A CIRCULAR BLOCK-RESAMPLING PROCEDURE FOR STATIONARY DATA

BY

DIMITRIS N. POLITIS and JOSEPH P. ROMANO

TECHNICAL REPORT NO. 370

March 1991

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT DMS89–05874

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

A Circular Block-Resampling Procedure for Stationary Data

By

Dimitris N. Politis and Joseph P. Romano

Technical Report No. 370

March 1991

Prepared Under the Auspices

Of

National Science Foundation Grant DMS89–05874

Also supported by National Science Foundation Grand DMS89-57217.

Department of Statistics
Stanford University
Stanford, California

A CIRCULAR BLOCK-RESAMPLING PROCEDURE FOR STATIONARY DATA

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

A block-resampling bootstrap for the sample mean of weakly dependent stationary observations has been recently introduced by Künsch (1989) and independently by Liu and Singh (1988). In Lahiri (1990) it was shown that the bootstrap estimate of sampling distribution is more accurate than the normal approximation, provided it is centered around the bootstrap mean, and not around the sample mean as customary. In this report, we introduce a variant of this block-resampling bootstrap that amounts to 'wrapping' the data around in a circle before blocking them. This 'circular' block-resampling bootstrap, has the additional advantage to be automatically centered around the sample mean. The consistency and asymptotic accuracy of the proposed method are demonstrated in comparison with the corresponding results for the block-resampling bootstrap.

AMS 1980 subject classifications: Primary 62G05; secondary 62M10.

Key words and phrases: Resampling schemes, bootstrap, time series, mixing sequences, weak dependence, nonparametric estimation.

1. Introduction

Suppose $X_1, \ldots, X_N$ are observations from the (strictly) stationary multivariate sequence $\{X_n, n \in \mathbf{Z}\}$, and the statistic of interest is the sample mean $\bar{X}_N = N^{-1}\sum_{i=1}^N X_i$. The sequence $\{X_n, n \in \mathbf{Z}\}$ is assumed to have a weak dependence structure. Specifically, the α-mixing (also called strong mixing) condition will be assumed, i.e. that $\alpha_X(k) \to 0$, as $k \to \infty$, where $\alpha_X(k) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|$, and $A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty$ are events in the σ-algebras generated by $\{X_n, n \leq 0\}$ and $\{X_n, n \geq k\}$ respectively.

The objective is to set confidence intervals for $\mu = EX_1$, for which an approximation to the sampling distribution of $\bar{X}_N$ is required. For this reason, a block-resampling bootstrap procedure has been introduced by Künsch (1989) and independently by Liu and Singh (1988). This method can be described as follows:

- Define $\mathcal{B}_i$ to be the block of b consecutive observations starting from X_i, that is $\mathcal{B}_i = (X_i, \ldots, X_{i+b-1})$, where $i = 1, \ldots, q$ and $q = N - b + 1$. Sampling with replacement from the set $\{\mathcal{B}_1, \ldots, \mathcal{B}_q\}$, defines a (conditional on the original data) probability measure P^*. If k is an integer such that $kb \sim N$, then letting $\xi_1, \ldots, \xi_k$ be drawn i.i.d. from P^*, it is seen that each ξ_i is a block of b observations $(\xi_{i,1}, \ldots, \xi_{i,b})$. If all $l = kb$ of the $\xi_{i,j}$'s are concatenated in one long vector denoted by $Y_1, \ldots, Y_l$, then the block-resampling bootstrap estimate of the variance of $\sqrt{N}\bar{X}_N$ is the variance of $\sqrt{l}\bar{Y}_l$ under P^*, and the bootstrap estimate of $P\{\sqrt{N}(\bar{X}_N - \mu) \leq x\}$ is $P^*\{\sqrt{l}(\bar{Y}_l - \bar{X}_N) \leq x\}$, where $\bar{Y}_l = \frac{1}{l}\sum_{i=1}^l Y_i$.

It is obvious that taking $b = 1$ makes the block-resampling bootstrap coincide with the classical (i.i.d.) bootstrap of Efron(1979) which has well-known optimality properties (cf. Singh(1981)). It can be shown (cf. Lahiri(1990)) that a slightly modified bootstrap estimate of sampling distribution turns out to be more accurate than the normal approximation, under some regularity conditions, resulting to more accurate confidence intervals for μ. The modification amounts to approximating the quantiles of $P\{\sqrt{N}(\bar{X}_N - \mu) \leq x\}$ by the corresponding quantiles of $P^*\{\sqrt{l}(\bar{Y}_l - E^*\bar{Y}_l) \leq x\}$, where $E^*\bar{Y}_l$ denotes the expected value of $\bar{Y}_l$ under the P^* probability (conditional on the original data). The need for re-centering the bootstrap distribution so as to have mean zero can also be traced back to Künsch's (1989) short calculation of the skewness of his block-resampling bootstrap.

The reason that the re-centered bootstrap provides a more accurate approximation is that $E^*\bar{Y}_l = q^{-1}\sum_{i=1}^{q} b^{-1}\sum_{j=i}^{i+b-1} X_j = \bar{X}_N + O_p(b/N)$, where, for consistency of the bootstrap in the dependent data setting, $b \to \infty$ as $N \to \infty$. In other words, the distribution of $\bar{Y}_l$ under P^* possesses a random bias of significant order. This bias is associated with the block-resampling bootstrap that assigns reduced weight to X_i's with $i < b$ or $i > N-b+1$. In other words, if we let P_i^* be the limit (almost sure in P^*) of the proportion l^{-1}(number of the Y_j's that equal X_i) as $l \to \infty$, (and assuming no ties among the X_i's), although $P_i^* = b/R$, with $R = b(N-b+1)$, for any i such that $b \leq i \leq N-b+1$, this proportion drops to $P_i^* = i/R$, for any $i < b$, and $P_i^* = (N-i+1)/R$, for any $i > N-b+1$.

2. A circular block-resampling bootstrap

A simple and 'automatic' way to have an unbiased bootstrap distribution is to 'wrap' the X_i's around in a 'circle', that is, to define (for $i > N$) $X_i \equiv X_{i_N}$, where $i_N = i(\text{mod}N)$, and $X_0 \equiv X_N$. This idea is closely associated with the definition of the circular autocovariance sequence of time series models. The 'circular' block-resampling bootstrap amounts to resampling whole 'arcs' of the circularly defined observations, and goes as follows.

- Define the blocks $\mathcal{B}_j$ as previously, that is, $\mathcal{B}_i = (X_i, \ldots, X_{i+b-1})$, but note that now for any integer b, there are N such $\mathcal{B}_j$, $j = 1, \ldots, N$. Sampling with replacement from the set $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$, defines a (conditional on the original data) probability measure $P^\star$. If k is an integer such that $kb \sim N$, then letting $\xi_1, \ldots, \xi_k$ be drawn i.i.d. from $P^\star$, it is seen that each ξ_i is a block of b observations $(\xi_{i,1}, \ldots, \xi_{i,b})$. If all $l = kb$ of the $\xi_{i,j}$'s are concatenated in one long vector denoted by $Y_1, \ldots, Y_l$, then the 'circular' block-resampling bootstrap estimate of the variance of $\sqrt{N}\bar{X}_N$ is the variance of $\sqrt{l}\bar{Y}_l$ under $P^\star$, and the 'circular' block-resampling bootstrap estimate of $P\{\sqrt{N}(\bar{X}_N - \mu) \leq x\}$ is $P^\star\{\sqrt{l}(\bar{Y}_l - \bar{X}_N) \leq x\}$, where $\bar{Y}_l = \frac{1}{l}\sum_{i=1}^{l} Y_i$.

The 'circular' construction is an integral part of a related resampling method in which blocks of random size are resampled (cf. Politis and Romano (1991)). It can also be immediately applied to bootstrapping general linear and nonlinear statistics, as in Künsch (1989), Liu and Singh (1988), and Politis and Romano (1989). Künsch's proposal of 'tapering' the observations in the $\mathcal{B}_j$ blocks can also be incorporated in the 'circular' construction without changing the first-order asymptotic results.

The following two theorems concern the consistency and asymptotic accuracy of the 'circular' block-resampling bootstrap. The theorems are stated for univariate sequences $\{X_n, n \in \mathbf{Z}\}$, although their extension to multivariate settings is straightforward.

Theorem 1 *Assume that* $E|X_t|^{6+\delta} < \infty$, *for some* $\delta > 0$, *and* $\sum_{k=1}^{\infty} k^2(\alpha_X(k))^{\frac{\delta}{6+\delta}} < \infty$. *As* $N \to \infty$, *let* $l/N \to 1$, *and let* $b \to \infty$, *but with* $bN^{-1} \to 0$. *Then* $\sigma_N^2 \equiv Var(\sqrt{N}\bar{X}_N)$ *has a finite limit* σ_∞^2, *and* $Var^\star(\sqrt{l}\bar{Y}_l) \xrightarrow{P} \sigma_\infty^2$, *where* $Var^\star(\sqrt{l}\bar{Y}_l)$ *is the variance of* $\sqrt{l}\bar{Y}_l$ *under* $P^\star$

conditional probability, as well as

$$\sup_x |P^\star\{\sqrt{l}(\bar{Y}_l - \bar{X}_N) \le x\} - P\{\sqrt{N}(\bar{X}_N - \mu) \le x\}| \to 0 \tag{1}$$

for almost all sample sequences $X_1, \ldots, X_N$.

Proof. The proof of Theorem 1 is immediate in view of the proof of the consistency of the block-resampling bootstrap of Künsch(1989). If we let $E^\star, E^*, Var^\star, Var^*$, represent expectation and variance under the $P^\star$ and P^* probabilities, then it is easily calculated that $E^\star \bar{Y}_l = \bar{X}_N$, and that

$$Var^\star(\sqrt{l}\bar{Y}_l) = \frac{b}{N}\sum_{i=1}^{N}(b^{-1}\sum_{j=i}^{i+b-1} X_i - \bar{X}_N)^2$$

$$= \frac{b}{N}\{\sum_{i=1}^{N-b+1}(b^{-1}\sum_{j=i}^{i+b-1} X_i - \bar{X}_N)^2 + \sum_{i=N-b+2}^{N}(b^{-1}\sum_{j=i}^{i+b-1} X_i - \bar{X}_N)^2\} = Var^*(\sqrt{l}\bar{Y}_l) + O_p(b/N)$$

where it was used that $E^*\bar{Y}_l = \bar{X}_N + O_p(b/N)$, and

$$Var^*(\sqrt{l}\bar{Y}_l) = \frac{b}{N-b+1}\sum_{i=1}^{N-b+1}(b^{-1}\sum_{j=i}^{i+b-1} X_i - E^*\bar{Y}_l)^2$$

$$= \frac{b}{N-b+1}\sum_{i=1}^{N-b+1}(b^{-1}\sum_{j=i}^{i+b-1} X_i - \bar{X}_N)^2 + O_p(b/N)$$

Since $Var^*(\sqrt{l}\bar{Y}_l) \xrightarrow{P} \sigma^2_\infty$, it is seen that the first two moments of $\sqrt{l}\bar{Y}_l$ under the $P^\star$ probability are asymptotically correct.

Finally, the moment and mixing conditions assumed are sufficient (cf. Hall and Heyde(1980)) to imply that $\sqrt{N}(\bar{X}_N - \mu)$ is asymptotically normal $N(0, \sigma^2_\infty)$. Noting that $\sqrt{l}(\bar{Y}_l - \bar{X}_N)$ is also asymptotically normal (conditionally on the data $X_1, \ldots, X_N$), the proof is concluded. □

It is noteworthy that to make the bias of the block-resampling bootstrap distribution to be of smaller order than its standard deviation, Künsch (1989) imposed the additional condition $bN^{-1/2} \to 0$ which is unnecessarily strong in our setting.

Theorem 2 *Assuming that the sequence* $\{X_n, n \in \mathbf{Z}\}$ *is defined on the probability space* $(\Omega, \mathcal{A}, P)$, *denote* $\mathcal{D}_n, n \in \mathbf{Z}$, *a sequence of sub σ-fields of* $\mathcal{A}$, *and* $\mathcal{D}_{n_1}^{n_2}$ *the σ-field generated by* $\mathcal{D}_{n_1}, \ldots, \mathcal{D}_{n_2}$. *Also assume that* $E|X_t|^{6+\delta} < \infty$, *for some* $\delta > 0$, *and, as* $N \to \infty$, *let*

$l/N \to 1$, and $b \to \infty$, but with $bN^{-1/3} \to 0$. Under the following regularity conditions:

(a_0) $\alpha_X(k)$ decreases geometrically fast;

(a_1) $\exists d > 0$ such that for all $k, n \in \mathbf{N}$, with $n > 1/d$, there exists a $\mathcal{D}_{k-n}^{k+n}$ measurable random variable $Z_{k,n}$, for which $E|X_k - Z_{k,n}| \le d^{-1}e^{-dn}$, and $E|Z_{k,n_k}|^{12} I(|Z_{k,n_k}| < k^{1/4}) < d^{-1}$, where n_k is a sequence of real numbers satisfying $\log k = o(n_k)$ and $n_k = O(\log k)^{1+d^{-1}}$, as $k \to \infty$;

(a_2) $\exists d > 0$ such that for all $k, n \in \mathbf{N}$, with $k > n > 1/d$, and all $t > d$,

$$E|E(e^{jt(X_{k-n}+X_{k-n+1}+\ldots+X_{k+n})}|\mathcal{D}_i, i \ne k)| \le e^{-d}$$

where j is used to denote the imaginary unit $\sqrt{-1}$;

(a_3) $\exists d > 0$ such that for all $k, n_1, n_2 \in \mathbf{N}$, and $A \in \mathcal{D}_{n_1-n_2}^{n_1+n_2}$,

$$E|P(A|\mathcal{D}_i, i \ne n_1) - P(A|\mathcal{D}_i, 0 < |n_1 - i| \le k + n_2)| \le d^{-1}e^{-dk};$$

the following is true (provided of course that $\sigma_\infty^2 > 0$),

$$\sup_x |P^\star\{\sqrt{l}\frac{\bar{Y}_l - \bar{X}_N}{\sqrt{Var^\star(\sqrt{l}\bar{Y}_l)}} \le x\} - P\{\sqrt{N}\frac{\bar{X}_N - \mu}{\sigma_\infty} \le x\}| = o_p(N^{-1/2}) \qquad (2)$$

Proof. As the proof of Theorem 1 relied on comparing the first two moments of $\sqrt{l}(\bar{Y}_l - \bar{X}_N)$ under $P^\star$ with the corresponding ones under P^*, the proof of Theorem 2 follows by looking at the third order moment. Specifically, under the regularity conditions we have assumed, first order Edgeworth expansions for $P^\star\{\sqrt{l}\frac{\bar{Y}_l - \bar{X}_N}{\sqrt{Var^\star(\sqrt{l}\bar{Y}_l)}} \le x\}$ and for $P\{\sqrt{N}\frac{\bar{X}_N - \mu}{\sigma_\infty} \le x\}$ are valid (cf. Lahiri(1990) where an extensive discussion relative to these regularity conditions can be found). Furthermore, equation (2) would be true, provided $b^2E^\star(U_1^\star - \bar{X}_N)^3 - N^2E(\bar{X}_N - \mu)^3 = o_p(1)$, where $U_1^\star = b^{-1}\sum_{j=1}^b \xi_{1,j}$, and the $\xi_{1,1}, \ldots, \xi_{1,b}$ are the elements of the first block-resample drawn from $P^\star$.

However, in Lahiri(1990) it was shown that $b^2E^*(U_1^* - E^*U_1^*)^3 - N^2E(\bar{X}_N - \mu)^3 = o_p(1)$, where $U_1^* = b^{-1}\sum_{j=1}^b \xi_{1,j}$, and the $\xi_{1,1}, \ldots, \xi_{1,b}$ are the elements of the first block-resample drawn from P^*. Finally, it is easily seen that

$$E^\star(U_1^\star - \bar{X}_N)^3 = \frac{1}{N}\sum_{i=1}^{N}(b^{-1}\sum_{j=i}^{i+b-1} X_i - \bar{X}_N)^3$$

$$= \frac{1}{N}\{\sum_{i=1}^{N-b+1}(b^{-1}\sum_{j=i}^{i+b-1} X_i - \bar{X}_N)^3 + \sum_{i=N-b+2}^{N}(b^{-1}\sum_{j=i}^{i+b-1} X_i - \bar{X}_N)^3\} = E^*(U_1^* - E^*U_1^*)^3 + O_p(b/N)$$

where it was used that $E^*U_1^* = \bar{X}_N + O_p(b/N)$, and

$$E^*(U_1^* - E^*U_1^*)^3 = \frac{1}{N-b+1} \sum_{i=1}^{N-b+1} (b^{-1} \sum_{j=i}^{i+b-1} X_i - E^*\bar{Y}_l)^3$$

$$= \frac{1}{N-b+1} \sum_{i=1}^{N-b+1} (b^{-1} \sum_{j=i}^{i+b-1} X_i - \bar{X}_N)^3 + O_p(b/N)$$

Hence, $E^*(U_1^* - \bar{X}_N)^3 - E^*(U_1^* - E^*U_1^*)^3 = O_p(b/N) = o_p(b^{-2})$, and the proof is concluded.□

References

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

[2] Hall, P. and Heyde, C. (1980), *Martingale Limit Theory and its Applications*, Academic Press, New York.

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

[4] Lahiri, S.N.(1990), Second order optimality of stationary bootstrap, Technical Report No.90-1, Department of Statistics, Iowa State University, (to appear in *Statist. Prob. Letters*).

[5] Liu, R.Y. and Singh, K. (1988), Moving Blocks Bootstrap and Jackknife Capture Weak Dependence, unpublished manuscript, Department of Statistics, Rutgers University.

[6] Politis D.N. and Romano, J.P. (1989), A General Resampling Scheme for Triangular Arrays of α-mixing Random Variables with application to the problem of Spectral Density Estimation, Technical Report No.338, Department of Statistics, Stanford University.

[7] Politis D.N. and Romano, J.P. (1991), The Stationary Bootstrap, Technical Report No. 365, Department of Statistics, Stanford University.

[8] Singh, K.(1981), On the asymptotic accuracy of Efron's bootstrap, *Ann.Statist.*, 9, 1187-1195.