - x| \le |c_n t|$. Thus

$$\mathbf{E}\, hg_n^{(p)}(x) = \frac{1}{c_n^p} \sum_{j=0}^{r-1} \frac{H\circ g^{(j+1)}(x)}{j!}\, c_n^j \int_{-1}^{+1} K(t)\, t^j dt$$

$$+ \frac{1}{c_n^p}\int_{-1}^{+1} K(t)\, \frac{H\circ g^{(r+1)}(\eta(t))}{r!}\, (c_n t)^r dt$$

$$= H\circ g^{(p+1)}(x) + O(\, c_n^{r-p}\,)$$

since $K \in \mathbf{K}_p$ and $H\circ g^{(r+1)}$ is bounded. This establishes (a).

By part (a) it follows that

$$\lim_{n\to\infty} c_n^{-(r-p)} [\, \mathbf{E}\, hg_n^{(p)}(x_n) - H\circ g^{(p+1)}(x_n)\,]$$

$$= \lim_{n\to\infty} \int_{-1}^{+1} K(t)\, \frac{H\circ g^{(r+1)}(\eta(t))}{r!}\, t^r dt$$

$$= H\circ g^{(r+1)}(x_o)\, \frac{1}{r!} \int_{-1}^{+1} K(t)\, t^r dt \quad \text{as } n \to \infty,$$

by the bounded convergence theorem and the continuity of $H\circ g^{(r+1)}$ at x_o. Thus (b) holds. □

Note that under the assumptions of Lemma 3.1, if $c_n = Cn^{-\gamma}$, where $\gamma = \frac{1}{2r+1}$, then A7 (a) holds. Then by Lemma 3.1 (b) one also has that

$$\lim_{n\to\infty} n^{(r-p)/(2r+1)} \left(\mathbf{E}\, hg_n^{(p)}(x_n) - H\circ g^{(p+1)}(x_n) \right)$$

$$= \lim_{n\to\infty} C^{(r-p)}\, c_n^{-(r-p)} \left(\mathbf{E}\, hg_n^{(p)}(x_n) - H\circ g^{(p+1)}(x_n) \right)$$

$$= C^{(r-p)}\, H\circ g^{(r+1)}(x_o)\, \frac{1}{r!} \int_{-1}^{+1} K(t)\, t^r dt$$

so that A8 (a) holds with $\Phi_1 = C^{(r-p)} \, Hog^{(r+1)}(x_o) \, \frac{1}{r!} \int_{-1}^{+1} K(t)\, t^r dt.$

The following lemma establishes results for $hg_n^{(p)}(x)$ similar to those contained in Lemma 3.2 of Frees(1983).

Lemma 3.2: Let $K \in \mathbf{K}_0$ and $hg_n(x)$ be defined by (3.1) with $p = 0$. Assume $Hog^{(1)}$ is a bounded function that is continuous at x_o; $\{x_n\}$ is a sequence such that $\lim_{n\to\infty} x_n = x_o$. Further assume that $c_n = o(1)$.

(a) If $t > -1$, then $\sup_x \mathbf{E}|\, hg_n(x)\,|^{t+1} = O(c_n^{-t})$.

(b) If $t \geq 0$ is an integer, then

$$\lim_{n\to\infty} c_n^t \, \mathbf{E}[\, hg_n(x_n)\,]^{t+1} = Hog^{(1)}(x_o) \int_{-1}^{1} [\, K(u)\,]^{t+1} \, du.$$

Proof: By the definition of hg_n, (2.5), and the boundedness of K and $Hog^{(1)}$

$$c_n^t \, \mathbf{E}|\, hg_n(x)\,|^{t+1} = \frac{1}{c_n} \mathbf{E}\left| \int_0^\infty K\left(\frac{g^{-1}(u) - x}{c_n}\right) dN_n(u) \right|^{t+1}$$

$$\leq \|K\|_\infty^{t+1} \frac{1}{c_n} \mathbf{E}[\, N_n(g(x+c_n)) - N_n(g(x-c_n))\,]^{t+1}$$

$$\leq K_1 \frac{1}{c_n} \mathbf{E}[\, N_n(g(x+c_n)) - N_n(g(x-c_n))\,]^{t+1}$$

$$\leq K_2$$

This establishes (a).

To show part (b), we condition on the number of renewals occurring in the interval $[g(x_n - c_n), g(x_n + c_n)]$. Let $Y_n = N_n(g(x_n+c_n)) - N_n(g(x_n-c_n))$. By (2.3) and the boundedness of K,

$$\sum_{j=2}^{\infty} \mathbf{E}\{ c_n^t [hg_n(x_n)]^{t+1} \mid Y_n = j \} \mathbf{P}\{Y_n = j \}$$

$$= \sum_{j=2}^{\infty} \mathbf{E}\{ \frac{1}{c_n} [\int_0^{\infty} K\left(\frac{g^{-1}(u) - x_n}{c_n}\right) dN_n(u)]^{t+1} \mid Y_n = j \} \mathbf{P}\{ Y_n = j \}$$

$$\leq \frac{1}{c_n} \sum_{j=2}^{\infty} \|K\|_{\infty}^{t+1} j^{t+1} \mathbf{P}\{ Y_n = j \} \longrightarrow 0 \text{ as } n \to \infty .$$

The following result will be used to find the conditional expectation when $Y_n = 1$. Let $X_1, X_2, \ldots$ be i.i.d. random variables with distribution function F which define a renewal process with renewals occurring at $S_n = X_1 + \ldots + X_n$, $n = 1, 2, \ldots$. Define the excess random variable at time t to be

$$\gamma(t) = S_{N(t)+1} - t,$$

so that $\gamma(t)$ is the remaining life of the part in use at time t. This random variable has distribution function

$$\mathbf{P}\{ \gamma(t) \leq x \} = \int_0^t [F(t-u+x) - F(t-u)] dH(u) + [F(t+x) - F(t)].$$

(See Barlow and Proschan(1965), Theorem 2.8.) Thus the distribution function and the density of the first renewal to occur after time $g(x_n - c_n)$ are given by

$$\mathbf{P}\{ S_{N(g(x_n - c_n))+1} \leq v \}$$

$$= \int_0^{g(x_n - c_n)} [F(v-u) - F(g(x_n - c_n) - u)] dH(u) + [F(v) - F(g(x_n - c_n))]$$

and

$$a(v) = h(v) - \int_{g(x_n - c_n)}^{v} f(v-u) dH(u),$$

respectively, for $v \geq g(x_n - c_n)$. Thus

$$\mathbf{P}\{ Y_n = 1 \} = \int_{g(x_n-c_n)}^{g(x_n+c_n)} a(v)\, [\, 1 - F(\, g(x_n+c_n) - v\,)\,]\, dv$$

and given that $Y_n = 1$, the conditional density of the lone renewal in the interval $[g(x_n - c_n), g(x_n+c_n)]$ is

$$b(v) = \frac{a(v)\, [\, 1 - F(g(x_n+c_n) - v)\,]}{\mathbf{P}\{ Y_n = 1 \}} .$$

Hence, by a change of variables,

$$\mathbf{E}\{ c_n^t\, [\, hg_n(x_n)\,]^{t+1} \mid Y_n = 1 \}\, \mathbf{P}\{ Y_n = 1 \}$$

$$= \mathbf{E}\left\{ \frac{1}{c_n} \int_{g(x_n-c_n)}^{g(x_n+c_n)} \left[K\left(\frac{g^{-1}(u) - x_n}{c_n} \right) \right]^{t+1} dN_n(u) \;\middle|\; Y_n = 1 \right\} \mathbf{P}\{ Y_n = 1 \}$$

$$= \frac{1}{c_n} \int_{g(x_n-c_n)}^{g(x_n+c_n)} \left[K\left(\frac{g^{-1}(u) - x_n}{c_n} \right) \right]^{t+1} b(u)\, du \;\; \mathbf{P}\{ Y_n = 1 \}$$

$$= \frac{1}{c_n} \int_{g(x_n-c_n)}^{g(x_n+c_n)} \left[K\left(\frac{g^{-1}(u) - x_n}{c_n} \right) \right]^{t+1} a(u)\, [\, 1 - F(g(x_n+c_n) - u\,)\,]\, du$$

$$= \int_{-1}^{+1} [\, K(v)]^{t+1}\, a(\, g(x_n+c_n v)\,)\, [1 - F(\, g(x_n+c_n) - g(x_n+c_n v)\,)\,]\, g^{(1)}(x_n+c_n v)\, dv$$

where

$$a(\, g(x_n+c_n v)\,) = h\circ g(x_n+c_n v) - \int_{g(x_n-c_n)}^{g(x_n+c_n v)} f(\, g(x_n+c_n v) - u\,)\, dH(u) .$$

Since K and f are bounded,

$$\int_{-1}^{+1} [K(v)]^{t+1} \int_{g(x_n-c_n)}^{g(x_n+c_nv)} f(g(x_n+c_nv)-u)\,dH(u)\,[1 - F(g(x_n+c_n)-g(x_n+c_nv))]\,g^{(1)}(x_n+c_nv)\,dv$$

$$\le 2\,\|K\|_\infty^{t+1}\,\|f\|_\infty\,\|g^{(1)}\|_\infty \int_{g(x_n-c_n)}^{g(x_n+c_n)} dH(u) \longrightarrow 0$$

by the continuity of F and g. By similar bounds

$$\lim_{n\to\infty} \int_{-1}^{+1} [K(v)]^{t+1}\, h\circ g(x_n+c_nv)\; F(g(x_n+c_n)-g(x_n+c_nv))\; g^{(1)}(x_n+c_nv)\,dv = 0$$

since $\lim_{\epsilon\to 0} F(\epsilon) = 0$. Thus,

$$\lim_{n\to\infty} \mathbf{E}\{ c_n^t\, [\, hg_n(x_n)\,]^{t+1} \mid Y_n = 1 \}\, \mathbf{P}\{ Y_n = 1 \}$$

$$= \lim_{n\to\infty} \int_{-1}^{+1} [K(v)]^{t+1}\, h\circ g(x_n+c_nv)\; g^{(1)}(x_n+c_nv)\,dv$$

$$= H\circ g^{(1)}(x_o) \int_{-1}^{+1} [K(v)]^{t+1}\,dv,$$

which completes the proof of the lemma. □

Suppose $c_n = Cn^{-\gamma}$ where $\gamma = \frac{1}{2r+1}$. Then Lemma 3.2 (a) implies that A7 (b) holds for the kernel estimator hg_n. Similarly, hg_n satisfies A8(b) by Lemma 3.2 (b).

Let the sequence $\{\phi_n\}$ of estimates of ϕ be defined recursively by (2.1), where in the definition (2.2) the estimator $hg_n(x)$ is redefined to be any estimator of $H\circ g^{(1)}(x)$. The following theorem establishes some convergence properties for ϕ_n assuming that

$hg_n(x)$ satisfies A7 and A8. The subsequent corollary addresses the special case when $hg_n(x)$ is of the form (3.1) with $p=0$.

<u>Theorem 3.1:</u> Let $\gamma = \frac{1}{2r+1}$, $c_n = C n^{-\gamma}$, and $\Gamma = A\frac{d}{dt}Dg(t)|_{t=\phi}$ for the positive constants A and C of A6.

(a) Assume A1–A4 and A7 hold. Then $\phi_n \to \phi$ a.s.

(b) Assume A1–A3, A6, A7, and A8 hold. Then $n^{(1-\gamma)/2}(\phi_n - \phi) \xrightarrow{D} N(\mu_o, \sigma_o^2)$ where

$$\mu_o = -\frac{A C_1 \Phi_1}{g(\phi)\,[\,\Gamma - (1-\gamma)/2\,]}$$

and

$$\sigma_o^2 = \frac{C_1^2 A^2 \Phi_{2,1}}{[\,g(\phi)\,]^2\,[\,2\Gamma - (1-\gamma)\,]}.$$

<u>Proof of Theorem 3.1 (a):</u> Again take $U_n = \phi_n - \phi$. By A7 (a) with $p=0$,

$$\sup_x |\mathbf{E}[hg_n(x)] - Hog^{(1)}(x)| = O(n^{-r/(2r+1)}).$$

This implies that

$$|\Delta_n| = |\,\mathbf{E}_{\mathcal{F}_n}[Dg_n(\phi_n) - Dg(\phi_n)]\,|$$

$$= C_1\,|\,\mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n) - Hog^{(1)}(\phi_n)\,| \Big/ g(\phi_n)$$

$$\le K_1\, n^{-r/(2r+1)} = K_1\, n^{-\gamma r}$$

By A7 (b) with $t=1$

$$\mathbf{E}_{\mathcal{F}_n}|hg_n(\phi_n)|^2 = O(n^{+1/(2r+1)}) = O(n^{+\gamma}),$$

so that by the proof of Lemma 2.2,

$$\mathbf{E}_{\mathcal{F}_n}[Dg_n(\phi_n)]^2 \le K_1 \frac{\mathbf{E}_{\mathcal{F}_n}|hg_n(\phi_n)|^2}{[\,g(\phi_n)\,]^2} + \frac{[\,K_2 + K_3\, \mathbf{E}_{\mathcal{F}_n}[N_n(g(\phi_n))]^2\,]}{[\,g(\phi_n)\,]^4}$$

$$\le K_4\, n^{+\gamma} + K_5$$

$$\le K_6\, n^{+\gamma}.$$

Thus from the proof of Theorem 2.1

$$\mathbf{E}_{\mathcal{F}_n} U_{n+1}^2$$

$$\le U_n^2 + a_n^2 \mathbf{E}_{\mathcal{F}_n} Dg_n^2(\phi_n) + 2a_n|U_n||\Delta_n| - 2a_n(\phi_n - \phi)Dg(\phi_n)$$

$$\le U_n^2 + K_6\, a_n^2\, n^{+\gamma} + 2a_n[K_7 + K_8|\phi_n - \phi|^2\,]n^{-\gamma r} - 2a_n(\phi_n - \phi)Dg(\phi_n)$$

$$\le U_n^2\{1 + 2K_8 a_n n^{-\gamma r}\} + \{K_6 a_n^2 n^{+\gamma} + 2K_7 a_n n^{-\gamma r}\} - 2a_n(\phi_n - \phi)Dg(\phi_n).$$

Since the required summations are finite by A4 with $c_n = C n^{-\gamma}$, $\phi_n \to \phi$ a.s. by an application of the Robbins-Siegmund result. □

The proof of Theorem 3.1 (b) will require the following lemma, similar to Lemma 2.5 (b). As in Section 2, let $V_n = c_n^{1/2}[Dg_n(\phi_n) - Dg(\phi_n) - \Delta_n]$.

Lemma 3.3: Assume A1–A3, A6(with $\gamma = 1/(2r+1)$), A7 and A8 hold. Then

$$\lim_{n\to\infty} \mathbf{E}_{\mathcal{F}_n} V_n^2 = C_1^2\, C\ \Phi_{2,1} \Big/ [\,g(\phi)\,]^2 .$$

Proof: By definition $V_n^2 = c_n C_1^2 [X_n - Y_n - Z_n]^2$ where

$$X_n = [g(\phi_n)]^{-1} [hg_n(\phi_n)],$$

$$Y_n = [g(\phi_n)]^{-1} E_{\mathcal{F}_n} [hg_n(\phi_n)] , \text{ and}$$

$$Z_n = [g(\phi_n)]^{-2} g^{(1)}(\phi_n) [N_n(g(\phi_n)) - H \circ g(\phi_n)].$$

As in Lemma 2.5, we consider separately the conditional expectations $E_{\mathcal{F}_n} X_n^2$, $E_{\mathcal{F}_n} X_n Y_n$, etc. Since the conditions of Theorem 3.1 (a) hold, $\phi_n \to \phi$ a.s.

(i) Since $\phi_n \to \phi$ a. s., by A8 (b) with t=1

$$c_n E_{\mathcal{F}_n} X_n^2 = C [g(\phi_n)]^{-2} n^{-1/(2r+1)} E_{\mathcal{F}_n} [hg_n(\phi_n)]^2$$

$$\longrightarrow C \Phi_{2,1} \Big/ [g(\phi)]^2 .$$

(ii) By A3 and A7 (b) with t=0

$$c_n E_{\mathcal{F}_n} X_n Y_n = [g(\phi_n)]^{-2} c_n [E_{\mathcal{F}_n} hg_n(\phi_n)]^2$$

$$\leq K_1 c_n O(1) \longrightarrow 0 .$$

(iii) By (2.6), A3, A7 (b) with t=1, and the conditional version of the Cauchy-Schwarz inequality,

$$c_n E_{\mathcal{F}_n} | X_n Z_n | \leq c_n [g(\phi_n)]^{-3} g^{(1)}(\phi_n) E_{\mathcal{F}_n} [hg_n(\phi_n)] \, | N_n(g(\phi_n)) - H \circ g(\phi_n) |$$

$$\leq c_n K_1 \{ E_{\mathcal{F}_n} [hg_n(\phi_n)]^2 \}^{1/2} \left\{ \frac{E_{\mathcal{F}_n} [N_n(g(\phi_n)) - H \circ g(\phi_n)]^2}{g(\phi_n)} \right\}^{1/2}$$

$$\leq c_n K_2 \{ O(n^{+1/(2r+1)}) \}^{1/2} \to 0.$$

(iv) Also by A3 and A7 (b) (with t=0),

$$c_n \mathbf{E}_{\mathcal{F}_n} Y_n^2 = c_n \, [g(\phi_n)]^{-2} \, [\, \mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n) \,]^2$$

$$\leq c_n \, K_1 \to 0.$$

(v) For all n, $\mathbf{E}_{\mathcal{F}_n} Y_n Z_n = 0$.

(vi) By part (vi) of the proof of Lemma 2.5, $\lim_{n\to\infty} c_n \, \mathbf{E}_{\mathcal{F}_n} Z_n^2 = 0$.

The result follows from (i)−(vi). □

<u>Proof of Theorem 3.1 (b):</u> Again we employ the theorem on asymptotic normality due to Fabian. From the proof of Theorem 2.2

$$U_{n+1} = U_n \, [\, 1 - n^{-1}\Gamma_n \,] + n^{(\frac{\gamma}{2}-1)} \Phi_n V_n + n^{-1} n^{-(1-\gamma)/2} T_n$$

where as before $\Gamma_n = A \, \frac{d}{dt} Dg(t)|_{t=\eta_n} \to \Gamma = A \, \frac{d}{dt} Dg(t)|_{t=\phi}$, $\Phi_n = \Phi = -AC^{-1/2}$, and $T_n = -An^{(1-\gamma)/2} \Delta_n$. The asymptotic behaviour of T_n now depends on how many derivatives of $H \circ g$ exist. Since $\phi_n \to \phi$ a.s., using A8 (a) with p=0 gives

$$\begin{aligned} T &= \lim_{n\to\infty} T_n \\ &= -\, AC_1 \lim_{n\to\infty} [\, g(\phi_n) \,]^{-1} \, n^{(1-\gamma)/2} \Big(\mathbf{E}_{\mathcal{F}_n}[hg_n(\phi_n)] - H \circ g^{(1)}(\phi_n) \Big) \\ &= -\, \frac{A \, C_1 \Phi_1}{g(\phi)} \,. \end{aligned}$$

By definition $\mathbf{E}_{\mathcal{F}_n} V_n = 0$. By Lemma 2.4

$$\mathbf{E}_{\mathcal{F}_n} V_n^2 \leq c_n \{\, K_1 + K_2 \mathbf{E}_{\mathcal{F}_n} [\, hg_n(\phi_n) \,]^2 + K_3 [\, \mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n) \,]^2 \,\},$$

which is bounded by A7 (b). By Lemma 3.3,

$$\lim_{n\to\infty} \mathbf{E}_{\mathcal{F}_n} V_n^2 = \frac{C_1^2\, C\, \Phi_{2,1}}{[\, g(\phi)\,]^2}\,.$$

We need to show that $\sigma^2_{n,r} = \mathbf{E}\{\, V_n^2\ \chi[\, V_n^2 \geq rn]\ \} \to 0$ as $n \to \infty$ for $r = 1, 2, \ldots$. To accomplish this using Lemma 2.7 it is first shown that for any $p > 2$, $\mathbf{E}_{\mathcal{F}_n} | c_n^{1/2} V_n |^p \to 0$. By Lemma 2.4,

$$c_n^{p/2} \mathbf{E}_{\mathcal{F}_n} | V_n |^p \leq c_n^p \left\{ K_1 + K_2 \mathbf{E}_{\mathcal{F}_n} | hg_n(\phi_n) |^p + K_3 |\, \mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n)\, |^p \right\}.$$

By A7 (b),

$$c_n^p\, \mathbf{E}_{\mathcal{F}_n} | hg_n(\phi_n) |^p = c_n^p\, O(\, n^{+(p-1)/(2r+1)})$$

$$= O(\, n^{-1/(2r+1)}) \longrightarrow 0$$

if $p > 2$. Now taking $t = 0$ in A7 (b)

$$c_n^p\ |\, \mathbf{E}_{\mathcal{F}_n} hg(\phi_n) |^p = c_n^p\ | O(1) | \to 0 \text{ as } n \to \infty.$$

For Lemma 2.7, $p, q > 0$ were such that $\frac{2}{p} + \frac{1}{q} = 1$. Thus, taking $p > 2$, Lemma 2.7 holds.

Taking $\alpha = 1$ and $\beta = \beta_+ = 1 - \gamma$, part (b) of Theorem 3.1 now follows from Fabian's theorem. □

Now let $K \in \mathbf{K}_0$ and define hg_n by (3.1) with p=0. Lemmas 3.1 and 3.2 apply to hg_n if the following assumption is met.

A9: For some integer $r \geq 1$, $Hog^{(1)}$ and $Hog^{(r+1)}$ exist and are bounded on the entire real line, and both functions are continuous in a neighborhood of ϕ.

Corollary 3.1: Assume A1–A3, A6, and A9 hold. Let $\gamma = \frac{1}{2r+1}$ in A6 (for the r of A9) and $\Gamma = A\frac{d}{dt}Dg(t)|_{t=\phi}$. Then

(a) $\phi_n \to \phi$ a.s. and

(b) $n^{(1-\gamma)/2}\,(\phi_n - \phi) \xrightarrow{D} N(\mu_1, \sigma_1^2)$ where

$$\mu_1 = -\frac{C_1\, A\, C^r Hog^{(r+1)}(\phi)\beta_1}{g(\phi)\,[\,\Gamma - (1-\gamma)/2\,]}$$

and

$$\sigma_1^2 = \frac{C_1^2\, A^2\, C^{-1}\, Hog^{(1)}(\phi)\, \beta_2}{[\,g(\phi)\,]^2\,[\,2\Gamma - (1-\gamma)\,]}$$

with $\beta_1 = \frac{1}{r!}\int_{-1}^{+1} u^r K(u)\,du$ and $\beta_2 = \int_{-1}^{+1} [\,K(u)\,]^2 du$.

Proof: By the discussions following Lemmas 3.1 and 3.2, A9 implies A7 and A8 hold for $hg_n(x)$. The corollary is then an immediate consequence of Theorem 3.1 with

$$\Phi_1 = \lim_{n\to\infty} n^{r/(2r+1)}\,[\,\mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n) - Hog^{(1)}(\phi_n)\,]$$

$$= C^r Hog^{(r+1)}(\phi)\ \frac{1}{r!}\int_{-1}^{+1} u^r\, K(u)\,du$$

and

$$\Phi_{2,1} = \lim_{n\to\infty} n^{-1/(2r+1)}\,\mathbf{E}_{\mathcal{F}_n}[\,hg_n(\phi_n)\,]^2$$

$$= C^{-1}\, Hog^{(1)}(\phi) \int_{-1}^{+1} [\,K(u)\,]^2\, du. \qquad \square$$

3.3 The Optimal Tuning Constants

In this section we consider the problem of choosing the constants A and C so as to minimize the asymptotic mean square error of ϕ_n. From this section onwards it is assumed that in the algorithm defined by (2.1) and (2.2) hg_n is the kernel estimator introduced in Section 3.2. From Corollary 3.1 it appears that the limiting distribution of ϕ_n depends on the function g as well as on the renewal function H. This dependence will be explored further to assess how important the choice of g is to the behaviour of the procedure.

Let ξ be a normal random variable with mean μ_1 and variance σ_1^2 , where μ_1 and σ_1^2 have been defined in Corollary 3.1. Then under the conditions assumed in Corollary 3.1,

$$n^{(1-\gamma)/2}(\phi_n - \phi) \xrightarrow{D} \xi.$$

Suppose the quadratic risk function $\mathbf{E}(\xi - \phi)^2$ is used as the criterion in estimating ϕ. Then once the rate of convergence has been determined by the kernel estimator hg_n, a measure of performance for the procedure is $E\xi^2$, the asymptotic mean square error of ϕ_n. One would naturally like to minimize this quantity, which depends in part on A and C. By taking partial derivatives of $\mathbf{E}\xi^2 = \mu_1^2 + \sigma_1^2$ with respect to A and C, the optimal choices of these constants are found to be

$$A_{opt} = \left[\frac{d}{dt} Dg(t) \,\Big|_{t=\phi} \right]^{-1}$$

and

$$C_{opt} = \left[\frac{H \circ g^{(1)}(\phi)\, \beta_2\, (1+\gamma)}{8r\, [\, H \circ g^{(r+1)}(\phi)\,]^2\, \beta_1^2} \right]^{1/(2r+1)} .$$

Thus the minimum possible mean square error (for the particular kernel K used in hg_n) is

$$\mathbf{E}\xi^2 = C_1^2 A_{opt}^2 C_{opt}^{-1} [g(\phi)]^{-2} \text{Hog}^{(1)}(\phi) \beta_2 [\gamma^2 4r(r+1)]^{-1}.$$

Unfortunately, the unknown constants $\text{Hog}^{(1)}(\phi)$, $\text{Hog}^{(2)}(\phi)$, and $\text{Hog}^{(r+1)}(\phi)$ partly determine A_{opt} and C_{opt}. In Section 3.4 an adaptive algorithm is introduced where A_{opt} and C_{opt} are estimated at each stage of the procedure so as to minimize the limiting mean square error. As in Section 2.2, let $\phi^* = g(\phi)$ be the optimal block replacement policy. Redefine ϕ_n^* so that $\phi_n^* = g(\phi_n)$, an estimate of ϕ^*. Since the real interest is in estimating ϕ^*, the remainder of this section examines to what extent the limiting distribution of ϕ_n^* depends on g.

Under the conditions of Corollary 3.1, $n^{(1-\gamma)/2} (\phi_n - \phi) \xrightarrow{D} N(\mu_1, \sigma_1^2)$. Since g is an appropriate differentiable function, by the delta-method

$$n^{(1-\gamma)/2} (\phi_n^* - \phi^*) \xrightarrow{D} N\Big(g^{(1)}(\phi)\mu_1, [g^{(1)}(\phi)]^2\sigma_1^2 \Big)$$

when the same conditions hold. Let $K_1 = C_1\beta_1/\phi^*$ and $K_2 = C_1^2 \beta_2/[\phi^*]^2$. Then

$$g^{(1)}(\phi)\mu_1 = - \frac{K_1 A C^r g^{(1)}(\phi) \text{Hog}^{(r+1)}(\phi)}{\Gamma - (1-\gamma)/2}$$

and

$$[g^{(1)}(\phi)]^2\sigma_1^2 = \frac{K_2 A^2 C^{-1} [g^{(1)}(\phi)]^2 \text{Hog}^{(1)}(\phi)}{2\Gamma - (1-\gamma)}.$$

Note that $\Gamma = A\frac{d}{dt}Dg(t)|_{t=\phi} = A D^{(1)}(\phi^*)[g^{(1)}(\phi)]^2$ (where D was defined in Section 2.2) since $D(\phi^*) = 0$. Thus the denominators of the above expressions may be freed from dependence on g by taking $A = A_o[g^{(1)}(\phi)]^{-2}$ for some positive constant A_o. This results in

$$g^{(1)}(\phi)\mu_1 = -\frac{K_1 A_o [g^{(1)}(\phi)]^{-1} C^r H \circ g^{(r+1)}(\phi)}{A_o D^{(1)}(\phi^*) - (1-\gamma)/2}$$

and

$$[g^{(1)}(\phi)]^2\sigma_1^2 = \frac{K_2 A_o^2 C^{-1} [g^{(1)}(\phi)]^{-1} h(\phi^*)}{2 A_o D^{(1)}(\phi^*) - (1-\gamma)} .$$

Letting $C = C_o[\, g^{(1)}(\phi)\,]^{-1}$ for some positive constant C_o,

$$g^{(1)}(\phi)\mu_1 = -\frac{K_1 A_o [g^{(1)}(\phi)]^{-(r+1)} C_o^r H \circ g^{(r+1)}(\phi)}{A_o D^{(1)}(\phi^*) - (1-\gamma)/2}$$

and

$$[\, g^{(1)}(\phi)\,]^2 \sigma_1^2 = \frac{K_2 A_o^2 C_o^{-1} h(\phi^*)}{2 A_o D^{(1)}(\phi^*) - (1-\gamma)} .$$

Thus the asymptotic bias of ϕ_n^* depends on g while the variance does not.

Since the asymptotic mean square error of ϕ_n^* is $[g^{(1)}(\phi)]^2\, [\mu_1^2 + \sigma_1^2]$, the best choices of A and C for estimating ϕ^* are the same constants A_{opt} and C_{opt}. By setting $A_{opt} = A_o[g^{(1)}(\phi)]^{-2}$ and $C_{opt} = C_o[g^{(1)}(\phi)]^{-1}$ it can be seen that one should let $A_o = [D^{(1)}(\phi^*)]^{-1}$ and $C_o = g^{(1)}(\phi)\, C_{opt}$. Therefore, for estimation of ϕ^* the optimal choice of C depends on g while the best choice of A does not.

3.4 An Adaptive Stochastic Approximation Algorithm

It was seen in Section 3.3 that by taking

$$A = A_{opt} = \left[\frac{d}{dt} Dg(t) \,|_{t=\phi} \right]^{-1}$$

and

$$C = C_{opt} = \left[\frac{Hog^{(1)}(\phi)\, \beta_2\, (1+\gamma)}{8r\, [\, Hog^{(r+1)}(\phi)\,]^2\, \beta_1^2} \right]^{1/(2r+1)}$$

the asymptotic mean square error of $n^{(1-\gamma)/2}(\phi_n - \phi)$ is minimized. The constants A_{opt} and C_{opt} of course depend on the unknown quantities $\frac{d}{dt}Dg(t)\,|_{t=\phi}$, $Hog^{(1)}(\phi)$, and $Hog^{(r+1)}(\phi)$. In this section consistent estimates of these quantities are defined which in turn yield consistent estimators A_n and C_n of A_{opt} and C_{opt}, respectively. Taking $a_n = A_n n^{-1}$ and $c_n = C_n n^{-\gamma}$ then results in the minimum asymptotic mean square error possible given the kernel K used to estimate $Hog^{(1)}(\phi)$. (The choice of K determines the constants β_1 and β_2 which appear in the limiting mean square error.) The problem of choosing K is not considered here.

Let $\{A_n^*\}$ and $\{C_n^*\}$ be sequences of estimates of A_{opt} and C_{opt}, as yet to be defined. Given positive constants Z_1, Z_2, Z_3, Z_4, a, b, c, and γ, let

$$A_n = (\, Z_1(\log n)^{-1} \vee A_n^* \,) \wedge Z_2 n^a \tag{3.2}$$

and

$$C_n = (\, Z_3 n^{-b} \vee C_n^* \,) \wedge Z_4 n^c. \tag{3.3}$$

These definitions ensure that A_n and C_n fall within known bounds. For large enough n, (3.4) $$Z_1(\log n)^{-1} \leq A_n \leq Z_2 n^a$$

and

(3.5) $$Z_3\, n^{-b} \leq C_n \leq Z_4\, n^c.$$

As in Section 3.2, let $\gamma = \frac{1}{2r+1}$. Assume the following restrictions on a, b, c, and γ are met.

A10: $a + rc < r\gamma$, $2a + b + \gamma < 1$, $3(b+\gamma) < 1$, and $b < r\gamma/(r+1)$.

With $c_n = C_n n^{-\gamma}$, define $hg_n(x)$ by (3.1) with $p=0$ and $K=K_0 \in \mathbb{K}_0$. Using this modified version of hg_n in Dg_n, let successive estimates of ϕ be defined by

(3.6) $$\phi_{n+1} = \phi_n - A_n n^{-1} Dg_n(\phi_n).$$

Before introducing the estimators A_n^* and C_n^* and thus completing the definitions of A_n and C_n, the modified SA procedure is shown to be consistent for ϕ. The following lemma provides a bound for the bias (Δ_n) of $Dg_n(\phi_n)$. The ensuing theorem provides sufficient conditions for ϕ_n to converge to ϕ with probability one.

Lemma 3.4: Suppose $Hog^{(r+1)}$ is bounded. Assume A1 and A3 hold. Then there exist positive constants K_1 and K_2 such that

$$|\Delta_n| = |E_{\mathcal{F}_n}[\, Dg_n(\phi_n) - Dg(\phi_n)\,]| \leq K_1 n^{r(c-\gamma)}.$$

Proof: By Lemma 3.1 and (3.3) ,

$$\sup_x |E_{\mathcal{F}_n} hg_n(x) - Hog^{(1)}(x)| = O(c_n^r) \leq K_1 n^{r(c-\gamma)}.$$

Thus

$$|\Delta_n| = C_1 \, |\mathbf{E}_{\mathcal{F}_n} h g_n(\phi_n) - H \circ g^{(1)}(\phi_n)| \Big/ g(\phi_n)$$

$$\leq K_2 n^{r(c-\gamma)}. \qquad \square$$

<u>Theorem</u> <u>3.2:</u> **Assume A1-A3, A9, and A10 hold. Then $\phi_n \to \phi$ a.s.**

<u>Proof:</u> Since Lemma 3.2 (a) holds, as in the proof of Theorem 3.1

$$\mathbf{E}_{\mathcal{F}_n}[\, Dg_n(\phi_n)\,]^2 \leq K_1 \frac{\mathbf{E}_{\mathcal{F}_n}|hg_n(\phi_n)|^2}{[\, g(\phi_n)\,]^2} + \frac{[\, K_2 + K_3\, \mathbf{E}_{\mathcal{F}_n}[\, N_n(g(\phi_n))\,]^2\,]}{[\, g(\phi_n)\,]^4}$$

$$\leq K_4 c_n^{-1} + K_5$$

$$\leq K_6 n^{b+\gamma}$$

for large enough n. From the proof of Theorem 3.1 and by Lemma 3.4, using $a_n = A_n n^{-1}$,

$$\mathbf{E}_{\mathcal{F}_n} U_{n+1}^2 \leq U_n^2 + a_n^2 \mathbf{E}_{\mathcal{F}_n}[Dg_n^2(\phi_n)] + 2a_n|U_n||\Delta_n| - 2a_n(\phi_n - \phi)Dg(\phi_n)$$

$$\leq U_n^2 + a_n^2 K_6 n^{b+\gamma} + 2a_n n^{r(c-\gamma)}[\, K_7 + K_8\, |\phi_n - \phi|^2\,] - 2a_n(\phi_n - \phi)Dg(\phi_n)$$

$$\leq U_n^2\{\, 1 + 2K_8 a_n n^{r(c-\gamma)}\,\} + \{\, K_6 a_n^2\, n^{b+\gamma} + 2K_7 a_n n^{r(c-\gamma)}\,\} - 2a_n(\phi_n - \phi)Dg(\phi_n).$$

By (3.2), $a_n = A_n n^{-1} \leq Z_2 n^{a-1}$ for large n. Thus by A10,

$$\sum_{n=1}^{\infty} a_n^2\, n^{b+\gamma} < \infty$$

and

$$\sum_{n=1}^{\infty} a_n n^{r(c-\gamma)} \leq Z_2 \sum_{n=1}^{\infty} n^{a-1+r(c-\gamma)} < \infty.$$

Since for large enough n, $a_n \geq Z_1(\log n)^{-1} n^{-1}$, it is the case that $\sum_{n=1}^{\infty} a_n = \infty$. Thus the conditions of the Robbins-Siegmund result are met and $\phi_n \to \phi$ a.s. □

An estimator of

$$\frac{d}{dt} Dg(t) \Big|_{t=\phi} = \frac{C_1 g(\phi) H \circ g^{(2)}(\phi) - g^{(2)}(\phi)[C_1 H(g(\phi)) + C_2]}{[\, g(\phi)\,]^2}$$

will be required to define the sequence of random variables $\{A_n^*\}$. Thus for $K_1 \in \mathbb{K}_1$ let $hg_n^{(1)}(x)$ be defined by (3.1) with $c_n = C_n n^{-\gamma}$ and $p=1$. Next, take

$$\alpha_n = \frac{C_1 g(\phi_n) hg_n^{(1)}(\phi_n) - g^{(2)}(\phi_n)[C_1 N_n(g(\phi_n)) + C_2]}{[\, g(\phi_n)\,]^2}$$

and

(3.7) $$A_{n+1}^* = [\, n^{-1} \sum_{j=1}^{n} \alpha_j \,]^{-1}.$$

Lemma 3.5 gives sufficient conditions for A_n defined by (3.2) to be a consistent estimator of A_{opt} with A_n^* defined by (3.7).

Lemma 3.5: Assume A1-A3, A9, and A10 hold. Further suppose $H \circ g^{(2)}$ is continuous in a neighborhood of ϕ. Then $A_n \to A_{opt}$ a.s.

Proof: By Lemma 3.1,

$$E_{\mathcal{F}_n} \alpha_n = \frac{C_1 g(\phi_n)\,[\, H \circ g^{(2)}(\phi_n) + O(\, c_n^{r-1})\,] - g^{(2)}(\phi_n)\,[\, C_1 H(g(\phi_n)) + C_2\,]}{[\, g(\phi_n)\,]^2}$$

$$\longrightarrow \frac{C_1 g(\phi) H \circ g^{(2)}(\phi) - g^{(2)}(\phi)\,[\, C_1 H(g(\phi)) + C_2 \,]}{[\, g(\phi)\,]^2}$$

as $n \to \infty$, since $c_n = o(1)$ and Theorem 3.2 holds. Thus $n^{-1} \sum_{j=1}^{n} \mathbf{E}_{\mathcal{F}_j} \alpha_j \to A_{opt}^{-1}$ a.s. as $n \to \infty$.

Let $M_n = \sum_{j=1}^{n} [\alpha_j - \mathbf{E}_{\mathcal{F}_j} \alpha_j]$. Then $\{ M_n, \mathcal{F}_{n+1} \}_{n=1}^{\infty}$ is a zero-mean martingale. Furthermore, by A1, A3 and (2.6),

$$\mathbf{E}_{\mathcal{F}_n}[\alpha_n - \mathbf{E}_{\mathcal{F}_n}\alpha_n]^2 \le \mathbf{E}_{\mathcal{F}_n}\alpha_n^2$$

$$\le K_1 \frac{\mathbf{E}_{\mathcal{F}_n}[\, hg_n^{(1)}(\phi_n)\,]^2}{[\, g(\phi_n)\,]^2} + \frac{K_2\, \mathbf{E}_{\mathcal{F}_n}[\, N_n(g(\phi_n))\,]^2 + K_3}{[\, g(\phi_n)\,]^4}$$

$$\le K_4\, \mathbf{E}_{\mathcal{F}_n}[\, hg_n^{(1)}(\phi_n)\,]^2 + K_5.$$

To bound $\mathbf{E}_{\mathcal{F}_n}[\, hg_n^{(1)}(\phi_n)\,]^2$ we may write

$$\mathbf{E}_{\mathcal{F}_n}[\, hg_n^{(1)}(\phi_n)\,]^2 = \frac{1}{c_n^4}\, \mathbf{E}_{\mathcal{F}_n}\left\{ \int_{g(\phi_n - c_n)}^{g(\phi_n + c_n)} K_1\!\left(\frac{g^{-1}(u) - \phi_n}{c_n}\right) dN_n(u) \right\}^2$$

$$\le \| K_1 \|_{\infty}^2\, \frac{1}{c_n^4}\, \mathbf{E}_{\mathcal{F}_n}[\, N_n(\, g(\phi_n + c_n)) - N_n(g(\phi_n - c_n))\,]^2$$

$$\le K_6\, c_n^{-3}$$

by (2.5). Thus, since $c_n^{-1} \le K_7 n^{b+\gamma}$ for large enough n, where $3(b+\gamma) < 1$ by A10,

$$\sum_{n=1}^{\infty} n^{-2}\, \mathbf{E}_{\mathcal{F}_n}\alpha_n^2 \;\le\; K_8 \sum_{n=1}^{\infty} n^{-2}\, n^{3(b+\gamma)} \;<\; \infty.$$

It follows by Chow(1965) that $[A^*_{n+1}]^{-1} \rightarrow A^{-1}_{opt}$ a.s. and thus $A_n \rightarrow A_{opt}$ a.s. by (3.2). □

As a first step in defining C^*_n an estimator of $H \circ g^{(1)}(\phi)$ is introduced. With the modified version of hg_n defined earlier in this section take

$$\Psi_n = n^{-1} \sum_{j=1}^{n} hg_n(\phi_n).$$

In Lemma 3.6 Ψ_n is shown to be a consistent estimator of $H \circ g^{(1)}(\phi)$.

Lemma 3.6: Assume A1-A3, A9, and A10 hold. Then $\Psi_n \rightarrow H \circ g^{(1)}(\phi)$ a.s.

Proof: By Lemma 3.1, (3.3), and Theorem 3.2 ,

$$\mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n) \;=\; H \circ g^{(1)}(\phi_n) \;+\; O(\,(C_n n^{-\gamma})^r\,) \;\rightarrow\; H \circ g^{(1)}(\phi)$$

as $n \rightarrow \infty$. Thus

$$n^{-1} \sum_{j=1}^{n} \mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n) \;\rightarrow\; H \circ g^{(1)}(\phi).$$

Letting $M_n = \Sigma_{j=1}^{n} [\, hg_j(\phi_j) - \mathbf{E}_{\mathcal{F}_j} hg_j(\phi_j)\,]$, it is again the case that $\{ M_n, \mathcal{F}_{n+1} \}_{n=1}^{\infty}$ is a zero-mean martingale. By Lemma 3.2 (a),

$$\mathbf{E}_{\mathcal{F}_n}[\, hg_n(\phi_n) - \mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n)\,]^2 \le \mathbf{E}_{\mathcal{F}_n}[\, hg_n(\phi_n)\,]^2$$

$$\le K_1 \,[\, C_n n^{-\gamma}\,]^{-1} .$$

This implies that

$$\sum_{n=1}^{\infty} \mathbf{E}_{\mathcal{F}_n}[\, hg_n(\phi_n) - \mathbf{E}_{\mathcal{F}_n} hg_n(\phi_n)\,]^2 \, n^{-2} \;<\; \infty$$

since $[C_n n^{-\gamma}]^{-1} \leq K_2 n^{1/3}$ for some constant K_2 by A10 and (3.3). Thus, again using Theorem 5 of Chow(1965), $\Psi_n \to Hog^{(1)}(\phi)$ a.s. □

An estimator of $Hog^{(r+1)}$ will be needed to define C_n^*. Let

$$\mathbf{K}_r = \left\{ K \in B \text{ such that } \frac{1}{j!}\int_{-1}^{+1} u^j K(u)\,du = \begin{array}{l} 1, \text{ if } j=r, \\ \\ 0, \text{ if } j=0, 1, \ldots, r-1. \end{array} \right\}$$

Next, for a $K_r \in \mathbf{K}_r$ define

$$hg_n^{(r)}(x) = \frac{1}{c_n^{r+1}} \int_{g(x-c_n)}^{g(x+c_n)} K_r\left(\frac{g^{-1}(u) - x}{c_n}\right) dN_n(u)$$

with $c_n = C_n n^{-\gamma}$. Let $\Lambda_n = n^{-1}\Sigma_{j=1}^n hg_j^{(r)}(\phi_n)$. The next lemma, similar in spirit to Lemmas 3.5 and 3.6, establishes that $\Lambda_n \to Hog^{(r+1)}(\phi)$ a.s.

Lemma 3.7: Assume A1-A3, A9, and A10 hold. Further suppose that for some $d>0$ and all real x, $Hog^{(r+1)}(x) = Hog^{(r+1)}(\phi) + O(|x-\phi|^d)$. Then $\Lambda_n \longrightarrow Hog^{(r+1)}(\phi)$ a.s.

Proof: First, by the usual change of variables and a Taylor expansion,

$$E_{\mathcal{F}_n} hg_n^{(r)}(x) = \frac{1}{c_n^{r+1}} \int_0^\infty K_r\left(\frac{g^{-1}(u) - x}{c_n}\right) h(u)\,du$$

$$= \frac{1}{c_n^r} \int_{-1}^{+1} K_r(t)\, Hog^{(1)}(x+c_n t)\,dt$$

$$= \frac{1}{c_n^r} \int_{-1}^{+1} K_r(t) \left[\sum_{j=0}^{r-1} \frac{Hog^{(j+1)}(x)\,(tc_n)^j}{j!} + \frac{Hog^{(r+1)}(\,\eta(t)\,)\,(tc_n)^r}{r!} \right] dt$$

where $|\eta(t) - x| \le c_n$ for all $t \in [-1, 1]$. Thus by the definition of K_r and the assumptions of the lemma,

$$\mathbf{E}_{\mathcal{F}_n} hg_n^{(r)}(x) = \int_{-1}^{+1} K_r(t) \frac{Hog^{(r+1)}(\eta(t))\, t^r}{r!} dt$$

$$= \frac{1}{r!} \int_{-1}^{+1} K_r(t)\, t^r \,[Hog^{(r+1)}(x) + O(|\eta(t) - x|^d)]\, dt$$

$$= Hog^{(r+1)}(x) + O(c_n^d)$$

which implies that $\mathbf{E}_{\mathcal{F}_n} hg_n^{(r)}(\phi_n) \to Hog^{(r+1)}(\phi)$ as $n \to \infty$. Thus

$$\lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} \mathbf{E}_{\mathcal{F}_j} hg_j^{(r)}(\phi_j) = Hog^{(r+1)}(\phi).$$

By the definition of $hg_n^{(r)}$, (2.5), and A3

$$\mathbf{E}_{\mathcal{F}_n}[\, hg_n^{(r)}(\phi_n)\,]^t \le \frac{1}{c_n^{t(r+1)}} \| K_r \|_\infty^t \, \mathbf{E}_{\mathcal{F}_n}[\, N_n(g(\phi_n + c_n)) - N_n(g(\phi_n - c_n))\,]^t$$

$$= K_1 \frac{1}{c_n^{t(r+1)}} \left\{ [g(\phi_n + c_n) - g(\phi_n - c_n)] + o(g(\phi_n + c_n) - g(\phi_n - c_n)) \right\}$$

$$\le K_2\, c_n^{1 - t(r+1)}.$$

Following Lemma 3.5 of Frees(1983) it can be shown that there exists $t \in (1, 2)$ such that

$$\text{(3.8)} \qquad \sum_{n=1}^{\infty} \mathbf{E}_{\mathcal{F}_n}[\, hg_n^{(r)}(\phi_n)\,]^t\, n^{-t} \le K_2 \sum_{n=1}^{\infty} c_n^{1 - t(r+1)}\, n^{-t} < \infty$$

which is sufficient for

$$\sum_{n=1}^{\infty} \mathbf{E}_{\mathcal{F}_n} | hg_n^{(r)}(\phi_n) - \mathbf{E}_{\mathcal{F}_j} hg_j^{(r)}(\phi_j) |^t \, n^{-t} < \infty,$$

when $t \in (1, 2)$. This will establish the lemma using Chow(1965). Since there exists a constant K_3 such that $c_n^{-1} \le K_3 \, n^{b+\gamma}$, for (3.8) to hold we need $t \in (1, 2)$ such that $-t + (b+\gamma)[\, t(r+1) - 1 \,] < -1$. That such a value of t exists when A10 holds is shown in the proof of Lemma 3.5, Frees(1983). It follows that (3.8) holds and thus $n^{-1}\Sigma_{j=1}^{n} \; [hg_j^{(r)}(\phi_j) - \mathbf{E}_{\mathcal{F}_j} hg_j^{(r)}(\phi_j)] \to 0$ a.s., from which the lemma follows by Chow(1965). □

To finish the definition of C_n given by (3.3), let

$$C_{n+1}^{*} = \left[\frac{\Psi_n \, \beta_2 \, (1+\gamma)}{8r \, \Lambda_n^2 \, \beta_1^2} \right]^{1/(2r+1)} .$$

Lemma 3.8: Assume A1-A3, A9, and A10 hold. Then $C_n \to C_{opt}$ a.s.

Proof: The proof is immediate from (3.3), Lemma 3.6, and Lemma 3.7. □

Condition A9 is strengthened before stating Theorem 3.3.

A9$'$: For some integer $r \ge 1$, $Hog^{(1)}$ and $Hog^{(r+1)}$ exist and are bounded on the entire real line, and both functions are continuous in a neighborhood of ϕ. Further suppose $Hog^{(2)}$ is continuous in a neighborhood of ϕ and that for some $d > 0$ and all real x, $Hog^{(r+1)}(x) = Hog^{(r+1)}(\phi) + O(\, |x - \phi|^d \,)$.

Theorem 3.3: Assume A1-A3, A9$'$, and A10 hold. Then $n^{(1-\gamma)/2}(\phi_n - \phi) \overset{D}{\to} N(\mu_2, \sigma_2^2)$ where

$$\mu_2 = -\frac{2\, C_1 A_{opt}\, C_{opt}^{r}\, H \circ g^{(r+1)}(\phi)\, \beta_1}{g(\phi)\, [\, 1+\gamma\,]}$$

and

$$\sigma_2^2 = \frac{C_1^2\, A_{opt}^2\, C_{opt}^{-1}\, H \circ g^{(1)}(\phi)\, \beta_2}{[\, g(\phi)\,]^2\, [\, 1+\gamma\,]}$$

with $\beta_1 = \frac{1}{r!} \int_{-1}^{+1} u^r\, K(u)\, du$ and $\beta_2 = \int_{-1}^{+1} [\, K(u)\,]^2\, du$.

Proof: Once again the proof is an application of the Fabian result. By definition

$$(C_n n^{-\gamma})^{-1/2}\, V_n = Dg_n(\phi_n) - Dg(\phi_n) - \Delta_n.$$

For the algorithm (3.6), as in the proof of Theorem 2.2 with $U_n = \phi_n - \phi$,

$$U_{n+1} = U_n \left[\, 1 - A_n n^{-1} \tfrac{d}{dt} Dg(t)|_{t=\eta_n} \,\right] - A_n\, n^{-1}\, C_n^{-1/2} n^{\gamma/2} V_n - A_n n^{-1} \Delta_n$$

$$= U_n \left[\, 1 - n^{-1} \Gamma_n \,\right] + n^{(\frac{\gamma}{2}-1)} \Phi_n V_n + n^{-1} n^{-(1-\gamma)/2} T_n$$

where $\Gamma_n = A_n \frac{d}{dt} Dg(t)|_{t=\eta_n} \rightarrow A_{opt} \frac{d}{dt} Dg(t)|_{t=\phi} = 1,$

$$\Phi_n = -A_n C_n^{-1/2} \rightarrow \Phi = -A_{opt} C_{opt}^{-1/2},$$

and $T_n = -A_n\, n^{(1-\gamma)/2} \Delta_n.$

The indicated limits hold by Lemmas 3.5 and 3.8. Take $\alpha = 1$ and $\beta = 1 - \gamma$. Then

$$U_{n+1} = U_n [1 - n^{-\alpha}\Gamma_n] - n^{-(\alpha+\beta)/2}\Phi_n V_n + n^{-\alpha-(\beta/2)}T_n.$$

Let

$$T = \lim_{n\to\infty} T_n = - C_1 \lim_{n\to\infty} A_n [g(\phi_n)]^{-1} n^{(1-\gamma)/2} [E_{\mathcal{F}_n} hg_n(\phi_n) - H\circ g^{(1)}(\phi_n)]$$

$$= - C_1 A_{opt} [g(\phi)]^{-1} C_{opt}^r\, H\circ g^{(r+1)}(\phi)\, \beta_1$$

by Lemma 3.5 and the proof of Corollary 3.1.

By definition $E_{\mathcal{F}_n} V_n = 0$. Lemmas 3.1 and 3.2 hold with $c_n = C_n n^{-\gamma}$ and thus by slight modifications to the proof of Lemma 3.3,

$$\sum = \lim_{n\to\infty} E_{\mathcal{F}_n} V_n^2 = C_1^2 [g(\phi)]^{-2} H\circ g^{(1)}(\phi)\, \beta_2.$$

To see that $E_{\mathcal{F}_n} V_n^2$ is bounded, note that by Lemma 2.4 and Lemma 3.2 (a),

$$E_{\mathcal{F}_n} V_n^2 \le c_n \{ K_1 + K_2 E_{\mathcal{F}_n} | hg_n(\phi_n) |^2 + K_3 | E_{\mathcal{F}_n} [hg_n(\phi_n)] |^2 \},$$

$$\le c_n \{ K_1 + K_2 O(c_n^{-1}) + K_3 O(1) \}.$$

Thus there exists a constant K such that $K > | E_{\mathcal{F}_n} V_n^2 - \Sigma | \to 0$.

It remains to be shown that for each r = 1, 2, 3,... $\lim_{n\to\infty} E\{ V_n^2 \ \chi[V_n^2 \ge rn] \} = 0$. By Lemmas 2.4 and 3.2 (a), for $t \ge 1$

$$(3.9) \quad E_{\mathcal{F}_n} | c_n^{1/2} V_n |^t \le c_n^t [K_1 + K_2 E_{\mathcal{F}_n} | hg_n(\phi_n) |^t + K_3 | E[hg_n(\phi_n)] |^t]$$

$$\le c_n^t [K_1 + K_2 O(c_n^{-(t-1)}) + K_3 O(1)]$$

$$= o(1).$$

Take $1 \le q \le 3$ and $p \ge 3$ such that $2/p + 1/q = 1$. Using (3.9), it can be shown (see the proof of Lemma 2.7, Frees(1983)) that

$$E\{ V_n^2 \chi[V_n^2 \geq rn]\} \leq \frac{o(1)}{[n^{1/q} c_n]^{p/2}}$$

$$\leq \frac{o(1)}{[n^{1/q-(b+\gamma)}]^{p/2}}$$

$$= o(1)$$

since $b+\gamma < 1/3$. This completes the proof of the theorem. □

3.5 A Fixed-Width Confidence Interval For The Optimal BRP

This section concerns itself with opimal stopping for the stochastic approximation algorithm of Section 3.2. The specific objective is a stopping time for fixed-width confidence interval estimation of ϕ. The principal results given here are an application of the work of Frees (1985) on fixed-width confidence intervals for stochastic approximation procedures. (See also Frees (1983).) As in Sections 3.3 and 3.4, it is assumed throughout that the kernel estimator hg_n defined by (3.1) with $p=0$ is used to estimate $Hog^{(1)}$ in the procedure given by (2.2) and (2.3). The constants A and C will once again be taken as fixed.

Frees (1985) studied the following general stochastic approximation procedure. Let f: $\mathbf{R} \rightarrow \mathbf{R}$ be a measurable function such that θ is the unique root of $f(x)=0$. Suppose β_n and ϵ_n are random variables. Define successive estimates of θ recursively by

(3.10) $$X_{n+1} = X_n - n^{-1}\{ f(X_n) + n^{\tau-1/2}\beta_n + n^{\tau}\epsilon_n \}$$

where the constant τ is such that $0 \leq \tau \leq 1/2$. As noted by Frees, the random variables β_n and ϵ_n may be interpreted as the (standardized) bias and error that arise from estimating $f(X_n)$.

Using the sequence of estimates $\{X_n\}$ derived from the algorithm (3.10), Frees(1985) defines a stochastic process W_n on $D[0, \infty)$, the space of all functions defined on $[0, \infty)$ having left-hand limits and continuous from the right. Weak convergence properties of this stochastic process are explored to more fully characterize weak convergence of the sequence $\{X_n\}$. For appropriate sequences of positive random variables $\{N_n\}$ the same weak convergence is shown to hold for the randomized version W_{N_n}. This in turn is used to show that for certain sequences of stopping times $\{N_n\}$ the randomly stopped estimator X_{N_n} is asymptotically normal. Frees then gives a stopping time and confidence interval for θ.

Let $1-\alpha$, $\alpha \in (0, \frac{1}{2})$, be the desired asymptotic coverage probability. We will need to refer to Frees' assumptions and some results, quoted here for convenience.

F1: Let $\eta > 0$ and $G > 1/2 - \tau$. Assume that $f(x) = G(x-\theta) + O(|x-\theta|^{1+\eta})$.

F2: Let $\rho > 0$ and $\beta \in \mathbf{R}$. Assume that $\beta_n = \beta + o(n^{-\rho})$.

F3: $\lim_{n \to \infty} X_n = \theta$ a.s.

F4: There exists a standard Brownian motion B on $[0, \infty)$ and constants σ, $\epsilon > 0$ such that

$$\sum_{k \leq t} \epsilon_k = \sigma B(t) + O(t^{1/2-\epsilon}).$$

F5: Let $N_n/m_n = N + o_p(1)$, where N is a positive random variable, $\{m_n\}$ are integers going to infinity, and $\{N_n\}$ is a sequence of random variables.

F6: Let $\{\beta'_n\}$, $\{\sigma_n\}$, and $\{G_n\}$ be sequences of random variables such that $\beta'_n = \beta + o(1)$, $\sigma_n = \sigma + o(1)$, and $G_n = G + o(1)$.

Theorem(Frees):

Consider the algorithm defined by (3.10) and assume F1-F4. Let

$$W_n(t) = [nt]^{1/2-\tau}(X_{[nt]+1} - \theta) + \frac{\beta}{(G-1/2+\tau)}, \quad n = 1, 2, \dots \tag{3.11}$$

where in this expression $[\cdot]$ denotes the greatest integer function. Then there exists a standard Brownian motion B on $[0, \infty)$ such that

$$W_n(t) \xrightarrow{D} Z(t),$$

where

$$Z(t) = \{ 2(G-1/2+\tau) \}^{-1/2} \sigma t^{-(G-1/2+\tau)} B(t^{2(G-1/2+\tau)}).$$

Further assume that F5 holds. Then

$$W_{N_n}(t) \xrightarrow{D} Z(t). \qquad \square$$

Denote the $(1-\alpha)$-th quantile of the standard normal distribution by z_α. Define the stopping time N_d for each $d>0$ by

$$N_d = \begin{cases} \inf\{ n\geq 1: d \geq z_{\alpha/2}\sigma_n n^{-(1/2-\tau)}[2(G_n-1/2+\tau)]^{-1/2} \} \\ \infty \text{ if no such n exists,} \end{cases} \tag{3.12}$$

where the random variables σ_n, and G_n are described in F6. Let

$$I_n = [X_n + n^{-(1/2-\tau)}\left(\frac{\beta'_n}{(G_n-1/2+\tau)} - z_{\alpha/2}\sigma_n[2(G_n-1/2+\tau)]^{-1/2} \right), \tag{3.13}$$

$$X_n + n^{-(1/2-\tau)} \left(\frac{\beta'_n}{(G_n - 1/2 + \tau)} + z_{\alpha/2}\sigma_n [\, 2(G_n - 1/2+\tau)\,]^{-1/2} \right) \,]$$

be a confidence interval for θ.

Corollary (Frees): For the algorithm defined by (3.10) assume F1-F4 and F6. Then

$$\lim_{d \to 0} \mathbf{P}(\theta \in I_{N_d}) = 1-\alpha. \qquad \square$$

Define the following quantities:

$$f(x) = A\, Dg(x), \tag{3.14}$$

$$G = A \frac{d}{dx} Dg(x)|_{x=\phi}$$

$$\beta_n = C^r A\, C_1 [\, g(\phi_n)\,]^{-1} c_n^{-r}\, \mathbf{E}_{\mathcal{F}_n} [\, hg_n(\phi_n) - H \circ g^{(1)}(\phi_n)\,],$$

$$\beta = C^r A\, C_1 [\, g(\phi)\,]^{-1} H \circ g^{(r+1)}(\phi)\, \frac{1}{r!} \int_{-1}^{+1} u^r K(u)\, du,$$

$$\epsilon_n = n^{-\tau} A [\, Dg_n(\phi_n) - \mathbf{E}_{\mathcal{F}_n} Dg_n(\phi_n)] = A\, C^{-1/2} V_n,$$

$\tau = \gamma/2$, and $\kappa = 1/2 - \tau$.

Lemma 3.9 gives sufficient conditions for F4 to hold. The proof relies on the following theorem due to Strassen(1967), given here as it appears in Sen(1981).

Theorem(Strassen): Let $\{X_n, \mathcal{F}_n\}_{n=1}^{\infty}$ be a martingale and $X_0 = \mathbf{E} X_1$. Define $X'_n = X_n - X_{n-1}$ for $n \geq 1$ so that $\{X'_n, \mathcal{F}_n\}_{n=1}^{\infty}$ is a martingale difference sequence. Let

$$Y_n = \sum_{k=1}^{n} \mathbf{E}_{\mathcal{F}_{k-1}} (X'_k)^2, \quad n \geq 1,$$

where it is assumed that $\mathbf{E}_{\mathcal{F}_n}(X'_n)^2$ exists a.s. for every $n \geq 1$. Define the stochastic process $S = \{\, S(t);\, t \geq 0 \,\}$ by $S(0) = X_0 = 0$, $S(Y_n) = X_n$ for $n \geq 1$, and

$$S(t) = S(Y_n) + \left[\frac{t - Y_n}{Y_{n+1} - Y_n} \right] X'_{n+1} \quad \text{for } Y_n \le t \le Y_{n+1},\ n \ge 1.$$

Let $f = \{f(t);\ t > 0\}$ be a nonnegative, nondecreasing function such that $f(t)$ is increasing and $t^{-1}f(t)$ is decreasing in $t > 0$. Assume $Y_n \to \infty$ a.s. as $n \to \infty$ and

$$(3.15) \qquad \sum_{n=1}^{\infty} [f(Y_n)]^{-1}\, \mathbf{E}_{\mathcal{F}_{n-1}}[\,(X'_n)^2\, \chi[\,(X'_n)^2 > f(V_n)\,]\,] < \infty \ \text{ a.s.}$$

Then there exists a Brownian motion process $W = \{\, W(t);\ t \ge 0\,\}$ such that

$$S(t) = W(t) + o\Big((\log t)\, [t\, f(t)]^{1/4} \Big) \ \text{ a.s. as } t \to \infty. \qquad \square$$

<u>Lemma</u> <u>3.9:</u> Assume A1-A3, A6 and A9 hold with $\gamma = \frac{1}{2r+1}$ in A6 (for the r of A9). Then F4 holds with

$$\sigma = \{\, \sigma_1^2\, [2\Gamma - (\,1 - \gamma)]\,\}^{1/2},$$

where σ_1^2 appears in Corollary 3.1.

<u>Proof:</u> By definition, $\{\epsilon_n, \mathcal{F}_{n+1}\}_{n=1}^{\infty}$ is a martingale difference sequence. Under the conditions of Corollary 3.1, by the limit indicated there for $\mathbf{E}_{\mathcal{F}_n} V_n^2$,

$$\lim_{n\to\infty} \mathbf{E}_{\mathcal{F}_n} \epsilon_n^2 = A^2 C^{-1} C_1^2\, [\, g(\phi)\,]^{-2}\, H \circ g^{(1)}(\phi) \int_{-1}^{+1} [\, K(u)\,]^2\, du = \sigma_1^2\, [2\Gamma - (\,1 - \gamma)\,].$$

Thus, as required by the Strassen result,

$$\lim_{n\to\infty} Y_n = \lim_{n\to\infty} \sum_{k \le n} \mathbf{E}_{\mathcal{F}_k} \epsilon_k^2 = \infty.$$

Let f be a nonnegative, nondecreasing function. To show that condition (3.15) holds, we first note that for $p > 2$,

$$[f(Y_n)]^{-1}\mathbf{E}_{\mathcal{F}_n}\{\epsilon_n^2\,\chi[\epsilon_n^2 > f(Y_n)]\} = [f(Y_n)]^{-(p-1)}\mathbf{E}_{\mathcal{F}_n}\Big\{\epsilon_n^2\,[f(Y_n)]^{p-2}\chi[\epsilon_n^2 > f(Y_n)]\Big\}$$

$$\leq [f(Y_n)]^{-(p-1)}\,\mathbf{E}_{\mathcal{F}_n}\,|\,\epsilon_n|^p.$$

From the proof of Theorem 3.1 (b), also using Lemma 3.2 (a),

$$c_n^{p/2}\,\mathbf{E}_{\mathcal{F}_n}|\,V_n|^p \leq c_n^p\{\,K_1 + K_2 n^{(p-1)\gamma} + K_3\,\}.$$

It follows that

$$\mathbf{E}_{\mathcal{F}_n}|\,\epsilon_n|^p \leq K_4\, n^{(p/2-1)\gamma}.$$

Thus, taking $\epsilon > 0$ and $f(t) = t^{1-\epsilon}$,

$$\sum_{n=1}^{\infty} [f(Y_n)]^{-1}\mathbf{E}_{\mathcal{F}_n}\{\,\epsilon_n^2\,\chi[\epsilon_n^2 > f(Y_n)]\,\} \;\leq \sum_{n=1}^{\infty} K_4\,[f(Y_n)]^{-(p-1)} n^{(p/2-1)\gamma}$$

$$= \sum_{n=1}^{\infty} K_4\,[\,(Y_n/n)^{1-\epsilon} n^{1-\epsilon}]^{-(p-1)} n^{(p/2-1)\gamma}$$

$$\leq \sum_{n=1}^{\infty} K_5\, n^{-(1-\epsilon)(p-1)} n^{(p/2-1)\gamma}.$$

This last summation is convergent for sufficiently small ϵ and large p. The remainder of the proof is identical to the latter portion of the proof of Lemma 5.4 of Frees(1983), applying the Strassen result to show that there exists a Brownian motion such that F4 holds. □

In Lemma 3.10 the assumptions F1 and F3 are shown to hold under conditions similar to those of Corollary 3.1. Assumption A9′ is first strengthened:

A9″: For some integer $r \geq 2$, $Hog^{(1)}$, $Hog^{(2)}$, $Hog^{(3)}$, and $Hog^{(r+1)}$ exist and are bounded on the entire real line. Furthermore, $Hog^{(1)}$, $Hog^{(2)}$, and $Hog^{(r+1)}$ are continuous in a neighborhood of ϕ. Also assume that for some $d > 0$ and all real x,

$$\mathrm{Hog}^{(r+1)}(x) = \mathrm{Hog}^{(r+1)}(\phi) + O(\,|x-\phi|^{d}\,).$$

Lemma 3.10: Assume A1-A3, A6, and A9$''$ hold. Then F1 and F3 also hold with $G = A\frac{d}{dx}Dg(x)|_{x=\phi} = \Gamma$.

Proof: B3 holds by Corollary 3.1 (a).

Expanding the function $f(x) = A\,Dg(x)$ about ϕ yields

$$f(x) = G(x-\phi) + \tfrac{1}{2}A\,Dg^{(2)}(\phi')(x-\phi)^2$$

for some ϕ', $|\phi' - \phi| \le |x - \phi|$. □

Lemma 3.11 verifies condition F2.

Lemma 3.11: If A1-A3, A6, and A9$''$ hold, then F2 holds.

Proof: By definition and the proof of Lemma 3.1 (b),

$$\beta_n - \beta =$$

$$A\,C^r C_1 \Big\{ [g(\phi_n)]^{-1} c_n^{-r}\, \mathbf{E}_{\mathcal{F}_n}[\, hg_n(\phi_n) - \mathrm{Hog}^{(1)}(\phi_n)\,] - [g(\phi)]^{-1}\mathrm{Hog}^{(r+1)}(\phi)\,\frac{1}{r!}\int_{-1}^{+1} u^r K(u)\,du \Big\}$$

$$= K_1\,[g(\phi_n)]^{-1}\,\frac{1}{r!}\int_{-1}^{+1} u^r\,K(u)\,[\,\mathrm{Hog}^{(r+1)}(\eta(u)) - \mathrm{Hog}^{(r+1)}(\,\phi\,)\,]\,du$$

where $|\,\eta(u) - \phi_n\,| \le c_n u$. Thus by A3, A9$''$, and the asymptotic normality of ϕ_n,

$$\beta_n - \beta = K_1 [g(\phi_n)]^{-1} \frac{1}{r!}\int_{-1}^{+1} u^r K(u) [Hog^{(r+1)}(\eta(u)) - Hog^{(r+1)}(\phi)] du$$

$$+ K_1\left\{ [g(\phi_n)]^{-1} - [g(\phi)]^{-1}\right\} Hog^{(r+1)}(\phi) \frac{1}{r!}\int_{-1}^{+1} u^r K(u) du$$

$$= O(| \eta(u) - \phi |^d) + O(| \phi_n - \phi |)$$

$$= O([O(n^{-r\gamma} + c_n)]^d) + O(| \phi_n - \phi |)$$

$$= O(n^{-d\gamma}) + O(n^{-r\gamma})$$

which establishes the lemma. □

Theorem 3.4: Assume A1-A3, A6, and A9$''$ hold. For $t \geq 0$ let the stochastic process W_n be defined by (3.11), with β, G, and κ defined by (3.12). Then there exists a standard Brownian motion B such that

$$W_n(t) = [nt]^{1/2-\tau}(X_{[nt]+1} - \theta) - \mu_1 \xrightarrow{D} Z(t) \quad \text{as } n \to \infty$$

where

$$Z(t) = \{ 2(G-\kappa) \}^{-1/2} \sigma t^{-(G-\kappa)} B(t^{2(G-\kappa)}).$$

The constant σ has been defined in Lemma 3.9.

Proof: By (3.14), $\frac{\beta}{G-\kappa} = \mu_1$. The theorem follows immmediately from the theorem due to Frees(1985) since the assumptions have been shown to hold in Lemmas 3.9, 3.10, and 3.11. □

Define $\Upsilon_n = n^{-1} \sum_{j=1}^{n} [g(\phi_n)]^{-1}$. Let α_n, Λ_n, and Ψ_n be as defined in Section 3.4. Also, define as estimates of G, β, and σ the random quantities

$$G_n = A\,n^{-1} \sum_{j=1}^{n} \alpha_j ,$$

$$\beta_n' = C^r A\, C_1 \beta_1 \Upsilon_n \Lambda_n,$$

and

$$\sigma_n = C_1 A\, \Upsilon_n \left[C^{-1} \beta_2 \Psi_n\right]^{1/2},$$

respectively.

Corollary 3.2: Assume A1-A3, A6, and A9'' hold. For $d > 0$, define N_d by (3.12) and I_n by (3.13). Then $\lim_{d \to 0} P(\phi \in I_{N_d}) = 1 - \alpha$.

Proof: By small changes to the proof of Lemma 3.5 it can be seen that $G_n \to G$ a.s. Since Corollary 3.1 holds, $\Upsilon_n \to [g(\phi)]^{-1}$ a.s. Thus $\beta_n' \to \beta$ a.s. by Lemma 3.7. Similarly, by Lemma 3.6, $\sigma_n \to \sigma$ a.s. Thus assumption F6 of Frees is satisfied. Since F1-F4 hold by Lemmas 3.9, 3.10, and 3.11, the result holds by the Frees corollary. □

CHAPTER 4

RENEWAL FUNCTION ESTIMATION AT A POINT

4.1 Introduction

Let X, X_1, $X_2,\ldots$ be independent and identically distributed random variables having distribution function F. Put $S_k = X_1+\ldots+X_k$ and denote by $F^{(k)}(t) = \mathbf{P}(S_k \le t)$ the k-fold convolution of F. The renewal function H can then be defined by

$$(4.1) \qquad H(t) = \sum_{k\ge1} F^{(k)}(t), \quad -\infty < t < \infty.$$

In Section 2 we discuss moment conditions on F sufficient for the finiteness of H(t), $-\infty < t < \infty$. If F is a lifetime distribution (i.e., $F(0^-)=0$), then H(t) is the expectation of the renewal random variable $N(t) = \max\{k\colon S_k \le t\}$. In general, H(t) is the expected number of partial sums S_k such that $S_k \le t$.

Motivated by (4.1), Frees (1986ii) considered estimating H(t) for $t\ge0$ by first estimating a finite number of convolutions of F. For $1\le k\le n$, let $X_{c1},\ldots,X_{ck}$ be a subset of k observations from the random sample $X_1,\ldots,X_n$. Denote by $\sum_c$ the summation over all $\binom{n}{k}$ such distinct subsets of k observations. A natural estimate of $F^{(k)}(t)$ is the U-statistic:

$$F_n^{(k)}(t) = \binom{n}{k}^{-1}\sum_c \chi(X_{c1}+\cdots+X_{ck} \le t).$$

Let $m=m(n)$ be a monotone increasing sequence of positive integers such that $m\le n$ and $m(n) \to \infty$ as $n\to\infty$. Define the renewal function estimator $H_n(t)$ by

$$H_n(t) = \sum_{k=1}^{m} F_n^{(k)}(t). \tag{4.2}$$

Let $X^- = \min(0, X)$. Assuming F has positive mean μ and finite variance σ^2, Frees showed that $H_n(t) \to H(t)$ a.s. if either

(4.3) for some $r \geq 3$, $\mathbf{E}|X^-|^r < \infty$ and $n = O(m^{r-2})$,

or

(4.4) for some $\theta_1 > 0$ and all $|\theta| < \theta_1$, $\mathbf{E}\exp(-\theta X^-) < \infty$ and $\log n = o(m)$.

The assumptions about the growth of the parameter m are such that the bias goes to zero at a fast enough rate for a.s. convergence. Frees also establishes that when $m(n) = n$, this convergence is uniform over bounded subsets of nonnegative real numbers: for fixed $T > 0$

$$\sup_{t \in [0,T]} |H_n(t) - H(t)| \to 0 \ \text{a.s.}$$

Under conditions similar to (4.3) and (4.4), it is proven that $n^{1/2}[H_n(t) - H(t)]$ is asymptotically normal with the limiting variance σ_t^2 dependent on t and F. Frees provides a weakly consistent estimator of σ_t^2, allowing the formulation of large sample confidence intervals for H(t). In Section 2 of this chapter weaker sufficient conditions for the consistency and asymptotic normality of $H_n(t)$ are given.

Large sample behavior of $H_n(t)$ for a single value of t is further characterized in Section 3 by an invariance principle useful for studying sequential versions of $H_n(t)$. In Section 4 two consistent estimators of the asymptotic variance σ_t^2 are given. One of these, considered by Frees, relies on an explicit expression for σ_t^2; the other is a jackknife estimator. A stopping time for a fixed-width confidence interval for H(t) is defined in Section 5. Asymptotic properties of the stopping time (as the desired interval width goes to zero) are then explored.

4.2 A Correction And Some Generalizations

A small error occurring in Frees (1986ii) is corrected before proceeding to note how some of Frees' results hold under more general conditions than given in that paper.

Theorem 2.1 of Frees (1986ii) gives sufficient conditions for

$$\lim_{n\to\infty} \int_0^\infty g_a(u)\,d\left(\sum_{k=1}^{m} k^a F_n^{(k)}(u)\right) = \int_0^\infty g_a(u)\,d\left(\sum_{k\geq 1} k^a F^{(k)}(u)\right) \text{ a.s.,} \tag{4.5}$$

where $a \in \mathbf{R}$ and $g_a\colon \mathbf{R}^+ \to \mathbf{R}$ is such that

$$\int_0^\infty |g_a(u)|\,d\left(\sum_{k\geq 1} k^a F^{(k)}(u)\right) < \infty.$$

Setting $a = 0$ and $g_a(u) = \chi(u \leq t)$ for some $t \in \mathbf{R}^+$, if (4.5) holds, then $H_n(t) - H_n(0^-) \to H(t) - H(0^-)$ a.s. This does not imply the consistency of $H_n(t)$ when $H(0^-) > 0$. Instead, let $g_a\colon \mathbf{R} \to \mathbf{R}$ be a function such that

$$\int_{-\infty}^\infty |g_a(u)|\,d\left(\sum_{k\geq 1} k^a F^{(k)}(u)\right) < \infty.$$

The trivial remedy is then to replace integration over the half-line $[0,\infty)$ by integration over the whole real line in Frees' Theorems 2.1 and 2.2.

The two results quoted here due to Heyde (1964) bear directly on the question of when $H(t)$ is finite. They will be used to indicate how Frees' estimator is more generally applicable.

Lemma (Heyde): Let $\mathbf{E}|X| < \infty$ and $\mathbf{E}X > 0$, or else $\mathbf{E}|X| = \infty$ and, in either case, $\mathbf{E}|X^-|^r < \infty$ for some integer $r \geq 1$. Then

$$\sum_{k\geq 1} k^{r-2}\, \mathbf{P}(S_k \leq x) < \infty, \quad -\infty < x < \infty.$$

Theorem (Heyde): Suppose $\mathbf{E}|X| < \infty$, $\mathbf{E}X > 0$, and let r be a nonnegative integer. A necessary and sufficient condition for the convergence of the series

$$\sum_{k\geq 1} k^{r}\, \mathbf{P}(S_k \leq x), \quad -\infty < x < \infty,$$

is that $\mathbf{E}|X^-|^{r+2} < \infty$.

Thus if $\mathbf{E}|X^-| < \mathbf{E}|X^+| \leq \infty$, then by the lemma a sufficient condition for $H(t)$ to be finite is that $\mathbf{E}|X^-|^2 < \infty$. By the theorem we see that when $0 < \mathbf{E}X < \infty$, the condition that $\mathbf{E}|X^-|^2 < \infty$ is also necessary.

The remainder of this section describes how the sufficient conditions for consistency and asymptotic normality of $H_n(t)$ given by Frees can be weakened. Frees assumed throughout his paper that F has positive mean μ and finite variance σ^2. Note that (4.3) and (4.4) both imply that $\mathbf{E}|X^-|^2 < \infty$ and thus that $H(t) < \infty$ for all $t \in \mathbb{R}$. If (4.3) holds, then Frees' proof of the consistency of $H_n(t)$ requires only that $0 \leq \mathbf{E}|X^-| < \mathbf{E}|X^+| \leq \infty$, the proof not relying on a finite mean μ or finite variance σ^2. Assuming (4.4), Frees uses the condition $0 < \mu < \infty$ to show that

(4.6) there exists a $\theta_2 > 0$ and $0 < p < 1$ such that $F^{(k)}(t) \leq \exp(\theta_2 t)p^k$ for all $k \geq 1$, for all $t \in \mathbf{R}$.

The remainder of the consistency proof assuming (4.4) relies solely on the inequality (4.6). Thus, in this case, the assumption of a finite variance can be dropped.

If $F(0^-) = 0$ and $F(0) < 1$, then it is easily shown that (4.6) holds with $\theta_2 = 1$, even when $\mathbf{E}X = \infty$. Taking $\log n = o(m)$, the consistency of $H_n(t)$ follows by the same arguments used by Frees when (4.4) holds and $\mu < \infty$.

Theorem 2.2 of Frees is now restated with the weaker conditions suggested by the preceding arguments. With the previously mentioned corrections to Section 2 of Frees' paper, it is not necessary that $t \geq 0$.

<u>Theorem A (Frees)</u>: Assume

(4.7) $\quad 0 \leq \mathbf{E}|X^-| < \mathbf{E}|X^+| \leq \infty$ and (4.3) holds,

or

(4.8) $\quad$ F has finite mean $\mu > 0$ and (4.4) holds,

or

(4.9) $\quad F(0^-) = 0, \;\; F(0) < 1$, and $\log n = o(m)$.

Then for each $t \in \mathbf{R}$, $H_n(t) \to H(t)$ a.s. as $n \to \infty$.

If F is continuous, it is not essential that $m(n) = n$ for there to be uniform convergence on finite intervals.

<u>Theorem:</u> Assume F is continuous. Further suppose that (4.7) or (4.8) or (4.9) holds. Then for fixed $T > 0$

$$\sup_{t \in [0,T]} |H_n(t) - H(t)| \to 0 \text{ a.s.}$$

<u>Proof:</u> If F is continuous, then H is also continuous. Thus $\{H_n\}$ is a sequence of monotone increasing functions converging pointwise to a continuous monotone increasing function on the interval [0, T] by Theorem A. It follows that the convergence is uniform on the interval. □

Frees' result on the asymptotic normality of $n^{1/2}[H_n(t) - H(t)]$ also holds under somewhat weaker conditions. These less stringent assumptions are identified separately here, since they will be used throughout Chapter 5 and 6.

B1: $0 \le \mathbf{E}|X^-| < \mathbf{E}|X^+| \le \infty$. For some $r \ge 5$, $\mathbf{E}|X^-|^r < \infty$ and $n = O(m^{2r-4})$.

B2: F has finite mean $\mu > 0$. For some $\theta_1 > 0$ and all $|\theta| < \theta_1$, $\mathbf{E}\exp(-\theta X^-) < \infty$ and $\log n = o(m)$.

B3: $F(0^-) = 0$, $F(0) < 1$, and $\log n = o(m)$.

Assumptions B2 and B3 are identical to (4.8) and (4.9).

A modified version of Theorem 3.1 of Frees is now given. Again, the result can be extended to include the case when t is negative.

Let X have distribution function F and

$$\xi_{rst}(c) = \mathrm{Cov}\{F^{(r-c)}(t-X), F^{(s-c)}(t-X)\}$$

for integer $r,s \ge 1$ and $1 \le c \le r \wedge s$. Define $\sigma_t^2 = \sum_{r,s=1}^{\infty} rs\,\xi_{rst}(1)$.

Theorem B (Frees): Assume B1 or B2 or B3 holds. Then for each $t \in \mathbf{R}$,

$$n^{1/2}[H_n(t) - H(t)] \xrightarrow{D} N(0, \sigma_t^2).$$

A careful reading of the proof of Frees' Theorem 3.1 shows that the weaker moment conditions of B1 and B2 still allow the indicated result. That B3 is a sufficient condition follows from Frees' proof by using the fact that B3 implies (4.6). (See also Section 5.3, Theorem 5.3 for a proof of this result.)

4.3 An Invariance Principle

For a fixed value of t, define the stochastic process $Y_n = \{Y_n(s), s \in [0,1]\}$ by

$$Y_n(s) = 0, \quad 0 \leq s < m(n)/n,$$

$$Y_n(s) = Y_n(k/n), \quad k/n \leq s < (k+1)/n, \quad k = m(n), \ldots, n-1,$$

where $$Y_n(k/n) = \frac{k[H_k(t) - H(t)]}{n^{1/2}\sigma_t}, \quad \text{and} \quad Y_n(1) = \frac{n^{1/2}[H_n(t) - H(t)]}{\sigma_t}.$$

Let D[0,1] be the space of all functions on [0, 1] that are right continuous and have left-hand limits. Then $Y_n \in D[0,1]$ for all n. It will be shown that Y_n converges weakly to a standard Brownian motion in the J_1-topology on D[0,1].

For the results of this section we will need a slightly stronger condition than B1. Thus let B1$'$ be as follows.

B1$'$: F is such that $0 \leq \mathbf{E}|X^-| < \mathbf{E}|X^+| \leq \infty$. For some $r \geq 7$, $\mathbf{E}|X^-|^r < \infty$ and $n = O(m^{2r-4})$.

We first consider a process closely related to Y_n. Let $Y'_n = \{Y'_n(s),\ s \in [0, 1]\}$ be defined by

$$Y'_n(s) = 0, \quad 0 \leq s < m(n)/n,$$

$$Y'_n(s) = Y'_n(k/n), \quad k/n \leq s < (k+1)/n, \quad k = m(n), \ldots, n-1,$$

where
$$Y'_n(k/n) = \frac{k[H_k(t) - \Sigma_{j=1}^{m(k)} F^{(j)}(t)]}{n^{1/2}\sigma_t}, \quad k = m(n), \ldots, n-1,$$

and
$$Y'_n(1) = \frac{n^{1/2}[H_n(t) - \Sigma_{j=1}^{m(n)} F^{(j)}(t)]}{\sigma_t}.$$

Lemma 4.1 states the asymptotic equivalence of Y_n and Y'_n.

<u>Lemma</u> <u>4.1</u>: Assume B1 or B2 or B3 holds. Then

$$\sup_{s \in [0,1]} |Y'_n(s) - Y_n(s)| \to 0 \qquad \text{as } n \to \infty.$$

<u>Proof</u>: Note that

$$\sup_{s \in [0,1]} |Y'_n(s) - Y_n(s)| = \max_{m(n) \leq k \leq n} \frac{k}{n^{1/2}\sigma_t} \sum_{j > m(k)} F^{(j)}(t)$$

$$\leq \frac{1}{\sigma_t} \max_{m(n) \leq k \leq n} k^{1/2} \sum_{j > m(k)} F^{(j)}(t)$$

$$\to 0 \quad \text{as } n \to \infty$$

by Lemma 3.1 of Frees (1986ii). □

In defining several other stochastic processes, use is made of a standard decomposition for U-statistics due to Hoeffding. For integer $\ell \geq 1$ let

$$g_\ell(x_1,\ldots,x_\ell) = \chi(x_1+\ldots+x_\ell \leq t)$$

and

$$g_{\ell,d}(x_1,\ldots,x_d) = \mathbf{E}\, g_\ell(x_1,\ldots,x_d,\ X_{d+1},\ldots,X_\ell)$$

$$= F^{(\ell-d)}(t-(x_1+\ldots+x_d)),\quad 0 \leq d \leq \ell.$$

Also, define the kernel

$$g^o_{\ell,d}(x_1,\ldots,x_d) = \sum_{h=0}^{d} (-1)^h S_{\ell,d,h},$$

where $S_{\ell,d,h} = \sum_c g_{\ell,(d-h)}(x_{j1},\ldots,x_{j(d-h)})$, the summation being over all $\binom{d}{d-h}$ subsets of $d-h$ elements drawn from $\{x_1,\ldots,x_d\}$. Then $F_k^{(\ell)}(t)$ may be written as

$$F_k^{(\ell)}(t) = F^{(\ell)}(t) + \sum_{d=1}^{\ell} \binom{\ell}{d} F^{(\ell)}_{k,d},$$

where

$$F^{(\ell)}_{k,d} = \binom{k}{d}^{-1} \sum_c g^o_{\ell,d}(X_{j1},\ldots,X_{jd}),$$

$\sum_c$ in this instance denoting summation over all subsets of d observations from $X_1,\ldots,X_k$.

Let $\zeta_{\ell,d} = \mathbf{Var}\{g_{\ell,d}(X_1,\ldots,X_d)\}$. A standard property of the $\zeta_{\ell,d}$ is that $0 \leq \zeta_{\ell,d} \leq \zeta_{\ell,\ell}$ for $0 \leq d \leq \ell$. Another fact is that

$$\mathbf{E}[g^o_{\ell,d}(X_{i1},\ldots,X_{id})\, g^o_{\ell,d}(X_{j1},\ldots,X_{jd})] = 0$$

unless $\{i_1,\ldots,i_d\} = \{j_1,\ldots,j_d\}$. Thus

$$\mathbf{E}[F^{(\ell)}_{k,d}]^2 = \binom{k}{d}^{-1}\mathbf{E}[g^{o}_{\ell,d}(X_1,\dots,X_d)]^2$$

$$\leq \binom{k}{d}^{-1}(2^d)^2 \max_{0\leq h\leq d} \mathbf{E}[g_{\ell,(d-h)}(X_1,\dots,X_{d-h})]^2$$

$$\leq \binom{k}{d}^{-1} 4^d F^{(\ell)}(t).$$

Another property of the decomposition given for $F^{(\ell)}_k(t)$ is that for $1 \leq d < d' \leq \ell$, $\mathbf{E}[F^{(\ell)}_{k,d} F^{(\ell)}_{k,d'}] = 0$. Hence, for future reference

$$\mathbf{E}\Big[\sum_{d=2}^{\ell} \binom{\ell}{d} F^{(\ell)}_{k,d}\Big]^2 \leq \sum_{d=2}^{\ell} \binom{\ell}{d}^2 \binom{k}{d}^{-1} 4^d F^{(\ell)}(t) \tag{4.10}$$

$$\leq K_1 \sum_{d=2}^{\ell} \frac{[\ell!]^2 (k-d)!}{[(\ell-d)!]^2 k!} F^{(\ell)}(t)$$

where the constant K_1 is independent of k, ℓ, and d.

We now define a projection of Y'_n in terms of i.i.d. random variables. Let $Y^{(1)}_n = \{Y^{(1)}_n(s), s \in [0, 1]\}$ with

$$Y^{(1)}_n(s) = 0, \quad 0 \leq s < 1/n,$$

$$Y^{(1)}_n(s) = Y^{(1)}_n(k/n), \quad k/n \leq s < (k+1)/n, \ 1 \leq k \leq n-1,$$

where $$Y^{(1)}_n(k/n) = \frac{k\,\Sigma_{j=1}^{m(k)}\, j\, F^{(j)}_{k,1}}{n^{1/2}\sigma_t} = \frac{\Sigma_{i=1}^{k}\,\Sigma_{j=1}^{m(k)}\, j\,[F^{(j-1)}(t-X_i) - F^{(j)}(t)]}{n^{1/2}\sigma_t},$$

$k = 1,\dots,n.$

Next, let $Y_n^{(2)} = \{ Y_n^{(2)}(s), s \in [0, 1]\}$ be defined by

$$Y_n^{(2)}(s) = 0, \quad 0 \le s < 1/n,$$

$$Y_n^{(2)}(s) = Y_n^{(2)}(k/n), \quad k/n \le s < (k+1)/n, \; 1 \le k \le n-1,$$

where $$Y_n^{(2)}(k/n) = \frac{k \, \Sigma_{j=1}^{m(n)} F_{k,1}^{(j)}}{n^{1/2}\sigma_t} = \frac{\Sigma_{i=1}^{k} \Sigma_{j=1}^{m(n)} j[F^{(j-1)}(t-X_i) - F^{(j)}(t)]}{n^{1/2}\sigma_t},$$

$$k = 1,\ldots,n.$$

We first show that under appropriate conditions $Y_n^{(2)} \xrightarrow{W} B$, where B is a standard Brownian motion in D[0,1]. Using $Y_n^{(1)}$, it is then shown that $\sup_{s\in[0,1]} |Y_n^{(2)}(s) - Y_n'(s)| \xrightarrow{P} 0$, implying by Lemma 3.1 that $Y_n \xrightarrow{W} B$.

<u>Lemma</u> <u>4.2</u>: Assume B1 or B2 or B3 holds. Then $Y_n^{(2)} \xrightarrow{W} B$ as $n \to \infty$ in the J_1-topology on D[0,1], where B is a standard Brownian motion.

<u>Proof</u>: Define the random variables Z_{ni} by

$$Z_{ni} = \frac{\Sigma_{j=1}^{m(n)} j[F^{(j-1)}(t-X_i) - F^{(j)}(t)]}{n^{1/2}\sigma_t}, \quad i = 1,\ldots,n,$$

and

$$Z_{n0} = 0.$$

Let $S_{nk} = \Sigma_{i=0}^{k} Z_{ni}$ and $\mathcal{F}_{nk} = \mathcal{F}_k = \sigma(X_1, \ldots, X_k)$, $k = 0,1,\ldots,n$. We note that

$$E_{\mathcal{F}_{n(i-1)}} Z_{ni} = 0, \quad i = 1,\ldots,n,$$

and

$$E_{\mathcal{F}_{n(i-1)}} Z_{ni}^2 = \frac{1}{n\,\sigma_t^2} \sum_{r,s=1}^{m(n)} rs\,\xi_{rst}(1), \quad i = 1,\ldots,n.$$

Thus, for any fixed $s \in [0, 1]$ and associated sequence of integers $\{k_n\}_{n=1}^{\infty}$ such that $k_n/n \le s < (k_n + 1)/n$ for all n, we have that

$$\sum_{i=1}^{k_n} E_{\mathcal{F}_{n(i-1)}} Z_{ni}^2 = \frac{k_n}{n\,\sigma_t^2} \sum_{r,s=1}^{m(n)} rs\,\xi_{rst}(1) \to s \quad \text{as } n \to \infty.$$

Also, by Lemma 3.2 of Frees (1986ii) or Lemma 5.3 of this work, the Lindeberg condition

$$\sum_{i=1}^{n} E_{\mathcal{F}_{n(i-1)}} [\, Z_{ni}^2\, \chi(|Z_{ni}| > \epsilon)] \xrightarrow{P} 0 \quad \text{as } n \to \infty$$

is satisfied. By a theorem of McLeish (1974) (see Sen(1981), Theorem 2.4.2), $Y_n^{(2)}$ converges weakly (in the J_1-topology on D[0,1]) to a standard Brownian motion in C[0,1]. □

<u>Lemma</u> <u>4.3</u>: Assume B1$'$ or B2 or B3 holds. Then for every $\epsilon > 0$

$$P\left\{ \max_{m(n) \le k \le n} |Y_n^{(2)}(k/n) - Y_n^{(1)}(k/n)| > \epsilon \right\} \to 0 \quad \text{as } n \to \infty.$$

<u>Proof</u>: Let

$$\eta_{nk} = \sum_{i=1}^{k} \frac{\Sigma_{j=m(k)+1}^{m(n)}\, j\, [F^{(j-1)}(t-X_i) - F^{(j)}(t)]}{n^{1/2}\sigma_t}$$

and $\mathcal{F}_k = \sigma(X_1,\ldots,X_k)$. Define $m^{-1}(\ell) = \inf\{k\colon m(k) \ge \ell\}$. Then for $\ell = 1,\ldots,m(n)$, $\{\eta_{nk};\ \mathcal{F}_k\}_{k=m^{-1}(\ell)}^{m^{-1}(\ell+1)-1}$ is a zero mean martingale. Thus

$$\mathbf{P}\Big\{\max_{m(n)\le k\le n}|Y_n^{(2)}(k/n)-Y_n^{(1)}(k/n)|>\epsilon\Big\}$$

$$=\mathbf{P}\Big\{\max_{m(n)\le k\le m^{-1}(m(n))-1}|\eta_{nk}|>\epsilon\Big\}$$

$$\le\sum_{\ell=m(m(n))}^{m(n)-1}\mathbf{P}\Big\{\max_{m^{-1}(\ell)\le k<m^{-1}(\ell+1)}|\eta_{nk}|>\epsilon\Big\}$$

$$=\sum_{\ell=m(m(n))}^{m(n)-1}\mathbf{P}\Big\{\max_{m^{-1}(\ell)\le k<m^{-1}(\ell+1)}|\eta_{nk}|^2>\epsilon^2\Big\}$$

$$=\sum_{\ell=m(m(n))}^{m(n)-1}\frac{m^{-1}(\ell+1)}{n\,\sigma_t^2\,\epsilon^2}\sum_{r,s=\ell+1}^{m(n)} rs\,|\xi_{rst}(1)|,$$

by the Kolmogorov inequality for submartingales.

Note that

(4.11) $$|\xi_{rst}(1)|\ \le\ F^{(r\vee s)}(t),$$

where $r\vee s$ is the maximum of r and s. Also,

(4.12) B2 and B3 both imply by (4.6) that for any fixed $a\in\mathbf{R}$

$$\sum_{r=1}^{\infty} r^a F^{(r)}(t)\ <\ \infty$$

for each $t\in\mathbf{R}$.

Thus, since $m^{-1}(m(n))\le n$,

$$\sum_{\ell=m(m(n))}^{m(n)-1} \frac{m^{-1}(\ell+1)}{n\,\sigma_t^2\,\epsilon^2} \sum_{r,s=\ell+1}^{m(n)} rs\,|\xi_{rst}(1)|$$

$$\le \frac{1}{\sigma_t^2\,\epsilon^2} \sum_{\ell=m(m(n))}^{\infty} \sum_{r,s=\ell+1}^{\infty} rs\,F^{(r\vee s)}(t)$$

$$\le K_1 \sum_{r=m(m(n))}^{\infty} r^4 F^{(r)}(t) \;\rightarrow\; 0$$

as $n \rightarrow \infty$ by (4.12) if B2 or B3 holds and by the Heyde lemma if B1$'$ holds. This completes the proof. □

Lemma 4.4: Assume B1$'$ or B2 or B3 holds. If $[m(n)]^2 \log n = o(n)$, then for every $\epsilon > 0$

$$\mathbf{P}\Big\{\max_{m(n)\le k\le n} |Y_n^{(1)}(k/n) - Y_n'(k/n)| > \epsilon\Big\} \;\rightarrow\; 0 \qquad \text{as } n \rightarrow \infty.$$

Proof: Using the previously defined decomposition of $F_k^{(\ell)}(t)$, we have that

$$\mathbf{P}\Big\{\max_{m(n)\le k\le n} |Y_n^{(1)}(k/n) - Y_n'(k/n)| > \epsilon\Big\}$$

$$= \mathbf{P}\Big\{\max_{m(n)\le k\le n} k\Big|\sum_{\ell=2}^{m(k)} \sum_{d=2}^{\ell} \binom{\ell}{d} F_{k,d}^{(\ell)}\Big| > n^{1/2}\epsilon\,\sigma_t\Big\}$$

$$\le \sum_{\ell=2}^{m(n)} \mathbf{P}\Big\{\max_{m^{-1}(\ell)\le k\le n} k\Big|\sum_{d=2}^{\ell} \binom{\ell}{d} F_{k,d}^{(\ell)}\Big| > \frac{n^{1/2}\epsilon\,\sigma_t}{m(n)}\Big\}.$$

Let $X_{k:1}, X_{k:2}, \ldots, X_{k:k}$ be the order statistics of the random sample of $X_1, \ldots, X_k$. Define $\mathcal{C}_k = \sigma\{X_{k:1}, X_{k:2}, \ldots, X_{k:k};\ X_{k+1}, \ldots)$, $k = 1, 2, \ldots$, the usual symmetric σ-fields associated with $X_1, X_2, \ldots$. Since the $F^{(\ell)}_{k,d}$ are U-statistics for $k \geq d$, $\{F^{(\ell)}_{k,d}, \mathcal{C}_k\}_{k \geq d}$ is a reverse martingale. Letting

$$\mathcal{N}_{\ell k} = \sum_{d=2}^{\ell} \binom{\ell}{d} F^{(\ell)}_{k,d}, \quad m^{-1}(\ell) \leq k < \infty, \quad \ell = 2,3,\ldots,$$

it follows that $\{\mathcal{N}_{\ell k}, \mathcal{C}_k\}_{k=m^{-1}(\ell)}^{\infty}$ is a reverse martingale for each $\ell = 2,3,\ldots$.

An upper bound for $\mathbf{E}\mathcal{N}^2_{\ell k}$ is now derived. By (4.10)

$$\mathbf{E}\mathcal{N}^2_{\ell k} \leq K_1 \sum_{d=2}^{\ell} \frac{[\ell!]^2 (k-d)!}{[(\ell-d)!]^2 k!} F^{(\ell)}(t)$$

$$\leq K_2 \frac{\ell^4}{k^2} F^{(\ell)}(t) \sum_{d=2}^{\ell} \frac{[(\ell-2)!]^2 (k-d)!}{[(\ell-d)!]^2 (k-2)!} .$$

Because $[m(n)]^2 \log n = o(n)$, ultimately $[m(n)]^2 \leq n$. Thus for large enough ℓ,

$$\ell^2 \leq [m(m^{-1}(\ell))]^2 \leq m^{-1}(\ell) \leq k$$

and

$$\mathbf{E}\mathcal{N}^2_{\ell k} \leq K_2 \frac{\ell^4}{k^2} F^{(\ell)}(t) \sum_{d=2}^{\ell} \frac{[(\ell-2)!]^2 (\ell^2-d)!}{[(\ell-d)!]^2 (\ell^2-2)!}$$

$$\leq K_2 \frac{\ell^4}{k^2} F^{(\ell)}(t) \sum_{d=2}^{\ell} \frac{[(\ell-2)!(\ell-3)!\ldots(\ell-d+1)!]^2}{(\ell^2-2)(\ell^2-3)\ldots(\ell^2-d+1)}$$

$$\leq K_2 \frac{\ell^5}{k^2} F^{(\ell)}(t).$$

Hence there exists a constant K_3 such that

$$\mathbf{E}\,\mathcal{N}_{\ell k}^{2} \leq K_3 \frac{\ell^5}{k^2} F^{(\ell)}(t) \quad \text{for } \ell = 2,3,\ldots \text{ and } k \geq m^{-1}(\ell).$$

Using the Kolmogorov inequality for reverse submartingales,

$$\sum_{\ell=2}^{m(n)} \mathbf{P}\Big\{\max_{m^{-1}(\ell)\leq k\leq n} k\Big|\sum_{d=2}^{\ell} \binom{\ell}{d} F_{k,d}^{(\ell)}\Big| > \frac{n^{1/2}\epsilon\,\sigma_t}{m(n)}\Big\}$$

$$= \sum_{\ell=2}^{m(n)} \mathbf{P}\Big\{\max_{m^{-1}(\ell)\leq k\leq n} k^2\,|\mathcal{N}_{\ell k}|^2 > \frac{n\epsilon^2\sigma_t^2}{[m(n)]^2}\Big\}$$

$$\leq \sum_{\ell=2}^{m(n)} \frac{[m(n)]^2}{n\epsilon^2\sigma_t^2}\Big\{[m^{-1}(\ell)]^2\,\mathbf{E}|\mathcal{N}_{\ell,m^{-1}(\ell)}|^2 + \sum_{k=m^{-1}(\ell)+1}^{n} (2k-1)\,\mathbf{E}|\mathcal{N}_{\ell k}|^2\Big\}$$

$$\leq K_4 \sum_{\ell=2}^{m(n)} \frac{[m(n)]^2}{n}\,\ell^5 F^{(\ell)}(t)\Big\{1 + 2\sum_{k=m^{-1}(\ell)+1}^{n} k^{-1}\Big\}$$

$$\leq \Big\{K_5 \sum_{\ell=2}^{\infty} \ell^5 F^{(\ell)}(t)\Big\}\frac{[m(n)]^2 \log n}{n}$$

$$\rightarrow 0 \quad \text{as } n \rightarrow \infty$$

since $\sum_{\ell=2}^{\infty} \ell^5 F^{(\ell)}(t) < \infty$ by either B1′ and the Heyde lemma or (4.12). This establishes the lemma. □

Lemma 4.5: Assume B1′ or B2 or B3 holds. If $m(n)=o(n)$, then for every $\epsilon>0$

$$\mathbf{P}\Big\{\max_{0\leq k<m(n)} |Y_n^{(2)}(k/n)| > \epsilon\Big\} \rightarrow 0 \qquad \text{as } n \rightarrow \infty.$$

Proof: By definition, for fixed n, $\{Y_n^{(2)}(k/n);\ \mathcal{F}_k\}_{k=1}^{m(n)}$ is a martingale. Thus, using the Kolmogorov inequality for submartingales,

$$P\left\{\max_{0\le k<m(n)} |Y_n^{(2)}(k/n)| > \epsilon\right\} \le P\left\{\max_{0\le k\le m(n)} |Y_n^{(2)}(k/n)|^2 > \epsilon^2\right\}$$

$$\le \frac{m(n)}{n\epsilon^2\sigma_t^2} \sum_{r,s=1}^{m(n)} rs\,\xi_{rst}(1) \to 0$$

as $n \to \infty$. □

Theorem 4.1. Assume B1′ or B2 or B3 holds. If $[m(n)]^2 \log n = o(n)$, then $Y_n \xrightarrow{W} B$ in the J_1-topology on D[0,1], where B is a standard Brownian motion.

Proof: For any $\epsilon > 0$

$$P\left\{\sup_{s\in[0,1]} |Y_n(s) - Y_n^{(2)}(s)| > \epsilon\right\}$$

$$\le P\left\{\max_{0\le k<m(n)} |Y_n^{(2)}(k/n)| > \epsilon\right\} + P\left\{\max_{m(n)\le k\le n} |Y_n(k/n) - Y_n^{(2)}(k/n)| > \epsilon\right\}$$

$$\le P\left\{\max_{0\le k<m(n)} |Y_n^{(2)}(k/n)| > \epsilon\right\} + P\left\{\max_{m(n)\le k\le n} |Y_n(k/n) - Y_n'(k/n)| > \epsilon/3\right\}$$

$$+ P\left\{\max_{m(n)\le k\le n} |Y_n'(k/n) - Y_n^{(1)}(k/n)| > \epsilon/3\right\}$$

$$+ P\left\{\max_{m(n)\le k\le n} |Y_n^{(1)}(k/n) - Y_n^{(2)}(k/n)| > \epsilon/3\right\}$$

$$\to 0 \quad \text{as } n \to \infty$$

by Lemma 4.1, 4.3, 4.4, and 4.5. Thus Y_n and $Y_n^{(2)}$ have the same limiting distribution, which is the distribution of a standard Brownian motion by Lemma 4.2. □

4.4 Estimating The Variance

By Theorem B of Section 4.2, for a large class of distribution functions F, $n^{1/2}[H_n(t) - H(t)]$ has asymptotic variance $\sigma_t^2 = \sum_{r,s=1}^{\infty} rs\,\xi_{rst}(1)$, where $\xi_{rst}(1) = \mathrm{Cov}\{F^{(r-1)}(t-X), F^{(s-1)}(t-X)\}$. Frees (1986ii) defined an estimate of σ_t^2 motivated explicitly by this definition; this estimator was shown to be weakly consistent with the goal of constructing a large sample confidence interval for H(t). In Section 4.4.1 sufficient conditions are given for Frees' estimator to be strongly consistent. A jackknife estimator considered in Section 4.4.2 also converges with probability one. Strong consistency is of importance here because subsequently these estimators will be used in defining stopping times for estimation of H(t).

Chapter 5 is concerned with estimating H(t) by $H_n(t)$ simultaneously for all $t \in [0,T]$, where $0 < T < \infty$. With this problem in mind, let W be a monotone increasing function of bounded variation on the interval $[0,T]$ and define $I_* = \int_0^T \sigma_u^2\, dW(u)$. I_* is then a (weighted) measue of expected error in estimating H on the interval [0, T]. For an application in Chapter 5 an estimate of I_* will be necessary. Thus this section treats the slightly more general problem of estimating I_*.

4.4.1 Frees' variance estimator

For $r,s \geq 1$ let $\xi_{rst} = \mathbf{E}\{F^{(r-1)}(t-X), F^{(s-1)}(t-X)\}$, $a(t) = \sum_{r,s=1}^{\infty} rs\,\xi_{rst}$, and $b(t) = \sum_{r,s=1}^{\infty} r\,F^{(r)}(t)$, so that

$$\sigma_t^2 = a(t) - [b(t)]^2.$$

Define the kernel

$$g_{rs}(t;\, x_1,\dots,x_{r+s-1}) = \chi(x_1+\dots+x_r \le t)\ \chi(x_1+x_{r+1}+\dots+x_{r+s-1} \le t)$$

and let $X_{i1},\dots,X_{i(r+s-1)}$ represent any permutation of $r+s-1$ observations drawn from the random sample $X_1,\dots,X_n$. Let $\sum_P$ denote the summation over all $\frac{n!}{[n-(r+s-1)]!}$ such permutations. Then

$$\xi_{nrst} = \frac{[n-(r+s-1)]!}{n!} \sum_P g_{rs}(t;\, X_{i1},\dots,X_{i(r+s-1)})$$

is an unbiased estimate of ξ_{rst}. Let $m_* = m_*(n)$ be a monotone increasing sequence of positive integers such that $m_*(n) \to \infty$ as $n \to \infty$. Define $a_n(t) = \sum_{r,s=1}^{m_*} rs\,\xi_{nrst}$ and $b_n(t) = \sum_{r=1}^{m_*} r\,F_n^{(r)}(t)$. Frees considered

$$\sigma^2_{nt} = a_n(t) - [b_n(t)]^2. \tag{4.13}$$

as an estimator of σ^2_t and gave sufficient conditions for σ^2_{nt} to be weakly consistent. Now let

$$J_n = \int_0^T \sigma^2_{nt}\, dW(t)$$

be an estimator of I_*. In Theorem 4.2 sufficient conditions are given for J_n to converge to I_* with probability one.

The following assumptions will be used.

B4: For some $p \ge 6$, $\mathbf{E}|X^-|^p < \infty$ and $n = O(m_*^{(p-5)})$.

B5: F has finite positive mean μ. For some $\theta_1 > 0$ and all $|\theta| < \theta_1$, $\mathbf{E}\exp(-\theta X^-) < \infty$, and $\log n = o(m_*)$.

B6: $F(0^-) = 0$, $F(0) < 1$, and $\log n = o(m_*)$.

Theorem 4.2: Suppose B4 or B5 or B6 holds. Then $J_n \to I_*$ a.s.

Proof: Again let $\mathcal{C}_n = \sigma(X_{n:1},...,X_{n:n};\ X_{n+1},X_{n+2},...)$, where $X_{n:1}, ..., X_{n:n}$ are the order statistics. It is first shown that $a_n(t) \to a(t)$ a.s. for fixed $t \in [0, T]$. Define

$$\alpha_n(t) = \sum_{r,s=1}^{\infty} rs\, \mathbf{E}_{\mathcal{C}_n}\{ g_{rs}(t;\ X_1,...,X_{r+s-1}) \}$$

$$= \sum_{r,s=1}^{m_*} rs\, \xi_{nrst} + \sum_{r \vee s > m_*} rs\, \mathbf{E}_{\mathcal{C}_n}\{ g_{rs}(t;\ X_1,...,X_{r+s-1}) \}.$$

For all $n \geq 1$, $\alpha_n(t)$ is $\mathcal{C}_n$ measurable. If $r+s-1 \leq n$, then $\mathbf{E}_{\mathcal{C}_n}\{g_{rs}(t;\ X_1,...,X_{r+s-1})\} = \xi_{nrst}$. Since $\{\mathcal{C}_n\}$ is a decreasing sequence of sigma-fields, $\mathbf{E}_{\mathcal{C}_n}\alpha_{n-1} = \alpha_n$. Hence $\{\alpha_n, \mathcal{C}_n\}_{n=1}^{\infty}$ is a reverse martingale and converges to its expectation $a(t)$ a.s.

Let

$$\alpha_{n1} = \sum_{r \vee s > m_*} rs\, \mathbf{E}_{\mathcal{C}_n}\{ g_{rs}(t;\ X_1,...,X_{r+s-1}) \}.$$

It remains to be shown that $\alpha_{n1} \to 0$ a.s. Let $m_*^{-1}(k) = \inf\{n:\ m_*(n) \geq k\}$. Using the fact that for any $t \in [0, T]$,

$$\mathbf{E}|\xi_{nrst}| \leq \mathbf{E}|g_{rs}(t;\ X_1,...,X_{r+s-1})|$$

$$\leq F^{(r \vee s)}(t),$$

it follows by the Markov inequality that for any $\epsilon > 0$,

$$\sum_{n \geq 1} \mathbf{P}\{|\alpha_{n1}(t)| > \epsilon\} \leq \epsilon^{-1} \sum_{n \geq 1} \sum_{r \vee s > m_*} rs\, \mathbf{E}|\xi_{nrst}|$$

$$\leq \epsilon^{-1} \sum_{n\geq 1} \sum_{r\vee s > m_*} rs\, F^{(r\vee s)}(t)$$

$$\leq K_1 \sum_{n\geq 1} \sum_{s=m_*(n)+1} s^3 F^{(r\vee s)}(t)$$

$$\leq K_1 \sum_{k=1}^{\infty} m_*^{-1}(k)\, k^3 F^{(k)}(t).$$

If B4 holds, then

$$m_*^{-1}(m_*(n)) \leq n \leq K_2[m_*(n)]^{p-5}.$$

In this case

$$\sum_{n\geq 1} \mathbf{P}\{|\alpha_{n1}(t)| > \epsilon\} \leq K_3 \sum_{k=1}^{\infty} k^{p-5} k^3 F^{(k)}(t) < \infty$$

by the Heyde lemma.

With $m^{-1}(k) = \inf\{n: m(n)=k\}$, Frees (1986ii) noted that if B2 or B3 holds, then $m^{-1}(n) \leq \exp(\delta_{m^{-1}(n)} n)$, where $\delta_n \to 0$ as $n \to \infty$. Similarly, B5 and B6 imply the existence of a sequence $\delta_n \to 0$ such that $m_*^{-1}(n) \leq \exp(\delta_{m_*^{-1}(n)} n)$. Since both B5 and B6 ensure there exists $K > 0$ and p, $0 < p < 1$, such that $F^{(k)}(t) \leq K p^k$ for all $k \geq 1$, we have that

$$\sum_{n\geq 1} \mathbf{P}\{|\alpha_{n1}(t)| > \epsilon\} \leq K_4 \sum_{k=1}^{\infty} \exp(\delta_{m_*^{-1}(k)} k)\, k^3 p^k$$

$$= K_4 \sum_{k=1}^{\infty} k^3 \exp(k[\delta_{m_*^{-1}(k)} + \log p])$$

$$< \infty.$$

Thus by the Borel-Cantelli lemma $\alpha_{n1}(t) \to 0$ a.s. and $a_n(t) \to a(t)$ a.s. when B4 or B5 or B6 holds.

By Theorem 2.1 of Frees(1986ii), $b_n(t) \to b(t)$ a.s. if $\Sigma_{k=1}^{\infty} [m_*^{-1}(k) - k] k F^{(k)}(t) < \infty$. This follows from the first part of the proof. Hence, $\sigma_{nt}^2 \to \sigma_t^2$ a.s. for each $t \in [0, T]$.

Next, since $a_n(t)$ is positive and monotone increasing,

$$\int_0^T a_n(t)\, dW(t) \le a_n(T) \int_0^T dW(t)$$

$$\to a(T) \int_0^T dW(t) \quad \text{a.s.}$$

Thus

$$\int_0^T a_n(t)\, dW(t) \to \int_0^T a(t)\, dW(t) \quad \text{a.s.}$$

Similarly, since $b_n(t)$ is also positive and monotone increasing,

$$\int_0^T [b_n(t)]^2 dW(t) \to \int_0^T [b(t)]^2 dW(t) \quad \text{a.s.}$$

Hence,

$$J_n = \int_0^T \{a_n(t) - [b_n(t)]^2\} dW(t)$$

$$\to \int_0^T \sigma_t^2\, dW(t) \quad \text{a.s.,}$$

completing the proof. □

For future reference, note that if F is continuous, then the functions a and b are both continuous on the interval $[0, T]$. Since both a_n and b_n are monotone increasing on $[0, T]$ for all $n \ge 1$, it follows that

$$\sup_{t\in[0,T]} |a_n(t) - a(t)| \to 0 \text{ a.s.}$$

and

$$\sup_{t\in[0,T]} |b_n(t) - b(t)| \to 0 \text{ a.s.}$$

Thus

$$\sup_{t\in[0,T]} |\sigma^2_{nt} - \sigma^2_t| \to 0 \text{ a.s.}$$

4.4.2 A jackknife estimator of variance

A jackknife estimator of σ^2_t is now defined. For $i = 1,\ldots,n$, $k = 1,\ldots,m(n)$, compute the statistics

$$F^{(k)}_{n(i)}(t) = \binom{n-1}{k}^{-1} \sum_{Ci} \chi(X_{j1},\ldots,X_{jk} \le t),$$

where Σ_{Ci} denotes summation over all $\binom{n-1}{k}$ distinct subsets of k observations which do not include X_i. (We will assume throughout that $m(n) \le n-1$ for all n.) Let $H_{n(i)}(t) = \Sigma^m_{k=1} F^{(k)}_{n(i)}(t)$. Take

$$V^*_n(t) = (n-1)\sum_{i=1}^{n} \{H_{n(i)}(t) - H_n(t)\}^2 \tag{4.14}$$

to be a jackknife estimator of σ^2_t. As an estimate of I_* define

$$I_{n*} = \int_0^T V^*_n(u)\,dW(u).$$

In Theorem 4.3 sufficient conditions are given for I_{n*} to converge to I_* with probability one.

Theorem 4.3: Assume B2 or B3 holds. Further suppose that $m(n) \leq n-1$ for all n and there exists a positive integer n_o such that $[m(n)]^2 \leq n - m(n)$ for all $n \geq n_o$. Then $I_{n*} \rightarrow I_*$ a.s.

Some preliminary definitions and lemmas precede the proof of Theorem 4.3.

Define $H^*_{n-1}(u) = \Sigma^m_{k=1} F^{(k)}_{n-1}(u)$. Then

$$V^*_n(u) = n(n-1)\mathbf{E}_{\mathcal{C}_n}[H^*_{n-1}(u) - H_n(u)]^2$$
$$= n(n-1)\,\mathbf{Var}_{\mathcal{C}_n}\{H^*_{n-1}(u) - H_n(u)\}$$

since $\mathbf{E}_{\mathcal{C}_n}\{H^*_{n-1}(u) - H_n(u)\} = 0$. Using the decomposition of $H_n(u)$ considered in Section 4.3, define

$$\gamma_{n1}(u) = \sum_{r=1}^{m} r\left(F^{(r)}_{n-1,1}(u) - F^{(r)}_{n,1}(u)\right)$$

and

$$\gamma_{n2}(u) = \sum_{r=1}^{m}\sum_{d=2}^{r}\binom{r}{d}\left(F^{(r)}_{n-1,d}(u) - F^{(r)}_{n,d}(u)\right).$$

Taking

$$I_{n1*} = \int_0^T n(n-1)\,\mathbf{E}_{\mathcal{C}_n}\gamma^2_{n1}(u)\,dW(u)$$

and

$$I_{n2*} = \int_0^T n(n-1)\mathbf{E}_{\mathcal{C}_n}\gamma^2_{n2}(u)\,dW(u)$$

we have that

$$I_{n*} = I_{n1*} + I_{n2*} + 2n(n-1)\int_0^T \mathbf{E}_{\mathcal{C}_n}\{\gamma_{n1}(u)\gamma_{n2}(u)\}\,dW(u). \tag{4.15}$$

In Lemmas 4.6 and 4.7 it is shown that $I_{n1*} \rightarrow I_*$ a.s. while $I_{n2*} \rightarrow 0$ a.s. The proof of Theorem 4.3 follows.

Lemma 4.6: Assume $m(n) \leq n-1$ for all n. Suppose B2 or B3 holds. Then $I_{n1*} \to I_*$ a.s.

Proof: Define $Y_{irs}(u) = [F^{(r-1)}(u-X_i) - F^{(r)}(u)][F^{(s-1)}(u-X_i) - F^{(s)}(u)]$, $i = 1,\ldots,n$, $r,s = 1,2,3,\ldots$. Noting that

$$F^{(r)}_{n-1,1}(u) - F^{(r)}_{n,1}(u) = (n-1)^{-1}[F^{(r)}_{n,1}(u) - F^{(r-1)}(u-X_n) + F^{(r)}(u)],$$

we can write

$$I_{n1*} = \frac{n}{(n-1)}\int_0^T \sum_{r,s=1}^{m} rs\left\{n^{-1}\sum_{i=1}^{n} Y_{irs}(u) - F^{(r)}_{n,1}(u)\,F^{(s)}_{n,1}(u)\right\}dW(u).$$

Next, define

$$\alpha_{n1} = \int_0^T \left\{\sum_{m<(r\vee s)\leq n} n^{-1}\sum_{i=1}^{n}[Y_{irs}(u) - \xi_{rsu}(1)]\right\}dW(u)$$

and

$$\alpha_{n2} = \int_0^T \left\{\sum_{(r\vee s)>n} rs[Y_{irs}(u) - \xi_{rsu}(1)]\right\}dW(u).$$

Then

$$\alpha_n = n^{-1}\sum_{i=1}^{n}\int_0^T \left\{\sum_{r,s=1}^{n} rs[Y_{irs}(u) - \xi_{rsu}(1)]\right\}dW(u) + \alpha_{n2}$$

is a zero-mean reverse martingale and hence converges to zero a.s.

It is first shown that $\alpha_{n1} \to 0$ a.s. Since α_{n1} is an average of zero-mean i.i.d.

random variables,

$$\mathbf{E}\,\alpha_{n1}^4 \le \frac{K_1}{n^2}\,\mathbf{E}\left[\int_0^T \left\{ \sum_{m<(r\vee s)\le n} rs\,[Y_{1rs}(u) - \xi_{rsu}(1)] \right\} dW(u)\right]^4$$

$$\le \frac{K_2}{n^2}\,\mathbf{E}\left[\sum_{r,s=1}^{\infty} rs \int_0^T |Y_{1rs}(u)|\,dW(u)\right]^4 + \frac{K_2}{n^2}\left[\sum_{r,s=1}^{\infty} rs \int_0^T |\xi_{rsu}(1)|\,dW(u)\right]^4$$

$$\le \frac{K_2}{n^2} \sum_{r_1,s_1=1}^{\infty} \cdots \sum_{r_4,s_4=1}^{\infty} \prod_{i=1}^{4} r_i s_i \int_0^T \cdots \int_0^T \mathbf{E} \prod_{i=1}^{4} |Y_{1rs}(u_i)|\,dW(u_1)\ldots dW(u_4)$$

$$+ \frac{K_2}{n^2}\left[\sum_{r,s=1}^{\infty} rs \int_0^T |\xi_{rsu}(1)|\,dW(u)\right]^4$$

$$\le \frac{K_2}{n^2} \sum_{r_1,s_1=1}^{\infty} \cdots \sum_{r_4,s_4=1}^{\infty} [\prod_{i=1}^{4} r_i s_i]\, F^{(\max(r_1,s_1,\ldots,r_4,s_4))}(T)\left[\int_0^T dW(u)\right]^4$$

$$+ \frac{K_2}{n^2}\left[\sum_{r,s=1}^{\infty} rs\, F^{(r\vee s)}(T)\right]^4 \left[\int_0^T dW(u)\right]^4$$

$$< \frac{K_3}{n^2},$$

by (4.12). Thus for any $\epsilon > 0$, by the Markov inequality, $\Sigma_{n=1}^{\infty} \mathbf{P}\{|\alpha_{n1}| > \epsilon\} < \infty$ and $\alpha_{n1} \to 0$ a.s. by the Borel-Cantelli Lemma.

To see that $\alpha_{n2} \to 0$ a.s., note that

$$\mathbf{E}|\alpha_{n2}| \;\leq\; \int_0^T \sum_{(r\vee s)>n} rs\left\{\mathbf{E}|Y_{1rs}(u)| + |\xi_{rsu}(1)|\right\} dW(u)$$

$$\leq \left[\int_0^T dW(u)\right] \sum_{(r\vee s)>n} 2rs\, F^{(r\vee s)}(T),$$

by (4.11). Again using the Markov inequality, for fixed $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbf{P}\left\{|\alpha_{n2}| > \epsilon\right\} \;\leq\; \tfrac{2}{\epsilon} \sum_{n=1}^{\infty} \sum_{(r\vee s)>n} rs\, F^{(r\vee s)}(T)$$

$$\leq\; K_1 \sum_{n\geq 1} \sum_{s\geq n} \sum_{r=1}^{s} rs\, F^{(s)}(T)$$

$$=\; K_2 \sum_{s\geq 1} s^4\, F^{(s)}(T) \;<\; \infty$$

by (4.12). Thus $\alpha_{n2} \to 0$ a.s., again using the Borel-Cantelli Lemma. Since α_n, α_{n1}, and α_{n2} all converge to zero with probability one, it follows that

$$n^{-1} \sum_{i=1}^{n} \int_0^T \sum_{r,s=1}^{m} Y_{irs}(u)\, dW(u) \;\to\; I_* \text{ a.s.}$$

The lemma will follow if it is shown that

$$\iota_n \;=\; \int_0^T \sum_{r,s=1}^{m} rs\, F_{n,1}^{(r)}(u)\, F_{n,1}^{(s)}(u)\, dW(u) \;\to\; 0 \text{ a.s.}$$

Since $F_{n,1}^{(r)}(u) \;=\; n^{-1} \sum_{i=1}^{n} \left[F^{(r-1)}(u - X_i) - F^{(r)}(u)\right],$

$$\mathbf{E}\iota_n^2 = \mathbf{E}\int_0^T\int_0^T \sum_{r_1,s_1,r_2,s_2=1}^{m} r_1 s_1 r_2 s_2 F_{n,1}^{(r_1)}(u) F_{n,1}^{(s_1)}(u) F_{n,1}^{(r_2)}(v) F_{n,1}^{(s_2)}(v)\, dW(u)\, dW(v)$$

$$\leq \frac{K_1}{n^2}\int_0^T\int_0^T \sum_{r_1,s_1,r_2,s_2=1}^{m} r_1 s_1 r_2 s_2 \; F^{(\max(r_1,s_1,r_2,s_2))}(T)\; dW(u)\, dW(v)$$

$$\leq \frac{K_2}{n^2}\,.$$

By again applying the Markov inequality and the Borel-Cantelli Lemma, $\iota_n \to 0$ a.s., which completes the proof. □

<u>Lemma</u> <u>4.7</u>: Assume B2 or B3 holds. If $m(n) = o(n^{1/2})$, then $I_{n2} \to 0$ a.s.

<u>Proof</u>: Using the notation defined in Section 4.3,

$$F_{n-1,d}^{(r)}(u) - F_{n,d}^{(r)}(u) \;=\; \frac{d}{n}\left[F_{n-1,d}^{(r)}(u) - \binom{n-1}{d-1}^{-1} \sum_{C_{d-1}} g_{r,d}^{o}(u;\, X_{i1},\ldots,X_{i(d-1)}, X_n)\right]$$

where the summation is over all $\binom{n-1}{d-1}$ subsets of d-1 observations from $X_1,\ldots,X_{n-1}$. Since

$$\mathbf{E}\, g_{r,d}^{o}(u;\, X_{i1},\ldots,X_{id})\, g_{r,d}^{o}(u;\, X_{j1},\ldots,X_{jd}) = 0$$

unless $\{i_1,\ldots,i_d\} = \{j_1,\ldots,j_d\}$, for large enough n and all $u \in [0, T]$,

$$\binom{r}{d}^4 \mathbf{E}\left[F_{n-1,d}^{(r)}(u)\right]^4 \leq \binom{r}{d}^4 \frac{K_1}{\binom{n-1}{d}^2} (2^d)^4 F^{(r)}(u)$$

$$\leq \frac{K_1 r^4}{(n-1)^2}\left[\frac{r^{2(d-1)}}{(n-2)(n-3)\ldots(n-d)}\right]^2 \left[\frac{4^d}{d!}\right]^2 F^{(r)}(T).$$

Thus, since $d \le r \le m(n)$ and $[m(n)]^2 \le n - m(n)$ (for large n) by assumption,

$$\binom{r}{d}^4 \mathbf{E}\left[F_{n-1,d}^{(r)}(u)\right]^4 \le \frac{K_2}{(n-1)^2} r^4 F^{(r)}(T).$$

Similarly, it can be shown that

$$\binom{r}{d}^4 \mathbf{E}\left[\binom{n-1}{d-1}^{-1} \sum_{C_{d-1}} g^o_{r,d}(u; X_{i1}, \ldots, X_{i(d-1)}, X_n)\right]^4 \le \frac{K_3}{(n-1)^2} r^8 F^{(r)}(T)$$

for all $u \in [0, T]$. Since B2 or B3 implies (4.6), for some p, $0 < p < 1$,

$$\mathbf{E}\,\gamma^4_{n2}(u)$$

$$\le \sum_{r_1,r_2,r_3,r_4=1}^{m} \sum_{d_1=2}^{r_1} \sum_{d_2=2}^{r_2} \sum_{d_3=2}^{r_3} \sum_{d_4=2}^{r_4} \prod_{i=1}^{4}\left[\binom{r_i}{d_i}^4 \mathbf{E}\left[F_{n-1,d_i}^{(r_i)}(u) - F_{n,d_i}^{(r_i)}(u)\right]^4\right]^{1/4}$$

$$\le K_1 \sum_{r_1,r_2,r_3,r_4=1}^{m} \sum_{d_1=2}^{r_1} \sum_{d_2=2}^{r_2} \sum_{d_3=2}^{r_3} \sum_{d_4=2}^{r_4} \prod_{i=1}^{4}\left[\frac{d_i^4}{n^4} \frac{K_2}{(n-1)^2} r_i^8 F^{(r_i)}(T)\right]^{1/4}$$

$$\le K_3 \sum_{r_1,r_2,r_3,r_4=1}^{m} \sum_{d_1=2}^{r_1} \sum_{d_2=2}^{r_2} \sum_{d_3=2}^{r_3} \sum_{d_4=2}^{r_4} \frac{1}{(n-1)^6} \prod_{i=1}^{4}\left[r_i^3 \quad p^{(r_i/4)}\right]$$

$$\le \frac{K_4}{(n-1)^6}.$$

By definition,

$$\mathbf{E}\,I_{n2*}^2 = n^2(n-1)^2\int_0^T\int_0^T \mathbf{E}\Big[\mathbf{E}_{\mathcal{C}_n}\gamma_{n2}^2(u)\cdot \mathbf{E}_{\mathcal{C}_n}\gamma_{n2}^2(v)\Big]dW(u)\,dW(v)$$

$$\leq n^2(n-1)^2\left[\int_0^T\Big\{\mathbf{E}[\,\mathbf{E}_{\mathcal{C}_n}\gamma_{n2}^2(u)\,]^2\Big\}^{1/2} dW(u)\right]^2$$

$$\leq n^2(n-1)^2\left[\int_0^T\Big\{\mathbf{E}\,\gamma_{n2}^4(u)\Big\}^{1/2} dW(u)\right]^2$$

$$\leq \frac{K_7}{n^2}.$$

By the Markov inequality, for every $\epsilon > 0$, $\mathbf{P}\{|I_{n2*}| > \epsilon\} \leq \frac{K_7}{n^2\epsilon^2}$. Thus $I_{n2*} \to 0$ by the Borel-Cantelli lemma. □

Proof of Theorem 4.3: By Hölder's inequality and the conditional version of the Cauchy-Schwarz inequality,

$$\left|2n(n-1)\int_0^T \mathbf{E}_{\mathcal{C}_n}\{\,\gamma_{n1}(u)\gamma_{n2}(u)\,\}\,dW(u)\right|$$

$$\leq 2n(n-1)\,\mathbf{E}_{\mathcal{C}_n}\left\{\left[\int_0^T\gamma_{n1}^2(u)\,dW(u)\right]^{1/2}\left[\int_0^T\gamma_{n2}^2(u)\,dW(u)\right]^{1/2}\right\}$$

$$\leq 2n(n-1)\left[\mathbf{E}_{\mathcal{C}_n}\left\{\int_0^T\gamma_{n1}^2(u)\,dW(u)\right\}\right]^{1/2}\left[\mathbf{E}_{\mathcal{C}_n}\left\{\int_0^T\gamma_{n2}^2(u)\,dW(u)\right\}\right]^{1/2}$$

$$= 2[I_{n1*}]^{1/2}[I_{n2*}]^{1/2} \to 0 \quad \text{a.s.}$$

The theorem follows by (4.15) and Lemmas 4.1 and 4.2. □

4.5 A Fixed-Width Confidence Interval For The Optimal Block Replacement Policy

The objective of this section is a sequential scheme for fixed-width confidence interval estimation of H(t). Let $1-\alpha$ be the desired coverage probability. For fixed $d>0$, define the random interval $I_n = [H_n(t)-d, H_n(t)+d]$. Ideally, one would like to be able to select the minimum sample size n_d^* such that

$$\mathbf{P}\{\ H(t) \in I_{n_d^*}\} \ \geq \ 1-\alpha.$$

Since this appears impossible in the present nonparametric setting, the approach taken is to define a stopping time N_d and show that under certain conditions

$$\lim_{d \downarrow 0} \mathbf{P}\{\ H(t) \in I_{N_d}\} \ = \ 1-\alpha.$$

Let $\tau_{\alpha/2}$ be the $(1-\alpha/2)$-th quantile of the standard normal distribution. Take n_d to be the non-random integer such that

$$n_d \ = \ \min\left\{k\colon k \geq n_o \text{ and } \frac{\sigma_t^2}{k}(\tau_{\alpha/2})^2 \leq d^2\right\},$$

where $n_o \geq 1$ is an initial sample size. Denote by s_n^2 a strongly consistent estimator of σ_t^2. Define the stopping time N_d by

$$N_d \ = \ \min\left\{k\colon k \geq n_o \text{ and } \frac{s_k^2}{k}(\tau_{\alpha/2})^2 \leq d^2\right\},$$

As is typically done, we will exploit the relationship between N_d and n_d to establish asymptotic properties of $H_{N_d}(t)$.

For stopping times defined like N_d one always has that

$$s^2_{N_d-1} > \frac{(N_d-1)d^2}{(\tau_{\alpha/2})^2} \geq \frac{(N_d-1)}{N_d} s^2_{N_d}$$

and thus

$$\frac{s^2_{N_d-1}}{\sigma^2_t} > \frac{(N_d-1)}{n_d} \cdot \frac{n_d d^2}{\sigma^2_t (\tau_{\alpha/2})^2} \geq \frac{(N_d-1)}{N_d} \frac{s^2_{N_d}}{\sigma^2_t} . \tag{4.16}$$

Since by definition $N_d \to \infty$ and $n_d \sim \dfrac{\sigma^2_t (\tau_{\alpha/2})^2}{d^2}$ as $d \downarrow 0$, we have that $\dfrac{N_d}{n_d} \to 1$ a.s. as $d \downarrow 0$.

Theorem B of Section 4.2 gave sufficient conditions for the asymptotic normality of $H_n(t)$. Lemma 4.8 asserts that the randomly stopped estimator has the same limiting distribution as $d \downarrow 0$.

<u>Lemma 4.8</u>: Assume B1$'$ or B2 or B3 holds. If $[m(n)]^2 \log n = o(n)$, then

$$N_d^{1/2} [H_{N_d}(t) - H(t)] \xrightarrow{D} N(0, \sigma^2_t).$$

<u>Proof:</u> Let $Z_n = n^{1/2}[H_n(t) - H(t)]$. The invariance principle of Theorem 4.1 can be used to show that the sequence $\{Z_n\}$ is uniformly continuous in probability. Anscombe's theorem then implies that Z_{N_d} and Z_{n_d} have the same limiting distribution as $d \downarrow 0$, establishing the lemma. (See Woodroofe (1982), Section 1.3 for the definition of a uniformly continuous in probability sequence and Anscombe's theorem.) □

For the remainder of this section let σ_{nt}^2 and $V_n^*(t)$ be defined by (4.13) and (4.14), respectively. If $s_n^2 = \sigma_{nt}^2$ and the conditions of Theorem 4.2 hold, or $s_n^2 = V_n^*(t)$ and the conditions of Theorem 4.3 hold, then $s_n^2 \to \sigma_t^2$ a.s. This follows from Theorems 4.2 and 4.3 by taking $W(u) = \chi(u \le t)$.

Theorem 4.4: Assume B1′ or B2 or B3 holds and that $[m(n)]^2 \log n = o(n)$. If either $s_n^2 = \sigma_{nt}^2$ and the conditions of Theorem 4.2 hold, or $s_n^2 = V_n^*(t)$ and the conditions of Theorem 4.3 hold, then

$$\lim_{d \downarrow 0} \mathbf{P}\{ H(t) \in I_{N_d} \} = 1 - \alpha.$$

Proof: By Lemma 4.8 and the strong consistency of s_n^2,

$$\frac{N_d^{1/2}[H_{N_d}(t) - H(t)]}{s_{N_d}} \xrightarrow{D} N(0,1) \quad \text{as } d \downarrow 0.$$

By (4.16), $\dfrac{N_d^{1/2} d}{s_{N_d}} \to \tau_{\alpha/2}$, which implies that

$$\mathbf{P}\left\{ H(t) \in I_{N_d} \right\} = \mathbf{P}\left\{ -\frac{N_d^{1/2} d}{s_{N_d}} \le \frac{N_d^{1/2}[H_{N_d} - H(t)]}{s_{N_d}} \le \frac{N_d^{1/2} d}{s_{N_d}} \right\}$$

$$\to 1 - \alpha \quad \text{as } d \downarrow 0. \qquad \square$$

CHAPTER 5

RENEWAL FUNCTION ESTIMATION ON AN INTERVAL

5.1 Introduction

This chapter considers simultaneously estimating the renewal function at all points t in the interval $[0, T]$, $0 < T < \infty$, using the estimator of Chapter 4.

In Section 5.2 the objective is minimum risk estimation, where the risk is a function of the sample size n. The risk function considered here is comprised of two summands. One term, essentially the asymptotic mean integrated square error, decreases as the sample size increases. The other term is intended to reflect the cost of the sample and consequently increases with the sample size. The stopping time proposed here is of the usual type for minimizing a risk function of this kind. We show that in a large sample situation this stopping time can give some indication of whether enough observations have been taken.

To help characterize the behaviour of the estimator when used over an interval, the process $\{ n^{1/2}[H_n(t) - H(t)],\ t \in [0, T]\}$ is shown to converge weakly to a Gaussian process under certain conditions. This result establishes the rate of convergence of the function estimate and is used to construct confidence bands for the renewal function. We outline bootstrap methodology for constructing fixed and variable width confidence bands in Sections 5.4 and 5.5, respectively. In both settings stopping times are introduced to guide the user in determining if the sample size is adequate. The confidence bands associated with these random sample sizes are shown to achieve their targeted confidence level in an asymptotic sense.

5.2 Minimum Risk Estimation

Let $c > 0$ be the cost of observing each of the random variables $X_1, X_2, \ldots, X_n$ to be used in estimating $H(t)$. As in Section 4.4, let W be a monotone increasing function of bounded variation on the interval $[0, T]$ for some fixed $T > 0$. For estimation of $H(t)$ over the interval $[0, T]$ consider the loss function

$$L(n; a, c) = a \int_0^T [H_n(u) - H(u)]^2 \, dW(u) + cn.$$

Here $a > 0$ is the "cost" associated with a unit of the weighted integrated squared error while cn is the cost of a random sample of n observations. The expected loss, or risk, is then

$$R^*(n; a, c) = a \int_0^T \mathbf{E}[H_n(u) - H(u)]^2 \, dW(u) + cn.$$

Ideally one would like to select the sample size n which minimizes $R^*(n; a, c)$. However, for any fixed t the form of the mean square error $\mathbf{E}[H_n(t) - H(t)]^2$ is complicated by terms which are asymptotically negligible. Since concern lies in large sample estimation of the renewal function, $R^*(n; a, c)$ is simplified before considering the problem of optimal stopping.

Letting $B_n(t)$ denote the bias of $H_n(t)$, it follows by definition that

$$B_n(t) = \sum_{k>m} F^{(k)}(t).$$

$R^*(n; a, c)$ can be rewritten as

$$R^*(n; a, c) = a \int_0^T [\mathbf{Var}\{H_n(u)\} + B_n^2(u)] \, dW(u) + cn.$$

Frees (1986ii) noted that

$$\mathbf{Var}\{H_n(t)\} = \sum_{r,s=1}^{m} \binom{n}{r}^{-1} \sum_{c=1}^{r} \binom{s}{c}\binom{n-s}{r-c} \xi_{rst}(c)$$

where as previously defined

$$\xi_{rst}(c) = \mathbf{Cov}\left\{ F^{(r-c)}(t - (X_1 + \dots + X_c)), F^{(s-c)}(t - (X_1 + \dots + X_c)) \right\}$$

for any r,s, and c, $1 \le c \le r \wedge s$. Since $\xi_{rst}(c) \le F^{(r \vee s)}(t)$, if B1 or B2 or B3 holds, then by Lemma 3.3 of Frees(1986ii),

$$\lim_{n\to\infty} n \int_0^T \left\{ \sum_{r,s=1}^{m} \binom{n}{r}^{-1} \sum_{c=2}^{r} \binom{s}{c}\binom{n-s}{r-c} \xi_{rsu}(c) \right\} dW(u)$$

$$\le [W(T) - W(0)] \lim_{n\to\infty} n \sum_{r,s=1}^{m} \binom{n}{r}^{-1} \sum_{c=2}^{r} \binom{s}{c}\binom{n-s}{r-c} F^{(r \vee s)}(T)$$

$$= 0.$$

Also, when B1 or B2 or B3 holds,

$$\lim_{n\to\infty} n \int_0^T B_n^2(u)\, dW(u) \le [W(T) - W(0)] \lim_{n\to\infty} [n^{1/2} B_n(T)]^2$$

$$= 0$$

by the following lemma.

<u>Lemma (Frees)</u>: Assume B1 or B2 or B3 holds. Then

(5.1)
$$\lim_{n\to\infty} n^{1/2} \sum_{k>m} F^{(k)}(T) = 0.$$

Proof: If B1 or B2 holds, the result is identical to Lemma 3.1 of Frees (1986ii). As previously mentioned, B3 implies (4.6), from which (5.1) follows. (See the proof of Lemma 3.1 of Frees (1986ii).) □

Thus

$$R^*(n;a,c) = a n^{-1} \int_0^T \left\{ \sum_{r,s=1}^{m} rs\, \xi_{rsu}(1) \right\} dW(u) + o(n^{-1}) + cn.$$

As in Section 4.4, let

$$I_* = \int_0^T \left\{ \sum_{r,s=1}^{\infty} rs\, \xi_{rsu}(1) \right\} dW(u).$$

Then the asymptotic risk may be written as

$$R(n;a,c) = an^{-1} I_* + cn.$$

$R(n;a,c)$ is a convex function of n in the sense that

$$\Delta^2 R(n;a,c) = R(n+2;\, a,c) - 2R(n+1;\, a,c) + R(n;\, a,c) \geq 0 \text{ for all n.}$$

Thus there exists a finite n_c such that

$$n_c = \min\left\{ n' \colon R(n';a,c) \leq R(n;\, a,c) \text{ for all } n \right\}.$$

By definition, $n_c \to \infty$ as the cost per observation $c \downarrow 0$. As one would hope, the stopping time N_c which is now introduced also depends on c in such a way that $N_c \to \infty$ as $c \downarrow 0$. Additionally, it will be shown that N_c enjoys a type of risk efficiency, namely that

$$\lim_{c \downarrow 0} \frac{R(n_c;a,c)}{E\, R(N_c;a,c)} = 1.$$

Suppose there exists known $\gamma > 0$ such that $I_* > \gamma$. Let $\hat{I}_n$ be a strongly consistent estimator of I_*. Define

$$N_c = \min\left\{ n \geq n_o \colon n \geq \left(\frac{a}{c}\right)^{1/2} [\hat{I}_n^{1/2} \vee \gamma] \right\}$$

where n_o is an initial sample size. By definition $N_c \to \infty$ a.s. as $c \downarrow 0$ and the relation

$$\left(\frac{a}{c}\hat{I}_n\right)^{1/2} \leq N_c \leq \left(\frac{a}{c}\right)^{1/2} [\hat{I}_n^{1/2} \vee \gamma] + 1$$

holds. Thus, since $n_c \sim \left(\frac{a}{c} I_*\right)^{1/2}$ as $c \downarrow 0$, $\frac{N_c}{n_c} \to 1$ a.s. as $c \downarrow 0$. Lemma 5.1 gives sufficient conditions for $\mathbf{E}\frac{N_c}{n_c} \to 1$ as $c \downarrow 0$.

<u>Lemma</u> <u>5.1</u>: Assume $\hat{I}_n \to I_*$ a.s. and there exists a positive constant K such that $\mathbf{E}\hat{I}_n \leq K$ for all n. Then $\mathbf{E}\frac{N_c}{n_c} \to 1$ as $c \downarrow 0$.

<u>Proof</u>: For $x \geq 1$ and some constant K_1,

$$\mathbf{P}\left\{ \frac{N_c}{n_c} > x \right\} \leq \mathbf{P}\left\{ x < K_1 [\hat{I}_{N_c - 1}^{1/2} \vee \gamma] \right\}$$

since $\left(\frac{a}{c}\right)^{1/2} = O(n_c)$. Thus, by the Markov inequality,

$$\mathbf{P}\left\{ \frac{N_c}{n_c} > x \right\} \leq \frac{K_2}{x^2},$$

since $\mathbf{E}\hat{I}_n \leq K$ for all n. This implies that independently of c, for $\eta > 1$,

$$\mathbf{E}\left\{ \frac{N_c}{n_c}\, \chi\left(\frac{N_c}{n_c} > \eta \right) \right\} \leq \int_\eta^\infty \frac{K_2}{u^2} du,$$

since for a nonnegative random variable X with distribution function F,

$$\mathbf{E}X = \int_0^\infty [1 - F(u)]\,d(u).$$

Hence

$$\lim_{\eta\to\infty} \sup_{c>0} \mathbf{E}\left\{\frac{N_c}{n_c}\,\chi\left(\frac{N_c}{n_c} > \eta\right)\right\} = 0,$$

implying that the family $\left\{\frac{N_c}{n_c},\ c>0\right\}$ of random variables is uniformly integrable and $\mathbf{E}\frac{N_c}{n_c} \to 1$ as $c\downarrow 0$. □

Theorem 5.1: Assume either $\hat{I}_n = J_n$ and the conditions of Theorem 4.2 hold, or $\hat{I}_n = I_{n*}$ and the conditions of Theorem 4.3 hold. If there exists a known $\gamma>0$ such that $I_* > \gamma$, then

$$\lim_{c\downarrow 0} \frac{R(n_c;\, a,c)}{\mathbf{E}R(N_c;\, a,c)} = 1.$$

Proof: Since $N_c \geq \gamma\left(\frac{a}{c}\right)^{1/2}$ and $n_c = O\left(\left(\frac{a}{c}\right)^{1/2}\right)$, for some constant $K_1>0$, $n_c/N_c \leq K_1/\gamma$. Thus $\mathbf{E}\left\{n_c/N_c\ \chi\left(n_c/N_c > \eta\right)\right\} = 0$ for all $\eta > K_1/\gamma$, for all $c > 0$. Thus $\left\{n_c/N_c,\ c > 0\right\}$ is a uniformly integrable class of random variables and $\mathbf{E}\,n_c/N_c \to 1$ as $c\downarrow 0$.

For the estimator J_n,

$$\mathbf{E}J_n \leq \int_0^T \{\, a(t) + [b(t)]^2 \,\}\,dW(t) < \infty$$

for all n. (This is when B4 or B5 or B6 holds.) If B5 or B6 holds, then it can be shown that for some $K>0$, $\mathbf{E}I_{n*} \leq K$ for all n. Using Lemma 5.2, the theorem follows. □

5.3 An Invariance Principle For Interval Estimation

Let $Y_n(t) = n^{1/2}[H_n(t) - H(t)]$ and define the stochastic process $Y_n = \{Y_n(t), t \in [0,T]\}$ for some fixed $T > 0$. Take D[0,T] to be the space of all functions on [0, T] that are right-continuous and have left-hand limits. Y_n is an element of D[0,T] for all $n \geq 1$. Let $\xrightarrow{W}$ denote weak convergence in the J_1-topology on D[0,T]. Theorem 5.2, stated immediately after the following definition and assumptions, provides a set of conditions under which $Y_n \xrightarrow{W} Y$, where Y is a zero-mean Gaussian process on $[0,T]$.

For $t_1, t_2 \in [0, T]$, define

(5.2) $$\xi_{rst_1t_2}(1) = \mathrm{Cov}\{F^{(r-1)}(t_1 - X_1), F^{(s-1)}(t_2 - X_1)\}.$$

Let $Y = \{Y(t), t \in [0,T]\}$ be a zero-mean Gaussian process defined on $[0,T]$ with covariance function

$$R(t_1,t_2) = \mathrm{Cov}\{Y(t_1), Y(t_2)\} = \sum_{r,s=1}^{\infty} rs\,\xi_{rst_1t_2}(1).$$

Assume F is absolutely continuous with density f and let $f^{(i)}$ denote the i-fold convolution of f.

B7: There exist positive constants K and ρ, $0 < \rho < 1$, such that for i = 1,2,...

$$f^{(i)}(t) \leq K\rho^i \text{ for all } t \in \mathbf{R}.$$

B8: $F(0) = 0$ and there exist positive constants K and ρ, $0 < \rho < 1$, such that for i = 1,2,...

$$f^{(i)}(t) \leq K\rho^i \text{ for all } t \leq T.$$

Theorem 5.2: Assume B1 or B2 or B3 holds. If B7 or B8 holds, then $Y_n \xrightarrow{W} Y$.

The theorem holds, for example, when the following simple and widely applicable condition is met: F is absolutely continuous with density f, $F(0)=0$, and f is bounded on the interval $[0,T]$. Clearly this implies that the assumption B3 is met. Furthermore, it can be shown that B8 also holds. (See, for example, Wold(1981), Theorem 2.3.3.) Thus $Y_n \overset{W}{\rightarrow} Y$.

By definition,

$$Y_n(t) = n^{1/2} \sum_{k=1}^{m} [F_n^{(k)}(t) - F^{(k)}(t)] - n^{1/2} \sum_{k>m} F^{(k)}(t).$$

Let

$$Y'_n(t) = n^{1/2} \sum_{k=1}^{m} [F_n^{(k)}(t) - F^{(k)}(t)]$$

and

$$Y'_n = \{ Y'_n(t),\ t \in [0, T] \}.$$

The lemma due to Frees (Section 5.2) asserts that

$$\sup_{t\in[0,T]} |Y_n(t) - Y'_n(t)| = n^{1/2} \sum_{k>m} F^{(k)}(T) \rightarrow 0$$

as $n \rightarrow \infty$. Thus Y_n and Y'_n have the same limiting distribution, a fact which is exploited here.

In Section 5.3.1 calculations verify that the finite dimensional distributions of Y'_n converge weakly to those of Y. Lemma 5.5 of Section 5.3.2 gives sufficient conditions for "tightness" of the sequence $\{Y'_n\}$. The proof of Theorem 5.2 appears at the end of Section 5.3.2.

5.3.1 Asymptotic normality of finite-dimensional distributions

For arbitrary positive integer ℓ, let $\underset{\sim}{t} = (t_1,\ldots,t_\ell)'$, where $t_1,\ldots,t_\ell$ are fixed nonnegative real numbers. The Cramér-Wold device is used to show that $[H_n(t_1),\ldots,H_n(t_\ell)]$, suitably standardized, has asymptotically a multivariate normal distribution. Thus, let $\underset{\sim}{a} = (a_1,\ldots,a_\ell)'$ be any ℓ-vector of real numbers. Define

$$D_n(\underset{\sim}{t}) = \sum_{i=1}^{\ell} a_i H_n(t_i)$$

and

$$D_n^*(\underset{\sim}{t}) = \sum_{i=1}^{\ell} a_i \sum_{k=1}^{m} F^{(k)}(t_i).$$

By a straightforward extension of the proofs found in Section 3 of Frees (1986ii), it is shown that $n^{1/2}[D_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})] = \Sigma_{i=1}^{\ell} a_i Y_n'(t_i)$ is asymptotically normal. The method is to first project $D_n(\underset{\sim}{t})$ onto the individual observations $X_1,\ldots,X_n$, yielding a statistic $\hat{D}_n(\underset{\sim}{t})$ which is the sum of i.i.d. random variables. Using an array central limit theorem $n^{1/2}[\hat{D}_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})]$ is shown to be asymptotically normal. The remainder $n^{1/2}[D_n(\underset{\sim}{t}) - \hat{D}_n(\underset{\sim}{t})]$ goes to zero in probability.

For $1 \leq c \leq r \vee s$, $i, j = 1, \ldots, \ell$, define

$$\xi_{rsij}(c) = \text{Cov}\{F^{(r-c)}(t_i - (X_1 + \ldots + X_c)),\ F^{(s-1)}(t_j - (X_1 + \ldots + X_c))\}$$

and

$$\xi_{rs\underset{\sim}{t}}(c) = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} a_i a_j \xi_{rsij}(c).$$

<u>Theorem</u> <u>5.3</u>: Assume B1 or B2 or B3 holds. Then

$$n^{1/2}[D_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})] \xrightarrow{D} N(0, \sigma_{\underset{\sim}{t}}^2)$$

where $\sigma_{\underset{\sim}{t}}^2 = \Sigma_{r,s=1}^{\infty} rs\, \xi_{rs\underset{\sim}{t}}(1)$, and thus the finite-dimensional distributions of Y_n'

converge weakly to the finite-dimensional distributions of Y.

Proof: The proof of Theorem 5.3 is begun by calculating the projection $\hat{D}_n(\underset{\sim}{t})$. Let

$$B_n^{(k)}(\underset{\sim}{t}) = \sum_{i=1}^{\ell} a_i F_n^{(k)}(t_i)$$

and

$$B^{(k)}(\underset{\sim}{t}) = \sum_{i=1}^{\ell} a_i F^{(k)}(t_i).$$

By the definition of $F_n^{(k)}(t)$ we have

$$\mathbf{E}\left(F_n^{(k)}(t)|X_1\right) = (k/n)F^{(k-1)}(t-X_1) + (1-k/n)F^{(k)}(t).$$

Adopting the convention that $(\underset{\sim}{t}-X)' = (t_1-X,\ldots,t_\ell-X)$, it follows that

$$\mathbf{E}\left(B_n^{(k)}(\underset{\sim}{t})|X_1\right) = \sum_{i=1}^{\ell} a_i\left[(k/n)F^{(k-1)}(t_i-X_1) + (1-k/n)F^{(k)}(t_i)\right]$$

$$= (k/n)B^{(k-1)}(\underset{\sim}{t}-X_1) + (1-k/n)B^{(k)}(\underset{\sim}{t}).$$

Define the projection

$$\hat{D}_n(\underset{\sim}{t}) = \left\{\sum_{j=1}^{n} \mathbf{E}\left(D_n(\underset{\sim}{t})|X_j\right)\right\} - (n-1)D_n^*(\underset{\sim}{t})$$

so that

$$\hat{D}_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t}) = n^{-1}\sum_{j=1}^{n}\sum_{k=1}^{m} k\left\{B^{(k-1)}(\underset{\sim}{t}-X_j) - B^{(k)}(\underset{\sim}{t})\right\}.$$

We next calculate $\mathbf{Var}\{D_n(\underset{\sim}{t})\}$, $\mathbf{Var}\{\hat{D}_n(\underset{\sim}{t})\}$, and $\mathbf{Cov}\{D_n(\underset{\sim}{t}),\hat{D}_n(\underset{\sim}{t})\}$.

By definition,

$$\mathbf{Cov}\{B^{(r-c)}(\underset{\sim}{t}-(X_1+\dots+X_c)),\ B^{(s-c)}(\underset{\sim}{t}-(X_1+\dots+X_c))\}$$

$$= \mathbf{Cov}\left\{\sum_{i=1}^{\ell} a_i F^{(r-c)}(t_i-(X_1+\dots+X_c)),\ \sum_{j=1}^{\ell} a_j F^{(s-c)}(t_j-(X_1+\dots+X_c))\right\}$$

$$= \xi_{rs\underset{\sim}{t}}(c).$$

Thus,

$$\mathbf{Var}\left\{\hat{D}_n(\underset{\sim}{t})\right\} = n^{-1}\mathbf{Var}\left\{\sum_{k=1}^{m} k[B^{(k-1)}(\underset{\sim}{t}-X_1) - B^{(k)}(\underset{\sim}{t})]\right\}$$

$$= n^{-1}\sum_{r,s=1}^{m} rs\,\xi_{rs\underset{\sim}{t}}(1).$$

By a direct calculation

$$\mathbf{Var}\left\{D_n(\underset{\sim}{t})\right\} = \sum_{r,s=1}^{m}\mathbf{Cov}\left\{B_n^{(r)}(\underset{\sim}{t}),\ B_n^{(s)}(\underset{\sim}{t})\right\}$$

$$= \sum_{r,s=1}^{m}\ \sum_{i,j=1}^{\ell} a_i a_j \binom{n}{r}^{-1}\sum_{c=1}^{r}\binom{s}{c}\binom{n-s}{r-c}\xi_{rsij}(c).$$

Finally,

$$\mathbf{Cov}\left\{D_n(\underset{\sim}{t}),\hat{D}_n(\underset{\sim}{t})\right\} = \mathbf{Cov}\left\{\sum_{r=1}^{m} B_n^{(r)}(\underset{\sim}{t}),\ n^{-1}\sum_{j=1}^{n}\ \sum_{s=1}^{m} s[B^{(s-1)}(\underset{\sim}{t}-X_j) - B^{(s)}(\underset{\sim}{t})]\right\}$$

$$= \sum_{r,s=1}^{m} s\ \mathbf{Cov}\left\{B_n^{(r)}(\underset{\sim}{t}),\ B^{(s-1)}(\underset{\sim}{t}-X_1)\right\}.$$

Since

$$\mathrm{Cov}\left\{B_n^{(r)}(\underset{\sim}{t}),\ B^{(s-1)}(\underset{\sim}{t}-X_1)\right\}$$

$$= \binom{n-1}{r-1}\binom{n}{r}^{-1}\mathrm{Cov}\left\{\sum_{i=1}^{\ell} a_i\,\chi(X_2+\dots+X_r \le t_i - X_1),\ \sum_{j=1}^{\ell} a_j F^{(s-1)}(t_j - X_1)\right\}$$

$$= n^{-1} r\, \xi_{rs\underset{\sim}{t}}(1),$$

we can conclude that

$$\mathrm{Cov}\left\{D_n(\underset{\sim}{t}),\ \hat{D}_n(\underset{\sim}{t})\right\} = n^{-1}\sum_{r,s=1}^{m} rs\,\xi_{rs\underset{\sim}{t}}(1).$$

<u>Lemma</u> <u>5.2</u>: Assume B1 or B2 or B3 holds. Then

$$\sigma_{\underset{\sim}{t}}^2 = \sum_{r,s=1}^{\infty} rs\,\xi_{rs\underset{\sim}{t}}(1) < \infty.$$

<u>Proof</u>: For every $1 \le c \le r \wedge s$, $i, j = 1,\dots,\ell$,

$$|\xi_{rsij}(c)| \le \min\left(F^{(r)}(t_i), F^{(s)}(t_j)\right). \tag{5.3}$$

Thus,

$$\sum_{r,s=1}^{\infty} rs\,|\xi_{rsij}(1)| \le \sum_{r=1}^{\infty}\sum_{s=1}^{r} rs\,F^{(r)}(t_i) + \sum_{s=1}^{\infty}\sum_{r=1}^{s} rs\,F^{(s)}(t_j)$$

$$< \infty$$

by the Heyde lemma (given in Section 4.2) if B1 holds or by (4.12) if B2 or B3 holds. Hence by definition

$$\sigma_{\underset{\sim}{t}}^2 \le \sum_{i,j=1}^{\ell} a_i a_j \sum_{r,s=1}^{\infty} rs\,|\xi_{rsij}(1)| < \infty. \qquad \square$$

Lemma 5.3: Assume B1 or B2 or B3 holds. Then

$$n^{1/2}[\hat{D}_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})] \xrightarrow{D} N(0, \sigma^2_{\underset{\sim}{t}}).$$

Proof: Let $X_{nj} = n^{-1/2}\Sigma_{k=1}^{m} k\left\{B^{(k-1)}(\underset{\sim}{t} - X_j) - B^{(k)}(\underset{\sim}{t})\right\}, j=1,...,n$. By definition,

$$n^{1/2}[\hat{D}_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})] = \sum_{j=1}^{n} X_{nj}.$$

Also, $\mathbf{E}X_{nj} = 0, j = 1,...,n$, and

$$\mathcal{S}_n^2 = \mathrm{Var}\left\{\sum_{j=1}^{n} X_{nj}\right\} = \sum_{r,s=1}^{m} rs\,\xi_{rs\underset{\sim}{t}}(1) \to \sigma^2_{\underset{\sim}{t}}.$$

For $\epsilon > 0$ let

$$S_n = \left\{u: \left|n^{-1/2}\sum_{k=1}^{m} k[B^{(k-1)}(\underset{\sim}{t}-u) - B^{(k)}(\underset{\sim}{t})]\right| > \epsilon\,\mathcal{S}_n\right\}.$$

We verify the Lindebergh condition

$$\lim_{n\to\infty} \sum_{j=1}^{n} \mathbf{E}\left\{X_{nj}^2\ \chi(|X_{nj}| > \epsilon\,\mathcal{S}_n)\right\}/\mathcal{S}_n^2$$

$$= \lim_{n\to\infty} \int \left\{\sum_{k=1}^{m} k[B^{(k-1)}(\underset{\sim}{t}-u) - B^{(k)}(\underset{\sim}{t})]\right\}^2 \chi(u\in S_n)\,dF(u)$$

$$= 0.$$

The first step is to bound the integrand by an integrable function of u:

$$\left\{\sum_{k=1}^{m} k[B^{(k-1)}(\underset{\sim}{t}-u) - B^{(k)}(\underset{\sim}{t})]\right\}^2$$

$$= \left\{\sum_{i=1}^{\ell} a_i \sum_{k=1}^{m} k\,[F^{(k-1)}(t_i-u) - F^{(k)}(t_i)]\right\}^2$$

$$\leq \sum_{i=1}^{\ell} a_i^2 \ell^2 \left\{\sum_{k=1}^{\infty} k \left|F^{(k-1)}(t_i-u) - F^{(k)}(t_i)\right|\right\}^2,$$

which is integrable with respect to the measure dF by Lemma 5.2. Since $\lim_{n\to\infty} \chi(u \in S_n) = 0$ for any finite u, the Lindebergh condition holds by dominated convergence. The lemma follows from a standard array central limit theorem. (See Serfling(1980), Section 1.9.2.) □

Lemma 5.4: Assume B1 or B2 or B3 holds. Then

$$n\,\mathbf{E}[D_n(\underset{\sim}{t}) - \hat{D}_n(\underset{\sim}{t})]^2 \rightarrow 0.$$

Proof: By prior calculations,

$$n\,\mathbf{E}[D_n(\underset{\sim}{t}) - \hat{D}_n(\underset{\sim}{t})]^2 = n\,\mathbf{Var}\{D_n(\underset{\sim}{t})\} + n\,\mathbf{Var}\{\hat{D}_n(\underset{\sim}{t}) - 2n\,\mathbf{Cov}\{D_n(\underset{\sim}{t}), \hat{D}_n(\underset{\sim}{t})\}$$

$$= \sum_{r,s=1}^{m} \left\{ n\binom{n}{r}^{-1} \sum_{c=1}^{s} \binom{s}{c}\binom{n-s}{r-c} \xi_{rs\underset{\sim}{t}}(c) - rs\,\xi_{rs\underset{\sim}{t}}(1) \right\}.$$

It is shown using dominated convergence that

(5.4) $$n \sum_{r,s=1}^{m} \binom{n}{r}^{-1} \sum_{c=2}^{s} \binom{s}{c}\binom{n-s}{r-c} \xi_{rs\underset{\sim}{t}}(c) \rightarrow 0$$

and

(5.5) $$\sum_{r,s=1}^{m} \left\{ n\binom{n}{r}^{-1} \binom{s}{1}\binom{n-s}{r-1} - rs \right\} \xi_{rs\underset{\sim}{t}}(1) \rightarrow 0,$$

which will imply the result.

First, $n\binom{n}{r}^{-1}\sum_{c=2}^{s}\binom{s}{c}\binom{n-s}{r-c} < rs$. Therefore,

$$\left|\sum_{r,s=1}^{m}\left\{n\binom{n}{r}^{-1}\sum_{c=2}^{r}\binom{s}{c}\binom{n-s}{r-c}\xi_{rs\underset{\sim}{t}}(c)\right\}\right|$$

$$\leq \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} a_i a_j \sum_{r,s=1}^{\infty} rs\min\left(F^{(r)}(t_i), F^{(s)}(t_j)\right)$$

$$< \infty.$$

For fixed r and s with $c \geq 2$, $n\binom{n}{r}^{-1}\binom{s}{c}\binom{n-s}{r-c} \rightarrow 0$. Thus (5.4) holds.

Next, $\left| n\binom{n}{r}^{-1}\binom{s}{1}\binom{n-s}{r-1} - rs \right| \leq 2rs$. We can now write

$$\left|\sum_{r,s=1}^{m}\left\{n\binom{n}{r}^{-1}\binom{s}{1}\binom{n-s}{r-1} - rs\right\}\xi_{rs\underset{\sim}{t}}(1)\right|$$

$$\leq \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} a_i a_j \sum_{r,s=1}^{\infty} 2rs\min\left(F^{(r)}(t_i), F^{(s)}(t_j)\right)$$

$$< \infty.$$

For fixed r and s, $\left| n\binom{n}{r}^{-1}\binom{s}{1}\binom{n-s}{r-1} - rs \right| \rightarrow 0$, so that (5.5) follows. □

<u>Proof of Theorem 5.3</u>: By Chebyshev's inequality and Lemma 5.5,

$$n^{1/2}\left(D_n(\underset{\sim}{t}) - \hat{D}_n(\underset{\sim}{t})\right) \rightarrow 0 \text{ in probability.}$$

It follows from Lemmas 5.3 and 5.4 that

$$n^{1/2}\left(D_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})\right) \xrightarrow{D} N(0, \sigma^2_{\underset{\sim}{t}}).$$

Now suppose $0 \le t_1,...,t_\ell \le T$. Then the finite-dimensional distributions of Y'_n are asymptotically multivariate normal by the Cramér-Wold device. By definition $\Sigma_{i=1}^{\ell} a_i Y(t_i)$ is a $N(0,\sigma_{\underset{\sim}{t}}^2)$ random variable. Since the choices of ℓ and $\underset{\sim}{a} = (a_1,...,a_\ell)'$ were arbitrary, we see that the finite-dimensional distributions of Y'_n converge weakly to the finite-dimensional distributions of Y. □

5.3.2 Tightness of the process

The assumptions B7 and B8 will be shown to be sufficient for "tightness" of the sequence $\{Y'_n\}$. Lemma 5.5 states the principal result of this subsection. Used in conjunction with Theorem 5.3 it establishes sufficient conditions for weak convergence of Y'_n to Y.

Lemma 5.5: Assume B7 or B8 holds. Then there exists a positive constant K such that for any $0 \le t_1 \le t_2 \le t_3 \le T$,

$$\mathbf{E}\left[Y'_n(t_2) - Y'_n(t_1)\right]^2\left[Y'_n(t_3) - Y'_n(t_2)\right]^2 \le K(t_3 - t_1)^2$$

for all $n \ge 1$.

The proof of Lemma 5.5 appears at the end of this subsection after several intermediate lemmas. Some additional notation will be required. For $i = 1,2,...$ let $p_{i1} = F^{(i)}(t_2) - F^{(i)}(t_1)$ and $p_{i2} = F^{(i)}(t_3) - F^{(i)}(t_2)$. Suppose $\{X_{c1},...,X_{ci}\}$ is a subset of i random variables from $X_1,...,X_n$ and let Σ_{ci} denote summation over all $\binom{n}{i}$ such subsets. Define $S_{ci} = X_{c1} + ... + X_{ci}$, $S_{ci}(1) = \chi(S_{ci} \in (t_1,t_2]) - p_{i1}$, and $S_{ci}(2) = \chi(S_{ci} \in (t_2,t_3]) - p_{i2}$.

With the new notation one may write

$$F_n^{(i)}(t) = \binom{n}{i}^{-1} \sum_{ci} \chi(S_{ci} \le t),$$

$$Y_n'(t_2) - Y_n'(t_1) = n^{1/2} \sum_{i=1}^{m} \binom{n}{i}^{-1} \sum_{ci} S_{ci}(1),$$

and

$$Y_n'(t_3) - Y_n'(t_2) = n^{1/2} \sum_{i=1}^{m} \binom{n}{i}^{-1} \sum_{ci} S_{ci}(2).$$

Thus

(5.6)
$$\mathbf{E}\Big[Y_n'(t_2) - Y_n'(t_1)\Big]^2 \Big[Y_n'(t_3) - Y_n'(t_2)\Big]^2$$

$$= n^2 \sum_{i,j,k,\ell=1}^{m} \binom{n}{i}^{-1} \binom{n}{j}^{-1} \binom{n}{k}^{-1} \binom{n}{\ell}^{-1} \sum_{ci} \sum_{cj} \sum_{ck} \sum_{c\ell} \mathbf{E}\Big[S_{ci}(1) S_{cj}(1) S_{ck}(2) S_{c\ell}(2)\Big].$$

Define

$$S_{nijk\ell} = \Big\{(S_{ci}, S_{cj}, S_{ck}, S_{c\ell}) \colon \mathbf{E}\Big[S_{ci}(1) S_{cj}(1) S_{ck}(2) S_{c\ell}(2)\Big] \ne 0\Big\}$$

and let $U_{nijk\ell} = \#(S_{nijk\ell})$ be the cardinality of $S_{nijk\ell}$. Upper bounds for $n^2\binom{n}{i}^{-1}\binom{n}{j}^{-1}\binom{n}{k}^{-1}\binom{n}{\ell}^{-1} U_{nijk\ell}$ and $\mathbf{E}\Big[S_{ci}(1) S_{ck}(2)\Big]^2$ are found in Lemmas 5.6 and 5.7, respectively. The proof of Lemma 5.5 follows immediately from these lemmas.

<u>Lemma 5.6</u>: With $U_{nijk\ell}$ as defined,

$$n^2\binom{n}{i}^{-1}\binom{n}{j}^{-1}\binom{n}{k}^{-1}\binom{n}{\ell}^{-1} U_{nijk\ell} \le 7ijk\ell.$$

Proof: Let $V_{nijk\ell} = n^2\binom{n}{i}^{-1}\binom{n}{j}^{-1}\binom{n}{k}^{-1}\binom{n}{\ell}^{-1}$. In order for $(S_{ci}, S_{cj}, S_{ck}, S_{c\ell}) \in S_n$, one of the following two possibilities must occur.

(a) One observation is shared by all four sums. An upper bound on the number of ways this can occur is $n\binom{n-1}{i-1}\binom{n-1}{j-1}\binom{n-1}{k-1}\binom{n-1}{\ell-1}$. It is easily shown that $V_{nijk\ell}\, n\binom{n-1}{i-1}\binom{n-1}{j-1}\binom{n-1}{k-1}\binom{n-1}{\ell-1} = n^{-1} ijk\ell$.

(b) One observation is shared by two sums, while a second observation is shared by the other two sums. In this case an upper bound on the number of ways to choose the four sums is $\binom{4}{2}n(n-1)\binom{n-1}{i-1}\binom{n-1}{j-1}\binom{n-1}{k-1}\binom{n-1}{\ell-1}$. One then has that $V_{nijk\ell}\binom{4}{2}n(n-1)\binom{n-1}{i-1}\binom{n-1}{j-1}\binom{n-1}{k-1}\binom{n-1}{\ell-1} \le 6ijk\ell$.

The lemma follows by adding the bounds obtained in cases (a) and (b). □

For the following lemma, let i, j, and c be nonnegative integers such that $i \ge 1$, $j \ge 1$, and $0 \le c \le i \wedge j$. Let $X_1, \ldots, X_{i+j-c}$ be i.i.d. random variables having distribution F. Define $S_i = X_1 + \ldots + X_i$ and $S_j = X_1 + \ldots + X_c + X_{i+1} + \ldots + X_{i+j-c}$, so that S_i and S_j have c random variables in common.

Lemma 5.7: Assume B7 or B8 holds. Then for i,j = 1,2,... there exists a constant K independent of i and j such that

$$\mathbf{E}\left\{\chi(S_i \in (t_1,t_2]) - p_{i1}\right\}^2 \left\{\chi(S_j \in (t_2,t_3]) - p_{j2}\right\}^2 \le K\rho^{i \vee j}(t_3 - t_1)^2$$

Proof: First suppose that $i \le j$ and $c < j$. If B7 or B8 holds, then

$$\mathbf{P}\left\{S_i \in (t_1,t_2],\ S_j \in (t_2,t_3]\right\}$$

$$= \int_{-\infty}^{\infty}\left[F^{(i-c)}(t_2-u) - F^{(i-c)}(t_1-u)\right]\left[F^{(j-c)}(t_3-u) - F^{(j-c)}(t_2-u)\right]dF^{(c)}(u)$$

$$\leq K\rho^{j-c}(t_3-t_2)\int_{-\infty}^{\infty}\left[F^{(i-c)}(t_2-u) - F^{(i-c)}(t_1-u)\right]dF^{(c)}(u)$$

$$\leq K^2\rho^{i+j-c}(t_3-t_2)(t_2-t_1)$$

$$\leq K^2\rho^{j}(t_3-t_1)^2.$$

If $c = i = j$, then $S_i = S_j$ and the inequality again holds. Thus it holds whenever $c \leq i \leq j$. Similarly, if $c \leq j \leq i$, then

$$\mathbf{P}\left\{S_i \in (t_1,t_2],\ S_j \in (t_2,t_3]\right\} \leq K^2\rho^{i}(t_3-t_1)^2.$$

Thus,

$$\mathbf{E}\left\{\chi(S_i \in (t_1,t_2]) - p_{i1}\right\}^2\left\{\chi(S_j \in (t_2,t_3]) - p_{j2}\right\}^2$$

$$\leq p_{i1}^2 p_{j2}^2 + p_{i1}^2 p_{j2} + p_{i1} p_{j2}^2 + \mathbf{P}\left\{S_i \in (t_1,t_2],\ S_j \in (t_2,t_3]\right\}$$

$$\leq 3\left[F^{(i)}(t_2) - F^{(i)}(t_1)\right]\left[F^{(j)}(t_3) - F^{(j)}(t_2)\right] + K^2(t_3-t_1)^2\rho^{i\vee j}$$

$$\leq 4K^2(t_3-t_1)^2\rho^{i\vee j}$$

which completes the proof of the lemma. □

Proof of Lemma 5.5: By Lemma 5.7 and the Cauchy-Schwarz inequality, there exists a constant K such that

$$\left|\mathbf{E}[\,S_{ci}(1)S_{cj}(1)S_{ck}(2)S_{c\ell}(2)\,]\right| \leq K(t_3-t_1)^2(\rho^{i\vee k})^{1/2}(\rho^{j\vee \ell})^{1/2}$$

for all i, j, k, and ℓ. Thus, by (5.6) and Lemma 5.6,

$$\mathbf{E}\left[Y_n'(t_2) - Y_n'(t_1)\right]^2\left[Y_n'(t_3) - Y_n'(t_2)\right]^2$$

$$\leq \sum_{i,j,k,\ell=1}^{\infty} 7ijk\ell K(t_3-t_1)^2(\rho^{i\vee k})^{1/2}(\rho^{j\vee \ell})^{1/2}$$

$$\leq K_1(t_3-t_1)^2$$

for all n.

Proof of Theorem 5.2: With F continuous, the process Y can be shown to be mean square continuous on the interval [0, T]. That is, for $t_1, t_2 \in [0,\ T]$, $\mathbf{E}[Y(t_1)-Y(t_2)]^2 \to 0$ as $t_1 \to t_2$. An immediate consequence of this is that $\mathbf{E}[Y(T)-Y(T-)]^2 = 0$ and thus $\mathbf{P}\{Y(T) \neq Y(T-)\} = 0$. That $Y_n' \xrightarrow{D} Y$ follows from Theorem 5.3 and Lemma 5.5, using Theorem 15.6 of Billingsley (1968). Thus $Y_n \xrightarrow{D} Y$ by the remarks at the beginning of this section. □

5.4 A Sequential Bootstrap Confidence Band

As in the previous section, let $Y_n(t) = n^{1/2}[H_n(t) - H(t)]$ and $Y_n = \{Y_n(t),\ t \in [0,T]\}$. Theorem 5.3 gives sufficient conditions for Y_n to converge weakly to Y, where Y is a zero-mean Gaussian process. Define the random variables

$$S_n = \sup_{t \in [0,T]} |Y_n(t)|,\ n = 1,2,\ldots \text{ and } S = \sup_{t \in [0,T]} |Y(t)|.$$

Since $Y_n \xrightarrow{D} Y$ and the L_∞ norm is continuous in the Skorohod metric on D[0,T], one has that $S_n \xrightarrow{D} S$. Let $G(x) = \mathbf{P}(S \leq x)$ for $x \in \mathbf{R}$.

If the distribution G were known, then a confidence band for $\{H(t),\ t \in [0,T]\}$ could be constructed in the following way. Let $G^{-1}(y) = \inf\{x\colon G(x) \geq y\}$. For a target confidence level of $1-\alpha$, take the confidence band

$$\left\{H_n(t) \pm \frac{G^{-1}(1-\alpha)}{n^{1/2}},\ t \in [0,T]\right\}. \tag{5.7}$$

Then

$$\mathbf{P}\left\{H_n(t) - \frac{G^{-1}(1-\alpha)}{n^{1/2}} \leq H(t) \leq H_n(t) + \frac{G^{-1}(1-\alpha)}{n^{1/2}},\ \text{for all } t \in [0,T]\right\}$$

$$= \mathbf{P}\left\{S_n \leq G^{-1}(1-\alpha)\right\}$$

$$\to \mathbf{P}\left\{S \leq G^{-1}(1-\alpha)\right\} = 1-\alpha$$

as $n \to \infty$, assuming G is continuous at $G^{-1}(1-\alpha)$. Thus if G were known, (5.7) would give a reasonable large sample confidence band for $\{H(t),\ t \in [0,T]\}$. In applications, the distribution G will be completely unknown. The remainder of this section describes how $G^{-1}(1-\alpha)$ and a corresponding confidence band can be estimated using a bootstrap procedure.

It is assumed that the following conditions on F and G hold.

B9: F is a distribution on the positive real line (i.e., $F(0) = 0$). Furthermore, F is absolutely continuous with density f such that f is continuous a.e. with respect to Lebesgue measure. There exists $L > 0$ such that $\sup_{x \in [0,T]} f(x) < L$.

B10: G is continuous and strictly increasing in a neighborhood of $G^{-1}(1-\alpha)$.

The estimator H_n is calculated from a random sample $X_1, \ldots, X_n$ of random variables having distribution F. Let $F_n = F_n^{(1)}$ be the ordinary sample distribution function. Bootstrap samples will be drawn from a smoothed version of F_n. The role played by the density f in the proof of Theorem 5.3 motivates the smoothing of F_n.

Let K be a real bounded function vanishing outside the interval (0,1) such that $\int K(y)dy = 1$. For $h_n > 0$ define the usual kernel density estimator

$$f_n(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), & x \geq 0. \end{cases}$$

One more assumption will be required.

B11: The sequence of positive constants $\{h_n\}_{n=1}^{\infty}$ is such that $h_n \to 0$ and $f_n(x) \to f(x)$ a.s. for every continuity point x of f.

For $t \in \mathbf{R}$, define

$$\tilde{F}_n(t) = \int_0^t f_n(u)\,du.$$

Finally, let $\tilde{F}_n^{(k)}$ be the k-th convolution of $\tilde{F}_n$ and $\tilde{H}_n$ be the renewal function associated with $\tilde{F}_n$, so that $\tilde{H}_n(t) = \Sigma_{k=1}^{\infty} \tilde{F}_n^{(k)}(t)$.

For some positive integer M let $X_{i1}^*, \ldots, X_{in}^*$, $i = 1, \ldots, M$, be i.i.d. random variables from the distribution $\tilde{F}_n$. For $i = 1, \ldots, M$ define

$$\tilde{H}_{ni}(t) = \sum_{k=1}^{m} \tilde{F}_{ni}^{(k)}(t),$$

where

$$\tilde{F}_{ni}^{(k)}(t) = \binom{n}{k}^{-1} \sum_{c} \chi\left(X_{i1c}^{*} + \dots + X_{ikc}^{*} \le t\right).$$

Here Σ_c denotes the summation over all subsets of k random variables drawn from $X_{i1}^{*},\dots,X_{in}^{*}$. (Thus the definition of $\tilde{H}_{ni}(t)$ in terms of $X_{i1}^{*},\dots,X_{in}^{*}$ is identical to the definition of $H_n(t)$ in terms of $X_1,\dots,X_n$.)

Letting $Y_{ni}^{*}(t) = n^{1/2}[\tilde{H}_{ni}(t) - \Sigma_{k=1}^{m} \tilde{F}_{n}^{(k)}(t)]$, define

$$Y_{ni}^{*} = \{Y_{ni}^{*}(t),\ t \in [0,T]\} \text{ and } S_{ni}^{*} = \sup_{t\in[0,T]} |Y_{ni}^{*}(t)|\ ,\ i=1,\dots,M.$$

For $x \in \mathbf{R}$, let

$$G_n(x) = \frac{1}{M} \sum_{i=1}^{M} \chi(S_{ni}^{*} \le x),$$

the sample distribution function of the S_{ni}^{*}'s. As was done for G, take $G_n^{-1}(y) = \inf\{x\colon G_n(x) \ge y\}$, $y \in [0,1]$. Emulating the form of (5.7), define a confidence band for H(t) on the interval $[0,T]$ by

$$\left\{H_n(t) \pm \frac{G_n^{-1}(1-\alpha)}{n^{1/2}},\ t \in [0,T]\right\}. \tag{5.8}$$

Theorem 5.4 partially answers the question of when $G_n^{-1}(1-\alpha)$ is consistent for $G^{-1}(1-\alpha)$.

<u>Theorem</u> <u>5.4</u>: Assume B9, B10 and B11 hold. Then $G_n^{-1}(1-\alpha) \to G^{-1}(1-\alpha)$ a.s. as $M, n \to \infty$.

Proof: In the appendix the results of Section 5.3 are modified to show that $Y^*_{ni} \xrightarrow{D} Y$ as $n \to \infty$, $i=1,...,M$. Thus $S^*_{ni} \xrightarrow{D} S$ as $n \to \infty$, $i=1,...,M$, by the continuity of the L_∞ norm in the Skorohod metric on D[0,T]. Given the smoothed sample distribution function $\tilde{F}_n$, the random variables $S^*_{n1}, S^*_{n2}, ..., S^*_{nM}$ are independent and identically distributed. Thus

$$G_n^{-1}(1-\alpha) \to G^{-1}(1-\alpha) \text{ a.s.}$$

as $M, n \to \infty$. □

Theorem 5.4 verifies that the bootstrap confidence band (5.8) provides the correct coverage probability asymptotically:

$$\mathbf{P}\left\{H_n(t) - \frac{G_n^{-1}(1-\alpha)}{n^{1/2}} \le H(t) \le H_n(t) + \frac{G_n^{-1}(1-\alpha)}{n^{1/2}}, \text{ for all } t \in [0,T]\right\}$$

$$= \mathbf{P}\left\{\sup_{t \in [0,t]} |n^{1/2}[H_n(t) - H(t)]| \le G_n^{-1}(1-\alpha)\right\}$$

$$\to \mathbf{P}\left\{S \le G^{-1}(1-\alpha)\right\} = 1 - \alpha,$$

as $M, n \to \infty$. This justifies the use of the bootstrap procedure when the sample size n and the number of bootstrap samples M are both large.

Section 4.5 yielded a sequential procedure for a fixed-width confidence interval when estimating H(t) at a single value of t. For estimation of H on the interval [0,T], the analogous problem is to develop a sequential version of the fixed-width confidence band

$$\{\ H_n(t) \pm d,\ t \in [0,T]\ \}$$

where $d > 0$.

Fix $d > 0$ and for some initial sample size n_o define the stopping time

$$N_d = \inf\left\{n \geq n_o:\ n^{1/2}d \geq G_n^{-1}(1-\alpha)\right\}.$$

Then $N_d \to \infty$ a.s. as $d \downarrow 0$, $M \to \infty$. By the relation

$$\left[G_{N_d}^{-1}(1-\alpha)\right]^2 \leq N_d d^2 < \left[G_{N_d-1}^{-1}(1-\alpha)\right]^2 + d^2,$$

under the conditions of Theorem 5.4, $N_d^{1/2} d \to G^{-1}(1-\alpha)$ a.s. as $d \downarrow 0$, $M \to \infty$. If $Y_{N_d} \xrightarrow{W} Y$ as $d \downarrow 0$, $M \to \infty$, then

$$P\left\{\sup_{t\in[0,T]}\left|H_{N_d}(t) - H(t)\right| \leq d\right\}$$

$$= P\left\{\sup_{t\in[0,T]}\left|N_d^{1/2}[H_{N_d}(t) - H(t)]\right| \leq N_d^{1/2} d\right\}$$

$$\to P\left\{S \leq G^{-1}(1-\alpha)\right\} = 1 - \alpha.$$

Thus the following theorem would hold.

Theorem 5.5: Assume B9, B10, and B11 hold. If $Y_{N_d} \xrightarrow{W} Y$ as $d \downarrow 0$, then

$$P\left\{\sup_{t\in[0,T]}\left|H_{N_d}(t) - H(t)\right| \leq d\right\} \to 1 - \alpha \qquad \text{as } d \downarrow 0,\ M \to \infty.$$

In Section 6.3 we outline one method for showing weak convergence of a randomly indexed stochastic process such as Y_{N_d} based on a two-dimensional process. Given small

$\delta > 0$, consider the random variables defined by

$$Z_n(s, t) = \frac{[ns]\left\{ H_{[ns]}(t) - H(t) \right\}}{n^{1/2}}, \quad s \in [0, 1+\delta],\ t \in [0,T].$$

(where $[\cdot]$ denotes the greatest integer function) and the associated process

$$Z_n = \{ Z_n(s,t);\ s \in [0, 1+\delta],\ t \in [0,T] \}.$$

Part of Section 6.3 describes how if Z_n converges weakly in an appropriate topology on $[0, 1+\delta] \times [0,T]$, then $Y_{N_d} \xrightarrow{W} Y$ as $d \downarrow 0$. That Z_n might converge weakly under appropriate conditions is suggested by Theorem 4.1 and 5.2

In practice one would only conduct the bootstrap procedure on some finite subset of points in the interval $[0,T]$. Thus rather than verify the weak convergence of Z_n, we show that for an arbitrary grid of points $t_1,\dots,t_\ell$ taken from $[0,T]$ one has that

$$(Y_{N_d}(t_1), \dots, Y_{N_d}(t_\ell)) \xrightarrow{D} (Y(t_1), \dots, Y(t_\ell)) \text{ as } d \downarrow 0.$$

Let $(a_1,\dots,a_\ell)$ be any vector of real numbers. By Theorem 5.2 and the Cramér-Wold device, $\Sigma_{i=1}^{\ell} a_i Y_n(t_i) \xrightarrow{D} \Sigma_{i=1}^{\ell} a_i Y(t_i)$. If one also assumes that $[m_n]^2 \log n = o(n)$, then as previously asserted Lemma 4.8 implies $\{Y_n(t_i)\}_{n \geq 1}$ is a uniformly continuous in probability sequence for $i = 1,\dots,\ell$. By Lemma 1.4 of Woodroofe (1982) the sequence $\{ \Sigma_{i=1}^{\ell} a_i Y_n(t_i) \}_{n \geq 1}$ is also uniformly continuous in probability. Since $N_d/d^{-2} \to c^2$ a.s. as $d \downarrow 0$, $\Sigma_{i=1}^{\ell} a_i Y_{N_d}(t_i) \xrightarrow{D} \Sigma_{i=1}^{\ell} a_i Y(t_i)$ by Anscombe's Theorem. Again using the Cramér-Wold device, it follows that $(Y_{N_d}(t_1), \dots, Y_{N_d}(t_\ell)) \xrightarrow{D} (Y(t_1), \dots, Y(t_\ell))$ as $d \downarrow 0$. (See Section 1.3 of Woodroofe (1982) for a discussion of uniformly continuous in probability sequences and Anscombe's Theorem.)

It should be noted that while we have taken the bootstrap sample size to be equal to the sample size n, this is not necessary in practice. If n_1 is the sample size of each of the M bootstrap samples, then it can be seen that all of the asymptotic results of this section hold if $n_1 \to \infty$ as $n \to \infty$. Another point is that one probably need not smooth the empirical distribution before resampling to get useful confidence bands. The convolution of F_n with the kernel K provided a means for showing weak convergence of the bootstrapped process. While the smoothing has not been shown unnecessary for the asymptotic results considered here, in the usual finite sample situation one could make the bandwidth h_n arbitrarily small, in effect resampling from the unsmoothed sample distribution function.

5.5 A Variable-Width Bootstrap Confidence Band

The supremum S of the limiting gaussian process Y is more likely to occur on subintervals of $[0,T]$ where the variance function σ_t^2 is relatively large, rather than on subintervals where σ_t^2 is comparatively small. The procedure of Section 5.4 ignores this fact and consequently yields a confidence band which is unduly conservative. The method presented in this section overcomes that defect by taking into account the variance function σ_t^2. The objective will be a confidence band for $\{H(t),\ t \in [T_1,T_2]\}$, where $0 < T_1 < T_2 < \infty$. It will be assumed that there exists a positive constant ϵ such that $\sigma_t^2 > \epsilon$ for all $t \in [T_1,T_2]$.

Define $\tilde{Y}_n(t) = n^{1/2}[H_n(t) - H(t)]/\sigma_t$ and $\tilde{Y}_n = \{\tilde{Y}_n(t),\ t \in [T_1,T_2]\}$. Then by Theorem 5.2, $\tilde{Y}_n \overset{W}{\to} \tilde{Y}$, where $\tilde{Y} = \{\tilde{Y}(t),\ t \in [T_1,T_2]\}$ is a zero-mean Gaussian process having covariance function

$$\tilde{R}(t_1,t_2) = \frac{R(t_1,t_2)}{\sigma_{t_1}\sigma_{t_2}}\ ,\ t_1,\ t_2 \in [T_1,T_2].$$

Here R is the covariance function of Y, defined by (5.2). Let $\tilde{\tilde{S}} = \sup_{t \in [T_1,T_2]} |\tilde{\tilde{Y}}(t)|$ and $\tilde{\tilde{G}}(x) = \mathbf{P}(\tilde{\tilde{S}} \leq x)$ for $x \in \mathbb{R}$.

Let σ_{nt}^2 be the estimator of σ_t^2 defined by (4.13). Since F is continuous,

$$\sup_{t \in [T_1,T_2]} |\sigma_{nt}^2 - \sigma_t^2| \rightarrow 0 \text{ a.s.}$$

by the remarks at the conclusion of Section 4.4.1. Since σ_t^2 is bounded away from zero,

$$\sup_{t \in [T_1,T_2]} |\sigma_{nt} - \sigma_t| \rightarrow 0 \text{ a.s.} \tag{5.9}$$

Thus σ_t can be replaced by σ_{nt} in the process $\tilde{\tilde{Y}}_n$ and the resulting process will converge weakly to the process $\tilde{\tilde{Y}}$. Thus for a large sample size n the confidence band

$$\left\{ H_n(t) \pm \frac{\tilde{\tilde{G}}^{-1}(1-\alpha)}{n^{1/2}} \sigma_{nt},\ t \in [T_1,T_2] \right\} \tag{5.10}$$

would have meaning in the sense that

$$\mathbf{P}\left\{ H_n(t) - \frac{\tilde{\tilde{G}}^{-1}(1-\alpha)}{n^{1/2}} \sigma_{nt} \leq H(t) \leq H_n(t) + \frac{\tilde{\tilde{G}}^{-1}(1-\alpha)}{n^{1/2}} \sigma_{nt},\ t \in [T_1,T_2] \right\}$$

$$= \mathbf{P}\left\{ \sup_{t \in [T_1,T_2]} \left| \frac{n^{1/2}[H_n(t) - H(t)]}{\sigma_{nt}} \right| \leq \tilde{\tilde{G}}^{-1}(1-\alpha) \right\}$$

$$\rightarrow \mathbf{P}\left\{ \tilde{\tilde{S}} \leq \tilde{\tilde{G}}^{-1}(1-\alpha) \right\} = 1-\alpha.$$

Unfortunately, as with the supremum distribution G of Section 5.4, the form of $\tilde{\tilde{G}}$ is completely unknown.

For the remainder of this section, it will be assumed that B11 holds, as well as the following modified versions of B9 and B10.

B9$'$: F is a distribution on the positive real line (i.e., $F(0) = 0$). Furthermore, F is absolutely continuous with density f such that f is continuous a.e. with respect to Lebesgue measure. There exists $L > 0$ such that $\sup_{x\in[0,T]} f(x) < L$. There exists a positive constant ϵ such that $\sigma_t^2 > \epsilon$ for all $t \in [T_1, T_2]$.

B10$'$: $\tilde{G}$ is continuous and strictly increasing in a neighborhood of $\tilde{G}^{-1}(1-\alpha)$.

Take M to be a positive integer and let Y^*_{ni}, $i=1,\ldots,M$ be the bootstrapped processes defined in Section 5.4. Define

$$\tilde{Y}^*_{ni} = \left\{ \frac{Y^*_{ni}(t)}{\sigma_{nt}}, \ t \in [T_1, T_2] \right\}$$

and

$$\tilde{S}^*_{ni} = \sup_{t\in[T_1,T_2]} |\tilde{Y}^*_{ni}(t)| ,$$

$i=1,\ldots,M$. By Theorem A1 of the appendix and (5.9), $\tilde{Y}^*_{ni} \xrightarrow{W} \tilde{Y}$ and $\tilde{S}^*_{ni} \xrightarrow{D} \tilde{S}$, $i=1,\ldots,M$. For $x \in \mathbf{R}$, let

$$\tilde{G}_n(x) = \frac{1}{M} \sum_{i=1}^{M} \chi(\tilde{S}^*_{ni} \le x) .$$

Following the form of (5.10), define the confidence band

$$\left\{ H_n(t) \pm \frac{\tilde{G}_n^{-1}(1-\alpha)}{n^{1/2}} \sigma_{nt}, \ t \in [T_1, T_2] \right\}. \tag{5.11}$$

Theorem 5.6 gives sufficient conditions for the confidence band (5.11) to have the correct asymptotic coverage probability.

Theorem 5.6: Assume B9′, B10′ and B11 holds. Then $\tilde{G}_n^{-1}(1-\alpha) \rightarrow \tilde{G}^{-1}(1-\alpha)$ a.s. as $M, n \rightarrow \infty$.

Proof: The proof is identical to that of Theorem 5.4, replacing Y^*_{ni}, S^*_{ni}, S, G_n^{-1}, and G^{-1} by $\tilde{Y}^*_{ni}$, $\tilde{S}^*_{ni}$, $\tilde{S}$, $\tilde{G}_n^{-1}$, and $\tilde{G}^{-1}$, respectively. □

By Theorem 5.6, as for the confidence band (5.8), we have that

$$\mathbf{P}\left\{ H_n(t) - \frac{\tilde{G}_n^{-1}(1-\alpha)}{n^{1/2}}\sigma_{nt} \le H(t) \le H_n(t) + \frac{\tilde{G}_n^{-1}(1-\alpha)}{n^{1/2}}\sigma_{nt},\ t \in [T_1,T_2] \right\}$$

$$= \mathbf{P}\left\{ \sup_{t\in[T_1,T_2]} \left| \frac{n^{1/2}[H_n(t)-H(t)]}{\sigma_{nt}} \right| \le \tilde{G}_n^{-1}(1-\alpha) \right\}$$

$$\rightarrow \mathbf{P}\left\{ \tilde{S} \le \tilde{G}^{-1}(1-\alpha) \right\} = 1-\alpha$$

as $M, n \rightarrow \infty$, which provides some justification for this bootstrap procedure.

To define a stopping time for the new procedure, the standard deviation σ_{nt} is scaled by a constant $d>0$; d is then allowed to go to zero. Fix $d>0$ and for some initial sample size n_o define the stopping time

$$\tilde{N}_d = \inf\left\{ n \ge n_o:\ n^{1/2}d \ge \tilde{G}_n^{-1}(1-\alpha) \right\}.$$

Upon stopping, take the confidence band

$$\left\{ H_{\tilde{N}_d}(t) \pm d\sigma_{\tilde{N}_d t},\ t \in [T_1,T_2] \right\}. \tag{5.12}$$

As for the stopping time N_d of Section 5.4, $\tilde{N}_d \rightarrow \infty$ and $\tilde{N}_d^{1/2}d \rightarrow \tilde{G}^{-1}(1-\alpha)$ a.s. as $d\downarrow 0$, $M \rightarrow \infty$. If $Y_{N_d} \xrightarrow{W} Y$ as $d\downarrow 0$, $M \rightarrow \infty$, it follows that $\tilde{Y}_{\tilde{N}_d} \xrightarrow{W} \tilde{Y}$ as $d\downarrow 0$, $M \rightarrow \infty$ and

$$\mathbf{P}\left\{ \left| H_{\tilde{N}_d}(t) - H(t) \right| \le d\,\sigma_{\tilde{N}_d t},\ t \in [T_1, T_2] \right\}$$

$$= \mathbf{P}\left\{ \sup_{t \in [T_1, T_2]} \frac{\left| \tilde{N}_d^{1/2} \left[H_{\tilde{N}_d}(t) - H(t) \right] \right|}{\sigma_{\tilde{N}_d t}} \le \tilde{N}_d^{1/2} d \right\}$$

$$\rightarrow \mathbf{P}\left\{ \tilde{S} \le \tilde{G}^{-1}(1-\alpha) \right\} = 1 - \alpha.$$

Thus the following theorem would hold, giving assurance that the confidence band (5.12) has the intended asymptotic coverage probability.

Theorem 5.7: Assume B9′, B10′, and B11 hold. If $Y_{N_d} \xrightarrow{W} Y$ as $d \downarrow 0$, then

$$\mathbf{P}\left\{ \sup_{t \in [T_1, T_2]} \left| H_{\tilde{N}_d}(t) - H(t) \right| \le d\,\sigma_{\tilde{N}_d t} \right\} \rightarrow 1 - \alpha \qquad \text{as } d \downarrow 0,\ M \rightarrow \infty.$$

CHAPTER 6

STOPPING TIMES FOR FIXED-WIDTH CONFIDENCE BANDS

6.1 Introduction

In Sections 5.4 and 5.5 the problem of renewal function estimation motivated the introduction of stopping times for fixed and variable width confidence bands. In this chapter we consider such stopping times from a more general point of view.

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having distribution function F. Suppose $R_F = \{R_F(t), t \in I\}$ is an unknown functional process of F, where $I \subseteq \mathbf{R}$ is either a bounded closed interval, a closed half-line, or $\mathbf{R}$ itself. There often exists an estimator $R_n(\cdot) = R_n(\cdot\,; X_1, \ldots, X_n)$ of $R_F(\cdot)$ and an associated process

$$r_n(\cdot) = n^{1/2}\{R_n(\cdot) - R_F(\cdot)\}$$

on I. Several authors have exploited weak convergence properties of r_n to construct large sample confidence bands for the functional process R_F. The methodology introduced in this section extends readily to the case where R_F is an unknown functional process of F defined on $I \subseteq \mathbf{R}^p$, $p > 1$. See Section 6.6 for an example.

In applications where I is an unbounded interval one can always find a monotone increasing function that maps I into a bounded closed interval I'. (The function estimator may then have to be appropriately defined at endpoints of I'.) Thus without loss of generality and for simplicity of presentation take I to be a closed bounded interval.

Denote by D(I) the space of functions on I that have left-hand limits and are right-continuous. Throughout this chapter we assume that R_F and the estimates r_n, $n \geq 1$, belong to the space D(I). Let ρ denote the usual Skorohod metric on D(I); weak convergence is then taken to be in the Skorohod J_1-topology on D(I). The notation $\overset{W}{\rightarrow}$ ($\overset{D}{=}$) will be used to denote weak convergence (equivalence in distribution) of processes while $\overset{D}{\rightarrow}$ will denote weak convergence of random variables.

Remark: If I were an unbounded closed interval, then one could let ρ be the metric k considered by Lindvall(1973) or any other metric which extends the J_1-topology to D(I); weak convergence could then be assumed to take place in this extended topology. The discussion in this section would then hold if the supremum norm were replaced by the norm $\rho(\cdot, \mathbf{0})$, where $\mathbf{0}$ is the function that is identically 0 on I.

A typical approach assumes that $r_n \overset{W}{\rightarrow} \mathcal{G}_F$, where $\mathcal{G}_F$ is a Gaussian process on the interval I such that $\mathbf{P}\{ \sup_{t\in I} |\mathcal{G}_F(t)| < \infty \} = 1$. Define

$$G_F(x) = \mathbf{P}\{ \sup_{t\in I} |\mathcal{G}_F(t)| \leq x \}, -\infty < x < \infty.$$

For a desired confidence level of $1-\alpha$, $0<\alpha<1$, let $c = \inf\{x\colon G_F(x) \geq 1-\alpha\}$. If c_n is a weakly consistent estimate of c, then a sensible large sample confidence band for R_F is given by

$$\{ R_n(t) \pm c_n n^{-1/2}, t\in I \}, \tag{6.1}$$

for then

$$\mathbf{P}\Big\{ R_n(t) - c_n n^{-1/2} \leq R_F(t) \leq R_n(t) + c_n n^{-1/2},\ t\in I \Big\} \rightarrow 1-\alpha, \tag{6.2}$$

assuming G_F is continuous at c.

In some special cases G_F is independent of F and explicitly known, so that one can take $c_n = c$ in (6.1). Estimation of a continuous distribution function F by the empirical distribution function is such a case, since then G_F is the distribution of the supremum of the Brownian bridge. Because G_F is usually completely unknown, bootstrap methods have been advocated for estimating c. (See, for example, Bickel and Freedman(1981) and Csörgő and Mason(1989); other work on bootstrapping empirical processes is summarized in the latter paper.) Sometimes the covariance function of $\mathcal{G}_F$ may be known, at least up to its dependence on F. Then a possible approach is to use either the covariance function or an estimate thereof to simulate independent copies of a Gaussian process hopefully close in distribution to $\mathcal{G}_F$; taking the supremum norm of each of these copies, G_F and c can be estimated empirically.

If the objective is to estimate R_F on I with some uniform prespecified precision, then the idea of a fixed-width confidence band is appealing. In practice one might want

$$\text{(6.3)} \qquad \mathbf{P}\left\{ |R_n(t) - R_F(t)| \leq d,\ t \in I \right\} \geq 1-\alpha,$$

where $d>0$ is a constant that reflects the desired accuracy. Typically the minimum sample size n such that (6.3) holds is unknown and some guidelines for deciding when an adequate sample size has been reached would be useful. The analogy with fixed-width confidence interval estimation of a real-valued parameter $\theta(F)$ is suggestive. Suppose θ_n is a consistent estimate of $\theta(F)$ and one would like the smallest sample size n, usually unknown, such that $\mathbf{P}\{ |\theta_n - \theta(F)| \leq d \} \geq 1-\alpha$. Further assume $n^{1/2}(\theta_n - \theta(F)) \xrightarrow{D} N(0, \sigma^2)$ as $n \to \infty$ and s_n^2 is a strongly consistent estimator of σ^2. For the fixed-width confidence interval problem, a number of authors have considered stopping times of the type

$$N'_d = \inf\left\{ n \geq n_o: \; n^{-1/2} s_n \tau_{\alpha/2} \leq d \right\},$$

where n_o is some initial sample size and $\tau_{\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal distribution. (See, for instance, Chow and Robbins(1965) and Sen(1981).) With appropriate conditions on the sequences $\{\theta_n\}$ and $\{s_n^2\}$ it has been shown that

$$\lim_{d \downarrow 0} \mathbf{P}\left\{ |\theta_{N'_d} - \theta(F)| \leq d \right\} = 1-\alpha,$$

so that for sufficiently small d the confidence band $(\theta_{N'_d} - d, \theta_{N'_d} + d)$ has approximately the desired coverage probability.

This chapter has as its primary purpose the definition of a stopping time N_d such that under certain conditions

$$\text{(6.4)} \qquad \lim_{d \downarrow 0} \mathbf{P}\left\{ |R_{N_d}(t) - R_F(t)| \leq d, \; t \in I \right\} = 1-\alpha.$$

Let c_n be a strongly consistent estimate of c. For fixed $d>0$ and some initial sample size n_o, define the stopping time

$$\text{(6.5)} \qquad N_d = \inf\left\{ n \geq n_o: \; n^{-1/2} c_n \leq d \right\};$$

upon stopping use the confidence band

$$\text{(6.6)} \qquad \{ R_{N_d}(t) \pm d, \; t \in I \}.$$

Thus the rule states that one should stop when the width of the confidence band determined by (6.1) is at most 2d.

The most basic properties of N_d follow directly from its definition. By the relation

$$\text{(6.7)} \qquad [c_{N_d}]^2 \leq N_d d^2 < [c_{N_d - 1}]^2 + d^2,$$

N_d is finite a.s. for fixed $d > 0$. If the quantile c of the distribution G_F were known, then for small $d > 0$ a reasonable fixed sample size would be

$$n_d = \inf\left\{ n \geq n_o : n^{-1/2} c \leq d \right\}, \tag{6.8}$$

where n_o is the initial sample size that appears in (6.5). Suppose $\mathbf{P}\{ c_n \leq 0 \} = 0$ for all $n \geq n_o$. Since $n_d \sim c^2/d^2$ as $d \downarrow 0$, one then has that $N_d/n_d \rightarrow 1$ as $d \downarrow 0$ by (6.7).

It is shown in Section 6.2 that if one also assumes that $r_{N_d} \overset{W}{\rightarrow} \mathcal{G}_F$ as $d \downarrow 0$, then (6.4) holds. Several methods for establishing weak convergence of r_{N_d} are given. In Section 6.3 sufficient conditions are given for the expectation of the stopping time to be finite for fixed $d > 0$, as well as for the asymptotic efficiency and normality of N_d as $d \downarrow 0$. Several bootstrap estimators of c are considered in Section 6.4, along with some resulting properties of the stopping time. In some applications one might want to allow confidence bands of variable width. Section 6.5 addresses the problem of stopping times for such modified confidence band procedures. An example in Section 6.6 illustrates how a stopping time of the form (6.5) can be used in the construction of a confidence region for a statistical function R_F defined on $\mathbb{R}^p$, $p \geq 2$.

6.2 Asymptotic Coverage Of The Confidence Band

The principal result of this section concerning the asymptotic coverage probability of the confidence band (1.6) rests on the assumption that $r_{N_d} \overset{W}{\rightarrow} \mathcal{G}_F$ as $d \downarrow 0$. Theorem 6.2 gives a single sufficient condition for the weak convergence of r_{N_d}; some discussion then indicates how one might verify this condition.

C1 summarizes the principal assumptions of the introduction.

C1: With previously defined notation, $r_n \overset{W}{\to} \mathcal{G}_F$ on I, where $\mathbf{P}\{\sup_{t\in I} |\mathcal{G}_F(t)| < \infty\} = 1$. $G_F(c) = 1-\alpha$ and G_F is continuous at c. There exists an estimator c_n of c such that $c_n \to c$ a.s. as $n\to\infty$ and $\mathbf{P}\{ c_n \leq 0 \} = 0$ for all $n \geq n_o$.

The estimator c_n can always be modified, without affecting its consistency, in order to meet the requirement that $\mathbf{P}\{c_n \leq 0\} = 0$ for all $n \geq n_o$. Tsirel'son(1975) showed that if $\mathcal{G}_F$ is a separable nondegenerate mean-zero Gaussian process which is almost surely bounded on I, then G_F must be continuous on the half-line $(0, \infty)$. Thus the continuity condition included in C1 is generally met.

<u>Theorem</u> <u>6.1:</u> Assume C1 and that $r_{N_d} \overset{W}{\to} \mathcal{G}_F$ as $d\downarrow 0$. Then (6.4) holds.

<u>Proof:</u> From the basic properties of N_d discussed in the introduction it follows that $c_{N_d} \to c$ a.s. and $N_d^{1/2} d \to c$ a.s. as $d\downarrow 0$. Thus the result holds by (6.2). □

In practice it will of course be necessary to verify that $r_{N_d} \overset{W}{\to} \mathcal{G}_F$ as $d\downarrow 0$. A sufficient condition for this is as follows; this condition generalizes the notion of a sequence of random variables which is uniformly continuous in probability. A sequence of random variables $\{X_n\}_{n\geq 1}$ is uniformly continuous in probability (u.c.i.p.) if given $\epsilon > 0$ there exists $\delta > 0$ such that for all $n \geq 1$

$$\mathbf{P}\{ \max_{0 \leq k \leq n\delta} |X_{n+k} - X_n| \geq \epsilon \} < \epsilon.$$

The condition we consider here is as follows.

Given $\epsilon > 0$ there exists $\delta > 0$ such that for all $n \geq 1$

$$\mathbf{P}\{ \max_{0 \leq k \leq n\delta} \rho(r_{n+k}, r_n) \geq \epsilon \} < \epsilon.$$

Anscombe's Theorem gives conditions for a weakly convergent sequence of random variables to have the same limiting distribution when the index is replaced by a random index. Theorem 6.2 is the analogous result in the present setting; the same type of argument employed in the proof of Anscombe's theorem is used here.

<u>Theorem 6.2:</u> Assume C1 and C2 hold. Then (6.4) holds.

<u>Proof:</u> Since $N_d/n_d \to 1$ a.s. as $d \downarrow 0$ by (6.7), for arbitrary fixed ϵ, $\epsilon' > 0$ one may choose $\delta > 0$ such that

$$\mathbf{P}\{\, \rho(r_{N_d}, r_{n_d}) \geq \epsilon \,\}$$

$$\leq \mathbf{P}\{\, |N_d/n_d - 1| > \delta \,\} + \mathbf{P}\{\, \max_{|n/n_d - 1| \leq \delta} \rho(r_{N_d}, r_{n_d}) \geq \epsilon \,\}$$

$$\leq \epsilon',$$

making use of C2. Thus $\rho(r_{N_d}, r_{n_d}) \xrightarrow{P} 0$ as $d \downarrow 0$ and r_{N_d} has the same limiting distribution as r_{n_d}, namely the distribution of $\mathcal{G}_F$. □

A convenient method for verifying condition C2 is to take recourse to a two-parameter process as introduced below. For each fixed $t \in I$ and given a small $\delta > 0$, define a process on the interval $[0, 1+\delta]$ by

$$Z_n(s, t) = \frac{[ns]\left\{ R_{[ns]}(t) - R_F(t) \right\}}{n^{1/2}}, \quad s \in [0, 1+\delta].$$

(Here $[\cdot]$ denotes the greatest integer function.) We consider weak convergence of the two-parameter process

$$(6.9) \qquad Z_n = \left\{ Z_n(s, t);\ s \in [0, 1+\delta], t \in I \right\}$$

as a condition which implies C2.

Endow $D[0, 1+\delta] \times D(I)$ with the S-topology considered by Bickel and Wichura(1971) and suppose that there exists a process Z on $D[0, 1+\delta] \times D(I)$ such that $Z_n \overset{W}{\to} Z$ in the S-topology. (Bickel and Wichura (1971), for example, give sufficient conditions for verifying weak convergence of Z_n.) For any n such that $|n/n_d - 1| < \delta$, choose s such that $n_d s = n$. Then for this fixed s

$$(6.10) \qquad \rho(r_n, r_{n_d})$$

$$\leq \rho(Z_{n_d}(s, \cdot), Z_{n_d}(1, \cdot)) + |s^{1/2} - 1| \sup_{t \in I} |n^{1/2}\{R_n(t) - R_F(t)\}|$$

and as a result

$$(6.11) \qquad P\{ \max_{|n/n_d - 1| \leq \delta} \rho(r_n, r_{n_d}) \geq \epsilon \}$$

$$\leq P\left\{ \sup_{s \in [1-\delta, 1+\delta]} \rho(Z_{n_d}(s, \cdot), Z_{n_d}(1, \cdot)) > \epsilon/2 \right\}$$

$$+ P\left\{ \max_{|n/n_d - 1|} \sup_{t \in I} |n^{1/2}\{R_n(t) - R_F(t)\}| > \epsilon/2|s^{1/2} - 1| \right\}.$$

Since $Z_n \overset{W}{\to} Z$, by Theorem 2 of Bickel and Wichura(1971),

$$(6.12) \qquad P\left\{ \sup_{s \in [1-\delta, 1+\delta]} \rho(Z_{n_d}(s, \cdot), Z_{n_d}(1, \cdot)) > \epsilon/2 \right\} \to 0 \text{ as } \delta \to 0.$$

Again using the weak convergence of Z_n, for some constant $K > 0$

$$(6.13) \qquad \mathbf{P}\Big\{ \max_{|n/n_d - 1|} \sup_{t\in I} |n^{1/2}\{R_n(t) - R_F(t)\}| > \epsilon/2|s^{1/2} - 1| \Big\}$$

$$\leq \mathbf{P}\Big\{ \sup_{s\in[1-\delta,\, 1+\delta]} \sup_{t\in I} |Z_n(s, t)| > K|s^{1/2} - 1|^{-1} \Big\} \to 0$$

as $\delta \to 0$.

Together (6.11), (6.12), and (6.13) imply C2 holds. This method for verifying C2 can also be used when R_F is defined on a compact set $I \subset \mathbf{R}^p$ by considering weak convergence of the process defined by (6.9) with $t \in \mathbf{R}^p$, again in the S-topology of Bickel and Wichura(1971).

Another potential approach for establishing C2 supposes that there exists a sequence of σ-fields $\{\mathcal{C}_n\}$ such that $\{R_n(\cdot) - R_F(\cdot);\ \mathcal{C}_n\}_{n=1}^{\infty}$ forms a reverse martingale process. Taking $\delta > 0$ and $N = [n\delta]$, one would then have that

$$\{R_{n+k}(\cdot) - R_{n+N}(\cdot);\ \mathcal{C}_{n+k}\}_{k=1}^{N}$$

is also a reverse martingale process. It follows from this that

$$\mathbf{E}_{\mathcal{C}_{n+k+1}} \{ \sup_{t\in I} |R_{n+k}(t) - R_{n+N}(t)| \}$$

$$\leq \sup_{t\in I} |\mathbf{E}_{\mathcal{C}_{n+k+1}}\{R_{n+k}(t) - R_{n+N}(t)\}|$$

$$= \sup_{t\in I} |R_{n+k+1}(t) - R_{n+N}(t)\}|$$

and consequently $\{ \sup_{t\in I} |R_{n+k}(t) - R_{n+N}(t)|;\ \mathcal{C}_{n+k} \}_{k=1}^{N}$ is a reverse submartingale. One could then try to bound the probability

$$\mathsf{P}\{ \max_{0 \leq k \leq n\delta} \rho(r_{n+k}, r_n) \geq \epsilon \}$$

using the Kolmogorov inequality for reverse submartingales.

6.3 Properties Of The Stopping Time

By making additional assumptions about the quantile estimator c_n, some questions about the expectation and rate of convergence of the stopping time can be answered. In the fixed-width confidence interval problem, the properties that will be considered here have been shown to hold for the stopping time N'_d under a variety of conditions on the variance estimates $\{s_n^2, n \geq n_o\}$. Due to the similarity in form between N'_d and N_d, by putting the same requirements on the sequence $\{c_n^2, n \geq n_o\}$, the same properties follow for N_d. Thus no proofs are given here, the necessary calculations being identical in spirit to those in the references cited. A problem is that conditions that are easily verified for variance estimators may not readily apply to quantile estimators, so discussion is restricted to those conditions which seem most relevant.

Among the basic properties established in the introduction were that $N_d < \infty$ a.s. for fixed $d > 0$ and $N_d/n_d \to 1$ a.s. as $d \downarrow 0$. Lemma 6.1 provides a condition which implies that for fixed $d > 0$ the expectation of the stopping time is finite, while as $d \downarrow 0$ the stopping time is asymptotically efficient in the sense that

$$\lim_{d \downarrow 0} n_d^{-1} \mathsf{E} N_d = 1. \tag{6.14}$$

The type of arguments needed to prove Lemma 6.1 may be found in Sen(1981), pp. 284 and 287.

Lemma 6.1: Assume C1 holds and that for fixed $d > 0$

(6.15) $$\mathbf{E}\{ \sup_{n \geq n_o} c_n^2 \} < \infty.$$

Then $\mathbf{E} N_d < \infty$ and (6.14) holds.

To show that the expectation of the stopping time is finite, another approach is to establish the equivalent condition that

(6.16) $$\sum_{n \geq n_o} \mathbf{P}\{ N_d > n \} < \infty.$$

Similarly, an alternative condition for verifying (6.14) is

(6.17) $$\lim_{d \downarrow 0} \; n_d^{-1} \sum_{n \geq n_d(1+\epsilon)} \mathbf{P}\{ N_d > n \} = 0.$$

(See, for example, Sen(1981), p.288.) In Section 6.4 it is demonstrated that if c_n is a particular type of bootstrap estimator it may be possible to verify (6.16) and (6.17).

In the fixed-width confidence interval case one may sometimes deduce a limiting distribution for N'_d as $d \downarrow 0$ when s_n^2 is asymptotically normal. Similarly, in the present setting asymptotic normality of c_n^2 may provide a rate of convergence for N_d as $d \downarrow 0$.

Let $\{ g(n) \}$ be a sequence of positive numbers such that $g(n)$ is increasing in n, $g(n) \to \infty$ as $n \to \infty$, and

(6.18) $$n_d^{-1} g(n_d) \to 0 \quad \text{as } d \downarrow 0.$$

Suppose that

(6.19) $$g(n_d)\{ c_{n_d}^2 / c^2 - 1 \} \xrightarrow{D} N(0, \gamma^2) \quad \text{as } d \downarrow 0,$$

where γ is a finite positive constant. Further assume that

$$(6.20) \qquad \{ g(n)[c_n^2/c^2 - 1],\ n \geq n_o \} \text{ is a u.c.i.p. sequence.}$$

(See Section 6.2.) Together with C1 these three conditions yield the following result, similar in derivation to Theorem 10.2.2 of Sen(1981).

Lemma 6.2: Assume C1, (6.18), (6.19), and (6.20) hold. Then

$$g(n_d)\{ N_d/n_d - 1 \} \xrightarrow{D} N(0, \gamma^2) \text{ as } d \downarrow 0.$$

6.4 Bootstrap Quantile Estimators

One method that has been suggested for constructing confidence bands of the form (6.1) is bootstrapping the process r_n. Given $X_1, \ldots, X_n$, let $X_1^*, \ldots, X_{m_n}^*$ be conditionally independent with the distribution $F_n(x) = n^{-1}\Sigma_{i=1}^n \chi(X_i \leq x)$, $-\infty < x < \infty$, where the bootstrap sample size $m_n \to \infty$ as $n \to \infty$. (Here χ denotes the indicator function.) Next, for $t \in I$ define $R_{m_n}^*(t) = R_{m_n}(t;\ X_1^*,\ \ldots,\ X_{m_n}^*)$; the bootstrapped process is then given by

$$r_{m_n}^*(\cdot) = m_n^{1/2}\{R_{m_n}^*(\cdot) - R_n(\cdot)\}$$

on I. One then typically makes the assumption that $r_{m_n}^* \xrightarrow{W} \mathcal{G}_F$ as n, $m_n \to \infty$ so that a natural estimator of the distribution G_F is defined by

$$G_{n,F_n}(x) = \mathbf{P}\{ \sup_{t \in I} |r_{m_n}^*(t)| \leq x \mid X_1, \ldots, X_n \},\ -\infty < x < \infty.$$

The bootstrap estimator of c, the $(1-\alpha)$-th quantile of G_F, is then defined to be the $(1-\alpha)$-th quantile of G_{n,F_n}.

For computational reasons the bootstrap estimate is often itself approximated by generating conditionally independent copies of $r^*_{m_n}$ as follows. For $i = 1,\ldots, M_n$ let $X^*_{i1},\ldots, X^*_{im_n}$ be a random sample from the distribution F_n, where $M_n \to \infty$ as $n \to \infty$. For $t \in I$ define $R^*_{m_n i}(t) = R_{m_n}(t; X^*_{i1},\ldots, X^*_{im_n})$. Then the processes

$$r^*_{m_n i}(\cdot) = m_n^{1/2}\{R^*_{m_n i}(\cdot) - R_n(\cdot)\}$$

on I, $i = 1,\ldots, M_n$, are conditionally independent and identically distributed. Estimate G_{n,F_n} empirically by

$$\hat{G}_{n,F_n}(x) = M_n^{-1}\Sigma_{i=1}^{M_n} \chi(\sup_{t\in I} |r^*_{m_n i}(t)| \leq x), \ -\infty < x < \infty.$$

The bootstrap estimate of c is then taken to be $c^*_n = \inf\{x: \hat{G}_{n,F_n}(x) \geq 1-\alpha\}$. Putting $c_n = c^*_n$ in (6.5) then defines a stopping time for terminating the bootstrap procedure.

Due to the conditional nature of the ordinary bootstrap estimator, the feasibility of establishing unconditional properties of the stopping time may depend on the nature of the function R_F and the estimator R_n. Showing that the stopping time has finite expectation for fixed $d > 0$ or has the asymptotic efficiency property (6.9) by confirming that conditions such as (6.10), (6.11), or (6.12) hold may be difficult or impossible. To gain some insight into this problem we first examine a condition on the random distribution function G_{n,F_n} which is sufficient for (6.11) and (6.12) to hold. While this condition will not typically be useful in practice, it does suggest how certain properties of the stopping time depend on the probability distribution of G_{n,F_n} and might not automatically follow when one uses an ordinary bootstrap estimator. It is then seen that by partitioning the sample into independent groups of observations one can instead estimate a distribution analogous to $G_n = \mathbf{E} G_{n,F_n}$ by bootstrapping; if the resulting quantile estimate is used to define the stopping time, the desired properties will hold.

Condition C1 is replaced by the following.

C3: With previously defined notation, $r_n \overset{W}{\to} \mathcal{G}_F$ and $r^*_{m_n} \overset{W}{\to} \mathcal{G}_F$ on I as n, $m_n \to \infty$, where $\mathbf{P}\{ \sup_{t\in I} |\mathcal{G}_F(t)| < \infty \} = 1$. $G_F(c) = 1-\alpha$ and G_F is continuous and strictly increasing in a neighborhood of c. $\mathbf{P}\{ \hat{G}_{n,F_n}(0) = 0 \text{ for all } n \geq n_o \} = 1$. The sequence $\{M_n\}$ is such that $M_n^{-1} \log n = o(1)$.

C3 implies that $c^*_n \to c$ a.s. Let $G_n(x) = \mathbf{E}\, G_{n,F_n}(x) = \mathbf{P}\{ \sup_{t\in I} |r^*_{m_n}(t)| \leq x \}$, $-\infty < x < \infty$. Lemmas 6.3 and 6.4 assume that $|G_{n,F_n}(x) - G_n(x)| \to 0$ completely as $n \to \infty$ in the following sense.

C4: For every $x \in \mathbf{R}$ and every $\delta > 0$, $\sum_{n \geq n_o} \mathbf{P}\{ |G_{n,F_n}(x) - G_n(x)| > \delta \} < \infty$.

<u>Lemma 6.3:</u> Assume C3 and C4 hold. Then for fixed $d > 0$, $\mathbf{E} N_d < \infty$.

<u>Proof:</u> For all $n \geq n_o$, $\mathbf{P}\{N_d > n\} \leq \mathbf{P}\{n^{-1/2} c^*_n > d\}$, so that

$$\sum_{n \geq n_o} \mathbf{P}\{ N_d > n \} \leq \sum_{n \geq n_o} \mathbf{P}\{ c^*_n > n^{1/2} d \}. \tag{6.21}$$

Since $G_n \overset{D}{\to} G_F$, for all n sufficiently large one has that $[1 - \alpha - G_n(n^{1/2} d)] < -\delta$ for some $\delta > 0$. Thus for large n

$$\mathbf{P}\{ c^*_n > n^{1/2} d \} = \mathbf{P}\{ \hat{G}_{n,F_n}(n^{1/2} d) < 1 - \alpha \}$$

$$\leq \mathbf{P}\left\{ \hat{G}_{n,F_n}(n^{1/2} d) - G_{n,F_n}(n^{1/2} d) < -\delta + G_n(n^{1/2} d) - G_{n,F_n}(n^{1/2} d) \right\}$$

$$\leq \mathbf{P}\left\{ \mid \hat{G}_{n,F_n}(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) \mid > \delta/2 \right\}$$

$$+ \mathbf{P}\left\{ \mid G_n(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) \mid > \delta/2 \right\}.$$

By a standard probability inequality for empirical distribution functions

$$\mathbf{P}\left\{ \sup_{x \in \mathbf{R}} \mid \hat{G}_{n,F_n}(x) - G_{n,F_n}(x) \mid > \delta/2 \right\} \leq C e^{-2 M_n (\delta/2)^2} \tag{6.22}$$

where C is a constant not depending on F_n or F. (See Serfling(1980), Section 2.1.3.) Hence by (6.21), C3, and C4

$$\sum_{n \geq n_o} \mathbf{P}\{ N_d > n \}$$

$$\leq \sum_{n \geq n_o} C e^{-2 M_n (\delta/2)^2} + \sum_{n \geq n_o} \mathbf{P}\left\{ \mid G_n(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) \mid > \delta/2 \right\}$$

$$< \infty,$$

which is equivalent to $\mathbf{E} N_d < \infty$. □

<u>Lemma 6.4:</u> Assume C3 and C4 hold. Then $\lim_{d \downarrow 0} n_d^{-1} \mathbf{E} N_d = 1$.

<u>Proof:</u> To show that (6.17) holds, we derive a bound for $\mathbf{P}\{ N_d > n \}$ which is independent of d for $n \geq n_d(1+\epsilon)$. For any $\delta > 0$ and all d sufficiently small, using the fact that $n \geq c^2(1+\epsilon)/d^2$,

$$\mathbf{P}\{\, c_n^* > n^{1/2}d \,\} \;=\; \mathbf{P}\Big\{\, \hat{G}_{n,F_n}(n^{1/2}d) < 1-\alpha \,\Big\}$$

$$\le\; \mathbf{P}\Big\{\, \hat{G}_{n,F_n}(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) < 1-\alpha - G_{n,F_n}((1+\epsilon)^{1/2}c)\Big\}$$

$$\le\; \mathbf{P}\Big\{\, \hat{G}_{n,F_n}(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) < -\delta\Big\} + \mathbf{P}\Big\{1-\alpha+\delta > G_{n,F_n}((1+\epsilon)^{1/2}c)\Big\}.$$

By C3, $G_n \xrightarrow{D} G_F$ and one can choose δ, $\delta'>0$ such that $(1-\alpha+\delta) - G_n((1+\epsilon)^{1/2}c) < -\delta'$ for all large n. For this δ and sufficiently small d, by C4,

$$\sum_{n \ge n_d(1+\epsilon)} \mathbf{P}\Big\{\, 1-\alpha+\delta > G_{n,F_n}((1+\epsilon)^{1/2}c) \,\Big\}$$

$$\le \sum_{n \ge n_d(1+\epsilon)} \mathbf{P}\Big\{\, |\, G_{n,F_n}((1+\epsilon)^{1/2}c) \;-\; G_n((1+\epsilon)^{1/2}c) \,|\, > \delta' \,\Big\}$$

$$< \infty$$

and using (6.22)

$$\sum_{n \ge n_d(1+\epsilon)} \mathbf{P}\{\, N_d > n \,\} \;\le\; \sum_{n \ge n_d(1+\epsilon)} \mathbf{P}\{\, c_n^* > n^{1/2}d \,\} \;<\; \infty.$$

Thus (6.17) holds and the result follows. □

For some statistical functions R_F and corresponding estimators R_n, if F_n is in some sense close to F then G_{n,F_n} may be close to G_n. This suggests the following condition which is sufficient for C4.

C5: [Hadamard-Lipschitz condition of order α.] There exists a real-valued function $K(x)$ such that for all n and $x \in \mathbf{R}$

$$|G_{n,F_n}(x) - G_n(x)| \le K(x)\,||F_n - F||^{\alpha} \text{ for some } \alpha > 0,$$

where $||\, f \,||$ denotes the supremum norm of the function f.

The inequality applied in (6.22) to $\hat{G}_{n,F_n}$ and G_{n,F_n} can now be applied to F_n and F to show that C5 implies C4:

$$\begin{aligned}\sum_{n \ge n_o} \mathbf{P}\{\, |\, G_{n,F_n}(x) - G_n(x)\,| > \delta \,\} &\le \sum_{n \ge n_o} \mathbf{P}\Big\{\, ||F_n - F|| > [\delta/K(x)]^{1/\alpha} \Big\} \\ &\le \sum_{n \ge n_o} C e^{-2n[\delta/K(x)]^{2/\alpha}} \\ &< \infty.\end{aligned}$$

While C5 seems a more natural condition than C4, in many situations neither of these assumptions will be verifiable. However, by modifying the procedure so that one unbiasedly estimates an unconditional distribution function analogous to G_n the desired properties for the stopping time will automatically follow.

In introducing the modified bootstrap procedure some notation is redefined. Now let $\{m_n\}$ and $\{M_n\}$ be sequences of positive integers such that $m_n \to \infty$ and $M_n \to \infty$ as $n \to \infty$ and $M_n m_n \le n$ for all n. Divide the sample into M_n groups of m_n observations each; denote the observations in the i-th group by $X_{i1}, \ldots, X_{im_n}$, $i = 1, \ldots, M_n$. One bootstrapped process will be generated from each of the M_n groups of observations; for convenience the bootstrap sample size will be taken to be m_n. For $i = 1, \ldots, M_n$ let $X^*_{i1}, \ldots, X^*_{im_n}$ now be a random sample drawn from the empirical distribution $F_{ni}(x) = m_n^{-1} \Sigma_{j=1}^{m_n} \chi(\, X_{ij} \le x\,)$, $-\infty < x < \infty$. For $t \in I$ define $R_{m_n i}(t) = R_{m_n}(t; X_{i1}, \ldots, X_{im_n})$ and then redefine

$$R^*_{m_n i}(t) = R_{m_n}(t;\, X^*_{i1},\ldots, X^*_{im_n})$$

Then the modified processes

$$r^*_{m_n i}(\cdot) = m_n^{1/2}\{R^*_{m_n i}(\cdot) - R_{m_n i}(\cdot)\}$$

on I, $i = 1,\ldots, M_n$, are unconditionally independent and identically distributed. Now redefine G_n by

$$G_n(x) = \mathbf{P}\{\ \sup_{t\in I} |r^*_{m_n i}(t)| \le x\ \},\ -\infty < x < \infty.$$

G_n can now be estimated in an unbiased fashion by

$$\hat{G}_n(x) = M_n^{-1}\Sigma_{i=1}^{M_n}\ \chi(\ \sup_{t\in I} |r^*_{m_n i}(t)| \le x\ ),\ -\infty < x < \infty.$$

C3 must be changed slightly.

C6: With previously defined notation, $r_n \overset{W}{\rightarrow} \mathcal{G}_F$ and $r^*_{m_n 1} \overset{W}{\rightarrow} \mathcal{G}_F$ on I as n, $m_n \rightarrow \infty$, where $\mathbf{P}\{\ \sup_{t\in I} |\mathcal{G}_F(t)| < \infty\ \} = 1$. $G_F(c) = 1-\alpha$ and G_F is continuous and strictly increasing in a neighborhood of c. $\mathbf{P}\{\ \hat{G}_n(0) = 0$ for all $n \ge n_o\ \} = 1$. The sequence $\{M_n\}$ is such that $M_n^{-1}\log n = o(1)$.

Let $c_n^* = \inf\{x\colon\ G_n(x) \ge 1-\alpha\}$ for which a natural estimate is $\hat{c}_n^* = \inf\{x\colon\ \hat{G}_n(x) \ge 1-\alpha\}$. By a standard probability inequality for deviations of sample quantiles, for any $\epsilon > 0$

$$\mathbf{P}\{\ |\ \hat{c}_n^* - c_n^*\ | > \epsilon\ \} \le 2e^{-2M_n\delta_n^2},$$

where $\delta_n = \min\{\ G_n(c_n^*+\epsilon) - (1-\alpha),\ (1-\alpha) - G_n(c_n^*-\epsilon)\ \}$. (See, for example, Serfling(1980), p.75.) Using the fact that $G_n \xrightarrow{D} G_F$ and G_F is continuous and strictly increasing in a neighborhood of c, it follows that $\liminf_{n\to\infty} \delta_n > 0$. If $M_n^{-1} \log n = o(1)$, then one has that $| \hat{c}_n^* - c_n^* | \to 0$ completely as $n \to \infty$ in the sense that for any $\epsilon > 0$, $\Sigma_{n=n_o}^{\infty} \mathbf{P}\{ | \hat{c}_n^* - c_n^* | > \epsilon \} < \infty$; consequently $\hat{c}_n^* \to c$ completely and a.s. as $n \to \infty$.

Now let the stopping time N_d be defined with $c_n = \hat{c}_n^*$. When C6 holds, one then has the properties indicated in Lemmas 6.3 and 6.4. This can be seen by replacing $\hat{G}_{n,F_n}$ by $\hat{G}_n$ and G_{n,F_n} by the new G_n in the proofs of Lemmas 6.3 and 6.4 and then making the appropriate simplifications.

The modified bootstrap algorithm results in essentially a grouped sequential procedure when used in conjunction with the stopping time (6.5). Given the sequences $\{M_n\}$ and $\{m_n\}$, the effective sample sizes are defined by the sequence $\{M_n m_n\}$. Even if the sample were increased a single observation at a time, one usually would not want to repeat the bootstrap procedure until either the number of subsamples M_n, the subsample size m_n, or both had been increased. Thus the stopping time would only take values in the sequence $\{M_n m_n\}$. This seems a disadvantageous feature.

Another point bears on the selection of M_n and m_n subject to the constraint $M_n m_n \leq n$. The number of subsamples M_n need only grow at a rate slightly faster than log n for complete convergence of $\hat{c}_n^*$ to c_n^*, so it seems prudent to allow the subset size m_n to increase relatively quickly to speed convergence of c_n^* to c. One might let M_n be of order n^λ for some λ, $0 < \lambda < 1$, where typically λ would not be greater than 1/2 and could be decreased as n becomes larger.

6.5 Variable-Width Confidence Bands

In some applications the need for a confidence band of variable width arises quite naturally. If the variance of $r_n(t)$ is large for some values of t and comparatively small for other values, then a confidence band of constant width such as (6.1) or (6.6) will often be unnecessarily wide for some values of t. This suggests scaling the confidence band width at a point t by an estimate of the standard deviation of the process at that point. (It should be noted that this may actually increase the width of the confidence band at some points.) In other applications a modified process of the form $\{r_n(t)/q(t),\ t \in I\}$ is known to converge weakly, where q is a known nonnegative weight function. These two situations are treated jointly here, since both lead to essentially the same modifications of the stopping time (6.5) and the confidence band (6.6).

Let $\{q_n\}$ be a sequence of possibly random nonnegative functions on I and q be a possibly unknown nonnegative function on I. Next suppose $J \subseteq I$ is a finite union of intervals and define the processes $\tilde{r}_n = \{\, r_n(t)/q_n(t),\ t \in J \,\}$, $n = 1, 2, \ldots$. Assume that $\tilde{r}_n \xrightarrow{W} \tilde{\mathcal{G}}_F$, where $\tilde{\mathcal{G}}_F$ is a zero mean Gaussian process on J such that $\mathbf{P}\{\, \sup_{t \in J} |\tilde{\mathcal{G}}_F(t)| < \infty \,\} = 1$.

In the case where $r_n \xrightarrow{W} \mathcal{G}_F$ on I and $q(t) = \mathbf{Var}\{\, \mathcal{G}_F(t) \,\}$, $J \subseteq I$ is necessarily such that q is bounded away from zero and infinity on J; if additionally q_n is a uniformly consistent estimator of q on J, then $\tilde{\mathcal{G}}_F(t) \overset{D}{=} \{\, \mathcal{G}_F(t)/\sigma(t),\ t \in J \,\}$. Alternatively, one might have that $q_n \equiv q$ for all n, where q is a known weight function that provides an appropriate normalization of r_n.

Set

$$\tilde{G}_F(x) = \mathbf{P}\{ \sup_{t \in J} |\tilde{\mathcal{G}}_F(t)| \le x \},\ -\infty < x < \infty,$$

and suppose that $\tilde{c}$ is such that $\tilde{G}_F(\tilde{c}) = 1-\alpha$ and $\tilde{G}_F$ is continuous at $\tilde{c}$. Given a weakly consistent estimator $\tilde{c}_n$ of $\tilde{c}$, a reasonable large sample confidence band for R_F on J is then given by

(6.23) $$\{ R_n(t) \pm \tilde{c}_n n^{-1/2} q_n(t),\ t \in J \}.$$

With the preceding assumptions,

(6.24) $$\mathbf{P}\Big\{ R_n(t) - \tilde{c}_n n^{-1/2} q_n(t) \le R_F(t) \le R_n(t) + \tilde{c}_n n^{-1/2} q_n(t),\ t \in J \Big\}$$

$$= \mathbf{P}\{ \sup_{t \in J} |\tilde{r}_n(t)| \le \tilde{c}_n \}$$

$$\to 1-\alpha.$$

Although the goal is no longer a fixed-width confidence band, by scaling the function q_n by a constant $d>0$ a stopping time can be introduced; the constant d is then let go to zero in order to study the asymptotic coverage probability. Assume $\tilde{c}_n \to \tilde{c}$ a.s. and define the stopping time

$$\tilde{N}_d = \inf\Big\{ n \ge n_o\colon n^{-1/2} \tilde{c}_n \le d \Big\},$$

where n_o is again an initial sample size. Let

(6.25) $$\{ R_{\tilde{N}_d}(t) \pm d\, q_{\tilde{N}_d}(t),\ t \in J \}$$

be the associated confidence band.

It follows from the strong consistency of $\tilde{c}_n$ that if $\mathbf{P}\{ \tilde{c}_n \le 0 \} = 0$ for all $n \ge n_o$, then $\tilde{c}_{\tilde{N}_d} \to \tilde{c}$ a.s. as $d \downarrow 0$. By a relation like (6.7), one then has that $[\tilde{N}_d]^{1/2} d \to \tilde{c}$ as

$d \downarrow 0$. Thus with the additional condition that $\tilde{r}_{\tilde{N}_d} \xrightarrow{W} \tilde{g}_F$ as $d \downarrow 0$,

$$(6.26) \qquad \lim_{d \downarrow 0} \mathbf{P}\left\{ |R_{\tilde{N}_d}(t) - R_F(t)| \le d\, q_{\tilde{N}_d}(t),\ t \in J \right\}$$

$$= \lim_{d \downarrow 0} \mathbf{P}\left\{ \sup_{t \in J} |\tilde{r}_{\tilde{N}_d}(t)| \le [\tilde{N}_d]^{1/2} d \right\}$$

$$= 1 - \alpha$$

by (6.24). Thus one can define a meaningful stopping time for the confidence band (6.23).

Letting $\tilde{n}_d = \inf\{ n \ge n_o : n^{-1/2} \tilde{c} \le d \}$, as for the stopping time N_d one has that $\tilde{N}_d$ is finite a.s. for fixed $d > 0$ and $\tilde{N}_d / \tilde{n}_d \to 1$ a.s. as $d \downarrow 0$. Other properties of the stopping time such as asymptotic efficiency and normality follow under conditions on $\tilde{N}_d$ and $\tilde{c}_n$ identical to those considered for N_d and c_n in Lemmas 3.1 and 3.2.

6.6 Example: Empirical Processes

Let $\underset{\sim}{X}, \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ be i.i.d. random vectors having continuous distribution function F defined on $\mathbf{R}^p$ for some $p \ge 1$. Define the empirical distribution function by

$$F_n(\underset{\sim}{x}) = n^{-1} \Sigma_{i=1}^n\, c(\underset{\sim}{x} - \underset{\sim}{X}_i), \quad \underset{\sim}{x} \in \mathbf{R}^p,$$

where $c(\underset{\sim}{u})$ is equal to 1 if all p coordinates of $\underset{\sim}{u}$ are nonnegative and is otherwise equal to 0. Letting $V_n(\underset{\sim}{x}) = n^{1/2}[F_n(\underset{\sim}{x}) - F(\underset{\sim}{x})]$, the usual empirical process is then defined by $V_n = \{ V_n(\underset{\sim}{x}), \underset{\sim}{x} \in \mathbf{R}^p \}$. Let $E^p = [0, 1]^p$ be the p-dimensional unit cube. With $\underset{\sim}{x}' = (x_1, \ldots, x_p)$ and $F_1, \ldots, F_p$ the marginal distributions of F, define the reduced empirical process by

$$U_n(\underset{\sim}{x}) = V_n(F_1^{-1}(x_1), \ldots, F_p^{-1}(x_p)), \ \underset{\sim}{x} \in E^p.$$

Since F is continuous, $U_n \overset{W}{\to} U$, where U is a tied-down zero-mean Gaussian process on E^p. (See Bickel and Wichura(1971).) Thus $\mathbf{P}\{\sup_{x \in E^p} |U(\underset{\sim}{x})| < \infty\} = 1$. For the remainder of this section, let

$$G_F(t) = \mathbf{P}\{\sup_{x \in E^p} |U(\underset{\sim}{x})| \leq t\}, \ -\infty < t < \infty.$$

If c is such that $G_F(c) = 1-\alpha$, then G_F is continuous and strictly increasing in a neighborhood of c.

As is well known, if $p=1$ then $U_n \overset{W}{\to} B$ as $n \to \infty$, where B is the Brownian bridge on $[0, 1]$. Thus when F is continuous G_F is the distribution of the supremum of the Brownian bridge. Bickel and Freedman (1981) have shown that bootstrapping U_n yields a strongly consistent estimate of c, the $(1-\alpha)$th quantile of G_F. Letting c_n be the bootstrap estimate and $r_n(\cdot) = U_n(\cdot)$ on $[0, 1]$, define the stopping time N_d by (6.5) and the associated confidence band by (6.6).

Now let $p \geq 1$. When $p \geq 2$ the corresponding G_F is usually unknown unless the p coordinates of $\underset{\sim}{X}$ are independent. Again suppose c is the $(1-\alpha)$th quantile of G_F and use the strongly consistent bootstrap estimator discussed by Beran(1984) to define the stopping time (1.5). Upon stopping, use the confidence region

$$\{ F_{N_d}(\underset{\sim}{x}) \pm d, \ \underset{\sim}{x} \in \mathbf{R}^p \} \tag{6.27}$$

for the distribution F. It then follows that if $U_{N_d} \overset{W}{\to} U$ as $d \downarrow 0$, then

(6.28) $$P\left\{ F_{N_d}(\underset{\sim}{x}) - d \leq F(\underset{\sim}{x}) \leq F_{N_d}(\underset{\sim}{x}) + d, \ \underset{\sim}{x} \in \mathbb{R}^P \right\}$$

$$= P\{ \sup_{\underset{\sim}{x} \in E^p} |U_{N_d}(\underset{\sim}{x})| \leq N_d^{1/2} d \}$$

$$\rightarrow 1-\alpha$$

as $d \downarrow 0$ by the assumptions and the inequality (6.7).

Bickel and Wichura (1971) note that if $\{N_n\}$ is a sequence of positive integer-valued random variables, $c_n \rightarrow \infty$, and N_n/c_n converges in probability to a positive random variable, then $U_{N_n} \overset{W}{\rightarrow} U$. Thus $U_{N_d} \overset{W}{\rightarrow} U$ since $N_d/n_d \rightarrow 1$ a.s. as $d \downarrow 0$. Alternatively, weak convergence could have been shown by using the process Z_n considered in Section 6.2 and the method considered there; weak convergence of Z_n in this instance has already been confirmed. (See the remarks following Theorem 6 of Bickel and Wichura(1971).)

In this chapter we have defined stopping times for several types of confidence band procedures and considered various properties of these stopping times. That these sequential procedures achieve the correct coverage probability asymptotically was seen in Sections 6.2 and 6.5. The multivariate empirical process example suggests how these sequential methods can be readily extended to confidence regions in p-dimensional Euclidean space, allowing a more general treatment than considered here.

APPENDIX

This appendix concerns the weak convergence of the bootstrapped process Y^*_{n1} defined in Section 5.4.

<u>Theorem</u> <u>AP1</u>: Assume B9 and B11 hold. Then $Y^*_{n1} \overset{W}{\to} Y$ as $n \to \infty$, where $Y = \{Y(t), t \in [0,T]\}$ is the zero-mean Gaussian process defined by (5.2).

The proof of Theorem AP1 may be found at the end of this appendix. As in Section 5.3, weak convergence will follow from asymptotic normality of the finite dimensional distributions of Y^*_{n1} (Theorem AP2) and a condition sufficient for tightness (Lemma AP3).

Assume the original random variables X_1, X_2, ... are defined on a probability space $(\Omega, \mathcal{F}, P)$ and fix $\omega \in \Omega$. By the definition of f_n, B9, and B11, there exists a constant $L(\omega)$ such that $f_n(t) \leq L(\omega)$ for all $t \leq T$, for all $n \geq 1$. This implies that for fixed $\omega \in \Omega$ there exist positive constants K and ρ, $0 < \rho < 1$, such that for i=1, 2, ... , and n=1, 2, ... ,

$$\text{(AP1)} \qquad f_n^{(i)}(t) \leq K\rho^i \text{ for all } t \leq T.$$

Here $f_n^{(i)}$ is the i-th convolution of f_n. See, for example, the proof of Theorem 2.1.3 of Wold(1981) for verification of A1. Thus for fixed $\omega \in \Omega$, i=1, 2, ... , and n=1, 2, ... ,

$$\text{(AP2)} \qquad F_n^{(i)}(t) \leq KT\rho^i \text{ for all } t \leq T.$$

To establish Theorem A1, it is shown that for fixed $\omega \in \Omega$, $Y^*_{n1} \overset{W}{\rightarrow} Y$; in order to accomplish this, frequent use is made of AP1 and AP2.

Asymptotic Normality of Finite-Dimensional Distributions

For arbitrary positive integer ℓ, let $\underset{\sim}{t} = (t_1,\ldots,t_\ell)'$, where $t_1,\ldots,t_\ell$ are fixed nonnegative real numbers. The Cramér-Wold device is used to show that $[\tilde{\tilde{H}}_{n1}(t_1),\ldots,\tilde{\tilde{H}}_{n1}(t_\ell)]$, suitably standardized, has asymptotically a multivariate normal distribution. Thus, let $\underset{\sim}{a} = (a_1,\ldots,a_\ell)'$ be any ℓ-vector of real numbers. Redefine

$$D_n(\underset{\sim}{t}) = \sum_{i=1}^{\ell} a_i \tilde{\tilde{H}}_{n1}(t_i)$$

and

$$D^*_n(\underset{\sim}{t}) = \sum_{i=1}^{\ell} a_i \sum_{k=1}^{m} \tilde{\tilde{F}}_n^{(k)}(t_i).$$

For $1 \le c \le r \vee s$, $i,j=1, \ldots, \ell$, let

$$\xi_{nrsij}(c) = \mathbf{Cov}_{\tilde{\tilde{F}}_n} \{ \tilde{\tilde{F}}_n^{(r-c)}(t_i - (X^*_{11} + \cdots + X^*_{1c})), \tilde{\tilde{F}}_n^{(s-1)}(t_j - (X^*_{11} + \cdots + X^*_{1c})) \}$$

and

$$\xi_{nrs\underset{\sim}{t}}(c) = \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} a_i a_j \xi_{nrsij}(c).$$

For all n,

(AP3) $$| \xi_{nrsij}(c) | \le \min(\tilde{\tilde{F}}_n^{(r)}(t_i), \tilde{\tilde{F}}_n^{(s)}(t_j))$$

$$\le \tilde{\tilde{F}}_n^{(r \wedge s)}(T)$$

$$\le K T \rho^{(r \wedge s)} .$$

Thus for some constant K

$$\sum_{r,s=1}^{\infty} rs\,|\xi_{nrs\underset{\sim}{t}}(1)| \le K$$

for all $n \ge 1$. Since $\xi_{nrs\underset{\sim}{t}}(1) \to \xi_{rs\underset{\sim}{t}}(1)$ a.s., by dominated convergence

$$\sum_{r,s=1}^{\infty} rs\,\xi_{nrs\underset{\sim}{t}}(1) \to \sigma^2_{\underset{\sim}{t}} \text{ a.s.}$$

<u>Theorem</u> <u>AP2</u>: Assume B9 and B11 hold. Then

$$n^{1/2}[D_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})] \xrightarrow{D} N(0, \sigma^2_{\underset{\sim}{t}})$$

where $\sigma^2_{\underset{\sim}{t}} = \Sigma_{r,s=1}^{\infty} rs\,\xi_{rs\underset{\sim}{t}}(1)$, and thus the finite-dimensional distributions of Y^*_{n1} converge weakly to the finite-dimensional distributions of Y.

The proof of Theorem AP2 is begun by calculating the projection $D_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})$ onto the individual observations $X^*_{11}, \dots, X^*_{1n}$. Let

$$B_{n1}^{(k)}(\underset{\sim}{t}) = \sum_{i=1}^{\ell} a_i \tilde{F}_{n1}^{(k)}(t_i)$$

and

$$B_n^{(k)}(\underset{\sim}{t}) = \sum_{i=1}^{\ell} a_i \tilde{F}_n^{(k)}(t_i).$$

By the definition of $\tilde{F}_n^{(k)}(t)$ we have

$$\mathbf{E}\left(\tilde{F}_{n1}^{(k)}(t)\,|\,X^*_{11}\right) = (k/n)\,\tilde{F}_n^{(k-1)}(t - X^*_{11}) + (1 - k/n)\,\tilde{F}_n^{(k)}(t).$$

Again using the convention that $(\underset{\sim}{t} - X)' = (t_1 - X, \dots, t_\ell - X)$, it follows that

$$\mathbf{E}\left(B_{n1}^{(k)}(\underset{\sim}{t})\,|\,X_{11}^*\right) = \sum_{i=1}^{\ell} a_i\left[(k/n)\,\tilde{F}_n^{(k-1)}(t_i - X_{11}^*) + (1-k/n)\,\tilde{F}_n^{(k)}(t_i)\right]$$

$$= (k/n)\,B_n^{(k-1)}(\underset{\sim}{t} - X_{11}^*) + (1-k/n)\,B_n^{(k)}(\underset{\sim}{t}).$$

Define the projection

$$\hat{D}_n(\underset{\sim}{t}) = \left\{\sum_{j=1}^{n} \mathbf{E}\left(D_n(\underset{\sim}{t})\,|\,X_{1j}^*\right)\right\} - (n-1)\,D_n^*(\underset{\sim}{t})$$

so that

$$\hat{D}_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t}) = n^{-1}\sum_{j=1}^{n}\sum_{k=1}^{m} k\left\{B_n^{(k-1)}(\underset{\sim}{t} - X_{1j}^*) - B_n^{(k)}(\underset{\sim}{t})\right\}.$$

We next calculate $\mathbf{Var}_{\tilde{F}_n}\{D_n(\underset{\sim}{t})\}$, $\mathbf{Var}_{\tilde{F}_n}\{\hat{D}_n(\underset{\sim}{t})\}$, and $\mathbf{Cov}_{\tilde{F}_n}\{D_n(\underset{\sim}{t}), \hat{D}_n(\underset{\sim}{t})\}$.

By definition,

$$\mathbf{Cov}_{\tilde{F}_n}\left\{B_n^{(r-c)}(\underset{\sim}{t} - (X_{11}^* + \dots + X_{1c}^*)),\ B_n^{(s-c)}(\underset{\sim}{t} - (X_{11}^* + \dots + X_{1c}^*))\right\}$$

$$= \mathbf{Cov}_{\tilde{F}_n}\left\{\sum_{i=1}^{\ell} a_i\,\tilde{F}_n^{(r-c)}(t_i - (X_{11}^* + \dots + X_{1c}^*)),\ \sum_{j=1}^{\ell} a_j\,\tilde{F}_n^{(s-c)}(t_j - (X_{11}^* + \dots + X_{1c}^*))\right\}$$

$$= \xi_{nrs\underset{\sim}{t}}(c).$$

Thus,

$$\mathbf{Var}_{\tilde{F}_n}\left\{\hat{D}_n(\underset{\sim}{t})\right\} \sum = n^{-1}\,\mathbf{Var}_{\tilde{F}_n}\left\{\sum_{k=1}^{m} k\,[B_n^{(k-1)}(\underset{\sim}{t} - X_{11}^*) - B_n^{(k)}(\underset{\sim}{t})]\right\}$$

$$\sum = n^{-1}\sum_{r,s=1}^{m} rs\,\xi_{nrs\underset{\sim}{t}}(1).$$

By a direct calculation

$$\mathbf{Var}_{\tilde{F}_n}\left\{D_n(\underset{\sim}{t})\right\} = \sum_{r,s=1}^{m} \mathbf{Cov}_{\tilde{F}_n}\left\{B_{n1}^{(r)}(\underset{\sim}{t}),\, B_{n1}^{(s)}(\underset{\sim}{t})\right\}$$

$$= \sum_{r,s=1}^{m} \sum_{i,j=1}^{\ell} a_i a_j \binom{n}{r}^{-1} \sum_{c=1}^{r} \binom{s}{c}\binom{n-s}{r-c} \xi_{nrsij}(c).$$

Finally,

$$\mathbf{Cov}_{\tilde{F}_n}\left\{D_n(\underset{\sim}{t}), \hat{D}_n(\underset{\sim}{t})\right\}$$

$$= \mathbf{Cov}_{\tilde{F}_n}\left\{\sum_{r=1}^{m} B_{n1}^{(r)}(\underset{\sim}{t}),\, n^{-1}\sum_{j=1}^{n}\ \sum_{s=1}^{m} s[B_n^{(s-1)}(\underset{\sim}{t} - X_{1j}^*) - B_n^{(s)}(\underset{\sim}{t})]\right\}$$

$$= \sum_{r,s=1}^{m} s\ \mathbf{Cov}_{\tilde{F}_n}\left\{B_{n1}^{(r)}(\underset{\sim}{t}),\, B_n^{(s-1)}(\underset{\sim}{t} - X_{11}^*)\right\}.$$

Since

$$\mathbf{Cov}_{\tilde{F}_n}\left\{B_{n1}^{(r)}(\underset{\sim}{t}),\, B_n^{(s-1)}(\underset{\sim}{t} - X_{11}^*)\right\}$$

$$= \binom{n-1}{r-1}\binom{n}{r}^{-1} \mathbf{Cov}_{\tilde{F}_n}\left\{\sum_{i=1}^{\ell} a_i\, \chi(X_{12}^* + \dots + X_{1r}^* \le t_i - X_{11}^*),\, \sum_{j=1}^{\ell} a_j \tilde{F}_n^{(s-1)}(t_j - X_{11}^*)\right\}$$

$$= n^{-1} r\, \xi_{nrs\underset{\sim}{t}}(1),$$

we can conclude that

$$\mathbf{Cov}_{\tilde{F}_n}\left\{D_n(\underset{\sim}{t}),\, \hat{D}_n(\underset{\sim}{t})\right\} = n^{-1} \sum_{r,s=1}^{m} rs\, \xi_{nrs\underset{\sim}{t}}(1).$$

<u>Lemma</u> <u>AP1</u>: Assume B9 and B11 hold. Then

$$n^{1/2}[\hat{D}_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})] \xrightarrow{D} N(0, \sigma_{\underset{\sim}{t}}^2).$$

Proof: Let $X^*_{nj} = n^{-1/2}\Sigma_{k=1}^{m} k\left\{ B_n^{(k-1)}(\underset{\sim}{t} - X^*_{1j}) - B_n^{(k)}(\underset{\sim}{t})\right\}$, $j=1,\ldots,n$. By definition,

$$n^{1/2}[\hat{D}_n(\underset{\sim}{t}) - D^*_n(\underset{\sim}{t})] = \sum_{j=1}^{n} X^*_{nj}.$$

Also, $\mathbf{E}_{\tilde{F}_n} X^*_{nj} = 0$, $j=1,\ldots,n$, and

$$\mathcal{I}_n^2 = \mathbf{Var}_{\tilde{F}_n}\left\{\sum_{j=1}^{n} X^*_{nj}\right\} = \sum_{r,s=1}^{m} rs\,\xi_{nrs\underset{\sim}{t}}(1) \rightarrow \sigma^2_{\underset{\sim}{t}}.$$

For $\epsilon > 0$ let

$$S_n = \left\{u: |n^{-1/2}\sum_{k=1}^{m} k[B_n^{(k-1)}(\underset{\sim}{t}-u) - B_n^{(k)}(\underset{\sim}{t})]| > \epsilon\mathcal{I}_n\right\}.$$

We verify the Lindebergh condition

$$\lim_{n\to\infty} \sum_{j=1}^{n} \mathbf{E}_{\tilde{F}_n}\left\{(X^*_{nj})^2 \chi(|X^*_{nj}| > \epsilon\mathcal{I}_n)\right\}/\mathcal{I}_n^2$$

$$= \lim_{n\to\infty} \int_{S_n} \left\{\sum_{k=1}^{m} k[B_n^{(k-1)}(\underset{\sim}{t}-u) - B_n^{(k)}(\underset{\sim}{t})]\right\}^2 d\tilde{F}_n(u)$$

$$= 0.$$

If $\mathcal{I}_n^2 \rightarrow \sigma^2_{\underset{\sim}{t}} = 0$, then $n^{1/2}[\hat{D}_n(\underset{\sim}{t}) - D^*_n(\underset{\sim}{t})]$ converges in mean square and hence in distribution to the degenerate random variable which is identically 0. Thus assume $\mathcal{I}_n^2 \rightarrow \sigma^2_{\underset{\sim}{t}} > 0$. Suppose $t_\ell = \max\{t_1, \ldots, t_\ell\}$. Then

$$|\sum_{k=1}^{m} k[B_n^{(k-1)}(\underset{\sim}{t}-u) - B_n^{(k)}(\underset{\sim}{t})]| \le \sum_{i=1}^{\ell} a_i \sum_{k=1}^{m} k[\tilde{F}_n^{(k-1)}(t_\ell) + \tilde{F}_n^{(k)}(t_\ell)]$$
$$\le K,$$

say, for all $n \geq 1$ by AP2. Hence there exists n_o such that for $n \geq n_o$

$$S_n = \left\{u: |\sum_{k=1}^{m} k[B_n^{(k-1)}(\underset{\sim}{t}-u) - B_n^{(k)}(\underset{\sim}{t})]| > n^{1/2}\epsilon\mathcal{I}_n\right\}$$

is identically the empty set. Thus the Lindebergh condition holds. The lemma follows from a standard array central limit theorem. (See Serfling(1980), Section 1.9.2.) □

Lemma AP2: Assume B9 and B11 hold. Then

$$n\,\mathbf{E}[D_n(\underset{\sim}{t})-\hat{D}_n(\underset{\sim}{t})]^2 \to 0.$$

Proof: By prior calculations,

$$n\,\mathbf{E}[D_n(\underset{\sim}{t})-\hat{D}_n(\underset{\sim}{t})]^2 = n\,\mathbf{Var}\{D_n(\underset{\sim}{t})\} + n\,\mathbf{Var}\{\hat{D}_n(\underset{\sim}{t})\} - 2n\,\mathbf{Cov}\{D_n(\underset{\sim}{t}), \hat{D}_n(\underset{\sim}{t})\}$$

$$= \sum_{r,s=1}^{m} \left\{ n\binom{n}{r}^{-1} \sum_{c=1}^{s} \binom{s}{c}\binom{n-s}{r-c} \xi_{nrs\underset{\sim}{t}}(c) - rs\,\xi_{nrs\underset{\sim}{t}}(1) \right\}.$$

It is shown using dominated convergence that

(AP4) $$n \sum_{r,s=1}^{m} \binom{n}{r}^{-1} \sum_{c=2}^{s} \binom{s}{c}\binom{n-s}{r-c} \xi_{nrs\underset{\sim}{t}}(c) \to 0$$

and

(AP5) $$\sum_{r,s=1}^{m} \left\{ n\binom{n}{r}^{-1} \binom{s}{1}\binom{n-s}{r-1} - rs \right\} \xi_{nrs\underset{\sim}{t}}(1) \to 0,$$

which will imply the result.

First, $n\binom{n}{r}^{-1} \sum_{c=2}^{s} \binom{s}{c}\binom{n-s}{r-c} < rs$. Therefore, by AP3, there exists a constant K independent of n such that

$$\left| \sum_{r,s=1}^{m} \left\{ n\binom{n}{r}^{-1} \sum_{c=2}^{r} \binom{s}{c}\binom{n-s}{r-c} \xi_{nrs\underset{\sim}{t}}(c) \right\} \right|$$

$$\leq \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} |a_i|\,|a_j| \sum_{r,s=1}^{\infty} rs \min\left(\tilde{\bar{F}}_n^{(r)}(t_i), \tilde{\bar{F}}_n^{(s)}(t_j)\right)$$

$$\leq K.$$

For fixed r and s with $c \geq 2$, $n\binom{n}{r}^{-1}\binom{s}{c}\binom{n-s}{r-c} \rightarrow 0$. Thus AP4 holds.

Next, $\left| n\binom{n}{r}^{-1}\binom{s}{1}\binom{n-s}{r-1} - rs \right| \leq 2rs$. Again by AP3, there exists a constant K independent of n such that

$$\left| \sum_{r,s=1}^{m} \left\{ n\binom{n}{r}^{-1}\binom{s}{1}\binom{n-s}{r-1} - rs \right\} \xi_{nrs\underset{\sim}{t}}(1) \right|$$

$$\leq \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} |a_i|\,|a_j| \sum_{r,s=1}^{\infty} 2rs \min\left(\tilde{\bar{F}}_n^{(r)}(t_i), \tilde{\bar{F}}_n^{(s)}(t_j)\right)$$

$$\leq K.$$

For fixed r and s, $\left| n\binom{n}{r}^{-1}\binom{s}{1}\binom{n-s}{r-1} - rs \right| \rightarrow 0$, so that AP5 follows. □

Proof of Theorem AP2: By Chebyshev's inequality and Lemma AP2,

$$n^{1/2}\left(D_n(\underset{\sim}{t}) - \hat{D}_n(\underset{\sim}{t})\right) \rightarrow 0 \text{ in probability.}$$

It follows from Lemmas AP1 and AP2 that

$$n^{1/2}\left(D_n(\underset{\sim}{t}) - D_n^*(\underset{\sim}{t})\right) \xrightarrow{D} N(0,\sigma_{\underset{\sim}{t}}^2).$$

Thus the finite-dimensional distributions of Y_{n1}^* are asymptotically multivariate normal by the Cramér-Wold device. By definition $\Sigma_{i=1}^{\ell} a_i Y(t_i)$ is a $N(0,\sigma_{\underset{\sim}{t}}^2)$ random variable. Since the choices of ℓ and $\underset{\sim}{a} = (a_1,\ldots,a_\ell)'$ were arbitrary, we see that the finite-dimensional distributions of Y^*_{n1} converge weakly to the finite-dimensional distributions of Y. □

Tightness

<u>Lemma AP3</u>: Assume B9 and B11 hold. Then there exists a positive constant K such that for any $0 \le t_1 \le t_2 \le t_3 \le T$,

$$\mathbf{E}\left[Y^*_{n1}(t_2) - Y^*_{n1}(t_1)\right]^2\left[Y^*_{n1}(t_3) - Y^*_{n1}(t_2)\right]^2 \le K(t_3 - t_1)^2$$

for all $n \ge 1$.

<u>Proof</u>: For $i = 1,2,\ldots$ let $p^*_{i1} = \tilde{F}^{(i)}_n(t_2) - \tilde{F}^{(i)}_n(t_1)$ and $p^*_{i2} = \tilde{F}^{(i)}_n(t_3) - \tilde{F}^{(i)}_n(t_2)$. Define $S^*_{ni} = X^*_{11} + \ldots + X^*_{1i}$ and $S^*_{nj} = X^*_{11} + \ldots + X^*_{1c} + X^*_{1(i+1)} + \ldots + X^*_{1(i+j-c)}$ for $0 \le c \le i \wedge j$. Using AP1, the proof of Lemma 5.8 is readily adapted to show that for i,j=1, 2, ... there exists a constant K such that

$$\mathbf{E}\left\{\chi(S^*_{ni} \in (t_1,t_2]) - p^*_{i1}\right\}^2\left\{\chi(S^*_{nj} \in (t_2,t_3]) - p^*_{j2}\right\}^2 \le K\rho^{i \vee j}(t_3 - t_1)^2$$

for all $n \ge 1$. Now let S^*_{ni}, S^*_{nj}, S^*_{nk}, and $S^*_{n\ell}$ represent arbitrary sums of i, j, k, and ℓ random variables, respectively, from the bootstrap sample $X^*_{11}, \ldots, X^*_{1n}$. Define $S^*_{ni}(1) = \chi(S^*_{ni} \in (t_1,t_2]) - p^*_{i1}$, $S^*_{ni}(2) = \chi(S^*_{ni} \in (t_2,t_3]) - p^*_{i2}$,

$$S^*_{nijk\ell} = \left\{(S^*_{ni},S^*_{nj},S^*_{nk},S^*_{n\ell}) : \mathbf{E}_{\tilde{F}_n}\left[S^*_{ni}(1)S^*_{nj}(1)S^*_{nk}(2)S^*_{n\ell}(2)\right] \ne 0\right\}$$

and $U^*_{nijk\ell} = \#(S^*_{nijk\ell})$, the cardinality of $S^*_{nijk\ell}$. Then, as in Lemma 5.7, it follows that

$$n^2\binom{n}{i}^{-1}\binom{n}{j}^{-1}\binom{n}{k}^{-1}\binom{n}{\ell}^{-1}U^*_{nijk\ell} \le 7\,ijk\ell.$$

Thus

$$|\,\mathbf{E}\left[S^*_{ni}(1)S^*_{nj}(1)S^*_{nk}(2)S^*_{n\ell}(2)\right]| \le |\,\mathbf{E}\left\{\mathbf{E}_{\tilde{F}_n}\left[S^*_{ni}(1)S^*_{nj}(1)S^*_{nk}(2)S^*_{n\ell}(2)\right]\right\}|$$

$$\le K(t_3-t_1)^2(\rho^{i\vee k})^{1/2}(\rho^{j\vee \ell})^{1/2}$$

for all i, j, k, and ℓ. Now using the cardinality arguments of Section 5.3.2 and mimicing the proof of Lemma 5.6,

$$\mathbf{E}\left[Y^*_{n1}(t_2) - Y^*_{n1}(t_1)\right]^2\left[Y^*_{n1}(t_3) - Y^*_{n1}(t_2)\right]^2$$

$$\le n^2\sum_{i,j,k,\ell=1}^{m}\binom{n}{i}^{-1}\binom{n}{j}^{-1}\binom{n}{k}^{-1}\binom{n}{\ell}^{-1}U^*_{nijk\ell}K(t_3-t_1)^2(\rho^{i\vee k})^{1/2}(\rho^{j\vee \ell})^{1/2}$$

$$= K_1(t_3-t_1)^2$$

for all $n\ge 1$. □

<u>Proof of Theorem AP1:</u> As noted in the proof of Theorem 5.2, $\mathbf{P}\{Y(T)\ne Y(T-)\} = 0$ when F is continuous. That $Y^*_{n1} \xrightarrow{D} Y$ follows from Theorem AP2 and Lemma AP3, again using Theorem 15.6 of Billingsley (1968). □

BIBLIOGRAPHY

Anscombe, F. (1952). Large sample theory of sequential estimation. Proc. Cambridge Philos. Soc. 48 600-607.

Barlow, R. E., Marshall, A. W., and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. Ann. Math. Statis. 34, 375-389.

Barlow, R. E. and Proschan, F. (1964). Comparison of replacement policies and renewal theory implications. Ann. Math. Statist. 35, 577-589.

Barlow, R. E. and Proschan, F. (1965) *Mathematical Theory of Reliability.* Wiley, New York.

Barlow, R. E. and Proschan, F. (1975) *Statistical Theory of Reliability and Life Testing.* Holt Reinhart and Winston, Inc., New York.

Beran, R. (1984). Bootstrap methods in statistics. Jber. Deutsch. Math.-Verein. 86 14-30.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. Ann. Statist. 9 1196-1217.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multi-parameter stochastic processes and some applications. Ann. Math. Statist. 42, 1656-1670.

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

Blum, J. R. (1954i). Approximation methods which converge with probability one. Ann. Math. Statis. 25, 382-386.

Blum, J. R. (1954ii). Multidimensional stochastic approximation methods. Ann. Math. Statis. 25, 737-744.

Chow, Y. S. (1965). Local convergence of martingales and the law of large numbers. Ann. Math. Statis. 36, 552-558.

Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. Ann. Math. Statist. 36 457-462.

Chung, K. L. (1954). On a stochastic approximation method. Ann. Math. Statis. 25, 463-483.

Csörgő, S. and Mason, D. M. (1989). Bootstrapping empirical functions. Ann. Statist. 17, in press.

Derman, C. (1956). An application of Chung's lemma. Ann. Math. Statis. 27, 532-536.

Fabian, V. (1968). On asymptotic normality in stochastic approximation. *Ann. Math. Statis.* 39, 1327-1332.

Frees, E. W. (1983). On construction of sequential age replacement policies via stochastic approximation. Institute of Statistics Mimeo Series #1525, the consolidated University of North Carolina.

Frees, E. W. (1985). Weak convergence of stochastic approximation processes with random indices. *Sequential Analysis* 4, 59-82.

Frees, E. W. (1986i). Warranty analysis and renewal function estimation. *Naval Res. Logist. Quart.* 33, 361-372.

Frees, E. W. (1986ii). Nonparametric renewal function estimation. *Ann. Statis.* 14, 1366-1378.

Heyde, C. (1964). Two probability theorems and their application to some first passage problems. *J. Austral. Math. Soc. Ser. B* 4, 214-222.

Kersting, G. (1977). Almost sure approximation of the Robbins-Monro process by sums of independent random variables. *Ann. Prob.* 5, 954-965.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Statis.* 23, 462-466.

Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0, \infty]$. *J. Applied Prob.* 10, 109-121.

McLeish, D. (1974). Dependent central limit theorems and invariance principles. *Ann. Prob.* 2, 620-628.

McLeish, D. L. (1976). Functional and random central limit theorems for the Robbins-Monro process. *J. Appl. Prob.* 13, 148-154.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statis.* 22, 400-407.

Robbins, H. and Siegmund, D. (1971). A convergence theorem for nonnegative almost supermartingales and some applications. *Optimizing Methods in Statistics* (J. S. Rustagi, Ed.), 233-257. Academic Press, New York.

Ruppert, D. (1982). Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise. *Ann. Prob.* 10, 178-187.

Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statis.* 29, 373-405.

Sen, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*, John Wiley and Sons, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, New York.

Sielken, R. L. (1973). Stopping times for stochastic approximation procedures. *Z. Wahrscheinlichkeitstheorie Verw. Geb.* 26, 67-75.

Singh, R. S. (1977) Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density. *Ann. Statist.* 5, 394-399.

Smith, W. L. (1954). Asymptotic renewal theorems. *Proc. Roy. Soc. Edinb. A* 64, 9-48.

Strassen, V. (1967). Almost sure behaviour of sums of independent random variables and martingales. *Proc. Fifth Berkeley Symp. Math. Stat. Prob.* (Ed. L. LeCam, et. al.), Los Angeles, University of California Press, Vol. 2, pp. 315-343.

Tsirel'son, J. L. (1975). The density of the distribution of the maximum of a Gaussian process. *Theory Probab. Appl.* 20 847-856.

Wold, D. E. (1981). On smooth estimation of the renewal density. Ph. D. dissertation, University of North Carolina-Chapel Hill.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, SIAM, Philadelphia.