THE STATIONARY BOOTSTRAP

by

Dimitris N. Politis and Joseph P. Romano

Dimitris N. Politis
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

Joseph P. Romano
Department of Statistics
Stanford University
Stanford, CA 94305

Technical Report # 91-03

Department of Statistics
Purdue University

January, 1991

△

THE STATIONARY BOOTSTRAP

by

Dimitris N. Politis
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

and

Joseph P. Romano
Department of Statistics
Stanford University
Stanford, CA 94305

ABSTRACT

In this article, 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. Previously, a technique based on resampling blocks of consecutive observations was introduced by Künsch (1989) and independently by Liu and Singh (1988) in order to construct confidence intervals for a parameter of the m-dimensional joint distribution of m consecutive observations, where m is fixed. This procedure has been generalized by Politis and Romano (1989, 1990) by constructing a 'blocks of blocks' resampling scheme that yields asymptotically valid procedures even for a multivariate parameter of the whole (infinite-dimensional) joint distribution of the stationary sequence of observations. These methods share the construction of resampling blocks of observations to form a pseudo time series, so that the statistic of interest may be recalculated based on the resampled data set. However, in the context of applying this method to stationary data, it is natural to require the resampled pseudo time series to be stationary (conditional on the original data) as well. While the aforementioned procedures lack this property, the stationary procedure developed here is indeed stationary and possesses other desirable properties. The stationary procedure is based on resampling blocks of random length, where the length of each block has a geometric distribution. In this article, fundamental consistency and weak convergence properties of the stationary resampling scheme are developed.

Key words: Approximate confidence limit; Bootstrap; Stationary; Time Series.

1. Introduction

The bootstrap of Efron (1979) has proven to be a powerful nonparametric tool for approximating the sampling distribution and variance of complicated statistics based on i.i.d. observations. Recently, Künsch (1989) and Liu and Singh (1988) have independently introduced nonparametric versions of the bootstrap and jackknife that are applicable to weakly dependent stationary observations. Their resampling technique amounts to resampling or deleting one by one whole blocks of observations, in order to obtain consistent procedures for a parameter of the the m-dimensional marginal distribution of the stationary series. Their resampling procedure has been generalized in Politis and Romano (1989, 1990) and Politis, Romano and Lai (1990) by resampling 'blocks of blocks' of observations to obtain asymptotically valid procedures even for multivariate parameters of the whole (infinite-dimensional) joint distribution of the stationary time series.

In this article, we introduce a new resampling method, called the stationary bootstrap, that is also generally applicable for stationary weakly dependent time series. Similar to the block resampling techniques, the stationary bootstrap involves resampling the original data to form a pseudo time series from which the statistic or quantity of interest may be recalculated; this resampling procedure is repeated to build up an approximation to the sampling distribution of the statistic. However, in contrast to the aforementioned block resampling methods, the pseudo time series generated by the stationary bootstrap method is actually a stationary time series. That is, conditional on the original data $X_1, \cdots, X_N$, a pseudo time series $X_1^*, \cdots, X_N^*$ is generated by an appropriate resampling scheme which is actually stationary. Hence, this procedure attempts to mimic the original model by retaining the stationarity property of the original series in the resampled pseudo time series. As will be seen, the pseudo time series is generated by resampling blocks of random size, where the length of each block has a geometric distribution. In contrast, the moving blocks procedure of Künsch and Liu and Singh is based on resampling blocks of fixed length.

In Section 2, the actual construction of the stationary bootstrap is presented and comparisons are made with the block resampling method of Künsch and Liu and Singh. Some theoretical properties of the method are investigated in Section 3 in the case of the mean. In Section 4, it is shown how the theory may be extended beyond the case of the mean to construct asymptotically valid confidence regions for general parameters.

2. The Stationary Bootstrap Resampling Scheme

Suppose $\{X_n, n \in \mathbf{Z}\}$ is a strictly stationary and weakly dependent time series, where the X_n are, for now, assumed to be real-valued. The degree of dependence will be quantified by Rosenblatt's α-mixing coefficient, but is left unspecified for now. Suppose μ is a parameter of the whole (infinite-dimensional) joint distribution of the sequence $\{X_n, n \in \mathbf{Z}\}$. For example, μ might be the mean of the process or the spectral distribution function. Given data $X_1, \cdots X_N$, the goal is to make inferences about μ based on some estimator $T_N = T_N(X_1, \cdots, X_N)$. In particular, we are interested in constructing a confidence region for μ 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. In general, we are led to considering a "root" or an approximate pivot $R_N = R_N(X_1, \cdots, X_N; \mu)$, which is just some functional depending on the data and possibly on μ as well. For example, R_N might be of the form $R_N = T_N - \mu$, or possibly a studentized version. The idea is that is the true sampling distribution of R_N were known, probability statements about R_N could be inverted to yield confidence statements about μ. The stationary bootstrap is a method that can be applied to approximate the distribution of R_N.

To describe the algorithm, let

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

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 $i = j(modN)$ and $X_0 = X_N$. Let p be a fixed number in $[0, 1]$. Independent of $X_1, \cdots, X_N$, let $L_1, L_2, \cdots$ 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 $m = 1, 2, \cdots$. Independent of the X_i and the L_i, let $I_1, I_2, \cdots$ be a sequence of independent and identically distributed variables which have the discrete uniform distribution on $\{1, \cdots, N\}$. Now, a pseudo time series $X_1^*, \cdots, X_N^*$ is generated in the following way. Sample a sequence of blocks of random length by the prescription $B_{I_1,L_1}, B_{I_2,L_2}, \cdots$. The first L_1 observations in the pseudo time series

$X_1^*, \cdots, X_N^*$ are determined by the first block B_{I_1,L_1} of observations $X_{I_1}, \cdots, X_{I_1+L_1-1}$, the next L_2 observations in the pseudo time series are the observations in the second sampled block B_{I_2,L_2}, namely $X_{I_2}, \cdots, X_{I_2+L_2-1}$. Of course, this process is stopped once N observations in the pseudo time series have been generated (though it is clear that the resampling method allows for time series of arbitrary length to be generated).

Once $X_1^*, \cdots, X_N^*$ has been generated, one can compute the quantity of interest $T_N(X_1^*, \cdots, X_N^*)$ or $R_N(X_1^*, \cdots, X_N^*; T_N)$ for the pseudo time series. The conditional distribution of $R_N(X_1^*, \cdots, X_N^*; T_N)$ given $X_1, \cdots, X_N$ is the stationary bootstrap approximation to the true (unconditional) sampling distribution of $R_N(X_1, \cdots, X_n, \mu)$. By repeatedly resampling and simulating a large number B of pseudo time series in the exact same manner, the true distribution of $R_N(X_1, \cdots, X_N; \mu)$ can be approximated by the empirical distribution of the B numbers $R_N(X_1^*, \cdots, X_N^*; T_N)$.

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.

The following proposition, though obvious, is of fundamental importance.

Proposition 1. Conditional on $X_1, \cdots, X_N$, the pseudo time series $X_1^*, X_2^*, \cdots, X_N^*$ is stationary.

Of course, much more is actually true. For example, if the original observations $X_1, \cdots, X_N$ are all distinct, then the new series $X_1^*, \cdots X_N^*$ is, conditional on $X_1, \cdots, X_N$, a stationary Markov chain. If, on the other hand, two of the original observations are identical and the remaining are distinct, then the new series $X_1^*, \cdots, X_N^*$ is a stationary second order Markov chain. An obvious generalization, depending on the number of identical subsequences of observations can be made. In fact, if m is the largest b such that, for some i distinct from j (and both i and j between 1 and N), $B_{i,b}$ and $B_{j,b}$ are identical (and $m = 0$ if all observations are distinct), then the series $X_1^*, \cdots, X_N^*$ is a $(m+1)$ order Markov chain.

The stationary bootstrap resampling scheme proposed here is distinct from that proposed by Künsch (1989) and Liu and Singh (1988). Their "moving blocks" method is described as follows. Suppose $N = kb$. Resample with replacement from the blocks $B_{1,b}, \cdots, B_{N-b+1,b}$ to get k resampled blocks, say $B_1^*, \cdots, B_k^*$. The first b observations in the pseudo time series are the sequence of b values in B_1^*, the next b observations in the pseudo time series are the b values in B_2^*, etc. In the case, N is not divisible by b, let k be the smallest integer satisfying $bk > N$. Resample k blocks as above to generate $X_1^*, \cdots, X_{bk}^*$. Now simply delete the observations X_j^* for $j > N$.

Some of the similarities and differences between the stationary bootstrap and the moving blocks bootstrap algorithms should be apparent. To begin, the pseudo time series generated by the moving blocks method is not stationary. Both methods involve resampling blocks of observations. In the moving blocks technique, the number of observations in each block is a fixed number b. In the stationary bootstrap method, the number of observations in each block is random and has a geometric distribution. The methods also differ in how they deal with end effects. For example, since there is no data after X_N, the moving blocks method does not define a block of length b beginning at X_N (if $b > 1$). In order to achieve stationarity for the resampled time series, the stationary bootstrap method "wraps" the data around in a "circle", so that X_1 "follows" X_N. More specific comparisons of mathematical properties will be made later.

Variants on the stationary bootstrap based on resampling blocks of random length are clearly possible. Instead of assuming the L_i have a geometric distribution, one can consider other possible distributions. Moreover, other distributions for the I_i can be employed as well. In this way, the moving blocks may be viewed as a special case. The choice of L_i having a geometric distribution and I_i the discrete uniform distribution was made so that the resampled series is stationary. Of course, there are other possible resampling schemes that achieve stationary for the resamples series. For example, one could take the series $X_1^*, \cdots, X_N^*$ as previously constructed and add an independent series $Z_1^*, \cdots, Z_N^*$ to it, as a "smoothing" device. For the sake of concreteness, attention will focus on the particular scheme initially proposed in this paper.

Another way to think about the difference between the moving blocks method and the stationary bootstrap is the following. For each fixed block size b, one can compute a bootstrap distribution or an estimate of standard error of an estimator. The stationary bootstrap method proposed here is essentially a weighted average of these moving blocks bootstrap distributions or estimates of standard error, where the weights are determined by a geometric distribution. It is important to keep in mind that a difficult aspect in applying these methods is how to choose b in the moving blocks scheme and how to choose p in the stationary scheme. Indeed, the issue becomes a "smoothing" problem.

3. The Mean

In this section, the special case of the sample mean is considered as a first step in order to justify the validity of the stationary bootstrap resampling scheme. Let $\mu = E(X_1)$ and set $T_N(X_1, \cdots, X_N) = \bar{X}_N = N^{-1} \sum_{i=1}^{N} X_i$. Note that, under stationarity, if σ_N^2 is defined to be the variance of $N^{1/2}\bar{X}_N$, then

$$\sigma_N^2 = var(X_1) + 2\sum_{i=1}^{N}(1 - \frac{i}{N})cov(X_1, X_{1+i}). \tag{2}$$

Under the assumption that

$$\sum_{j=1}^{\infty} |cov(X_1, X_j)| < \infty,$$

which is implied by typical assumptions of weak dependence, it follows that $\sigma_N^2 \to \sigma_\infty^2$ as $N \to \infty$, where

$$\sigma_\infty^2 = var(X_1) + 2\sum_{i=1}^{\infty} cov(X_1, X_{1+i}). \tag{3}$$

Moreover, we typically have that $R_N(X_1, \cdots, X_N; \mu) \equiv N^{1/2}(\bar{X}_N - \mu)$ tends in distribution to the normal distribution with mean 0 and variance σ_∞^2. A primary goal of this section is to establish the validity of the stationary bootstrap approximation defined by the conditional distribution of $R_N(X_1^*, \cdots, X_N^*; \bar{X}_N)$ given the data.

As a first step toward this end, and of interest in its own right, we first consider the mean and variance of $N^{1/2}\bar{X}_N^*$ (conditional on the data), where $\bar{X}_N^* = N^{-1} \sum_{i=1}^{N} X_i^*$. Since $E(X_1^*|X_1, \cdots, X_N) = \bar{X}_N$, a trivial consequence of stationarity is $E(\bar{X}_N^*|X_1, \cdots, X_N) = \bar{X}_N$. Since the true distribution of $N^{1/2}(\bar{X}_N - \mu)$ has mean 0, it follows that the bootstrap approximation to the sampling distribution of $N^{1/2}(\bar{X}_N - \mu)$ has the same mean, since both are identically zero.

Remark 1. For the moving blocks scheme, it is not the case that $E(\bar{X}_N^*|X_1, \cdots, X_N) = \bar{X}_N$. To see why, let $A_{i,b}$ be the average of the observations in $B_{i,b}$ defined in (1). Then, except in the uninteresting case $b = N$,

$$E(\bar{X}_N^*|X_1, \cdots, X_N) = (N - b + 1)^{-1} \sum_{i=1}^{N-b+1} A_{i,b} = \frac{1}{(N - b + 1)b} \sum_{i=1}^{N-b+1} \sum_{j=1}^{b} X_{i+j-1}$$

$$= \frac{\sum_{i=1}^{b-1} i(X_i + X_{N-i+1}) + b\sum_{j=b}^{N-b+1} X_j}{(N-b+1)b}. \tag{4}$$

Thus, if $b/N \to 0$ as $N \to \infty$,

$$E(\bar{X}_N^* | X_1, \cdots, X_N) = \bar{X}_N + O_P(b/N).$$

To see why, simply calculate the mean and variance of $E(\bar{X}_N^*|X_1,\cdots,X_N) - \bar{X}_N$ with the aid of (4) or see the proof of (iii) in Theorem 6 of Liu and Singh (1988). In summary, the moving blocks bootstrap approximation to the sampling distribution of $N^{1/2}(\bar{X}_N - \mu)$ does not have mean 0; instead, it has a mean which is $O_P(b/N^{1/2})$ as $N \to \infty$ and $b/N \to 0$. As demonstrated in Liu and Singh (1988), in order to achieve consistency of the moving blocks bootstrap estimate of variance of $N^{1/2}\bar{X}_N$, it is necessary that $b \to \infty$ as $N \to \infty$. Moreover, Künsch (1989) proves that the choice $b \propto N^{1/3}$ is optimal in order to minimize the mean squared error of the moving blocks bootstrap estimate of variance. For such a choice, the moving blocks bootstrap distribution is centered at a location which is $O_P(b/N^{1/2}) = O_P(N^{-1/6})$, which tends to zero quite slowly and may be quite large in finite samples. This order of bias is too large to expect the bootstrap approximation to the sampling distribution of $N^{1/2}(\bar{X}_N - \mu)$ to improve over a normal approximation, for example. Thus, one cannot expect the moving blocks bootstrap to possess any second order optimality properties, at least not without correcting for the bias by recentering the bootstrap distribution. One possibility is to approximate the distribution of $N^{1/2}(\bar{X}_N - \mu)$ by the (conditional) distribution of $N^{1/2}[\bar{X}_N^* - E(\bar{X}_N^*|X_1,\cdots,X_N)]$. Such an approach may be satisfactory in the case of the mean, but it weakens the claim that the bootstrap is supposed to be a general purpose "automatic" technique. Moreover, this approach would not work as well outside the case of the mean. That is, in the general context of estimating a parameter μ by some estimator $T_N = T_N(X_1,\cdots,X_N)$, consider the general approximation to the sampling distribution of $N^{1/2}(T_N - \mu)$ by the (conditional) distribution of $N^{1/2}[T_N^* - E(T_N^*|X_1,\cdots,X_N)]$, where $T_N^* = T_N(X_1^*,\cdots,X_N^*)$. In this case, the approximating bootstrap distribution necessarily has mean 0, and hence does not account for the bias of T_N as an estimator of μ (except in the case where T_N has zero bias).

Remark 2. In fact, if we consider the more general (possibly nonstationary) resampling scheme where the $L_i's$ are i.i.d. with a common (possibly nongeometric) distribution, but the $I_i's$ are i.i.d. uniform on $\{1, \cdots, N\}$, then the conditional mean of $\bar{X}_N^*$ is $\bar{X}_N$. In particular, a close cousin of the moving blocks bootstrap scheme that yields the correct (conditional) mean for the corresponding bootstrap distribution is obtained by letting L_i to be the distribution assigning mass one to a fixed b.

Next, we consider the stationary bootstrap estimate of variance of $N^{1/2}\bar{X}_N$ defined by

$$\hat{\sigma}_{N,p}^2 \equiv var(N^{1/2}\bar{X}_N^*|X_1, \cdots, X_N).$$

In Lemma 1 below,, a formula for $\hat{\sigma}_{N,p}^2$ is obtained. Hence, in this case, the stationary bootstrap estimate of variance, $\hat{\sigma}_{N,p}^2$, may be calculated without actually resampling. In the Lemma, $\hat{\sigma}_{N,p}^2$ is given in terms of the circular autocovariances, defined by

$$\hat{C}_N(i) = \frac{1}{N}\sum_{j=1}^{N}[(X_j - \bar{X}_N)(X_{j+i} - \bar{X}_N)],$$

and the usual covariance estimates

$$\hat{R}_N(i) = \frac{1}{N}\sum_{j=1}^{N-i}[(X_j - \bar{X}_N)(X_{j+i} - \bar{X}_N)].$$

Lemma 1.

$$\hat{\sigma}_{N,p}^2 = \hat{C}_N(0) + 2\sum_{i=1}^{N-1}(1 - \frac{i}{N})(1-p)^i\hat{C}_N(i). \tag{5}$$

Alternatively,

$$\hat{\sigma}_{N,p}^2 = \hat{R}_N(0) + 2\sum_{i=1}^{N-1} b_N(i)\hat{R}_N(i), \tag{6}$$

where

$$b_N(i) = (1 - \frac{i}{N})(1-p)^i + \frac{i}{N}(1-p)^{N-i} \qquad (7).$$

Proof. For purposes of the proof, it is understood that all expectations and covariances are conditional on $X_1, \cdots, X_N$. First, recall L_1 in the construction of the stationary resampling scheme. Then,

$$E(X_1^*X_{1+i}^*) = E(X_1^*X_{1+i}^*|L_1 > i)P(L_1 > i) + E(X_1^*X_{1+i}^*|L_1 \le i)P(L_1 \le i)$$

$$= N^{-1} \sum_{j=1}^{N} X_j X_{j+i}(1-p)^i + \bar{X}_N^2[1-(1-p)^i]$$

because X_1^* and X_{1+i}^* are (conditional on $X_1, \cdots, X_N$) independent given that $L_1 \le i$, but the product $X_1^* X_{1+i}^*$ is equally likely to be any of the N values $X_j X_{j+i}$, $j = 1, \cdots N$. Hence,

$$cov(X_1^*, X_{1+i}^*) = \hat{C}_N(i)(1-p)^i.$$

Therefore, by stationarity,

$$\hat{\sigma}_{N,p}^2 = var(X_1^*) + 2\sum_{i=1}^{N-1}(1-\frac{i}{N})cov(X_1^*, X_{1+i}^*)$$

$$= \hat{C}_N(0) + 2\sum_{i=1}^{N-1}(1-\frac{i}{N})(1-p)^i \hat{C}_N(i),$$

yielding (5). To get (6), note $\hat{R}_N(0) = \hat{C}_N(0)$, and for $i = 1, \cdots, N-1$,

$$\hat{C}_N(i) = \hat{R}_N(i) + \hat{R}_N(N-i).$$

Therefore, by (5),

$$\hat{\sigma}_{N,p}^2 = \hat{R}_N(0) = 2\sum_{i=1}^{N-1}(1-\frac{i}{N})(1-p)^i \hat{R}_N(i) + 2\sum_{i=1}^{N-1}(1-\frac{i}{N})(1-p)^i \hat{R}_N(N-i).$$

Letting $j = N - i$ in the last sum yields

$$\hat{\sigma}_{N,p}^2 = \hat{R}_N(0) + 2\sum_{i=1}^{N-1}(1-\frac{i}{N})(1-p)^i \hat{R}_N(i) + 2\sum_{j=1}^{N-1}\frac{j}{N}(1-p)^{N-j}\hat{R}_N(j)$$

and (6) clearly follows.

Evidently, Lemma 1 tells us that the bootstrap estimate of variance $\hat{\sigma}_{N,p}^2$, given by (6), is closely related to a lag window spectral density estimate of $f(0)$, where $f(\cdot)$ is the spectral density of the original process. Note that the spectral density, defined by

$$f(\omega) = \frac{1}{2\pi}\sum_{r=-\infty}^{\infty} cos(r\omega)R(r), \tag{8}$$

clearly exists if it is assumed that $\sum_r |R(r)| < \infty$. Then, $f(0)$ is simply $\sigma_\infty^2/2\pi$, where σ_∞^2 is given by (3). Hence, it is clear that, accounting for the factor $1/2\pi$, estimating σ_∞^2 is equivalent to estimating $f(0)$. Moreover, estimating σ_N^2 defined by (2) is equivalent to estimating σ_∞^2, in a first order asymptotic sense. The main point here is the bootstrap estimate of variance, $\hat{\sigma}_{N,p}^2$, corresponds to estimating $2\pi f(0)$ by $2\pi \hat{f}(0)$, where $\hat{f}(0)$ is a general lag window spectral density estimator of the form

$$\hat{f}(0) = \frac{1}{2\pi} \sum_{s=-(N-1)}^{N-1} \lambda_N(s) \hat{R}_N(s), \tag{9}$$

where $\lambda_N(s)$ is given by $b_N(s)$ in (7). Such estimates were initially proposed by Grenander and Rosenblatt (1953). In much of the spectral density estimation literature, $\hat{R}_N(s)$ is replaced by $\hat{R}_{N,0}(s)$, where $\hat{R}_{N,0}(s) = \sum_{i=1}^{N-s} X_i X_{i+s}/N$ is appropriately used when it is known that $E(X_j) = 0$. However, in our case, only $\hat{R}_N(s)$ is appropriate. Many other choices for the weight functions $\lambda_N(s)$ have been proposed; see Chapter 6 of Priestley (1981). The bootstrap estimate $\hat{\sigma}_{N,p}^2$ corresponds to a particular form and depends on a choice of p. Thus, the parameter p may be regarded as a smoothing parameter.

We now prove a basic consistency property of $\hat{\sigma}_{N,p}^2$. While many authors have developed general theorems on the consistency properties of spectral estimates, such as Priestley (1981), Zurbenko (1986), and Brillinger (1981), none easily fits in our framework. For example, Priestley (1981) assumes the underlying process is a linear process. Hence, we include a direct proof under weak dependence assumptions on the process. In the theorem below, $\kappa_4(s,r,v)$ is the fourth joint cumulant of the distribution of $(X_j, X_{j+r}, X_{j+s}, X_{j+s+r+v})$. The assumptions of the theorem are similar to those used by Brillinger (1981) and Rosenblatt (1984) in establishing consistency for spectral density estimates.

Theorem 1. Let $X_1, X_2, \cdots$ be a strictly stationary process with covariance function $R(\cdot)$ satisfying $R(0) + \sum_r |rR(r)| < \infty$. Assume

$$\sum_{u,v,w} |\kappa_4(u,v,w)| = K < \infty. \tag{10}$$

Assume $p = p_N \to 0$ and $Np_N \to \infty$. Then, the bootstrap estimate of variance $\hat{\sigma}^2_{N,p_N}$ tends to σ^2_∞ in probability.

Proof. For purposes of the proof, we may assume $E(X_i) = 0$. Let

$$s_N^2 = s_{N,p_N}^2 = \hat{R}_{N,0}(0) + 2\sum_{i=1}^{N-1} b_N(i)\hat{R}_{N,0}(i), \tag{11}$$

where $\hat{R}_{N,0}(i) = \sum_{j=1}^{N-i} X_j X_{j+i}/N$ and $b_N(i)$ depends on $p = p_N$ and is given in (7). Then, it is easily verified by (5) that

$$\hat{\sigma}^2_{N,P} = s^2_{N,P} - \bar{X}_N^2 - 2\bar{X}_N^2 \sum_{i=1}^{N-1} b_N(i).$$

Under the assumed conditions, $\bar{X}_N \to 0$ in probability, and actually $\bar{X}_N = O_P(N^{-1/2})$. Also, $\sum_{i=1}^{N-1} b_N(i) \leq 1/p_N$, which implies

$$\bar{X}_N^2 \sum_{i=1}^{N-1} b_N(i) = O_P(1/Np_N) = o_P(1).$$

Hence, it suffices to show the estimator s_N^2 in (11) satisfies $s_N^2 \to \sigma^2_\infty$ in probability. To accomplish this, we show the bias and variance of s_N^2 tend to zero. By (7) and $E[\hat{R}_{N,0}(i)] = \frac{N-i}{N}R(i)$, it follows that

$$E(s_N^2) = R(0) + 2\sum_{i=1}^{N-1}(1 - \frac{i}{N})b_N(i)R(i)$$

$$= R(0) + 2\sum_{i=1}^{N-1}(1-\frac{i}{N})^2(1-p_N)^i R(i) + 2\sum_{i=1}^{N-1}(1-\frac{i}{N})\frac{i}{N}(1-p_N)^{N-i}R(i).$$

The absolute value of the last term is bounded above by $2\sum_{i=1}^{\infty}|iR(i)|/N = O(N^{-1})$. To handle the first summation, use the approximation $(1-p_N)^i \approx 1 - ip_N$ to get this term is

$$2\sum_{i=1}^{N-1} R(i) - 2p_N \sum_{i=1}^{N-1} iR(i) + o(p_N). \tag{12}$$

Hence, $E(s_N^2) = \sigma^2_\infty + O(p_N)$. To calculate the variance of s_N^2, by the result (5.3.21) of Priestley (1981) originally due to Bartlett (1946),

$$cov[\hat{R}_{N,0}(i), \hat{R}_{N,0}(j)]$$

$$= \frac{1}{N} \sum_{-(N-i)+1}^{N-j-1} [1 - \frac{\eta(m)+j}{N}][R(m)R(m+j-i) + R(m+j)R(m-i) + \kappa_4(m,i,j-i)],$$

where $\eta(m) = m$ if $m > 0$, $\eta(m) = -m-(j-i)$ for $-(N-i)+1 \le m < -(j-i)$, and $\eta(m) = 0$ otherwise. Note that, for the values of m and j considered, $|1 - \frac{\eta(m)+j}{N}| \le 1$. Also, $|R(j)| \le R(0)$ for all j. Hence,

$$cov[\hat{R}_{N,0}(i), \hat{R}_{N,0}(j)] \le \frac{2R(0)}{N} \sum_{m=-\infty}^{\infty} R(m) + \frac{K}{N} = \frac{S}{N},$$

where $S = 2R(0)\sum_{m=-\infty}^{\infty} R(m) + K$. Now,

$$var(s^2_{N,p}) = \sum_{i=-(N-1)}^{N-1} \sum_{-(N-1)}^{N-1} b_N(i) b_N(j) cov[\hat{R}_{N,0}(i), \hat{R}_{N,0}(j)]$$

$$\le \frac{S}{N} \sum_{i=-(N-1)}^{N-1} \sum_{j=-(N-1)}^{N-1} b_N(i) b_N(j) \le \frac{S}{N} \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} [p_N^i + (1-p_N)^i][p_N^j + (1-p_N)^j]$$

$$= \frac{S}{N} \cdot \frac{1 - p_N + p_N^2}{p_N(1-p_N)} \to 0$$

if $Np_N \to \infty$ and $p_N \to 0$. Thus, the result is proved.

In fact, with only slightly more effort, it can be shown that, under the same conditions of Theorem 1, $\hat{\sigma}^2_{N,p_N}$ tends to σ^2_∞ in the sense $E(\hat{\sigma}^2_{N,p_N} - \sigma^2_\infty)^2 \to 0$. The proof actually shows much more. In particular (see (12)),

$$E(\hat{\sigma}^2_{N,p_N}) = \sigma^2_N - 2p_N \sum_{i=1}^{\infty} iR(i) + o(p_N) \tag{13}$$

and $var(\hat{\sigma}^2_{N,p_N}) = O(1/Np_N)$. Consequently, if the goal is to choose $p = p_N$ so that the mean squared error of $\hat{\sigma}^2_{N,p_N}$ as an estimator of σ^2_N is minimized, then the order of the squared bias, p_N^2, should be the same order as the variance, $(Np_N)^{-1}$. This occurs if $p_N \propto N^{-1/3}$. The calculation also points toward the difficulty in choosing p optimally. For if the goal remains minimizing the mean squared error of $\hat{\sigma}^2_{N,p}$, then p_N should satisfy $N^{1/3}p_N \to c$, where the constant c depends on intricate properties of the original process, such as $\sum_i iR(i)$. Estimation of this constant c appears difficult. Fortunately, fundamental consistency properties of the bootstrap are unaffected by not choosing p in an optimal fashion. It appears that it is important to have p tending to 0 at the proper rate to achieve second order properties, but getting the constant c right seems to enter in third order properties.

Note that, the assumption in Theorem 1 of strict stationarity of the series $X_1, \cdots, X_N$ was not used in full force. In fact, the proof shows that the bias of $\hat{\sigma}^2_{N,p_N}$ will tend to 0 if the process is second order stationary, and the variance of $\hat{\sigma}^2_{N,p_N}$ will tend to zero if the process is fourth order stationary.

Remark 3. We now compare the stationary bootstrap estimate of variance, $\hat{\sigma}^2_{N,p}$, with the moving blocks bootstrap estimate of variance. Suppose, for simplicity, that $N = kb$. Then, the moving blocks bootstrap estimate of variance is $k/N \cdot var(X_1^* + \cdots + X_b^* | X_1, \cdots X_N)$, where $(X_1^*, \cdots, X_b^*)$ is a block of fixed length b chosen at random from $B_{1,b}, \cdots B_{N-b,b}$. Except for end effects, the moving blocks bootstrap estimate of variance is equivalent to $m^2_{N,b} \equiv b^{-1} var(S_{I,b} | X_1, \cdots, X_N)$, where $S_{i,b}$ is the sum of the observations in $B_{i,b}$ defined in (1),, and I is chosen at random from $\{1, \cdots, N\}$. By an argument similar to Lemma 1,

$$m^2_{N,b} = \hat{C}_N(0) + 2 \sum_{i=1}^{b-1} (1 - \frac{i}{b}) \hat{C}_N(i). \tag{14}$$

Comparing $m^2_{N,b}$ with $\hat{\sigma}^2_{N,p}$ in (5), the two are quite close, in view of the approximation $(1 - iN^{-1})(1-p)^i \approx 1 - ip$, provided p^{-1} is approximately b. Intuitively, the stationary bootstrap scheme samples blocks of random length $1/p$, so the two approaches are roughly the same if the expected number of observations in each resampled block is the same for both methods. In fact, (14) shows that the moving blocks and stationary bootstrap variance estimates are approximately equivalent to a lag window spectral estimate using Bartlett's kernel. A perhaps more interesting way to view the two variance estimates is the following. One can compute $m^2_{N,b}$ defined by (14) for each b and then average over a distribution of b values. In particular, compute $E(m^2_{N,B})$, where B (independently) has a geometric distribution with mean p_N^{-1}, yielding

$$E(m^2_{N,B}) = \hat{C}_N(0) + 2 \sum_{b=1}^{\infty} \sum_{i=1}^{b-1} (1 - \frac{i}{b})(1 - p_N)^{b-1} p_N \hat{C}_N(i)$$

$$= \hat{C}_N(0) + 2\sum_{i=1}^{\infty}\sum_{b=i+1}^{\infty}(1-\frac{i}{b})(1-p_N)^{b-1}p_N\hat{C}_N(i)$$

$$= \hat{C}_N(0) + 2\sum_{i=1}^{\infty}\tilde{b}_N(i)\hat{C}_n(i),$$

where $\tilde{b}_N(i) = (1-p_N)^i[1-(1-p_N)^{-1}ip_N log(p_N)]$. Because $1-iN^{-1} \approx 1$ and $p_N log(p_N) \to 0$ as $p_N \to 0$, $\tilde{b}_N(i) \approx b_N(i)$, where $b_N(i)$ is given in (7). Hence, the stationary bootstrap estimate of variance may be viewed as a weighted average over b of estimates of variance based on resampling blocks of fixed length b, suggesting that the choice of p in the stationary scheme is less crucial than the choice of b in the moving blocks scheme. Moreover, it may be shown by an argument similar to Theorem 1 that

$$var(\hat{\sigma}^2_{N,p_N} - m^2_{N,b_N}) \to 0$$

if $b = b_N = 1/p_N$, and the conditions of Theorem 1 are satisfied. The same claim can be made if $m^2_{N,b}$ is replaced by the exact moving blocks estimate of variance. To further substantiate the claim that $m^2_{N,b} \approx \hat{\sigma}^2_{N,p}$ if $p = 1/b$, note that Künsch's expansion for the bias of the moving blocks estimate of variance exactly coincides with (13).

We now take up the problem of estimating the distribution of $N^{1/2}(\bar{X}_N - \mu)$, with the goal of constructing confidence intervals for μ. A strong mixing assumption on the original process will be in force. That is, it is assumed that data $X_1, \cdots, X_N$ are observed from an infinite sequence $\{X_n, n \in Z\}$. Let $\alpha_X(k) = sup_{A,B}|P(AB) - P(A)P(B)|$, where A and B vary over events in the σ-fields generated by $\{X_n, n \le 0\}$ and $\{X_n, n \ge k\}$, respectively.

The bootstrap approximation to the sampling distribution of $N^{1/2}(\bar{X}_N - \mu)$ is the distribution of $N^{1/2}(\bar{X}^*_N - \bar{X}_N)$, conditional on $X_1, \cdots, X_N$. The fundamental consistency result is the following.

Theorem 2. Let $X_1, X_2, \cdots$ be a strictly stationary process with covariance function $R(\cdot)$ satisfying $R(0) + \sum_r |rR(r)| < \infty$. Assume (10) in Theorem 1. Assume, for some $d > 0$, that $E|X_i|^{d+2} < \infty$ and $\sum_k [\alpha_X(k)]^{d/(2+d)}$. Then, σ^2_∞ given in (3) is finite. Moreover, if

$\sigma_\infty > 0$, then

$$sup_x|P\{N^{1/2}(\bar{X}_N - \mu) \le x\} - \Phi(x/\sigma_\infty)| \to 0, \tag{15}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Assume $p_N \to 0$ and $Np_N \to \infty$. Then, the bootstrap distribution is uniformly close to the true sampling distribution in the sense:

$$sup_x|P\{N^{1/2}(\bar{X}_N^* - \bar{X}_N) \le x|X_1, \cdots, X_N\} - P\{N^{1/2}(\bar{X}_N - \mu) \le x\}| \to 0 \tag{16}$$

in probability.

Proof. For purposes of the proof, we may assume without loss of generality that $\mu = 0$. The result (15) follows immediately from Corollary 5.1 of Hall and Heyde (1980). We must prove a central limit theorem for the bootstrap distribution, namely the distribution of $N^{1/2}(\bar{X}_N^* - \bar{X}_N)$, conditional on $X_1, \cdots, X_N$.

For now, assume the following three convergences hold for the sequence $X_1, X_2, \cdots$.

(C1). $N\bar{X}_N^2/(Np_N) \to 0$.

(C2). $\hat{C}_N(0) + 2\sum_{i=1}^{\infty}(1-p_N)^i\hat{C}_n(i) \to \sigma_\infty^2$.

(C3).

$$\frac{p_N}{N^{1+\delta/2}}\sum_{r=1}^{\infty}\sum_{i=1}^{N}|S_{i,r} - \frac{\bar{X}_N}{p_N}|^{2+\delta}(1-p_N)^{r-1}p_N \to 0.$$

In (C3), $S_{i,b}$ is defined to be the sum of observations in $B_{i,b}$ defined in (1). The following claim is the basis of the proof.

Claim. The distribution of $N^{1/2}(\bar{X}_N^* - \bar{X}_N)$, conditional on $X_1, \cdots, X_N$, tends weakly to the normal distribution with mean 0 and variance σ_∞^2, for every sequence $X_1, X_2, \cdots$ satisfying (C1), (C2),and (C3).

The proof of this claim will be given below in five steps. In the proof, all calculations referring to this bootstrap distribution will be assumed conditional on $X_1, \cdots, X_N$. The following terminology will be used. Set

$$E_{N,m} = \frac{1}{N} \cdot (S_{I_1,L_1} + \cdots + S_{I_m,L_m}), \tag{17}$$

where, as in the construction of the stationary bootstrap resampling scheme, the $I_1, I_2, \cdots$ are i.i.d. uniform on $\{1, \cdots, N\}$ and the $L_1, L_2, \cdots$ are i.i.d. geometric with mean $1/p_N$. Let M be the random variable equal to the smallest integer m such that $L_1 + \cdots + L_m \geq N$. Also, let $J_1 = L_1 + \cdots L_{M-1}$ and $J = L_M + J_1$. Then, $\bar{X}_N^*$ is approximately $E_{N,M}$. In fact, the difference $E_{N,M} - \bar{X}_N^*$ is just N^{-1} times the sum of the observations in B_{I_M,L_M}, after deleting the first $N - J_1$ of them. Let R_1 be the exact number of observations required from block B_{I_M,L_M} so that N observations from the M blocks have been sampled; that is, $R_1 = N - J_1$. Also, let $R = L_M - R_1$. A key observation is R, conditional (and unconditional) on (R_1, J_1), has a geometric distribution with mean $1/p_N$. This follows from the "memoryless" property of the geometric distribution. Hence, $E_{N,M} - \bar{X}_N^*$ is equal in distribution to $N^{-1}S_{I,R}$, where I is uniform on $\{1, \cdots N\}$ and R is geometric with mean $1/p_N$.

Step 1. Show that $N^{1/2}(E_{N,M} - \bar{X}_N^*) \to 0$ in (conditional) probability. By the above observation, it is enough to show the mean and variance of $N^{-1}S_{I,R}$ tends to 0. But, $E[S_{I,R}|R] = R\bar{X}_N$, so that $N^{-1/2}E(S_{I,R}) = N^{1/2}\bar{X}_N/(Np_N) \to 0$. Now,

$$N^{-1}var(S_{I,R}) = N^{-1}E[var(S_{I,R}|R] + N^{-1}var[E(S_{I,R}|R)]. \tag{18}$$

But, $var[E(S_{I,R}|R)] = var(R\bar{X}_N) = \bar{X}_N^2(1 - p_N)/p_N^2$. Thus, by (C1) and $Np_N \to \infty$, $N^{-1}E[var(S_{I,R})|R] \to 0$, yielding $N^{-1}var(S_{I,R}) \to 0$ as well.

Step 2. Show that, for any fixed sequence $m = m_N$ satisfying $Np_N/m_N \to 1$, the distribution of

$$N^{1/2}(E_{N,m_N} - \frac{m_N\bar{X}_N}{Np_N}) \tag{19}$$

tends to the normal distribution with mean 0 and variance σ_∞^2. First, note $E(S_{I_i,L_i}) = \bar{X}_N\, p_N$. For $1 \leq i \leq m_N$, let $Y_{N,i} = m_N^{1/2}S_{I_i,L_i}/N^{1/2}$. Then, (19) is $m_N^{1/2}[\bar{Y}_{m_N} - E(\bar{Y}_{m_N})]$, and $\bar{Y}_{m_N} = \sum_{i=1}^{m_N} Y_{N,i}/m_N$ is the average of i.i.d. variables. But, as in step 1, $var(Y_{N,i})$ is the same as the variance of m_N/N times the variance of $S_{I,R}$, where I is uniform on $\{1, \cdots, N\}$ and R is geometric with mean p_N. Again, apply the relationship

$$var(S_{I,R}) = E[var(S_{I,R}|R)] + var[E(S_{I,R}|R)]. \tag{20}$$

The second term on the right side of (20) is $var(\bar{X}_N/p_N) = 0$. Also, $r^{-1}var(S_{I,R}|R = r)$ is, in fact, given by $m_{N,r}^2$ defined in (14). Thus,

$$\frac{m_N}{N}var(S_{I,R}) = \frac{m_N}{N}E(Rm_{N,R}^2) = \frac{m_N}{N}\{\sum_{r=1}^{\infty}[\hat{C}_N(0) + 2\sum_{i=1}^{r-1}(1 - \frac{i}{r}\hat{C}_N(i)]r(1-p_N)^{r-1}p_N\}$$

$$= \frac{m_N}{Np_N}\hat{C}_N(0) + \frac{2m_N}{Np_N}\sum_{i=1}^{\infty}(1-p_N)^i\hat{C}_N(i).$$

By the assumption $Np_N/m_N \to 1$ and (C2), it follows that $var(Y_{N,i}) \to \sigma_\infty^2$. To complete step 2, by Katz's (1963) Berry-Esseen bound, it suffices to show

$$m_N^{-\delta/2}E|Y_{N,i} - E(Y_{N,i})|^{2+\delta} \to 0 \tag{21}$$

as $m_N \to \infty$. But, the left side of (21) is (by conditioning on R) equal to

$$\frac{m_N}{N^{1+\delta/2}}E|S_{I,R}|^{2+\delta} = \frac{m_N}{N^{2+\delta/2}}\sum_r\sum_i|S_{i,r} - \frac{\bar{X}_N}{p_N}|^{2+\delta}(1-p_N)^{r-1}p_N,$$

which tends to 0 by (C3).

Step 3. The distribution of $N^{1/2}(E_{N,m_N} - \bar{X}_N)$ tends to normal with mean 0 and variance σ_∞^2. This follows by step 2 and (C1).

Step 4. The distribution of $N^{1/2}(E_{N,M} - \bar{X}_N)$ tends to normal with mean 0 and variance σ_∞^2. To see why, if $\tilde{M}$ is any random variable (sequence) satisfying $\tilde{M}/Np_N \to 1$ in probability, then $N^{1/2}(E_{N,\tilde{M}} - \bar{X}_N)$ tends to normal with mean 0 and variance σ_∞^2. This essentially follows by an extension of Theorem 7.3.2. (to a triangular array setting) of Chung (1974). In our case, $M = Np_N + O_P(N^{1/2}p_N^{1/2})$.

Step 5. Combine steps 1 and 4 to prove the claim.

Now, to deduce (16), by a subsequence argument, it suffices to show the convergences (C1), (C2) and (C3) hold in probability for the original sequence $X_1, X_2, \cdots$. First, (C1) holds in probability because $N\bar{X}_N^2$ is order one in probability and $Np_N \to \infty$. Second, the convergence (C2) holds in probability by an argument very similar to Theorem 1. Finally, to show (C3) holds in probability, write the term in question as

$$\frac{p_N}{N^{\delta/2}} E|S_{I,R} - \frac{\bar{X}_N}{p_N}|^{2+\delta}. \tag{22}$$

It suffices to show that (22) raised to the power $(2+\delta)^{-1}$ tends to 0 in probability, which by Minkowski's inequality is bounded above by

$$\left(\frac{p_N}{N^{\delta/2}}\right)^{\frac{1}{2+\delta}} [E|S_{I,R} - R\bar{X}_N|^{2+\delta}]^{\frac{1}{2+\delta}} + \left(\frac{p_N}{N^{\delta/2}}\right)^{\frac{1}{2+\delta}} \bar{X}_N E[|R - p_N^{-1}|^{2+\delta}]^{\frac{1}{2+\delta}}. \tag{23}$$

The second term in (23) is of order $\bar{X}_N N^{1/2}[Np_N]^{-(1+\delta)/(2+\delta)}$, which tends to 0 in probability. It now suffices to show

$$\frac{p_N}{N^{\delta/2}} E|S_{I,R} - R\bar{X}_N|^{2+\delta} \to 0$$

in probability, or that its expectation tends to zero; that is,

$$\frac{p_N}{N^{1+\delta/2}} \sum_r \sum_i E[|S_{i,r} - r\bar{X}_N|^{2+\delta}](1-p_N)^{r-1} p_N \to 0. \tag{24}$$

In order to bound $E|S_{i,r} - r\bar{X}_N|^{2+\delta}$, note that if $1 \le i \le i+r-1 \le N$, then Yokoyama's (1980) moment inequality applies, yielding $E|S_{i,r}|^{2+\delta} \le Kr^{1+\frac{\delta}{2}}$, where the constant K depends only on the mixing sequence $\{\alpha(k)\}$. Thus, by Minkowski's inequality and then Yokoyama's inequality, we have

$$E|S_{i,r} - r\bar{X}_N|^{2+\delta} \le \left[K^{\frac{1}{2+\delta}} r^{(1+\frac{\delta}{2})(\frac{1}{2+\delta})} + \left(E|r\bar{X}_N|^{2+\delta}\right)^{\frac{1}{2+\delta}}\right]^{2+\delta}$$

$$\le \left[K^{\frac{1}{2+\delta}} r^{(1+\frac{\delta}{2})(\frac{1}{2+\delta})} + \frac{r}{N} K N^{(1+\frac{\delta}{2})(\frac{1}{2+\delta})}\right]^{2+\delta} \le (2K)^{2+\delta} r^{\frac{1+\delta}{2}}.$$

In the case $i+r-1 > N$ but $r < N$, write $S_{i,r} = (X_i + \cdots + X_N) + (X_1 + \cdots + X_{i+r-1-N})$. Apply Minkowski's inequality and Yokoyama's inequality to get $E|S_{i,r}| \le 2^{2+\delta} K r^{1+\frac{\delta}{2}}$. Then, arguing as above we find $E|S_{i,r} - r\bar{X}_N|^{2+\delta} \le (3K)^{2+\delta} r^{1+\frac{\delta}{2}}$. In the general case, suppose $r + N(j-1) + \tilde{r}$, where $1 \le \tilde{r} \le N$. Then, $S_{i,r} = (j-1)N\bar{X}_N + S_{i,\tilde{r}}$. So,

$$E|S_{i,r} - r\bar{X}_N|^{2+\delta} = E|S_{i,\tilde{r}} - \tilde{r}\bar{X}_N|^{2+\delta},$$

and the general bound $(3K)^{2+\delta}r^{1+\frac{\delta}{2}}$ applies. Hence, (24) is bounded above by

$$\frac{p_N}{N^{\delta/2}}\sum_{r=1}^{\infty}(3K)^{2+\delta}r^{1+\delta/2}(1-p_N)^{r-1}p_N = O\left(\frac{p_N}{N^{\delta/2}}\cdot\frac{1}{p_N^{1+\frac{\delta}{2}}}\right) = o(1),$$

and the proof is complete.

Remark 4. In Theorems 1 and 2, the condition (10) is implied by $E|X_i|^{6+\epsilon} < \infty$ and $\sum_k k^2[\alpha(k)]^{[\epsilon/(6+\epsilon)]} < \infty$. To appreciate why, see (A.1) of Künsch (1989). Hence, the conditions for Theorem 2 may be expressed solely in terms of a mixing condition and moment condition, without referring to cumulants. In summary, assume for some $\epsilon > 0$ that $E|X_i|^{6+\epsilon} < \infty$. Then, the mixing conditions are implied by the single mixing condition $\alpha_X(k) = O(k^{-r})$ for some $r > 3(6+\epsilon)/\epsilon$. This condition also implies $\sum_r |rR(r)| < \infty$.

The immediate application of Theorem 2 lies in the construction of confidence intervals for μ. For example, let $\hat{q}_N(1-\alpha)$ be obtained from the bootstrap distribution by

$$P\{\bar{X}_N^* - \bar{X}_N \leq \hat{q}_N(1-\alpha)\} = 1-\alpha. \tag{25}$$

Due to possible discreteness or uniqueness problems, $\hat{q}_N(1-\alpha)$ should be defined to be the $1-\alpha$ quantile of the (conditional) distribution of $\bar{X}_N^* - \bar{X}_N$; in general, take as definition the $1-\alpha$ quantile of an arbitrary distribution G to be $inf\{q : G(q) \geq 1-\alpha\}$. Then, it immediately follows that the bootstrap interval

$$[\bar{X}_N - \hat{q}_N(1-\frac{\alpha}{2}), \bar{X}_N - \hat{q}_N(\frac{\alpha}{2})]$$

has asymptotic coverage $1-\alpha$. Indeed, the theorem implies $\hat{q}_N(1-\alpha) \to \sigma_\infty\Phi^{-1}(1-\alpha)$ in probability.

Other bootstrap confidence intervals may similarly be shown to be asymptotically valid in the sense of having the correct asymptotic coverage, such as a simple percentile method. The bootstrap-t method would require approximating the distribution of $(\bar{X}_N - \mu)/\sigma_N(X_1,\cdots,X_N)$, by the (conditional) distribution of $(\bar{X}_N^* - \bar{X}_N)/\sigma_N(X_1^*,\cdots,X_N^*)$, where $\sigma_N(X_1,\cdots,X_N)$ is some estimate of σ_N such as the bootstrap estimator $\hat{\sigma}_{N,p}$.

Because of Theorem 2, the validity of the bootstrap would follow once it is shown that, conditional on $X_1, X_2, \cdots$, $\sigma_N(X_1^*, \cdots, X_N^*)$ converges in probability to σ_∞. In fact, an argument similar to Theorem 1 could establish this result.

In practice, it seems inevitable that a data-based choice for p would be made. For example, as previously mentioned, if p is chosen to minimize the mean squared error of $\hat{\sigma}^2_{N,p}$, then p should satisfy $N^3 p_N \to C$. The constant C will depend on the spectral density and can be estimated consistently, say by some sequence $\hat{C}_N$. One could then choose $\hat{p}_N = N^{-1/3}\hat{C}_N$. In fact, with some additional effort, Theorem 2 can be generalized to consider a data-based choice for p. One would expect that a data-based choice, $\hat{p}_N$, for p would have to satisfy $N\hat{p}_N \to \infty$ in probability and $\hat{p}_N \to 0$ in probability, in order for the bootstrap central limit theorem to remain valid. This assumption is clearly satisfied for the above "optimal" construction for $\hat{p}_N$. Subsequent work will focus on a proper choice of p. At this stage, it is clear that as long as p satisfies $p \to 0$ and $Np \to \infty$, the choice of p will not enter into first order properties, such as coverage error, of the stationary bootstrap procedure. Getting the right rate for p to tend to 0 will undoubtedly enter into second order properties, but getting "optimal" constants correct will be a third order consideration. Such an investigation, though of vital importance, is beyond the scope of the present work. A step toward understanding second order properties is presented in Lahiri (1990) in the case of moving blocks bootstrap.

4. Extensions.

In this section, we extend the results in Section 3 to more general parameters of interest. A basic theme in this section is that results about the sample mean readily imply results for much more complicated statistics.

4.1. Multivariate Mean. Suppose the X_i take values in $\mathbf{R}^d$, with jth component denoted $X_{i,j}$. Interest focuses on the mean vector, $\mu = E(X_i)$, having jth component $\mu_j = E(X_{i,j})$. The definition of $\alpha_X(\cdot)$ readily applies to the multivariate case. As before, the stationary resampling algorithm is the same, yielding a pseudo multivariate time series $X_1^*, \cdots, X_N^*$ with mean vector $\bar{X}_N^*$.

Theorem 3. Suppose, for some $\epsilon > 0$, that $E|X_{i,j}|^{6+\epsilon} < \infty$. Assume that $\alpha_X(k) = O(k^{-r})$ for some $r > 3(6+\epsilon)/\epsilon$. Then, $N^{1/2}(\bar{X}_N - \mu)$ tends in distribution to the multivariate Gaussian distribution with mean 0 and covariance matrix $\Sigma = (\sigma_{i,j})$, where

$$\sigma_{i,j} = cov(X_{1,i}, X_{1,j}) + 2\sum_{k=1}^{\infty} cov(X_{1,i}, X_{1+k,j}).$$

Then, if $p_N \to 0$ and $Np_N \to \infty$,

$$sup_s|P^*\{\|\bar{X}_N^* - \bar{X}_N\| \leq s\} - P\{\|\bar{X}_N - \mu\| \leq s\}| \to 0 \tag{26}$$

in probability, where $\|\cdot\|$ is any norm on $\mathbf{R}^d$ and P^* refers to a probability conditional on the original series.

Proof. The proof follows immediately by considering linear combinations of the components and applying Theorem 2, which is applicable by Remark 4. Then, (26) follows by the continuous mapping theorem (since a norm is almost everywhere continuous with respect to a Gaussian measure).

The immediate application of the theorem is the construction of joint confidence regions for $\mu = (\mu_1, \cdots, \mu_d)$. Various choices for the norm yield different shaped regions.

Notice how easily the bootstrap handles the problem of constructing simultaneous confidence regions. An asymptotic approach would involve finding the distribution of the norm of a multivariate Gaussian random variable having a complicated (unknown) covariance structure. The resampling approach avoids such a calculation and handles all norms with equal facility.

4.2. Smooth Function of Means.

Again, suppose the X_i take values in $\mathbf{R}^d$. Suppose $\theta = (\theta_1, \cdots, \theta_p)$, where $\theta_j = E[h_j(X_i)]$. Interest focuses on θ or some function f of θ. Let $\hat{\theta}_N = (\hat{\theta}_{N,1}, \cdots, \hat{\theta}_{N,j})$, where $\hat{\theta}_{N,j} = \sum_{i=1}^N h_j(X_i)/N$. Assume moment conditions of the h_j and mixing conditions on the X_i. Then, by the multivariate case, the bootstrap approximation to the distribution of $N^{1/2}(\hat{\theta}_N - \theta)$ is appropriately close in the sense

$$d\left(P\{N^{1/2}(\hat{\theta}_N - \theta) \le x\}, P^*\{N^{1/2}(\hat{\theta}_N^* - \hat{\theta}_N) \le x\}\right) \to 0 \tag{27}$$

in probability, where d is any metric metrizing weak convergence in $\mathbf{R}^p$. Moreover,

$$d\left(P\{N^{1/2}(\hat{\theta}_N - \theta) \le x\}, P\{Z \le x\}\right) \to 0, \tag{28}$$

where Z is multivariate Gaussian with mean 0 and covariance matrix Σ having (i,j) component

$$cov(Z_i, Z_j) = cov[h_i(X_1), h_j(X_1)] + 2\sum_{k=1}^{\infty} cov[h_i(X_1), h_j(X_{1+k})].$$

To see why, define Y_i to be the vector in $\mathbf{R}^p$ with jth component $h_j(X_i)$. Then, the Y_i are weakly dependent if the original X_i are weakly dependent; in fact, $\alpha_Y(k) \le \alpha_X(k)$. Hence, with a moment assumption on the h_i, we are exactly back in the multivariate case. Now, suppose f is an appropriately smooth function from $\mathbf{R}^p$ to $\mathbf{R}^q$, and interest now focuses on the parameter $\mu = f(\theta)$. Assume $f = (f_1, \cdots, f_q)$, where $f_i(y_1, \cdots, y_p)$ is a real-valued function from $\mathbf{R}^p$ having a nonzero differential at $(y_1, \cdots, y_p) = (\theta_1, \cdots, \theta_p)$. Let D be the $p \times q$ matrix with (i,j) entry $\partial f(y_1, \cdots, y_p)/\partial y_j$ evaluated at $(\theta_1, \cdots, \theta_p)$. Then, the following is true.

Theorem 4. Suppose f satisfies the above smoothness assumptions. Assume, for some $\epsilon > 0$, $E[h_j(X_1)]^{6+\epsilon} < \infty$, and $\alpha_X(k) = O(k^{-r})$ for some $r > 3(6+\epsilon)/\epsilon$. Then, if $p_N \to 0$ and $Np_N \to \infty$, (27) and (28) hold. Moreover,

$$d\left(P\{N^{1/2}[f(\hat{\theta}_N) - f(\theta)] \le x\}, P^*\{N^{1/2}[f(\hat{\theta}_N^*) - f(\hat{\theta}_N)] \le x\}\right) \to 0 \qquad (29)$$

in probability and

$$sup_s \left| P\{\|f(\hat{\theta}_N) - f(\theta)\| \le s\} - P^*\{\|f(\hat{\theta}_N^*) - f(\hat{\theta}_N)\| \le s\} \right| \to 0 \qquad (30)$$

in probability.

The proof follows as (27) and (28) are immediate from Theorem 3, and the smoothness assumptions on f imply $N^{1/2}[f(\hat{\theta}_N) - f(\theta)]$ has a limiting multivariate Gaussian distribution with mean 0 and covariance matrix $D\Sigma D'$; see Theorem A of Serfling (1980), p.122.

As an immediate application, consider the problem of constructing uniform confidence bands for $(R(1), \cdots, R(q))$, where $R(i) = cov(X_1, X_{1+i})$. (To apply the previous theorem, let $W_i = (X_i, \cdots, X_{i+q})$, for $1 \le i \le N' = N - q$.) While even asymptotic distribution theory for even Gaussian data seems formidable, the stationary bootstrap resampling approach handles the problem easily. The only caveat is to note that q is fixed as $N \to \infty$.

4.3. Differentiable Functionals.

For simplicity, assume the X_i are real-valued with common continuous distribution function F. Suppose the parameter of interest μ is some functional T of F. A sensible estimate of F is $T(\hat{F}_N)$, where $\hat{F}_N$ is the empirical distribution of $X_1, \cdots, X_N$. Assume T is Frechet differentiable; that is, suppose

$$T(G) = T(F) + \int h_F d(G - F) + o(\|G - F\|), \qquad (31)$$

for some (influence) function h_F, centered so $\int h_F dF = 0$. For concreteness, suppose $\|\cdot\|$ is the supremum norm, but this can be generalized. Then,

$$N^{1/2}[T(\hat{F}_N) - T(F)] = N^{-1/2} \sum_{i=1}^{N} h_F(X_i) + o(N^{1/2}\|\hat{F}_N - F\|). \qquad (32)$$

If, for some $d \geq 0$, $E[h_F(X_1)]^{2+d} < \infty$, and $\sum_k [\alpha_X(k)]^{d/(2+d)}$, then $N^{-1/2} \sum_i h_F(X_i)$ is asymptotically normal with mean 0 and variance

$$E[h_F^2(X_i)] + 2 \sum_{k=1}^{\infty} cov[h_F(X_1), h_F(X_{1+k})]. \tag{33}$$

To handle the remainder term in (32), Deo (1973) has shown that if

$$\sum_k k^2 [\alpha_X(k)]^{\frac{1}{2}-\tau} < \infty$$

for some $0 < \tau < 1/2$, then $N^{1/2}[\hat{F}_N(\cdot) - F(\cdot)]$, regarded as a random element of the space of cadlag functions endowed with the supremum norm, converges weakly to $Z(\cdot)$, where $Z(\cdot)$ is a Gaussian process having continuous paths, mean 0, and

$$cov[Z(t), Z(s)] = E[g_s(X_1) g_t(X_1)] + \sum_{k=1}^{\infty} E[g_s(X_1) g_t(X_{1+k})] + \sum_{k=1}^{\infty} E[g_s(X_{1+k}) g_t(X_1)],$$

where $g_t(x) = I_{[0,t]}(x) - F(t)$. Hence, Deo's result implies $N^{1/2}[T(\hat{F}_N) - T(F)]$ is asymptotically normal with mean 0 and variance given by (33).

The bootstrap approximation to the distribution of $N^{1/2}[T(\hat{F}_N) - T(F)]$ is the distribution, conditional on $X_1, \cdots, X_N$, of $N^{1/2}[T(\hat{F}_N^*) - T(\hat{F}_N)]$, where $\hat{F}_N^*$ is the empirical distribution of $X_1^*, \cdots, X_N^*$ obtained by the stationary resampling procedure. If the error terms in the differential approximation of $T(\hat{F}_N^*)$ are negligible, it is clear that the bootstrap will behave correctly, because Theorem 2 is essentially applicable. The key to formally justifying negligibility of error terms is to show that

$$\rho\left(N^{1/2}[\hat{F}_N^*(\cdot) - \hat{F}_N(\cdot)], Z(\cdot)\right) \to 0$$

in probability, where ρ is any metric metrizing weak convergence in the assumed function space. By Theorem 3, it is clear that the finite dimensional distributions of $N^{1/2}[\hat{F}_N^*(\cdot) - F(\cdot)]$ will appropriately converge to those of $Z(\cdot)$. The only technical difficulty is showing tightness of the bootstrap empirical process. In fact, by an argument similar to Deo's, tightness can be shown if $Np_N^2 \to \infty$. The technical details will appear elsewhere.

In fact, the above sketchy argument actually applies if T is only assumed compactly differentiable. For example, asymptotic validity for quantile functionals follows.

4.4. Linear Statistics Defined on Subseries.

Assume, $X_i \in \mathbf{R}^d$. In this section, we discuss how the stationary bootstrap may be applied to yield valid inferences for a parameter $\mu \in \mathbf{R}^D$ which may depend on the whole infinite-dimensional distribution of the process.

Consider the subseries $S_{i,M,L} = (X_{(i-1)L+1}, \cdots, X_{(i-1)L+M})$. These subseries can be obtained from the $\{X_i\}$ by a "window" of width M "moving" at lag L. Suppose $T_{i,M,L}$ is an estimate of μ based on the subseries $S_{i,M,L}$, so $T_{i,M,L} = \phi_M(S_{i,M,L})$, for some function ϕ_M from $\mathbf{R}^{dM}$ to $\mathbf{R}^D$. Let $\bar{T}_N = \sum_{i=1}^{Q} T_{i,M,L}/Q$, where $Q = [\frac{N-M}{L}] + 1$; here, $[\cdot]$ is the greatest integer function. To apply resampling to approximate the distribution of $\bar{T}_N$, just regard $(T_{1,M,L}, \cdots, T_{Q,M,L})$ as a time series in its own right. Note that M, L, and Q may depend on N. Weak dependence properties of the original series readily translate into weak dependence properties of this new series. Hence, we are essentially back into the sample mean setting. A small technical complication is that we are dealing with the average of the Nth row of a triangular array of variables, so that Theorem 2 needs to be generalized slightly.

By taking this point of view, one can establish consistency and weak convergence properties of the stationary bootstrap. Indeed, this approach has been applied fruitfully in the moving blocks resampling scheme in Politis and Romano (1989, 1990). The technical details of this approach, as applied to the stationary bootstrap, will appear in a future report. It is fully expected that the results established for moving blocks will have immediate counterparts for the stationary bootstrap.

To appreciate the applicability of this approach, consider the problem of estimating the spectral density $f(\omega)$. Suppose $T_{i,M,L}(\omega)$ is the periodogram evaluated at ω based on data $S_{i,M,L}$. Then, in fact, $\bar{T}_N(\omega)$ is approximately equal to Bartlett's kernel estimate of $f(\omega)$. Other kernel estimators can be (approximately) obtained by appropriate tapering of the individual periodogram estimates. A great advantage of the resampling approach is it easily yields simultaneous confidence regions for the spectral density over some finite grid of ω values. Other examples falling in this framework are the spectral measure and cross-spectrum, where asymptotic approximations to sampling distributions are particularly intractable.

4.5. Future Work. Subsequent work will focus on three important problems. First, establish theoretical results to construct uniform confidence bands for the spectral measure. The discussion in Section 4.4 will readily allow one to construct confidence bands for the spectral measure over a finite grid of ω values, but this is theoretically unsatisfying. By constructing uniform confidence bands over the whole continuous range of ω, a basis for goodness of fit procedures can be established. Second, higher order asymptotics are necessary, especially to compare procedures, just as in the i.i.d. case. Finally, the practical implementation, especially the choice of p, and the finite sample validity based on simulations will be addressed.

References

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *J. Roy. Statist. Soc. Suppl.* **8**, 27-41.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco.

Chung, K. (1974). *A Course in Probability Theory*, second edition, Academic Press, New York.

Deo, C. (1973). A note on empirical processes of strong-mixing sequences. *Ann. Probab.* **1**, 870-875.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.* **7**, 1-26.

Grenander, U. and Rosenblatt, M. (1953). Statistical spectral analysis arising from stationary stochastic processes. *Ann. Math. Statist.* **24**, 537-558.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York.

Katz, M. (1963). Note on the Berry-Esseen theorem. *Ann. Math. Statist.*, **34**, 1107-1108.

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

Lahiri, S. (1990). Second order optimality of stationary bootstrap. Department of Statistics, Iowa State University, unpublished manuscript.

Liu, R. Y. and Singh, K. (1988). Moving blocks jackknife and bootstrap capture weak dependence. Unpublished manuscript, Department of Statistics, Rutgers University.

Politis, D. and Romano, J. (1989). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation.

Technical Report 338, Department of Statistics, Stanford University.

Politis, D. and Romano, J. (1990). A nonparametric resampling procedure for multivariate confidence regions in time series analysis. Technical Report 348, Department of Statistics, Stanford University.

Politis, D., Romano, J. and Lai, T. (1990). Bootstrap confidence bands for spectra and cross-spectra. Technical Report 342, Department of Statistics, Stanford University.

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

Rosenblatt, M. (1984). Asymptotic normality, strong mixing, and spectral density estimates. *Ann. Probab.* **12**, 1167-1180.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Birkhauser, Boston.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley, New York.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch. verw. Gebiete*, **52**, 45-57.

Zurbenko, I. G. (1986). *The Spectral Analysis of Time Series.* North-Holland, Amsterdam.