Statistica Sinica 4(1994), 461-476

LIMIT THEOREMS FOR WEAKLY DEPENDENT HILBERT SPACE VALUED RANDOM VARIABLES WITH APPLICATION TO THE STATIONARY BOOTSTRAP

Dimitris N. Politis and Joseph P. Romano

Purdue University and Stanford University

Abstract: In this article, some weak convergence results are developed for approximate sums of weakly dependent stationary Hilbert space valued random variables in a triangular array setting. The motivation for such results lies in understanding the weak convergence properties of estimators which are smooth functionals of the empirical process. By regarding the empirical process as an element of an appropriate Hilbert space, the asymptotic distributional properties can be deduced. The results are designed to be strong enough to handle the study of estimators under the stationary bootstrap resampling plan. In Politis and Romano (1991), a resampling procedure, called the stationary bootstrap, is introduced as a means of calculating standard errors of estimators and constructing confidence regions for parameters based on weakly dependent stationary observations. The results derived here support the asymptotic validity of the stationary bootstrap method for a broad class of estimators. In particular, minimum distance estimators, whose robustness properties have been well-established by Millar (1981, 1984), are shown to have robustness of validity in the sense that confidence intervals constructed by the stationary bootstrap method based on such estimators are asymptotically valid even when the usual independence assumption often used in robustness studies is seriously violated.

Key words and phrases: Approximate confidence limit, bootstrap, differentiability, Hilbert space, minimum distance estimators, stationary, time series.

1. Introduction

In this paper, some weak convergence results are proved for sums and approximate sums of dependent Hilbert space valued random variables. Some motivation for the present work is the following. Suppose $\xi_1, \ldots, \xi_n$ are observations from a real-valued stationary time series with empirical distribution function $\hat{F}_n$. Many estimators can be regarded as smooth functionals of the empirical process Z_n defined by $Z_n(\cdot) = n^{1/2}[\hat{F}_n(\cdot) - F(\cdot)]$, where $F(\cdot)$ is the cumulative distribution function of ξ_1. Hence, to understand the behavior of such estimators, it suffices to understand the behavior of the empirical process. But, $Z_n(\cdot) = n^{-1/2}\Sigma_i X_i(\cdot)$ itself is a normalized sum of stationary processes, where $X_i(t) = 1(\xi_i \leq t) - F(t)$

and $1(\xi_i \leq t)$ is the indicator function of the event $\{\xi_i \leq t\}$. We regard X_i as elements of a certain Hilbert space, so that the problem is reduced to studying the behavior of a sum of stationary Hilbert space valued random variables. A very general class of estimators where this approach is fruitful is the class of minimum distance estimators, in which the distance is defined by a Hilbert norm.

A goal of this paper is to study the stationary bootstrap resampling method as a means of constructing confidence intervals based on such estimators. The stationary bootstrap, introduced in Politis and Romano (1991), is a variant of the moving blocks bootstrap developed by Künsch (1989) and Liu and Singh (1992), and the blocks of blocks bootstrap developed by Politis and Romano (1992a, b). All these methods are designed to construct nonparametric confidence intervals in the setting of stationary time series where even asymptotic distribution theory is often intractable. Specifically, in order to approximate the distribution of $\hat{\theta}_n(\xi_1, \ldots, \xi_n) - \theta$, we consider the distribution (conditional on the data $\xi_1, \ldots, \xi_n$) of $\hat{\theta}_n(\xi_1^*, \ldots, \xi_n^*) - \hat{\theta}_n(\xi_1, \ldots, \xi_n)$, where $\xi_1^*, \ldots, \xi_n^*$ is a new pseudo time series generated by the stationary resampling scheme. This algorithm is presented more clearly later. For now, the motivating remark is that the behavior of this approximating distribution can be deduced by studying the bootstrap empirical process Z_n^* defined by $Z_n^*(\cdot) = n^{1/2}[\hat{F}_n^*(\cdot) - \hat{F}_n(\cdot)]$, where $\hat{F}_n^*$ is the empirical distribution of $\xi_1^*, \ldots, \xi_n^*$. Again, we are led to consider the sum of stationary Hilbert space random variables, although in this case the underlying stochastic mechanism changes with n and is in fact random because it depends on the original data.

The above considerations motivate the need for the results presented in Section 2. In particular, we develop simple conditions for tightness of a sum of stationary Hilbert space valued random variables in a triangular array setup. Tightness plus examination of finite dimensional projections allows one to deduce weak convergence results. The conditions presented in Theorem 2.1 are particularly simple to apply because they do not involve a particular choice of basis and because all that is required is a means of getting a handle on second order moments of the underlying processes.

The paper culminates in Section 3. Here, the stationary bootstrap resampling algorithm is presented. A general bootstrap central limit theorem is proved under relatively mild dependence assumptions. The motivation for studying the empirical process and bootstrap empirical process is elaborated to deduce asymptotic validity of bootstrap confidence intervals. Some effort is made to show how these results immediately imply asymptotic distributional results for minimum distance estimators. Minimum distance estimators have been well studied in Millar (1981, 1984), where a certain robustness of these estimators is impressively demonstrated from a decision theoretic point of view. In short, the efficiency of the estimators does not deteriorate if the underlying model is misspecified. Our goal is to show that the stationary bootstrap method for constructing confidence regions based on such estimators has a certain robustness of validity even in the sense that the usual independence assumption imposed in robustness studies can be seriously violated.

2. Basic Theorem

Throughout, H denotes a separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. Equip H with the usual Borel σ-field. A sequence of H-valued random variables Z_n converges in distribution to Z if $E[f(Z_n)] \to E[f(Z)]$ for all real-valued bounded continuous functions f. The basic theory of weak convergence of Hilbert space valued random variables may be found in Parthasarathy (1967) and Bergstrom (1982). Suppose Z takes values in H and has probability distribution μ, so that $\mu(E)$ is the probability Z falls in E. If $E(\|Z\|^2) < \infty$, then the covariance operator of Z (or μ) is a continuous, linear, symmetric, positive operator S from H to H satisfying

$$\langle Sh, h\rangle = \int_{x\in H} \langle x, h\rangle\langle x, h\rangle\mu(dx) = E(\langle Z, h\rangle\langle Z, h\rangle).$$

Z has mean vector $m \in H$ if $E(\langle Z, h\rangle) = \langle m, h\rangle$ for all $h \in H$.

Theorem 2.1. *Let $X_{n,1},\dots,X_{n,n}$ be H-valued, stationary, mean zero random variables such that $E(\|X_{n,i}\|^2)<\infty$. Assume, for any integer $k\geq 1$, that $(X_{n,1},\dots,X_{n,k})$, regarded as a random element of H^k, converges in distribution to $(X_1,\dots,X_k)$, say. Moreover, assume, $E[\langle X_{n,1}, X_{n,k}\rangle] \to E[\langle X_1, X_k\rangle]$ as $n\to\infty$ and*

$$\lim_{n\to\infty}\sum_{k=1}^{n} E(\langle X_{n,1}, X_{n,k}\rangle) = \sum_{k=1}^{\infty} E(\langle X_1, X_k\rangle) < \infty, \tag{2.1}$$

where the last series is assumed absolutely summable. Let $Z_n = n^{-1/2}\Sigma_{i=1}^{n} X_{n,i}$. For any $h\in H$, let $\sigma_{n,h}^2$ denote the variance of $\langle Z_n, h\rangle$. Assume

$$\sigma_{n,h}^2 \to \sigma_h^2 \equiv \mathrm{Var}(\langle X_1, h\rangle) + 2\sum_{i=1}^{\infty}\mathrm{Cov}(\langle X_1, h\rangle, \langle X_{1+i}, h\rangle). \tag{2.2}$$

Then Z_n is weakly compact.

Proof. Fix a complete orthonormal basis for H, denoted by $e_1, e_2, \dots$. Let $r_N^2(x) = \Sigma_{j=N}^{\infty}|\langle x, e_j\rangle|^2$. By Theorem 1.13 of Prokhorov (1956), it is sufficient to show $\lim_N \sup_n E[r_N^2(Z_n)] = 0$. Because $r_N^2(x)$ monotonically decreases to 0 for every x, it is sufficient to show $\lim_{N\to\infty}\lim_{n\to\infty} E[r_N^2(Z_n)] \to 0$. Now,

$E[r_N^2(Z_n)] = \Sigma_{j=N}^{\infty} E[|\langle Z_n, e_j\rangle|^2]$. First, check

$$E[r_1^2(Z_n)] \to E\|X_1\|^2 + 2\sum_{i=1}^{\infty} E[\langle X_1, X_{1+i}\rangle] = L < \infty,$$

by the assumptions. Hence, it suffices to show $E[r_1^2(Z_n) - r_N^2(Z_n)] \to L$ as $N, n \to \infty$. But,

$$\begin{aligned} &E[r_1^2(Z_n) - r_N^2(Z_n)] \\ &= \sum_{j=1}^{N-1} E[\langle X_{n,1}, e_j\rangle\langle X_{n,1}, e_j\rangle] + 2\sum_{j=1}^{N-1}\sum_{i=1}^{n}\Big(1-\frac{i}{n}\Big)E[\langle X_{n,1}, e_j\rangle\langle X_{n,1+i}, e_j\rangle]. \end{aligned}$$

We claim for every i and j,

$$E[\langle X_{n,1}, e_j\rangle\langle X_{n,1+i}, e_j\rangle] \to E[\langle X_1, e_j\rangle\langle X_{1+i}, e_j\rangle] \tag{2.3}$$

as $n \to \infty$. To show this, first consider the case $i = 0$. Note,

$$\liminf_n E|\langle X_{n,1}, e_j\rangle|^2 \geq E|\langle X_1, e_j\rangle|^2 \tag{2.4}$$

for every j. But by assumption,

$$E\|X_{n,1}\|^2 = \sum_{j=1}^{\infty} E[\langle X_{n,1}, e_j\rangle\langle X_{n,1}, e_j\rangle] \to \sum_{j=1}^{\infty} E[\langle X_1, e_j\rangle\langle X_1, e_j\rangle] = E\|X_1\|^2.$$

Hence, the inequality in (2.4) must be an equality for every j. To prove (2.3) for general i, by the continuous mapping theorem, $\langle X_{n,1} + X_{n,1+i}, e_j\rangle$ converges in distribution to $\langle X_1 + X_{1+i}, e_j\rangle$. By the assumptions, $E\|X_{n,1} + X_{n,1+i}\|^2 \to E\|X_1 + X_{1+i}\|^2$. So, the previous argument with $i = 0$ implies

$$E|\langle X_{n,1} + X_{n,1+i}, e_j\rangle|^2 \to E|\langle X_1 + X_{1+i}, e_j\rangle|^2.$$

Expanding the square and applying the case $i = 0$ (and noting $X_{n,1+i}$ has the same distribution as $X_{n,1}$) yields (2.3). We now conclude (using Assumption (2.2)) that

$$\lim_n E[r_1^2(Z_n) - r_N^2(Z_n)] = \sum_{j=1}^{N-1} E[\langle X_1, e_j\rangle\langle X_1, e_j\rangle] + 2\sum_{i=1}^{\infty}\sum_{j=1}^{N-1} E[\langle X_1, e_j\rangle\langle X_{1,1+i}, e_j\rangle].$$

As $N \to \infty$, this last expression tends to L (using the absolute summability of the $E[\langle X_1, X_{1+i}\rangle]$).

Remark 2.1. The assumptions in Theorem 2.1 of the weak convergence of $(X_{n,1},\dots,X_{n,k})$ for each fixed k actually guarantees the existence of a single limiting process $X_1, X_2,\dots$ that satisfies the conditions of the theorem. So, there is no abuse of notation in the theorem as stated. To see why a single process satisfies the conditions, apply a variant of Kolmogorov's Consistency Theorem, such as Theorem 2, p.42 in Bergstrom (1982).

Before writing down limit results, we recall some standard notation. For a stationary series $X = \{X_n, n \in \mathbf{Z}^+\}$, define the α-mixing coefficient (see Rosenblatt (1956)) by $\alpha_X(j) = \sup_{A,B} |P(AB) - P(A)P(B)|$, where A and B vary in the σ-fields generated by $\{X_n, n \le k\}$ and $\{X_n, n \ge j+k\}$, respectively for any $k \ge 1$. The sequence X is said to be α-mixing if $\alpha_X(j) \to 0$ as $j \to \infty$.

Remark 2.2. The assumption $\Sigma_k E(\langle X_1, X_k\rangle) < \infty$ in Theorem 2.1 follows if the process $X_1, X_2, \dots$ is essentially bounded and has α-mixing coefficients $\alpha_X(\cdot)$ that satisfy $\Sigma_j \alpha_X(j) < \infty$. Indeed, a basic inequality (3.1) of Dehling (1983) states (assuming $\|X_j\| \le 1$ almost surely) that $E(\langle X_k, X_{k+j}\rangle) \le 10\alpha_X(j)$. Alternatively, if the process is only assumed to satisfy $E(\|X_1\|^{2+\delta}) < \infty$ for some $\delta > 0$, then a sufficient condition for $\Sigma_k E(\langle X_1, X_k\rangle) < \infty$ is $\Sigma_j [\alpha_X(j)]^{\delta/(2+\delta)} < \infty$; (see (3.2) of Dehling (1983).)

Remark 2.3. The assumptions in Theorem 2.1 yield, for any m, $\Sigma_{k=1}^m E(\langle X_{n,1}, X_{n,k}\rangle) \to \Sigma_{k=1}^m E(\langle X_1, X_k\rangle)$. Hence, to prove Condition (2.1) holds, it suffices to show, given any $\epsilon > 0$, there exists N so that $\lim_{n\to\infty} \Sigma_{k=N}^\infty |E(\langle X_{n,1}, X_{n,k}\rangle)| < \epsilon$. For example, if $\|X_{n,1}\| \le 1$ for all n, then a sufficient condition becomes

$$\lim_{N\to\infty} \limsup_{n\to\infty} \sum_{j=N}^{\infty} \alpha_n(j) = 0, \tag{2.5}$$

where $\alpha_n(j)$ is the jth α-mixing coefficient of the sequence $X_{n,1}, X_{n,2}, \dots$. If $E(\|X_{n,1}\|)^{2+\delta} \le B < \infty$ for some $\delta > 0$, then $\alpha_n(j)$ in (2.5) should be replaced by $[\alpha_n(j)]^{\delta/(2+\delta)}$.

In order to apply Theorem 2.1 to prove a limit result for Z_n, the projections of Z_n must be examined. But, $\langle Z_n, h\rangle = n^{-1/2}\Sigma_{i=1}^n \langle X_{n,i}, h\rangle$ is a sum of weakly dependent, stationary real-valued random variables. Moreover, $\langle Z_n, h\rangle$ has mean 0 and variance

$$\sigma_{n,h}^2 = E|\langle Z_n, h\rangle|^2 = \mathrm{Var}(\langle X_{n,1}, h\rangle) + 2\sum_{i=1}^{n}(1 - \frac{i}{n})\mathrm{Cov}(\langle X_{n,1}, h\rangle, \langle X_{n,1+i}, h\rangle),$$

which, by the assumptions of Theorem 2.1, tends to σ_h^2 defined in (2.2). Hence, Z_n should tend weakly to Z, where Z is asymptotically Gaussian, mean 0, and covariance operator S satisfies $\langle Sh, h\rangle = \sigma_h^2$. However, in order to conclude the limiting distribution of $\langle Z_n, h\rangle$ is actually Gaussian, further assumptions are required. Most of the existing Central Limit Theorems for the mean of weakly dependent stationary real-valued random variables employ mixing conditions and moments assumptions. Moreover, there is a tradeoff between the mixing conditions and the moment assumptions in the sense that if higher moments are assumed, the conditions on the mixing coefficients can be less stringent. In addition, the existing results are typically not stated for triangular arrays. Some of the results that do apply to triangular arrays (e.g. Withers (1981)) assume too strong mixing conditions that are not applicable when we discuss the stationary bootstrap. One could easily adapt proofs of the asymptotic normality of sample means of weakly dependent real-valued sequences of α-mixing variables (or other types of mixing) to sample means in a triangular array setting. This could be accomplished by imposing conditions on $\alpha(j) \equiv \sup_n \alpha_n(j)$, where $\alpha_n(\cdot)$ is the mixing sequence corresponding to the nth row of the triangular array. The conditions imposed on $\alpha(\cdot)$ would then be such that asymptotic normality holds for the sample mean of a dependent sequence with mixing coefficients $\alpha(\cdot)$. Unfortunately, taking such an approach would *not* be useful when we study the stationary bootstrap because, in essence, we will be studying triangular arrays satisfying $\sup_n \alpha_n(j) = 1$ for every j. For now, appealing to Corollary 1 of Withers (1975) does yield the following useful result.

Theorem 2.2. *Assume the $X_{n,j}$ satisfy the conditions of Theorem 2.1, and the nth row of variables $X_{n,1}, X_{n,2}, \ldots$ has α-mixing coefficients denoted by $\alpha_n(\cdot)$. Assume, for all n, $\|X_{n,1}\| \le B$ with probability one, and $\Sigma_{i=1}^{j} i^2 \alpha_n(i) \le K j^r$ for all $1 \le j \le n$ and n and some $r < 3/2$. Then Z_n converges weakly to Z, where Z is Gaussian with mean 0 and covariance operator S satisfies $\langle Sh, h\rangle = \sigma_h^2$, and σ_h^2 is given by (2.2).*

In the special case when the $X_{n,j} = X_j$ form a stationary sequence, the following is true.

Theorem 2.3. *Assume $X_1, X_2, \ldots$ is a stationary sequence of H-valued random variables with mean m and mixing sequence $\alpha_X(\cdot)$. Let $Z_n = n^{-1/2}\Sigma_{i=1}^{n}(X_i - m)$.*
(i) *If $E(\|X_1\|^{2+\delta}) < \infty$ for some $\delta > 0$ and $\Sigma_j [\alpha_X(j)]^{\delta/(2+\delta)} < \infty$, then Z_n converges weakly to a Gaussian measure with mean 0 and covariance operator S satisfying $\langle Sh, h\rangle = \sigma_h^2$, where σ_h^2 is given by (2.2).*
(ii) *If the X_i are essentially bounded and $\Sigma_j \alpha_X(j) < \infty$, then Z_n converges weakly to a Gaussian measure with mean 0 and covariance operator S.*

Proof. Tightness follows in each case by Theorem 2.1 and Remark 2.2. The convergence of finite dimensional distributions follows, for example, from Corollary

8 of Carlstein (1986) under assumptions (i) and Corollary 9 of Carlstein (1986) under assumptions (ii).

Remark 2.4. Almost sure invariance principles for partial sums of H-valued stationary random variables have been obtained by Dehling and Philipp (1982) and Dehling (1983). While these results are stronger than our Theorem 2.3, they depend on heavier assumptions on the mixing sequence $\alpha_X(\cdot)$.

3. The Stationary Bootstrap

3.1. The stationary resampling algorithm

Suppose $\{X_j, j \in \mathbf{Z}\}$ is a strictly stationary and weakly dependent time series, where the X_j may take values in an arbitrary space S. In the mathematical theory developed in this section, S is a separable Hilbert space H, but the stationary resampling scheme applies more generally. Let P_0 be the marginal distribution of X_1. Interest focuses on a parameter $T(P_0)$, where T is some functional of P_0. The case where T is a functional of the m-dimensional marginal distribution of $(X_1, \ldots, X_m)$ can also be considered by a simple reduction to the previous case; just consider a new series Y_i defined by $Y_i = (X_i, \ldots, X_{i+m-1})$. Given data $X_1, \ldots, X_n$, the goal is to make inferences about $T(P_0)$ based on the estimator $T_n = T(\hat{P}_n)$, where $\hat{P}_n$ is the empirical measure constructed from $X_1, \ldots, X_n$. In particular, we are interested in constructing a confidence region for $T(P_0)$ or constructing an estimate of the standard error of the estimator T_n. Typically, an estimate of the sampling distribution of T_n is required, and the stationary bootstrap method proposed here is developed for this purpose. This resampling algorithm is similar to that of Künsch (1989) and Liu and Singh (1992), and has been introduced in Politis and Romano (1991). In general, we are led to considering a "root" or an approximate pivot $r_n = r_n(X_1, \ldots, X_n; T(P_0))$, which is just some functional depending on the data and possibly on $T(P_0)$ as well. For example, r_n might be of the form $r_n = T_n - T(P_0)$. The idea is that if the true sampling distribution of r_n were known, probability statements about r_n could be inverted to yield confidence statements about $T(P_0)$. The stationary bootstrap is a method that can be applied to approximate the distribution of r_n.

To describe the algorithm, let

$$B_{i,b} = \{X_i, X_{i+1}, \ldots, X_{i+b-1}\} \tag{3.1}$$

be the block consisting of b observations starting from X_i. In the case $j > n$, X_j is defined to be X_i, where $i = j \pmod n$ and $X_0 = X_n$. Let p be a fixed number in $[0, 1]$. Independent of $X_1, \ldots, X_n$, let $L_1, L_2, \ldots$ be a sequence of independent and identically distributed random variables having the geometric distribution, so that the probability of the event $\{L_i = m\}$ is $(1-p)^{m-1}p$ for $m = 1, 2, \ldots$.

Independent of the X_i and the L_i, let $I_1, I_2, \ldots$ be a sequence of independent and identically distributed variables which have the discrete uniform distribution on $\{1, \ldots, n\}$. Now, a pseudo time series $X_1^*, \ldots, X_n^*$ is generated in the following way. Sample a sequence of blocks of random length by the prescription $B_{I_1,L_1}, B_{I_2,L_2}, \ldots$. The first L_1 observations in the pseudo time series $X_1^*, \ldots, X_n^*$ are determined by the first block B_{I_1,L_1} of observations $X_{I_1}, \ldots, X_{I_1+L_1-1}$, the next L_2 observations in the pseudo time series are the observations in the second sampled block B_{I_2,L_2}, namely $X_{I_2}, \ldots, X_{I_2+L_2-1}$. This process is stopped once n observations in the pseudo time series have been generated. Once $X_1^*, \ldots, X_n^*$ has been generated, one can compute $r_n(X_1^*, \ldots, X_n^*; T_n)$ for the pseudo time series . The conditional distribution of $r_n(X_1^*, \ldots, X_n^*; T_n)$ given $X_1, \ldots, X_n$ is the stationary bootstrap approximation to the true (unconditional) sampling distribution of $r_n(X_1, \ldots, X_n, T(P_0))$. By repeatedly resampling and simulating a large number B of pseudo time series in the exact same manner, the true distribution of $r_n(X_1, \ldots, X_n; T(P_0))$ can be approximated by the empirical distribution of the B numbers $r_n(X_1^*, \ldots, X_n^*; T_n)$.

An alternative and perhaps simpler description of the resampling algorithm is the following. Let X_1^* be picked at random from the original n observations, so that $X_1^* = X_{I_1}$. With probability p, let X_2^* be picked at random from the original n observations; with probability $1-p$, let $X_2^* = X_{I_1+1}$ so that X_2^* would be the "next" observation in the original time series following X_{I_1}. In general, given that X_i^* is determined by the Jth observation X_J in the original time series, let X_{i+1}^* be equal to X_{J+1} with probability $1-p$ and picked at random from the original n observations with probability p. This description makes it clear that, conditional on $X_1, \ldots, X_n$, the new process $X_1^*, \ldots, X_n^*$ is indeed stationary.

For example, suppose $X_t = \xi_t$ is a real-valued stationary series with correlation ρ at lag one. Let $\hat{\rho}_n = \hat{\rho}_n(\xi_1, \ldots, \xi_n)$ be the usual sample correlation at lag one. Then, the stationary bootstrap approximation to the sampling distribution of $r_n(\xi_1, \ldots, \xi_n; \rho) \equiv \hat{\rho}_n(\xi_1, \ldots, \xi_n) - \rho$ is the distribution, conditional on $\xi_1, \ldots, \xi_n$, of $r_n(\xi_1^*, \ldots, \xi_n^*; \hat{\rho}_n) = \hat{\rho}_n(\xi_1^*, \ldots, \xi_n^*) - \hat{\rho}_n(\xi_1, \ldots, \xi_n)$. This estimated distribution can be approximated by simulation, as described above.

In anticipation of asymptotic results, the parameter p used in the above construction of the resampling scheme will depend on n and be denoted p_n. Denote by $\hat{\alpha}_n(k)$ the mixing sequence associated with the series $X_1^*, X_2^*, \ldots$ based on the parameter p_n. Let us be clear that the probabilities required in calculating $\hat{\alpha}_n(k)$ are conditional on $X_1, \ldots, X_n$. Specifically,

$$\hat{\alpha}_n(k) = \sup_{A,B} |P(AB|X_1, \ldots, X_n) - P(A|X_1, \ldots, X_n)P(B|X_1, \ldots, X_n)|,$$

where events A and B vary in the σ-fields generated, respectively, by $\{X_i^*, i \le j\}$ and $\{X_i^*, i \ge j+k\}$ for any $j \ge 1$.

Proposition 3.1. *Conditional on $X_1, \ldots, X_n$, the pseudo time series $X_1^*, X_2^*, \ldots, X_n^*$ is stationary. Moreover, $\hat{\alpha}_n(k) \leq 4(1-p_n)^k$.*

The stationarity of the X_i^* sequence should be apparent. The bound for $\hat{\alpha}_n(k)$ will not be used in the sequel, and so a proof is omitted.

3.2. The bootstrap central limit theorem

The main theorem of this paper is the following. In the theorem, it is assumed that the X_i are essentially bounded. As in Theorem 2.3, this can easily be weakened, but the case where the X_i are essentially bounded will suffice for our purposes. Indeed, in the statistical applications motivating the problem, the X_i actually represent empirical distribution functions, possibly constructed from a real-valued series ξ_i.

Theorem 3.1. *Let $X_1, \ldots, X_n$ be a stationary sequence of H-valued random variables with mean m and mixing sequence $\alpha_X(\cdot)$. Assume the X_i are essentially bounded and $\Sigma_j \alpha_X(j) < \infty$. Let $\bar{X}_n = n^{-1}\Sigma_{i=1}^n X_i$ and $Z_n = n^{1/2}(\bar{X}_n - m)$; also, let $L(Z_n)$ denote the law of Z_n. Conditional on $X_1, \ldots, X_n$, let $X_1^*, \ldots, X_n^*$ be generated according to the stationary resampling scheme with $p = p_n$ satisfying $p_n \to 0$ and $np_n^2 \to \infty$ as $n \to \infty$. The bootstrap approximation to $L(Z_n)$ is the distribution, conditional on $X_1, \ldots, X_n$, of Z_n^*, where $Z_n^* = n^{1/2}(\bar{X}_n^* - \bar{X}_n)$ and $\bar{X}_n^* = n^{-1}\Sigma_{i=1}^n X_i^*$; denote this distribution by $L(Z_n^*|X_1, \ldots, X_n)$. Then $\rho(L(Z_n), L(Z_n^*|X_1, \ldots, X_n)) \to 0$ in probability, where ρ is any metric metrizing weak convergence on H.*

Proof. Assume without loss of generality that $m = 0$. Observe that, conditional on $X_1, X_2, \ldots$, the variables $X_1^*, \ldots, X_n^*$ are really part of a triangular array of variables; as such, they should perhaps be called $X_{n,1}^*, \ldots, X_{n,n}^*$ in keeping with the notation of Theorem 2.1. However, this notation is not used without risk of confusion.

By Theorem 2.3(ii), Z_n converges weakly to the law of Z, denoted by $L(Z)$, where Z is a Gaussian H-valued random variable with mean 0 and $E(\langle Z, h\rangle)^2 = \sigma_h^2$, where σ_h^2 is given by (2.2). Consider, for any $h \in H$, the projection $\langle Z_n^*, h\rangle = n^{-1/2}\Sigma_{i=1}^n[\langle X_i^*, h\rangle - \langle \bar{X}_n, h\rangle]$, a normalized sum of weakly dependent real-valued stationary random variables. In Politis and Romano (1991), the stationary bootstrap approximation for means is shown to be valid under our conditions. Specifically, the bootstrap approximation, $L(\langle Z_n^*, h\rangle|X_1, \ldots, X_n)$, to the distribution of $n^{-1/2}[\langle \bar{X}_n, h\rangle - \langle m, h\rangle]$ satisfies $\rho_1(L(\langle Z_n^*, h\rangle|X_1, \ldots, X_n), L(\langle Z, h\rangle)) \to 0$ in probability, where ρ_1 is any metric metrizing weak convergence of probability measures on the real line.

We now consider the issue of tightness of the distribution of Z_n^* (conditional on $X_1, \ldots, X_n$) by appealing to Theorem 2.1. Of course, the variables in Theorem 2.1 are centered to have mean 0, so that the triangular array of variables we really consider is $X_1^* - \bar{X}_n, \ldots, X_n^* - \bar{X}_n$. Note that $\bar{X}_n \to 0$ almost surely under the assumed mixing conditions. So, depending on the condition in Theorem 2.1 we need to verify, we may not need to worry about recentering.

First, note that for any fixed $k \geq 1$, the conditional distribution of $(X_1^*, \ldots, X_k^*)$ converges weakly to the distribution of $(X_1, \ldots, X_k)$ for almost all sample sequences $X_1, X_2, \ldots$. To see why, let $\hat{F}_{n,k}$ be the empirical distribution of the $(n-k+1)$ H^k-valued random variables $B_i = (X_i, X_{i+1}, \ldots, X_{i+k-1})$ for $1 \leq i \leq n-k+1$. Then, it is easy to see that the conditional distribution of $(X_1^*, \ldots, X_k^*)$ is of the form $\epsilon_n \hat{F}_{n,k} + (1-\epsilon_n) G_{n,k}$ for some distribution $G_{n,k}$, where

$$\epsilon_n = \Big(1 - \frac{k-1}{n}\Big)(1-p_n)^{k-1}.$$

To see why, X_1^* is X_I with probability $1/n$, where I is chosen at random from $1, \ldots, n$. Then, given $I \leq n-k+1$ and $X_1^* = X_I$, we have $(X_1^*, \ldots, X_k^*) = (X_I, X_{I+1}, \ldots, X_{I+k-1})$ with probability $(1-p_n)^{k-1}$. Since, $p_n \to 0$ and k is fixed here, $\epsilon_n \to 1$ as $n \to \infty$. In the above, $G_{n,k}$ is clearly the distribution of $(X_1^*, \ldots, X_k^*)$ conditional on $X_1, \ldots, X_n$, $I \leq n-k+1$ and $(X_1^*, \ldots, X_k^*) = (X_I, \ldots, X_{I+k-1})$. So, it suffices to show $\hat{F}_{n,k}$ converges weakly to the distribution of $(X_1, \ldots, X_k)$ for almost all sample sequences $X_1, X_2, \ldots$. But, for fixed k, $B_1, \ldots, B_{n-k+1}$ is a stationary sequence of H^k-valued random variables with mixing sequence tending to 0. Hence (by applying the Ergodic theorem or the inequalities of Roussas and Ioannides (1987) which hold under the mixing assumptions of our theorem), if E is any (measurable) subset of H^k, $\hat{F}_{n,k}(E) \to F_k(E) \equiv P((X_1, \ldots, X_k) \in E)$ almost surely. Of course, the exceptional set where this fails may depend on E, but the above holds for all E in some countable collection of sets, and the set where it does not hold for all E in such a countable collection has probability zero. In particular, consider all sets E which are finite intersections of spheres centered at x and radius r, where x varies over a dense subset of H^k and r is rational. By the separability of H^k and Corollary 1 of Billingsley (1968, p.29), this entails the weak convergence of $\hat{F}_{n,k}$ to F_k for almost all sample sequences $X_1, X_2, \ldots$.

Second, in order to invoke Theorem 2.1, we show

$$E^*[\langle X_1^*, X_k^* \rangle] \to E[\langle X_1, X_k \rangle] \tag{3.2}$$

with probability one, where the starred expectation denotes expectation conditional on $X_1, \ldots, X_n$. Now, recall L_1 in the construction of the stationary resampling scheme. Also, with n fixed, let X_j for $j > n$ be defined to be X_{j-n}.

Set $\hat{R}_{n,k} = n^{-1}\Sigma_{j=1}^{n-k+1}\langle X_j - \bar{X}_n, X_{j+k} - \bar{X}_n\rangle$. Then,

$$\begin{aligned} &E^*(\langle X_1^*, X_k^*\rangle) \\ &= E^*(\langle X_1^*, X_k^*\rangle | L_1 \geq k)P(L_1 \geq k) + E^*(\langle X_1^*, X_k^*\rangle | L_1 < k)P(L_1 < k) \\ &= n^{-1}\sum_{j=1}^{n}\langle X_j X_{j+k-1}\rangle(1-p_n)^{k-1} + \|\bar{X}_n^2\|[1-(1-p_n)^{k-1}]. \end{aligned} \tag{3.3}$$

So,

$$\begin{aligned} E^*(\langle X_1^* - \bar{X}_n, X_k^* - \bar{X}_n\rangle) &= n^{-1}(1-p_n)^{k-1}\sum_{j=1}^{n}\langle X_j - \bar{X}_n, X_{j+k-1} - \bar{X}_n\rangle \\ &= (1-p_n)^{k-1}(\hat{R}_{n,k-1} + \hat{R}_{n,n-k+1}). \end{aligned}$$

For fixed k, $|\hat{R}_{n,n-k+1}| \leq k/n \to 0$, and $(1-p_n)^{k-1} \to 1$. Also, $\hat{R}_{n,k-1} \to R_{k-1} \equiv E(\langle X_1, X_k\rangle)$ almost surely, because $\langle X_1, X_k\rangle, \langle X_2, X_{k+1}\rangle, \ldots$ is a stationary strong mixing sequence. Therefore, the convergence (3.2) holds almost surely.

Finally, to show tightness of Z_n^* for almost all sample sequences, it would suffice to show

$$\lim_{n\to\infty}\sum_{k=1}^{n} E^*(\langle X_1^* - \bar{X}_n, X_k^* - \bar{X}_n\rangle) = \sum_{k=1}^{\infty} R_{k-1} \tag{3.4}$$

almost surely. We now show this convergence holds at least in probability. First, note that $\Sigma_{k=1}^{n}(1-p_n)^{k-1}\|\bar{X}_n^2\| \to 0$ in probability, because $n\|\bar{X}_n^2\|$ is tight and $n^{-1}\Sigma_{k=1}^{n}(1-p_n)^{k-1} \leq (np_n)^{-1} \to 0$ by assumption on p_n. This observation, in conjunction with (3.3), shows that (3.4) holds in probability provided

$$\sum_{k=1}^{n}(1-p_n)^{k-1}(C_{n,k-1} + C_{n,n-k+1}) \to \sum_{k=1}^{\infty} R_{k-1}, \tag{3.5}$$

where $C_{n,k} = n^{-1}\Sigma_{j=1}^{n-k+1}\langle X_j, X_{j+k-1}\rangle$. The left side of (3.5) has mean

$$\sum_{k=1}^{n}(1-p_n)^{k-1}\Big[\Big(1-\frac{k}{n}\Big)R_{k-1} + \frac{k}{n}R_{n-k+1}\Big]. \tag{3.6}$$

It is easy to see the first term in (3.6) tends to $\Sigma_k R_{k-1}$, so it suffices to show the second term is negligible. But, the second term can be rewritten (by letting $j = n-k+1$) as

$$\sum_{j=1}^{n}(1-p_n)^{n-j}(n-j+1)R_j/n \leq \sum_{j=1}^{n}(1-p_n)^{n-j}R_j/n.$$

This, in turn can be rewritten as $\Sigma_{j=1}^{J}(1-p_n)^{n-j}R_j/n+\sum_{j=J+1}^{n}R_j/n$. For fixed ϵ, J could be chosen to make the second term less than ϵ. Then, for fixed J, the first term tends to zero because $(1-p_n)^{n-j}\leq(1-p_n)^{n-J}\to 0$ if $np_n\to\infty$, which holds under the assumptions. So, to show (3.4) holds in probability, it suffices to show the variance of the left hand side of (3.5) tends to 0; that is, we must show $\mathrm{Var}(\Sigma_{k=0}^{n-1}b_{n,k}C_{n,k})\to 0$, where $b_{n,k}=(1-p_n)^k+(1-p_n)^{n-k}$. Now,

$$\mathrm{Var}\Big(\sum_{k=0}^{n-1}b_{n,k}C_{n,k}\Big)=\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}b_{n,k}b_{n,j}\mathrm{Cov}(C_{n,k},C_{n,j}) \tag{3.7}$$

and

$$\begin{aligned}\bar{n}^2\mathrm{Var}(C_{n,j})=&\sum_{i=1}^{n-j}\mathrm{Var}(\langle X_i,X_{i+j}\rangle)+2\sum_{i=1}^{n-j}\sum_{l=i+1}^{i+j}\mathrm{Cov}(\langle X_i,X_{i+j}\rangle,\langle X_l,X_{l+j}\rangle)\\&+2\sum_{i=1}^{n-j}\sum_{l=i+j+1}^{n-j}\mathrm{Cov}(\langle X_i,X_{i+j}\rangle,\langle X_l,X_{l+j}\rangle).\end{aligned} \tag{3.8}$$

But, for $i<l\leq i+j$,

$$\begin{aligned}|\mathrm{Cov}(\langle X_i,X_{i+j}\rangle,\langle X_l,X_{l+j}\rangle)|&=|E(\langle X_i,X_{i+j}\langle X_l,X_{l+j}\rangle\rangle)-E^2(\langle X_i,X_{i+j}\rangle)|\\&\leq 10\alpha_X(l-i)+100\alpha_X^2(j)\end{aligned}$$

by repeated use of Dehling's inequality; (see Dehling (1983), Lemma 3.1). Similarly, if $l>i+j$,

$$|\mathrm{Cov}(\langle X_i,X_{i+j}\rangle,\langle X_l,X_{l+j}\rangle)|\leq 10\alpha_X(l-i-j)+100\alpha_X^2(j).$$

Substituting into (3.8) and calling $B=\Sigma_i\alpha_X(i)$ yields $\mathrm{Var}(C_{n,j})\leq n^{-1}(1+40B+400j\alpha_X^2(j))$. But, the summability assumption and monotonicity of the $\alpha_X(\cdot)$ sequence implies $j\alpha_X(j)\to 0$ as $j\to\infty$. (To appreciate why, think of the $\alpha_X(\cdot)$ sequence as tail probabilities $P(X\geq j)$ for some nonnegative, integrable random variable Z; then, a variant of Chebychev's inequality implies $jP(Z\geq j)\to 0$ as $j\to\infty$.) Hence, there is a constant D independent of n and j (which depends only on the $\alpha_X(\cdot)$ sequence) such that $\mathrm{Var}(C_{n,j})\leq D/n$. Hence, applying Cauchy-Schwarz to (3.7) results in

$$\mathrm{Var}\Big(\sum_{k=0}^{n-1}b_{n,k}C_{n,k}\Big)\leq Dn^{-1}\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}b_{n,k}b_{n,j}=O((np_n^2)^{-1})\to 0$$

as $n\to\infty$.

Now, because the convergence (3.4) has been shown to hold in probability and not almost surely, we cannot deduce tightness of Z_n^* for almost all sample sequences. However, given any subsequence n_j, there exists a further subsequence n_{j_k} such that the convergence in (3.4) holds almost surely along this subsequence. This implies $Z^*_{n_{j_k}}$ is tight for almost all sample sequences.

To show $\rho(L(Z^*_n|X_1,\ldots,X_n),L(Z)) \to 0$ in probability, it suffices to show that given any subsequence n_j, there exists a further subsequence where this convergence holds almost surely. But, by the above, there exists a further subsequence such that $Z^*_{n_{j_k}}$ is tight with probability one. Moreover, if necessary, we could extract yet a further subsequence $n_{j_{k_l}}$ satisfying

$$\rho_1(L(\langle Z^*_{n_{j_{k_l}}},h\rangle|X_1,\ldots,X_{n_{j_{k_l}}}),L(\langle Z,h\rangle)) \to 0$$

almost surely. In fact, the exceptional set can be taken so this holds for all h in a countable dense subset of H. But, tightness and convergence of a dense subset of projections entails weak convergence. Thus, $\rho(L(Z^*_{n_{j_{k_l}}}|X_1,\ldots,X_n),L(Z)) \to 0$ almost surely, and the result now follows.

3.3. Confidence limits for stationary time series

By assuming the X_i take values in $H = \mathbf{R}^k$, the previous results imply a bootstrap central limit theorem for a multivariate mean. Using standard arguments, this implies a bootstrap central limit theorem for estimators that are smooth functions of means, such as the construction of joint confidence bands for the first k autocorrelations of the time series. Now, we focus on differentiable functionals. Let $\xi_1,\ldots,\xi_n$ be real-valued observations from a stationary series with mixing sequence $\alpha_\xi(\cdot)$. Let F be the marginal cumulative distribution function (cdf) of ξ_1. Interest now focuses on some functional $T(\cdot)$ of F, as F varies in some class $\mathbf{F}$. (The case where interest focuses on some functional of $(\xi_1,\ldots,\xi_m)$ with m fixed can be handled similarly.) Let $\hat{F}_n$ be the empirical distribution of $\xi_1,\ldots,\xi_n$. Let $\xi^*_1,\ldots,\xi^*_n$ be generated according to the stationary bootstrap resampling scheme, with empirical cdf $\hat{F}^*_n$. We regard $\hat{F}_n - F$ and $\hat{F}^*_n - \hat{F}_n$ as elements of a certain Hilbert space, namely $H = L^2(\nu)$, where ν is a finite measure on the real line (though extensions to the case where ν is only sigma finite can be handled). Assume T is Fréchet differentiable in the sense that, for fixed F, and as G varies in $\mathbf{F}$, $T(G) - T(F) = \langle\psi, G-F\rangle + o(\|G-F\|)$ as $\|G-F\| \to 0$, where $\langle\cdot,\cdot\rangle$ is the $L^2(\nu)$ inner product and $\psi \in L^2(\nu)$.

Corollary 3.1. *Under the above setup, assume* $\Sigma_k\alpha_\xi(k) < \infty$, $np_n^2 \to \infty$, *and* $p_n \to 0$. *Let* L_n *be the (true) distribution function of* $n^{1/2}(T(\hat{F}_n)-T(F))$, *and let* $\hat{L}_n$ *be the distribution function (conditional on* $\xi_1,\ldots,\xi_n$*) of* $n^{1/2}(T(\hat{F}^*_n)-T(\hat{F}_n))$. *Then,* $\rho_1(L_n,\hat{L}_n) \to 0$ *in probability, where* ρ_1 *is any metric metrizing weak convergence of distribution functions on the real line. Moreover,* L_n *converges weakly to a Gaussian distribution with mean* 0 *and variance* σ^2_ψ, *where* σ^2_ψ *is* *given by* (2.2) *with* $X_i(\cdot) = 1(\xi_i \leq \cdot)$. *Let* $\hat{c}_n(1-\alpha) = \inf\{t : \hat{L}_n(t) \geq 1-\alpha\}$. *Then, the asymptotic coverage of the interval* $(T(\hat{F}_n) - n^{-1/2}\hat{c}_n(1-\alpha), \infty)$ *is* $1-\alpha$.

Proof. Apply Theorem 3.1 with $X_i(t) = 1(\xi_i \leq t)$. The rest of the argument is then routine.

3.4. Minimum distance estimation

An important class of estimators that satisfy the assumptions imposed in Subsection 3.3 is the class of minimum distance estimators, where the distance is defined by a Hilbert norm. To define the estimators, let $\{F_\theta, \theta \in \Theta\}$ be a parametric family of distributions on the line, indexed by Θ, an open subset of $\mathbf{R}^d$. Given data $\xi_1, \ldots, \xi_n$ with empirical cdf $\hat{F}_n$, let $\hat{\theta}_n$ satisfy $\inf_\theta \|F_\theta - \hat{F}_n\| = \|F_{\hat{\theta}_n} - \hat{F}_n\|$; here, $\|\cdot\|$ is a certain Hilbert norm. For now, we do not dwell on issues of existence or uniqueness of $\hat{\theta}_n$. The statistical problem is to estimate θ based on some assumed parametric model for the distribution of the ξ_i. Of course, if the true distribution F is not in the parametric family, we are led to estimating a certain minimum distance functional $\theta(F)$. The properties of these estimators have been systematically developed in Millar (1981, 1984). By taking an appropriate decision theoretic point of view, Millar has established desirable robustness properties of such minimum distance estimators. The goal here then is to show that confidence intervals resulting from minimum distance estimators have a certain robustness of validity. That is, confidence intervals constructed by the stationary resampling scheme are asymptotically valid when the data are stationary and weakly dependent. In summary, a statistician using such minimum distance estimators has guarded against both a misspecified model and a certain lack of independence in the data.

Rather than repeat the asymptotic arguments presented in Millar for the i.i.d. case, we summarize the key assumptions needed to reduce the problem to one of studying the empirical process. In this way, it is clear that the key mathematical results needed to extend the distributional results of Millar's to the stationary case are already developed in Theorem 3.1 and Corollary 3.1. Here, we only consider the case $d = 1$ for simplicity, though Corollary 3.1 is strong enough to handle more general situations. Assume the identifiability hypothesis: if $\|F_{\theta_n} - F_\theta\| \to 0$, then $\theta_n \to \theta$. Assume the differentiability hypothesis: there exists a real-valued function η of a real variable such that $\eta \in L^2(\nu)$ and $\|F_\theta - F_{\theta_0} - (\theta - \theta_0)\eta\| = o(|\theta|)$ as $\theta \to \theta_0$. As before, let $Z_n(\cdot) = n^{1/2}[\hat{F}_n(\cdot) - F(\cdot)]$, where F is the true marginal distribution function of ξ_1. Then, the asymptotic arguments presented in Millar show that

$$n^{1/2}(\hat{\theta}_n - \theta) = \langle Z_n, \eta\rangle / \|\eta\|^2 + o_P(1). \tag{3.9}$$

(See (2.8) of Millar (1981) or (2.11) of Millar (1984).) Similarly, one can argue that

$$n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) \approx \langle Z_n^*, \eta \rangle / \|\eta\|^2,$$

where Z_n^* is the corresponding stationary bootstrap process. By Corollary 3.1, both Z_n and Z_n^* behave asymptotically like Z, an $L^2(\nu)$-valued mean 0, Gaussian random variable with the same covariance structure. It follows that the stationary bootstrap approximation to the sampling distribution of $n^{1/2}(\hat{\theta}_n - \theta)$ is asymptotically valid and confidence intervals based on this approximation are asymptotically of the nominal level. (In actuality, a slight generalization of Theorem 3.1 and Corollary 3.1 is required to handle the case that F falls outside of the parametric model.) The expansion (3.9) allows us to identify the so-called influence function of the estimator. First, specialize further to the case of a location model where $F_\theta(t) = F_0(t - \theta)$. Assume F_0 has a density f with respect to Lebesgue measure. Under weak conditions, $\eta(t) = -f(t)$. Then (see Section (3B) of Millar (1981)), $\hat{\theta}_n = n^{-1}\Sigma IC(\xi_i)$, where $IC(t) = b[G(t) - c]$, $\int G(t) = \int_{-\infty}^{t} f(s)\nu(ds)$, $c = \int G(t)F_0(dt)$ and $b^{-1} = \int f^2(t)\nu(dt)$. Thus, if ν is a finite measure, the influence curve $IC(\cdot)$ is monotone and bounded. If ν is nonatomic, then $IC(\cdot)$ is continuous. Also, if ν and f are symmetric about 0, then $IC(\cdot)$ is odd. By varying the choice of ν, Millar (1981) has shown that one can recover the influence curves of famous estimators, such as the class of trimmed means or the Hodges-Lehmann estimator. Thus, even in this specialized model, our asymptotic justification of the stationary bootstrap method applies to a broad range of estimators.

In summary, in order to deduce asymptotic distributional properties of minimum distance estimators defined by a Hilbert norm, it is necessary to develop limit theorems for the empirical process. This has been accomplished in Theorem 3.1 for general H. Hence, for other choices of Hilbert norm considered in Millar (1984), the asymptotic distribution of the resulting estimators can be deduced even when the data are stationary. In the abstract setup considered by Millar (1984), we have verified his convergence hypothesis (2.5) under the assumption the data are stationary and weakly dependent for a large class of estimation problems. Thus, very similar considerations allow us to deduce analogous results for other minimum distance procedures, such as minimum chi-squared methods, minimum Hellinger methods, and other Hilbert distances.

References

Bergstrom, H. (1982). *Weak convergence of measures.* Academic Press, New York.

Billingsley, P. (1968). *Convergence of probability measures.* John Wiley, New York.

Carlstein, E. (1986). Asymptotic normality for a general statistic from a stationary sequence. *Ann. Probab.* **14**, 1371-1379.

De Acosta, A. (1981). Inequalities for B-valued random vectors with applications to the strong law of large numbers. *Ann. Probab.* **9**, 157-161.

Dehling, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. verw. Gebiete* **63**, 393-432.

Dehling, H. and Philipp, W. (1982). Almost sure invariance principles for weakly dependent vector-valued random variables. *Ann. Probab.* **10**, 689-701.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**, 1217-1241.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap* (Edited by LePage and Billard), 225-248, John Wiley, New York.

Millar, P. W. (1981). Robust estimation via minimum distance methods. *Z. Wahrsch. verw. Gebiete* **55**, 73-89.

Millar, P. W. (1984). A general approach to the optimality of minimum distance estimators. *Trans. Amer. Math. Soc.* **286**, 377-418.

Parthasarathy, K. R. (1967). *Probability measures on metric spaces.* Academic Press, New York.

Politis, D. and Romano, J. (1991). The stationary bootstrap. Technical Report 365, Department of Statistics, Stanford University.

Politis, D. N. and Romano, J. P. (1992a). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.* **20**, 1985-2007.

Politis, D. and Romano, J. (1992b). A nonparametric resampling procedure for multivariate confidence regions in time series analysis. *Proceedings of the 22nd Symposium on the Interface* (Edited by C. Page and R. LePage), 98-103, Springer-Verlag, New York.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, New York.

Prokhorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1**, 157-214.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.*, U.S.A. **42**, 43-47.

Roussas, G. G. and Ioannides, D. (1987). Moment inequalities for mixing sequences of random variables. *Stochastic Anal. Appl.* **5**, 61-120.

Varadarajan, V. S. (1958). Weak convergence of measures on separable metric spaces. *Sankyā Ser. A* **19**, 15-22.

Withers, C. S. (1975). Convergence of empirical processes of mixing rv's on $[0, 1]$. *Ann. Statist.* **3**, 1101-1108.

Withers, C. S. (1981). Central limit theorems for dependent variables. I. *Z. Wahrsch. verw. Gebiete* **57**, 509-534.

Department of Statistics, Purdue University, West Lafayette, IN 47907, U.S.A.
Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

(Received January 1993; accepted March 1994)