journal of
statistical planning
and inference

Journal of Statistical Planning and
Inference 79 (1999) 179–190

www.elsevier.com/locate/jspi

Weak convergence of dependent empirical measures with application to subsampling in function spaces

Dimitris Politis[a], Joseph P. Romano[b,*], Michael Wolf[c]

[a] U.C., San Diego, USA
[b] Department of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305, USA
[c] U.C., Los Angeles, USA

Received 7 July 1998

Abstract

Consider the problem of inference for a parameter of a stationary time series, where the parameter takes values in a metric space (such as a function space). In this paper, we develop asymptotic theory based on subsampling to approximate the distribution of estimators for such parameters. The reason for this level of abstraction is to be able to consider parameters that take values in a function space. For example, we consider the estimation of the distribution of the empirical process and the spectral process. In order to accomplish this, we provide a general result based on simple arguments. The main technical result relies on the weak convergence of triangular arrays of dependent empirical measures, where the variables making up the arrays can take values in a (possibly nonseparable) metric space. This approach based on subsampling is quite powerful in that it leads to straightforward arguments where corresponding results based on the moving blocks bootstrap are much harder to obtain.

AMS classification: 60B10; 62G30; 62G09

Keywords: Empirical process; Spectral measure; Stationary time series; Subsampling; Weak convergence

1. Introduction

Let $X_1, \ldots, X_n$ denote a realization of a stationary time series. Suppose the infinite-dimensional distribution of the infinite sequence is denoted Q. The problem we consider is inference for a parameter $\theta(Q)$. The focus of the present paper is the case when the parameter space Θ is a metric space. The reason for considering such generality is to be able to consider the case when the parameter of interest is an unknown function,

* Corresponding author. Tel.: +1 415 723 6326; fax: +1 415 725 8977
E-mail address: romano@stat.stanford.edu (J.P. Romano)

PII: S0378-3758(98)00174-8

such as the marginal distribution of the process or the spectral distribution function of the process.

Our approach to the problem is to estimate the distribution of an estimator by subsampling. In Politis and Romano (1994), a general theory is given to show how subsampling leads to asymptotically valid confidence intervals for a real parameter. Here, we extend the arguments to cover the more general case.

In order to accomplish this task, we first develop a general result on the closeness of the empirical measure based on triangular arrays of dependent random variables, where the random variables take values in a (possibly nonseparable) metric space. In Section 2, we present such a result, which may be viewed as a generalization of the classical result of Varadarajan (1958), who considered a sequence of i.i.d. variables in a separable metric space. The threefold generalization to triangular arrays, to dependent variables, and to nonseparable metric spaces are all required for the statistical applications.

In Section 3, we apply the result of Section 2 to obtain a general result for subsampling. In Section 4, this result is applied to the special case of estimating the distribution of the empirical process of a stationary time series. The argument is seen to be quite simple and direct. Comparable results based on the (moving blocks) bootstrap are much more involved and they rely on heavier assumptions. In Section 5, the result is immediately applied to estimating the distribution of the spectral process (where no bootstrap counterpart has been established).

2. The basic theorem

Throughout this section, S denotes a (possibly nonseparable) metric space, equipped with a metric d. S is endowed with a σ-field $\boldsymbol{A}$, which will be assumed large enough to contain all closed balls, but perhaps not as large as the Borel σ-field. The general problem considered concerns the closeness of the empirical distribution of S-valued observations to the underlying law. The observations are assumed stationary (though this can be generalized) and weakly dependent.

Weak dependence is quantified in terms of strong mixing coefficients. Specifically, let $\{X_t,\ t\in T\}$ denote a collection of random variables defined on some common probability space, and assume T is some subset of the integers. Let F_k denote the σ-field generated by $\{X_t,\ t\leqslant k\}$ and let G_j denote the σ-field generated by $\{X_t,\ t\geqslant j\}$. Define Rosenblatt's α-mixing coefficients by

$$\alpha_X(j)=\sup|P(AB)-P(A)P(B)|\colon\ A\in F_k,\ B\in G_{k+j},\ k=1,2,\ldots\}. \tag{2.1}$$

Closeness of measures is described in terms of a metric metrizing weak convergence. Specifically, the bounded-Lipschitz metric ρ_L is defined as follows. Let $\boldsymbol{L}$ be the class of $\boldsymbol{A}$-measurable functions f satisfying $|f(x)-f(y)|\leqslant d(x,y)$ and $\sup_{x\in S}|f(x)|\leqslant 1$. For probability laws P and Q, define

$$\rho_L(P,Q)=\sup\{|Pf-Qf|\colon\ f\in\boldsymbol{L}\}.$$

Convergence of $\rho_L(P_n, P)$ to zero and P concentrating on a separable set implies weak convergence; see Pollard (1984) (p. 74).

The following theorem is a generalization of a classical result of Varadarajan (1958), who considered the case of i.i.d. observations in a separable metric space. Beran et al. (1987) and Bickel and Millar (1992) extended his result to triangular arrays of i.i.d. variables in possibly nonseparable metric spaces. The result here covers the dependent case, which is needed in the next section for inference in time series.

Theorem 2.1. *Let $Y_{n,1}, \ldots, Y_{n,j_n}$ be S-valued stationary observations with strong mixing sequence $\alpha_n(\cdot)$. Assume $j_n \to \infty$ as $n \to \infty$. Denote by P_n the marginal distribution of $Y_{n,1}$, and let $\hat{P}_n$ denote the empirical measure of the $Y_{n,i}$, $1 \leqslant i \leqslant j_n$. Assume $\{P_n\}$ is δ-tight, that is, for every $\varepsilon > 0$, there exists a compact set K and $\delta_n \downarrow 0$ so that $P_n(K^{\delta_n}) > 1 - \varepsilon$ $\forall n$, where $K^\delta = \{x \in S\colon\ d(x, K) < \delta\}$. Assume the mixing coefficients satisfy $\sum_{i=1}^{j_n} \alpha_n(i)/j_n \to 0$ as $n \to \infty$. Then, $\rho_L(\hat{P}_n, P_n) \to 0$ in probability.*

Remark 2.1. The issue of measurability cannot be ignored because $\rho_L(\hat{P}_n, P_n)$ need not even be measurable (because the sup of an uncountable collection of random variables need not be measurable). Instead of worrying about whether measurability holds, the proof shows the result is true if we interpret convergence in probability to mean convergence in outer probability; see van der Vaart and Wellner (1996) (Section 1.9).

Remark 2.2. The result can clearly be generalized to nonstationary observations if P_n is replaced by the expectation of $\hat{P}_n$.

Remark 2.3. The result holds for any metric metrizing weak convergence (by a subsequence argument).

Remark 2.4. The argument can be strengthened to yield an almost sure result when all the variables are defined on a common probability space.

Remark 2.5. The tightness assumption cannot be removed, even in the i.i.d. case; a counterexample is given in Beran et al. (1987).

Proof of Theorem 2.1. For ease of notation, we assume $j_n = n$. Now, for any real-valued measurable function f defined on S which is uniformly bounded by one, $\hat{P}_n f - P_n f \to 0$ in probability. Indeed, $\hat{P}_n f - P_n f$ has mean zero and variance

$$
\begin{aligned}
\sigma_n^2(f) &= n^{-1} P_n f(f - P_n f) + 2n^{-1} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \operatorname{cov}[f(Y_{n,1}), f(Y_{n,1+i})] \\
&\leqslant n^{-1} + 8n^{-1} \sum_{i=1}^{n} \alpha_n(i) \equiv b_n \to 0, \qquad (2.2)
\end{aligned}
$$

by the standard strong mixing inequality for uniformly bounded variables. The difficulty in establishing the theorem lies in showing this convergence is uniform over an uncountable collection of functions f.

Now, to show

$$\sup_{f\in \boldsymbol{L}} |\hat{P}_n f - P_n f| \to 0$$

in probability, it suffices to show

$$\sup_{f\in \boldsymbol{L}_n} |\hat{P}_n f - P_n f| \to 0$$

in probability, where $\boldsymbol{L}_n$ are the functions of the form $f(x)I(x\in K^{\delta_n})$, f is a function in $\boldsymbol{L}$ and δ_n is as in the statement of the theorem. To appreciate why, first note that $P_n(K^{\delta_n})\geqslant 1-\varepsilon$ by δ-tightness. Second, by the above variance calculation for general f, $\hat{P}_n(K^{\delta_n})$ has variance bounded by b_n. So, by Chebychev,

$$\begin{aligned}
\mathrm{pr}\{1-\hat{P}_n(K^{\delta_n})\geqslant 2\varepsilon\} &= \mathrm{pr}\{\hat{P}_n(K^{\delta_n})-P_n(K^{\delta_n})\leqslant 1-P_n(K^{\delta_n})-2\varepsilon\}\\
&\leqslant \mathrm{pr}\{\hat{P}_n(K^{\delta_n})-P_n(K^{\delta_n})\leqslant -\varepsilon\}\\
&\leqslant \mathrm{pr}\{|\hat{P}_n(K^{\delta_n})-P_n(K^{\delta_n})|\geqslant \varepsilon\}\\
&\leqslant \mathrm{var}[\hat{P}_n(K^{\delta_n})]/\varepsilon^2 \leqslant b_n/\varepsilon^2 \to 0.
\end{aligned}$$

So,

$$\sup_{f\in \boldsymbol{L}} |\hat{P}_n f - P_n f| \leqslant \sup_{f\in \boldsymbol{L}_n} |\hat{P}_n f - P_n f| + \sup_{f\in \boldsymbol{H}_n} |\hat{P}_n f - P_n f|,$$

where $\boldsymbol{H}_n$ is the collection of functions $\{f(\cdot)[1-I(\cdot\in K^{\delta_n})]:\ f\in \boldsymbol{L}\}$. But, the last term

$$\sup_{f\in \boldsymbol{H}_n} |\hat{P}_n f - P_n f| \leqslant \sup_{f\in \boldsymbol{H}_n} |\hat{P}_n f| + \sup_{f\in \boldsymbol{H}_n} |P_n f|$$

is small because

$$\sup_{f\in \boldsymbol{H}_n} |P_n f| \leqslant \varepsilon$$

and

$$\sup_{f\in \boldsymbol{H}_n} |\hat{P}_n f| \leqslant 1-\hat{P}_n(K^{\delta_n}) \leqslant 2\varepsilon$$

with probability tending to one, by the above. Hence, it is sufficient to show $\sup_{f\in \boldsymbol{L}_n} |\hat{P}_n f - P_n f|$ tends to zero in probability.

Fix $\varepsilon>0$. Next, let $\{f_1,\ldots,f_{m_n}\}$ be an ε-net (where the metric is sup norm) for the collection of functions $\boldsymbol{L}_K=\{f(\cdot)I(\cdot\in K)\}$. That is, if $f\in \boldsymbol{L}$, there is an i such that

$$\sup_{x\in K} |f(x)-f_i(x)| < \varepsilon.$$

Note that the number m_ε of approximating functions in the ε-net is finite by the Arzela–Ascoli Theorem. At this point, we cannot assume the approximating functions f_i are bounded-Lipschitz on all of S. However, by the Kirszbraun–McShane extension theorem (see Theorems 6.1.1 and 11.2.3 of Dudley, 1989), the functions f_i can be assumed to be in $\boldsymbol{L}$. Now, if $x \in K^{\delta_n}$, then there exists $\tilde{x} \in K$ so that $d(x,\tilde{x}) < \delta_n$. Then, for any $f \in \boldsymbol{L}$,

$$|f(x) - f(\tilde{x})| \leqslant d(x,\tilde{x}) < \delta_n,$$

where we have used the fact that f is Lipschitz. This inequality is also true for the approximating functions f_i. So,

$$f(x) \leqslant f(\tilde{x}) + \delta_n \leqslant f_i(\tilde{x}) + \varepsilon + \delta_n \leqslant f_i(x) + \varepsilon + 2\delta_n.$$

Hence, for any $f \in \boldsymbol{L}_n$, there exists an $i \leqslant m_\varepsilon$ satisfying

$$\sup_{x \in K^{\delta_n}} |f(x) - f_i(x)| \leqslant \varepsilon + 2\delta_n. \tag{2.3}$$

Given an f, let $\tilde{f}$ denote the approximating function f_i satisfying Eq. (2.3). Then,

$$\begin{aligned}
\sup_{f \in \boldsymbol{L}_n} |\hat{P}_n f - P_n f| &\leqslant \sup_{f \in \boldsymbol{L}_n} |(\hat{P}_n - P_n)(f - \tilde{f})| + \max_{1 \leqslant i \leqslant m_\varepsilon} |(\hat{P}_n - P_n) f_i| \\
&\leqslant 2\varepsilon + 4\delta_n + \max_{1 \leqslant i \leqslant m_\varepsilon} |(\hat{P}_n - P_n) f_i|.
\end{aligned}$$

To show the left-hand side tends to 0 in probability, fix any $\eta > 0$ and let $\varepsilon = \eta/8$. Then,

$$\begin{aligned}
\mathrm{pr}\left\{ \sup_{f \in \boldsymbol{L}_n} |\hat{P}_n f - P_n f| > \eta \right\} &\leqslant \mathrm{pr}\left\{ \sup_{f \in \boldsymbol{L}_n} |(\hat{P}_n - P_n)(f - \tilde{f})| > \eta/2 \right\} \\
&\quad + \mathrm{pr}\left\{ \max_{1 \leqslant i \leqslant m_\varepsilon} |(\hat{P}_n - P_n) f_i| > \eta/2 \right\} \\
&\leqslant \mathrm{pr}\left\{ \max_{1 \leqslant i \leqslant m_\varepsilon} |(\hat{P}_n - P_n) f_i| > \eta/2 \right\}
\end{aligned}$$

as soon as $2\varepsilon + 4\delta_n \equiv \eta/4 + 4\delta_n$ is less than $\eta/2$, or equivalently when $\eta > 16\delta_n$. Finally, the last term tends to zero because it can be bounded by $m_\varepsilon 4b_n^2/\eta^2$, where b_n is defined in Eq. (2.2).

3. A general theorem on subsampling

In this section, we consider the problem of constructing asymptotically valid confidence regions for a parameter of a stationary time series $\{X_t,\ t = 0, \pm 1, \pm 2, \ldots\}$, whose joint distribution will be denoted Q. The variables X_t are all defined on some common probability space (Ω_1, F_1, μ_1) and take values in some general measure space (Ω_2, F_2), though Ω_2 is usually assumed to be the real line. Hence, the joint law of the X_t variables, denoted by Q, is a probability on the product space which is the countable product of Ω_2 endowed with the product σ-field.

Attention focuses on a parameter $\theta(Q)$ that takes values in a parameter space Θ. At this point, nothing is assumed about Θ. The goal is to construct an asymptotically valid confidence set for $\theta(Q)$ based on $X_1,\ldots,X_n$. The whole motivation of the present paper is to present a general result for the case when Θ is quite large, such as an infinite-dimensional function space.

Suppose an estimator, $\hat\theta_n = \hat\theta_n(X_1,\ldots,X_n)$, of $\theta(Q)$ is given. In order to construct a confidence region for $\theta(Q)$, some knowledge of the sampling distribution of $\hat\theta_n(X_1,\ldots,X_n)$ is required. More generally, let $R_n(X_1,\ldots,X_n;\theta(Q))$ be a root (a term used by Beran, 1984), which is just some function of $X_1,\ldots,X_n$, $\theta(Q)$ and n, taking values in a metric space S (endowed with a σ-field $\boldsymbol{A}$). For example, in the case where Θ is a linear space, we might take

$$R_n(X_1,\ldots,X_n;\theta(Q)) = \tau_n[\hat\theta_n(X_1,\ldots,X_n) - \theta(Q)], \tag{3.1}$$

in which case $S=\Theta$; here, $\{\tau_n\}$ is just some normalizing sequence in anticipation of our asymptotic results. Alternatively, if Θ is a normed linear space with norm denoted $\|\cdot\|$, we might take

$$R_n(X_1,\ldots,X_n;\theta(Q)) = \tau_n\|\hat\theta_n(X_1,\ldots,X_n) - \theta(Q)\|,$$

so that S is the real line and possibly distinct from Θ. In any case, the idea is that if the sampling distribution of $R_n(X_1,\ldots,X_n;\theta(Q))$ under Q where known, this information could be used to obtain a confidence region for $\theta(Q)$.

To fix ideas, consider the following examples. In all these examples, assume the time series consists of stationary real-valued observations. If Q denotes the joint distribution of the process, let Q_1 denote the marginal distribution of X_1. Let $\theta(Q)=Q_1$, a parameter which takes values in the space of distribution functions. Several choices for S exists, one being $D[-\infty,\infty]$ equipped with the uniform metric. Here, we could take $\hat\theta$ to be the empirical distribution function of the data. Alternatively, one might be interested in a simple real-valued functional of Q_1, but our theory is intentionally general enough to handle general parameters. Certainly, if we can handle inference for Q_1, we should be able to handle functionals of Q_1. A more important example in the context of modelling time series is the spectral distribution function, which again can be assumed to take values in a suitable function space. Note that in both of these examples, a function space equipped with the supremum norm is nonseparable. These examples will be developed in the next two sections.

Let $J_n(Q)$ denote the law of $R_n(X_1,\ldots,X_n;\theta(Q))$, regarded as a random element of a metric space S. We are implicitly assuming S is endowed with an appropriate σ-field so that R_n is measurable. The subsampling approximation to $J_n(Q)$, denoted $\hat J_{n,b}$, is the empirical distribution of the $n-b+1$ values of $R_b(X_i,\ldots,X_{i+b-1};\hat\theta_n)$ as i ranges from 1 to $n-b+1$. The main assumption we will need is the following:

Assumption A. Assume $J_n(Q)$ converges weakly to a limit law $J(Q)$ which concentrates on a separable subset of S.

As a preliminary step to analyzing $\hat J_{n,b}$, we first analyze $L_{n,b}$, which is defined to be the empirical distribution of the $n-b+1$ values of $R_b(X_i,\ldots,X_{i+b-1};\theta(Q))$.

Proposition 3.1. *Assume Assumption A. Let $X_1,\dots,X_n$ be a stationary time series with α-mixing sequence $\alpha_X(\cdot)$. Assume $\alpha_X(i)\to 0$ as $i\to\infty$. Also, assume $b/n\to 0$ and $b\to\infty$ as $n\to\infty$. Then,*

$$\rho_L(L_{n,b},J_n(Q))\to 0$$

in probability.

Proof. Apply Theorem 2.1 with $Y_{n,i}=R_b(X_i,\dots,X_{i+b-1};\theta(Q))$ and $j_n=n-b+1$. Note that $L_{n,b}$ is the empirical distribution of S-valued stationary observations with exact distribution $J_b(Q)$. So, P_n is the distribution of $Y_{n,1}$ and the tightness assumption follows from Assumption A because $P_n=J_b(Q)$. Also, the mixing sequence of the nth row $Y_{n,1},\dots,Y_{n,j_n}$ satisfies $\alpha_n(i)\leqslant\alpha_X(i-b)$ if $i-b\geqslant 0$. Bound $\alpha_n(i)$ by 1 otherwise. Then,

$$\begin{aligned}\sum_{i=1}^{j_n}\alpha_n(i)/j_n &\leqslant \frac{1}{n-b+1}\left[b+\sum_{i=b+1}^{n-b+1}\alpha_X(i-b)\right]\\ &\leqslant \frac{b}{n-b+1}+\frac{1}{n-b+1}\sum_{i=1}^{n-b+1}\alpha_X(i)\to 0,\end{aligned}$$

by assumptions on $\alpha_X(\cdot)$ and b. $\square$

The previous result cannot in general be used for inference because the construction of L_n involves the unknown $\theta(Q)$. In order to obtain a result for the subsampling law $\hat{J}_{n,b}$, we specialize a little. In particular, assume R_n takes the form (3.1) so that $\Theta=S$ and S is a normed linear space with norm $\|\cdot\|$ and $d(x,y)=\|x-y\|$ if x and y are in S.

Theorem 3.1. *Under the assumptions of Proposition* 3.1 *and the additional assumption that $\tau_b/\tau_n\to 0$ as $n\to\infty$,*

$$\rho_L(\hat{J}_{n,b},J_n(Q))\to 0$$

in probability.

Proof. By the triangle inequality, it suffices to show $\rho_L(\hat{J}_{n,b},L_{n,b})\to 0$ in probability. Make the same identifications as in the proof of Proposition 3.1. Since $\hat{J}_{n,b}$ is the empirical distribution of the values

$$\tau_b[\hat{\theta}_b(X_i,\dots,X_{i+b-1})-\hat{\theta}_n]=\tau_b[\hat{\theta}_b(X_i,\dots,X_{i+b-1})-\theta(Q)]+\tau_b[\theta(Q)-\hat{\theta}_n],$$

we see that $\hat{J}_{n,b}$ is just $L_{n,b}$ *shifted* by $\tau_b[\theta(Q)-\hat{\theta}_n]$. Now, if $f\in\boldsymbol{L}$ is a (measurable) bounded Lipschitz function,

$$|\hat{J}_{n,b}f-L_{n,b}f|\leqslant\frac{1}{n-b+1}\sum_{i=1}^{n-b+1}d(Y_{n,i},Y_{n,i}-\tau_b[\theta(Q)-\hat{\theta}_n])=\|\tau_b[\theta(Q)-\hat{\theta}_n]\|,$$

where $Y_{n,i} = \tau_b[\hat{\theta}_b(X_i,\dots,X_{i+b-1}) - \theta(Q)]$. So,

$$\rho_L(\hat{J}_{n,b}, L_{n,b}) \leqslant \|\tau_b[\theta(Q) - \hat{\theta}_n]\| = \frac{\tau_b}{\tau_n}\|\tau_n[\hat{\theta}_n - \theta(Q)]\| \to 0$$

in probability by Assumption A.

4. Subsampling the empirical process

One of the great successes of Efrons (1979) i.i.d. bootstrap is it can be used to accurately approximate the distribution of the empirical process, and hence certain functionals of the empirical process. There has been considerable interest in obtaining similar results for dependent data using the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992). For example, see Bühlmann (1993, 1994) and Naik–Nimbalkar and Rajarshi (1994). Here, we obtain similar results based on subsampling using much simpler arguments. The aforementioned results rely on intricate chaining arguments and exponential inequalities. Our argument completely avoids such calculations if the original process of interest converges (which it is known to), whereas convergence of the moving blocks bootstrap process appears to require these arguments to be generalized and repeated. In fact, our simple arguments weaken the assumptions made on the underlying marginal distribution of the process, the choice of block size, and the mixing coefficients.

As before, let $X_1,\dots,X_n$ denote a stretch of a stationary process, whose entire joint distribution is denoted Q. Let P denote the marginal distribution of X_1 and let $\hat{P}_n$ denote the empirical measure. Consider the empirical process $Z_n(\cdot)$ indexed by a class of functions $f \in \boldsymbol{F}$, defined by

$$Z_n(f) = n^{1/2}[\hat{P}_n f - Pf].$$

Here, if the X_i take values in a space Ω_2, the functions f are assumed to be real-valued with domain Ω_2. Regard Z_n as a random element of the metric space $L_\infty(\boldsymbol{F})$, the metric space of real-valued bounded functions on $\boldsymbol{F}$ with sup norm denoted $\|\cdot\|_\infty$. The goal is to approximate the distribution of Z_n, which we will denote by $J_n(Q)$ in agreement with the general notation of Section 3.

Assumption A$_1$. Assume the law of Z_n converges weakly to a limiting process Z which concentrates on a separable subset of $L_\infty(\boldsymbol{F})$.

Remark 4.1. To avoid measurability problems, we simply assume $\boldsymbol{F}$ to be a permissible class of functions, as in Pollard (1984). Also note that one needs to endow $L_\infty(\boldsymbol{F})$ with an appropriate σ-field or, alternatively, understand that weak convergence to be in the sense of Hoffman–Jorgensen; the latter approach is fully developed in Part 1 of van der Vaart and Wellner (1996). All our assumptions are wrapped in the assumption that Z_n converges weakly. Arcones and Yu (1994) have given a sufficient condition for this assumption to hold. They assume $\boldsymbol{F}$ is a VC graph class with envelope function F satisfying $PF^p < \infty$ for some $p > 2$ and β-mixing coefficients satisfying

$$\sum_{i=1}^{\infty} [\beta(i)]^{(p-2)/p} < \infty$$

and $i^{p/(p-2)} \log(i)\beta(i) \to 0$ as $i \to \infty$. Then, Z is a mean 0 Gaussian process with

$$\mathrm{cov}[Z(f), Z(g)] = \sum_{j=-\infty}^{\infty} \mathrm{cov}[f(X_0), f(X_j)].$$

In the special case of the real line when $\boldsymbol{F}$ is the usual class of intervals, Deo (1973) has obtained a sufficient α-mixing condition and Yoshihara (1975) considers the multidimensional case. Since no best sufficient mixing condition exists for the weak convergence of Z_n, we simply take Assumption A_1 as given.

As in Section 3, the subsampling approximation to $J_n(Q)$ is $\hat{J}_{n,b}$, the empirical distribution of the $(n - b + 1)$ values $Z_{n,1}, \ldots, Z_{n,n-b+1}$, where

$$Z_{n,i}(f) = b^{1/2}[\hat{P}_{n,i} f - \hat{P}_n f],$$

and $\hat{P}_{n,i}$ is the empirical measure based on $X_i, \ldots, X_{i+b-1}$.

Theorem 4.1. *Assume Assumption A_1 and that the X process is α-mixing. Then, if $b \to \infty$ and $b/n \to 0$, we have*

$$\rho_L(J_n(Q), \hat{J}_{n,b}) \to 0$$

in probability.

Proof. Apply Theorem 3.1 with $R_n = n^{1/2}[\hat{P}_n(\cdot) - P(\cdot)]$, so that $S = \Theta = L_\infty(\boldsymbol{F})$. Here $\tau_n = n^{1/2}$ and Assumption A_1 is Assumption A specialized to the empirial process.

Remark 4.2. We have aimed for a simple result by elementary but general methods. Even so, the assumptions are remarkably weak. In the case of the moving blocks bootstrap, Naik–Nimbalkar and Rajarshi (1994) assume the blocksize b of order n^p for some $p \in (0, \frac{1}{4})$, but they obtain an almost sure convergence result (in the special case of the real line). We could also obtain an almost sure convergence result simply by strengthing our arguments a little, though the statistical uses for the stronger result are not clear enough to warrant doing this at this time. Bühlmann (1994) assumes $b = n^p$ for $p \in (0, \frac{1}{2})$. These papers assume the X_i's are real and vector-valued, respectively, and that the marginal distributions are continuous, which we do not need. Bühlmann (1993) also considers the general empirical process case which we consider here, but considerably more effort is required for the consistency of the moving blocks bootstrap.

Our mixing assumption is, of course, quite weak, with the real mixing assumption wrapped up in the verification of Assumption A_1.

Remark 4.3. Needless to say, it follows that any continuous functional of the empirical process which has an unknown distribution that can be consistently estimated by its subsampling counterpart. By looking at the supremum of the empirical process, asymptotically valid confidence bands for the unknown measure ensue.

Remark 4.4. Our result can actually be used to prove the corresponding result for the moving blocks bootstrap. By exploiting the linear structure of the empirical process, one sees that the moving blocks distribution can be obtained from the subsampling distribution by an appropriate normalized convolution operation. Since the subsampling distribution is an approximate Gaussian process, so must be a normalized convolution. This approach would simplify the arguments of Naik–Nimbalkar and Rajarshi (1994) and Bühlmann (1994).

5. Subsampling the spectral measure

The general theory in Sections 2 and 3 was motivated by the problem of approximating the distribution of the spectral process, which we define below. Let $F(\cdot)$ denote the spectral distribution function of a real-valued stationary time series, assumed to have a finite second moment. Here, $\theta = F(\cdot)$. Borrowing notation from Dahlhaus (1985), let $I_n(\lambda)$ denote the periodogram with tapered data, defined by

$$I_n(\lambda) = [2\pi H_{n,2}(0)]^{-1} d_n(\lambda) d_n(-\lambda),$$

where

$$d_n(\lambda) = \sum_{t=1}^{n} h[t/(n+1)] X_t \exp[-\mathrm{i}\lambda t]$$

and

$$H_{n,k}(\lambda) = \sum_{t=1}^{n} h^k[t/(n+1)] \exp[-\mathrm{i}\lambda t].$$

The data taper h is assumed of bounded variation and square integrable on $[0,1]$. Let $\hat{F}_n(\cdot)$ be the corresponding integrated periodogram given by

$$\hat{F}_n(\lambda) = \frac{2\pi}{n} \sum_{0 < 2\pi s/n \leqslant \lambda} I_n(2\pi s/n).$$

Take $\tau_n = n^{1/2}$ and regard $S_n(\cdot) = n^{1/2}[\hat{F}_n(\cdot) - F(\cdot)]$ as a random element of $D[0,\pi]$ endowed with the sup norm $\|\cdot\|$. Under suitable weak dependence conditions, the process $S_n(\cdot)$ converges weakly to a mean zero Gaussian process $S(\cdot)$ with covariance

$$\mathrm{cov}[S(\lambda), S(\mu)] = 2\pi G(\min\{\lambda, \mu\}) + 2\pi F_4(\lambda, \mu),$$

where

$$G(\lambda) = \int_0^{\lambda} f^2(\beta)\,\mathrm{d}\beta$$

and

$$F_4(\lambda, \mu) = \int_0^{\lambda}\int_0^{\mu} f_4(\alpha, -\alpha, -\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta;$$

here, f is the spectral density and f_4 is the fourth order cumulant spectrum (see e.g., Brillinger, 1975). For various sets of conditions for this weak convergence to hold, see Anderson (1993), Brillinger (1975), and Dahlhaus (1985). Since the limit distribution is that of a certain Gaussian process whose covariance structure depends on intricate fourth order properties of the underlying stationary process, analytical approximations to this limit law would be difficult to obtain; but, see Anderson (1993). In summary, the weak convergence of the process S_n has been well-studied and holds quite generally under weak dependence. As in Section 4, we assume this convergence as fundamental.

Assumption A$_2$. Assume the law of S_n converges weakly to a limiting process S whose paths concentrate on a separable subset of $D[0, \pi]$.

Letting $J_n(Q)$ denote the law of S_n, we immediately have the following result.

Theorem 5.1. *Assume Assumption A_2 and that the X process is α-mixing. Then, if $b \to \infty$ and $b/n \to 0$, we have*

$$\rho_L(J_n(Q), \hat{J}_{n,b}) \to 0$$

in probability.

Remark 5.1. The above arguments apply to the case where θ is the standardized spectral distribution function. Consider the process $Z_n(\cdot) = n^{1/2}[\hat{F}_n(\cdot)/\hat{F}_n(\pi) - F(\cdot)/F(\pi)]$. The weak convergence properties of Z_n can be deduced from that of Y_n, so that Assumption A holds here as well.

Remark 5.2. Remark 4.3 applies here as well. Thus, we can get asymptotically valid approximations to the distribution of the supremum of the spectral process, yielding a confidence band with asymptotic coverage probability equal to the nominal level. Actually, one needs to know a little more to claim the limiting coverage probability, namely that the limiting distribution in Assumption A$_2$ is continuous. But, if the limit process is Gaussian with mean 0 and continuous sample paths, as it is here, this continuity property follows by a general result of Tsirel'son (1975). Some simulation results of uniform confidence bands by subsampling for this example are presented in Politis et al. (1993).

Remark 5.3. In fact, the argument can be generalized to get uniform confidence bands for the spectral density itself, which is a harder problem. Here, assumption A must be weakened so that it is assumed $\tau_n\|\hat{\theta}_n - \theta(Q)\| - c_n$ has a limit distribution for some c_n. This assumption holds for spectral density estimates; see Woodroofe and VanNess (1967).

References

Anderson, T., 1993. Goodness of fit tests for spectral distributions. Ann. Statist. 21, 830–847.

Arcones, M., Yu, B., 1994. Central limit theorems for emprical and U-processes of stationary mixing sequences. J. Theoret. Probab. 7, 47–71.

Beran, R., 1984. Bootstrap methods in statistics. Jber. d. Dt. Math.-Verein 86, 14–30.

Beran, R., LeCam, L., Millar, P., 1987. Convergence of stochastic empirical measures. J. Multivariate Anal. 23, 159–168.

Bickel, P., Millar, P., 1992. Uniform convergence of probability measures on classes of functions. Statist. Sinica 2, 1–15.

Brillinger, D., 1975. Time Series: Data Analysis and Theory. Holt, Rinehart and Winston, New York.

Bühlmann, P., 1993. The blockwise bootstrap in time series and empirical processes. Dissertation at the Swiss Federal Institute of Technology No. 10354, Zurich.

Bühlmann, P., 1994. Block bootstrapped empirical process for stationary sequences. Ann. Statist. 22, 995–1012.

Dahlhaus, R., 1985. On the asymptotic distribution of Bartlett's U_p-statistic. J. Time Ser. Anal. 6, 213–227.

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

Dudley, R., 1989. Real Analysis and Probability. Wadsworth, Belmont, CA.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1–26.

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

Liu, R.Y., Singh, K., 1992. Moving blocks jackknife and bootstrap capture weak dependence. In: LePage, Billard (Eds.), Exploring the Limits of Bootstrap. Wiley, New York.

Naik–Nimbalkar, U., Rajarshi, M., 1994. Validity of blockwise bootstrap for empirical processes with stationary observations. Ann. Statist. 22, 980–994.

Politis, D., Romano, J., 1994. A general theory for large sample confidence regions based on subsamples under minimal assumptions. Ann. Statist. 22, 2031–2050.

Politis, D., Romano, J., You, L., 1993. Uniform confidence bands for the spectrum based on subsamples. In: Tarter, M., Lock, M. (Eds.), Computing Science and Statistics, Proc. 25th Symp. on the Interface. The Interface Foundation of North America, pp. 346–351.

Pollard, D., 1984. Convergence of Stochastic Processes. Springer, New York.

Tsirel'son, V., 1975. Density of distribution of the maximum of a gaussian process. Theoret. Probab. Appl. 20, 847–856.

van der Vaart, A., Wellner, J., 1996. Weak Convergence and Empirical Processes. Springer, New York.

Varadarajan, V.S., 1958. Weak convergence of measures on separable metric spaces. Sankhya, Ser. A, 19, 15–22.

Woodroofe, M., VanNess, J., 1967. The maximum deviation of sample spectral densities. Ann. Math. Statist. 38, 1558–1569.

Yoshihara, K., 1975. Weak convergence of multidimensional empirical processes for strong mixing sequences of stochastic vectors. Z. Wahrsch. Gebiete 33, 133–137.