(#2018)

ON STOPPING TIMES FOR FIXED-WIDTH CONFIDENCE BANDS

John I. Crowell and P. K. Sen

Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

Keywords and Phrases: confidence regions; bootstrap.

ABSTRACT

Analogous to the usual stopping time for fixed-width confidence interval estimation of a real-valued parameter $\theta(F)$, a stopping time for fixed-width confidence band estimation of a statistical functional process R_F is considered. The proposed confidence band is shown to have the correct asymptotic coverage probability as the band width goes to zero. Sufficient conditions for asymptotic efficiency and normality of the stopping time are given. Some implications of using the stopping time in conjunction with a bootstrap confidence band procedure are investigated. An example illustrates how essentially the same stopping time may be applicable when estimating a confidence region for a statistical function R_F defined on $\mathbb{R}^p$, $p \geq 2$. Stopping times for the construction of variable width confidence bands are also considered.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having distribution function F. Suppose $R_F = \{R_F(t), t \in I\}$ for some $I \subseteq \mathbb{R}$ is an unknown functional process of F. There often exists a natural 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.

A typical approach assumes that $r_n \overset{W}{\to} \mathcal{G}_F$, where $\mathcal{G}_F$ is a Gaussian process on the interval I such that $\mathbf{P}\{\sup_{t\in I}|\mathcal{G}_F(t)|<\infty\}=1$. (Throughout this paper $\overset{W}{\rightarrow}$ ($\overset{D}{=}$) will be used to denote weak convergence (equivalence in distribution) of processes while $\overset{D}{\rightarrow}$ will denote weak convergence of random variables.) Define

$$G_F(x)=\mathbf{P}\{\sup_{t\in I}|\mathcal{G}_F(t)|\le x\},\ -\infty<x<\infty.$$

For a desired confidence level of $1-\alpha$, $0<\alpha<1$, let $c=\inf\{x\colon G_F(x)\ge 1-\alpha\}$. If c_n is a weakly consistent estimate of c, then a sensible large sample confidence band for R_F is given by

$$\{\, R_n(t)\pm c_n n^{-1/2},\ t\in I\,\}, \tag{1.1}$$

for then

$$\mathbf{P}\Big\{\, R_n(t)-c_n n^{-1/2}\le R_F(t)\le R_n(t)+c_n n^{-1/2},\ t\in I\,\Big\}\rightarrow 1-\alpha, \tag{1.2}$$

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 (1.1). 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

$$\mathbf{P}\Big\{\,|R_n(t)-R_F(t)|\le d,\ t\in I\,\Big\}\ge 1-\alpha, \tag{1.3}$$

where $d>0$ is a constant that reflects the desired accuracy. Typically the minimum sample size n such that (1.3) 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 $\mathbf{P}\{\,|\theta_n-\theta(F)|\le d\,\}\ge 1-\alpha$. Assume $n^{1/2}(\theta_n-\theta(F))\overset{D}{\rightarrow}N(0,\sigma^2)$ as $n\rightarrow\infty$ and s_n^2 is a strongly consistent estimator of σ^2. For the fixed-width confidence band problem, a number of authors have considered stopping times of the type

$$N'_d = \inf\left\{ n \ge n_o:\ n^{-1/2} s_n \tau_{\alpha/2} \le d \right\},$$

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

$$\lim_{d \downarrow 0} \mathbf{P}\left\{ |\theta_{N'_d} - \theta(F)| \le d \right\} = 1 - \alpha,$$

so that for sufficiently small d the confidence band $(\theta_{N'_d} - d,\ \theta_{N'_d} + d)$ has approximately the desired coverage probability.

This paper has as its primary purpose the definition of a stopping time N_d such that under certain conditions

$$\lim_{d \downarrow 0} \mathbf{P}\left\{ |R_{N_d}(t) - R_F(t)| \le d,\ t \in I \right\} = 1 - \alpha. \tag{1.4}$$

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

$$N_d = \inf\left\{ n \ge n_o:\ n^{-1/2} c_n \le d \right\}; \tag{1.5}$$

upon stopping use the confidence band

$$\{ R_{N_d}(t) \pm d,\ t \in I \}. \tag{1.6}$$

Thus the rule states that one should stop when the width of the confidence band determined by (1.1) is at most 2d.

The most basic properties of N_d follow directly from its definition. By the relation

$$[c_{N_d}]^2 \le N_d d^2 < [c_{N_d - 1}]^2 + d^2, \tag{1.7}$$

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

$$n_d = \inf\left\{ n \ge n_o:\ n^{-1/2} c \le d \right\}, \tag{1.8}$$

where n_o is the initial sample size that appears in (1.5). Suppose $\mathbf{P}\{ c_n \le 0 \} = 0$ for all $n \ge n_o$. Since $n_d \sim c^2/d^2$ as $d \downarrow 0$, one then has that $N_d/n_d \to 1$ as $d \downarrow 0$ by (1.7).

It is shown in Section 2 that if one also assumes that $r_{N_d} \xrightarrow{W} \mathcal{G}_F$ as $d \downarrow 0$, then (1.4) holds. A simple criterion for establishing weak convergence of r_{N_d} is stated. In Section 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 4, along with some resulting properties of the stopping time. In some applications one might want to allow confidence bands of variable width. Section 5 addresses the problem of stopping times for such modified confidence band procedures. An example in Section 6 illustrates how a stopping time of the form (1.5) can be used in the construction of a confidence region for a statistical function R_F defined on $\mathbb{R}^p$, $p \geq 2$.

2. ASYMPTOTIC COVERAGE OF THE CONFIDENCE BAND

Theorem 2.1 contains the principal result of the section regarding the correct asymptotic coverage probability of the fixed-width confidence band (1.6) when $r_{N_d} \xrightarrow{W} \mathcal{G}_F$ as $d \downarrow 0$. Corollary 2.1 gives a single sufficient condition for the weak convergence of r_{N_d} which is easily verified in some cases.

A1 summarizes the principal assumptions of the introduction.

A1: With previously defined notation, $r_n \xrightarrow{W} \mathcal{G}_F$ 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 A1 is generally met.

Theorem 2.1: Assume A1 and that $r_{N_d} \xrightarrow{W} \mathcal{G}_F$ as $d \downarrow 0$. Then (1.4) holds.
Proof: 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 (1.2). □

In practice it will of course be necessary to verify that $r_{N_d} \xrightarrow{W} \mathcal{G}_F$ as $d \downarrow 0$. Given the stopping time N_d and that $r_n \xrightarrow{W} \mathcal{G}_F$ on I, we consider a single sufficient condition for weak convergence of R_{N_d} based on the notion of a uniformly continuous in probability sequence. 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 $\mathbf{P}\{ \max_{0\leq k \leq n\delta} |X_{n+k} - X_n| \geq \epsilon \} < \epsilon$ for all $n \geq 1$. The condition considered here is as follows.

A2: For each $t \in I$, $\{r_n(t), n \geq 1\}$ is a u.c.i.p. sequence.

Corollary 2.1: Assume A1 and A2 hold. Then (1.4) holds.
Proof: By Theorem 2.1 it is sufficient to verify that the finite-dimensional distributions of r_{N_d} converge weakly to those of $\mathcal{G}_F$. For arbitrary positive integer m, let $(a_1, \ldots, a_m)'$ be any m-

vector of real numbers and $\{t_1, \ldots, t_m\} \subseteq I$. Define $Y_n = \Sigma_{i=1}^m a_i r_n(t_i)$. By the weak convergence of r_n and the Cramér-Wold device, $Y_n \xrightarrow{D} Y$, where $Y \overset{D}{=} \Sigma_{i=1}^m a_i \mathcal{G}_F(t_i)$ is a normal random variable. Anscombe's theorem is used to show that Y_{N_d} has the same limiting distribution as $d \downarrow 0$.

By the proof of Theorem 2.1, $N_d/d^{-2} \to c^2$ a.s. as $d \downarrow 0$. Since for each $t \in I$, $\{r_n(t), n \geq 1\}$ is a u.c.i.p. sequence, $\{Y_n, n \geq 1\}$ is also a u.c.i.p. sequence. (See Lemma 1.4 of Woodroofe(1982).) Thus by Anscombe's theorem $Y_{N_d} \xrightarrow{D} Y$ as $d \downarrow 0$, which implies the finite-dimensional distributions of r_{N_d} converge to those of $\mathcal{G}_F$. □

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 3.1 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

$$\lim_{d \downarrow 0} n_d^{-1} \mathbf{E} N_d = 1. \tag{3.1}$$

The type of arguments needed to prove Lemma 3.1 may be found in Sen(1981), pp. 284 and 287.

<u>Lemma 3.1:</u> Assume A1 holds and that for fixed $d > 0$

$$\mathbf{E}\{ \sup_{n \geq n_o} c_n^2 \} < \infty. \tag{3.2}$$

Then $\mathbf{E} N_d < \infty$ and (3.1) holds.

To show that the expectation of the stopping time is finite, another approach is to establish the equivalent condition that

$$\sum_{n \geq n_o} \mathbf{P}\{ N_d > n \} < \infty. \tag{3.3}$$

Similarly, an alternative condition for verifying (3.1) is

$$\lim_{d \downarrow 0} \quad n_d^{-1} \sum_{n \geq n_d(1+\epsilon)} \mathbf{P}\{ N_d > n \} = 0. \tag{3.4}$$

(See, for example, Sen(1981), p.288.) In Section 4 it is demonstrated that if c_n is a particular type of bootstrap estimator it may be possible to verify (3.3) and (3.4).

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

$$n_d^{-1} g(n_d) \to 0 \quad \text{as } d \downarrow 0. \tag{3.5}$$

Suppose that

$$g(n_d)\{ c_{n_d}^2 / c^2 - 1 \} \xrightarrow{D} N(0, \gamma^2) \quad \text{as } d \downarrow 0, \tag{3.6}$$

where γ is a finite positive constant. Further assume that

$$\{ g(n)[c_n^2/c^2 - 1], n \geq n_o \} \text{ is a u.c.i.p. sequence.} \tag{3.7}$$

Together with A1 these three conditions yield the following result, similar in derivation to Theorem 10.2.2 of Sen(1981).

Lemma 3.2: Assume A1, (3.5), (3.6), and (3.7) hold. Then

$$g(n_d)\{ N_d/n_d - 1 \} \xrightarrow{D} N(0, \gamma^2) \quad \text{as } d \downarrow 0.$$

4. BOOTSTRAP QUANTILE ESTIMATORS

One method that has been suggested for constructing confidence bands of the form (1.1) 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 $r^*_{m_n} \stackrel{W}{\to} \mathcal{G}_F$ as n, $m_n \to \infty$ so that a natural estimator of the distribution G_F is defined by

$$G_{n,F_n}(x) = \mathbf{P}\{ \sup_{t\in I} |r^*_{m_n}(t)| \le x \mid X_1,\dots, X_n \}, \ -\infty < x < \infty.$$

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 $i = 1,\dots, M_n$ let $X^*_{i1},\dots, X^*_{im_n}$ be a random sample from the distribution F_n, where $M_n \to \infty$ as $n \to \infty$. For $t \in I$ define $R^*_{m_n i}(t) = R_{m_n}(t; X^*_{i1},\dots, X^*_{im_n})$. Then the processes

$$r^*_{m_n i}(\cdot) = m_n^{1/2}\{ R^*_{m_n i}(\cdot) - R_n(\cdot)\}$$

on I, $i = 1,\dots, M_n$, are conditionally independent and identically distributed. Estimate G_{n,F_n} empirically by

$$\hat{G}_{n,F_n}(x) = M_n^{-1} \Sigma_{i=1}^{M_n} \chi(\sup_{t\in I} |r^*_{m_n i}(t)| \le x), \ -\infty < x < \infty.$$

The bootstrap estimate of c is then taken to be $c^*_n = \inf\{ x\colon \hat{G}_{n,F_n}(x) \ge 1-\alpha\}$. Putting $c_n = c^*_n$ in (1.5) 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 (3.1) by confirming that conditions such as (3.2), (3.3), or (3.4) 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 (3.3) and (3.4) 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 A1 is replaced by the following.

A3: 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. $\mathbf{P}\{ \hat{G}_{n,F_n}(0) = 0$ for all $n \geq n_o \} = 1$. The sequence $\{M_n\}$ is such that $M_n^{-1} \log n = o(1)$.

A3 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 4.1 and 4.2 assume that $|G_{n,F_n}(x) - G_n(x)| \to 0$ completely as $n \to \infty$ in the following sense.

A4: For every $x \in \mathbb{R}$ and every $\delta > 0$, $\sum_{n \geq n_o} \mathbf{P}\{ | G_{n,F_n}(x) - G_n(x) | > \delta \} < \infty$.

<u>Lemma</u> <u>4.1:</u> Assume A3 and A4 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

$$\sum_{n \geq n_o} \mathbf{P}\{ N_d > n \} \leq \sum_{n \geq n_o} \mathbf{P}\{ c^*_n > n^{1/2} d \}. \tag{4.1}$$

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

$$\mathbf{P}\{ c^*_n > n^{1/2} d \} = \mathbf{P}\{ \hat{G}_{n,F_n}(n^{1/2}d) < 1 - \alpha \}$$

$$\leq \mathbf{P}\Big\{ \hat{G}_{n,F_n}(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) < -\delta + G_n(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) \Big\}$$

$$\leq \mathbf{P}\Big\{ | \hat{G}_{n,F_n}(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) | > \delta/2 \Big\} + \mathbf{P}\Big\{ |G_n(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) | > \delta/2 \Big\}.$$

By a standard probability inequality for empirical distribution functions

$$\mathbf{P}\Big\{ \sup_{x \in \mathbb{R}} | \hat{G}_{n,F_n}(x) - G_{n,F_n}(x) | > \delta/2 \Big\} \leq C e^{-2M_n(\delta/2)^2} \tag{4.2}$$

where C is a constant not depending on F_n or F. (See Serfling(1980), Section 2.1.3.) Hence by (4.1), A3, and A4

$$\sum_{n \geq n_o} \mathbf{P}\{ N_d > n \}$$

$$\leq \sum_{n \geq n_o} C e^{-2M_n(\delta/2)^2} + \sum_{n \geq n_o} \mathbf{P}\Big\{ |G_n(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) | > \delta/2 \Big\}$$

$< \infty,$

which is equivalent to $\mathbf{E}\,N_d < \infty$. □

Lemma 4.2: Assume A3 and A4 hold. Then $\lim_{d \downarrow 0} n_d^{-1}\,\mathbf{E}\,N_d = 1$.

Proof: To show that (3.4) 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 $n \geq c^2(1+\epsilon)/d^2$,

$$\mathbf{P}\{\, c_n^* > n^{1/2}d \,\} = \mathbf{P}\Big\{\, \hat{G}_{n,F_n}(n^{1/2}d) < 1-\alpha \,\Big\}$$

$$\leq \mathbf{P}\Big\{\, \hat{G}_{n,F_n}(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) < 1-\alpha - G_{n,F_n}((1+\epsilon)^{1/2}c)\Big\}$$

$$\leq \mathbf{P}\Big\{\, \hat{G}_{n,F_n}(n^{1/2}d) - G_{n,F_n}(n^{1/2}d) < -\delta\Big\} + \mathbf{P}\Big\{1-\alpha+\delta > G_{n,F_n}((1+\epsilon)^{1/2}c)\Big\}.$$

By A3, $G_n \xrightarrow{D} G_F$ and one can choose $\delta, \delta' > 0$ 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 A4,

$$\sum_{n \geq n_d(1+\epsilon)} \mathbf{P}\Big\{\, 1-\alpha+\delta > G_{n,F_n}((1+\epsilon)^{1/2}c) \,\Big\}$$

$$\leq \sum_{n \geq n_d(1+\epsilon)} \mathbf{P}\Big\{\, |\, G_{n,F_n}((1+\epsilon)^{1/2}c) \,-\, G_n((1+\epsilon)^{1/2}c)\,| > \delta' \,\Big\}$$

$$< \infty$$

and using (4.2)

$$\sum_{n \geq n_d(1+\epsilon)} \mathbf{P}\{\, N_d > n \,\} \leq \sum_{n \geq n_d(1+\epsilon)} \mathbf{P}\{\, c_n^* > n^{1/2}d \,\} \,<\, \infty.$$

Thus (3.4) 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 A4.

A5: [Hadamard-Lipschitz condition of order α.] There exists a real-valued function K(x) such that for all n and $x \in \mathbf{R}$

$$|G_{n,F_n}(x) - G_n(x)| \leq K(x) \| F_n - F \|^{\alpha} \text{ for some } \alpha > 0,$$

where $\| f \|$ denotes the supremum norm of the function f.

The inequality applied in (4.2) to $\hat{G}_{n,F_n}$ and G_{n,F_n} can now be applied to F_n and F to show that A5 implies A4:

$$\sum_{n \geq n_o} \mathbf{P}\{ \mid G_{n,F_n}(x) - G_n(x) \mid > \delta \} \leq \sum_{n \geq n_o} \mathbf{P}\Big\{ \| F_n - F \| > [\delta / K(x)]^{1/\alpha} \Big\}$$

$$\leq \sum_{n \geq n_o} C e^{-2n[\delta/K(x)]^{2/\alpha}}$$

$$< \infty.$$

While A5 seems a more natural condition than A4, 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 $M_n m_n \leq n$ 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 $F_{ni}(x) = m_n^{-1} \Sigma_{j=1}^{m_n} \chi(X_{ij} \leq x)$, $-\infty < x < \infty$. 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

$$G_n(x) = \mathbf{P}\{ \sup_{t \in I} | r^*_{m_n i}(t) | \leq x \}, \; -\infty < x < \infty.$$

G_n can now be estimated in an unbiased fashion by

$$\hat{G}_n(x) = M_n^{-1}\Sigma_{i=1}^{M_n} \chi(\sup_{t\in I} |r^*_{m_n i}(t)| \leq x), \ -\infty < x < \infty.$$

A3 must be changed slightly.

A6: With previously defined notation, $r_n \overset{W}{\to} \mathcal{G}_F$ and $r^*_{m_n 1} \overset{W}{\to} \mathcal{G}_F$ 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. $\mathbf{P}\{ \hat{G}_n(0) = 0$ for all $n \geq n_o \} = 1$. The sequence $\{M_n\}$ is such that $M_n^{-1} \log n = o(1)$.

Let $c^*_n = \inf\{x: G_n(x) \geq 1-\alpha\}$ for which a natural estimate is $\hat{c}^*_n = \inf\{x: \hat{G}_n(x) \geq 1-\alpha\}$. By a standard probability inequality for deviations of sample quantiles, for any $\epsilon > 0$

$$\mathbf{P}\{ | \hat{c}^*_n - c^*_n | > \epsilon \} \leq 2e^{-2M_n\delta_n^2},$$

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 $G_n \overset{D}{\to} G_F$ 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 A6 holds, one then has the properties indicated in Lemmas 4.1 and 4.2. 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 4.1 and 4.2 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 (1.5). 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.

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 (1.1) or (1.6) 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 (1.5) and the confidence band (1.6).

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 $\tilde{r}_n \overset{W}{\to} \tilde{\mathcal{G}}_F$, 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 $r_n \overset{W}{\to} \mathcal{G}_F$ 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

$$\{ R_n(t) \pm \tilde{c}_n n^{-1/2} q_n(t),\ t \in J \}. \tag{5.1}$$

With the preceding assumptions,

$$\mathbf{P}\Big\{ R_n(t) - \tilde{c}_n n^{-1/2} q_n(t) \ \le\ R_F(t) \ \le\ R_n(t) + \tilde{c}_n n^{-1/2} q_n(t),\ t \in J \Big\} \tag{5.2}$$

$$= \mathbf{P}\{ \sup_{t \in J} |\tilde{r}_n(t)| \le \tilde{c}_n \}$$

$$\to 1 - \alpha.$$

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

$$\tilde{N}_d = \inf\left\{ n \geq n_o: n^{-1/2}\tilde{c}_n \leq d \right\},$$

where n_o is again an initial sample size. Let

$$\{ R_{\tilde{N}_d}(t) \pm d\, q_{\tilde{N}_d}(t),\ t \in J \} \tag{5.3}$$

be the associated confidence band.

It follows from the strong consistency of $\tilde{c}_n$ that if $\mathbf{P}\{ \tilde{c}_n \leq 0 \} = 0$ for all $n \geq n_o$, then $\tilde{c}_{\tilde{N}_d} \to \tilde{c}$ a.s. as $d \downarrow 0$. By a relation like (1.7), 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 $d \downarrow 0$,

$$\lim_{d \downarrow 0} \mathbf{P}\left\{ | R_{\tilde{N}_d}(t) - R_F(t) | \leq d\, q_{\tilde{N}_d}(t),\ t \in J \right\} \tag{5.4}$$

$$= \lim_{d \downarrow 0} \mathbf{P}\left\{ \sup_{t \in J} | \tilde{r}_{\tilde{N}_d}(t) | \leq [\tilde{N}_d]^{1/2} d \right\}$$

$$= 1 - \alpha$$

by (5.2). Thus one can define a meaningful stopping time for the confidence band (5.1).

As in Corollary 2.1, the additional condition that for each $t \in J$, $\{\tilde{r}_n(t),\ n \geq 1\}$ is a u.c.i.p. sequence is then sufficient for (5.4). Letting $\tilde{n}_d = \inf\{ n \geq n_o: n^{-1/2}\tilde{c} \leq d \}$, 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. EXAMPLE: EMPIRICAL PROCESSES

Let $\underset{\sim}{X}, \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ be i.i.d. random vectors having distribution function F defined on $\mathbb{R}^p$ for some $p \geq 1$. Define the empirical distribution function by

$$F_n(\underset{\sim}{x}) = n^{-1}\Sigma_{i=1}^n\, c(\underset{\sim}{x} - \underset{\sim}{X}_i), \quad \underset{\sim}{x} \in \mathbb{R}^p,$$

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 $V_n = \{ V_n(\underset{\sim}{x}),\ \underset{\sim}{x} \in \mathbb{R}^p \}$. For the remainder of this section, let

$$G_F(t) = \mathbf{P}\{ \sup_{x \in \mathbb{R}^p} | V(\underset{\sim}{x}) | \leq t \},\ -\infty < t < \infty,$$

whenever V_n converges weakly to a zero-mean Gaussian process V on $\mathbb{R}^p$. In addition, make the assumption that $\mathbf{P}\{\sup_{x \in \mathbb{R}^p} |V(\underset{\sim}{x})| < \infty\} = 1$, c is such that $G_F(c) = 1-\alpha$, and G_F is continuous and strictly increasing in a neighborhood of c.

As is well known, if $p=1$ then $V_n \overset{W}{\to} B \circ F$ as $n \to \infty$, where B is the Brownian bridge on [0, 1] and $B \circ F(x) = B(F(x))$, $-\infty < x < \infty$. Thus when F is continuous G_F is the distribution of the supremum of the Brownian bridge, which is known, while for discontinuous F

$$\mathbf{P}\{\ \sup_{x \in \mathbb{R}} |B \circ F(x)| \le t\ \} \ge \mathbf{P}\{\ \sup_{x \in \mathbb{R}} |B(x)| \le t\ \} \quad \text{for all } t \in \mathbb{R}, \tag{6.1}$$

so that one can give an upper bound for any quantile of G_F. (See Kolmogorov(1941) and Noether(1963).) Bickel and Freedman(1981) have shown that bootstrapping the univariate empirical process V_n yields a strongly consistent estimate of c, the $(1-\alpha)$th quantile of G_F. Letting c_n be the bootstrap estimate and $r_n(\cdot) = V_n(\cdot)$ on $\mathbb{R}$, define the stopping time N_d by (1.5) and the associated confidence band by (1.6).

For each $x \in \mathbb{R}$ the statistic $V_n(x)$ is a normalized sum of independent random variables whose distribution has mean zero and finite variance; thus by Example 1.8 of Woodroofe(1982) the sequence $\{V_n(x),\ n \ge 1\}$ is u.c.i.p. for each $x \in \mathbb{R}$ and by Corollary 2.1 the stopping time (1.5) and confidence band (1.6) provide the correct asymptotic coverage probability as $d \downarrow 0$. If one instead implemented the bootstrap method of Section 4, by the arguments presented there the stopping time would have finite expectation and (3.1) would hold. Finally, modifications of the sort considered in Section 5 resulting in a variable width confidence band could be made. Thus while in the continuous case this sequential methodology is not necessary to calculate a useful confidence band for F, if F has a discrete component one can improve on the potentially conservative confidence band suggested by (6.1).

Now let $p \ge 2$. In this case $V_n \overset{W}{\to} V$, where V is a tied-down zero-mean Gaussian process on $\mathbb{R}^p$. 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

$$\{\ F_{N_d}(\underset{\sim}{x}) \pm d,\ \underset{\sim}{x} \in \mathbb{R}^p\ \}$$

for the multivariate distribution F. It then follows that if $V_{N_d} \overset{W}{\to} V$ as $d \downarrow 0$, then

$$\mathbf{P}\Big\{\ F_{N_d}(\underset{\sim}{x}) - d \le\ F(\underset{\sim}{x})\ \le\ F_{N_d}(\underset{\sim}{x}) + d,\ \ \underset{\sim}{x} \in \mathbb{R}^P \Big\} \tag{6.2}$$

$$= \mathbf{P}\{\ \sup_{\underset{\sim}{x} \in \mathbb{R}^p} |V_{N_d}(\underset{\sim}{x})| \le N_d^{1/2} d\ \}$$

$\to 1-\alpha$

as $d \downarrow 0$ by the assumptions and the inequality (1.7). As in the univariate case, for each $\underset{\sim}{x} \in \mathbb{R}^p$ the sequence $\{ V_n(\underset{\sim}{x}), n \geq 1 \}$ is u.c.i.p.; by an argument similar to the proof of Corollary 2.1 it follows that $V_{N_d} \overset{W}{\to} V$ as $d \downarrow 0$ so that (6.2) holds.

In this paper 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 2 and 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.

BIBLIOGRAPHY

Anscombe, F. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* 48 600-607.

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.

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.

Csörgő, S. and Mason, D. M. (1989). Bootstrapping empirical functions. *Ann. Statist.* 17, in press.

Kolmogorov, A. N. (1941). Confidence limits for an unknown distribution function. *Ann. Math. Statis.* 12 461-463.

Noether, G. E. (1963). Note on the Kolmogorov statistic in the discrete case. *Metrika* 7 115-116.

Sen, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*, John Wiley and Sons, New York.

Tsirel'son, J. L. (1975). The density of the distribution of the maximum of a Gaussian process. *Theory Probab. Appl.* 20 847-856.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, SIAM, Philadelphia.