Local Block Bootstrap

Efstathios Paparoditis [a], Dimitris N. Politis [b].

[a] Department of Mathematics and Statistics, University of Cyprus, P. O. Box 20537, Nicosia, Cyprus. E-mail: stathisp@ucy.ac.cy

[b] Department of Mathematics, University of California–San Diego, La Jolla, CA 92093-0112, USA. E-mail: politis@euclid.ucsd.edu

Abstract

For time series that are not stationary, the block bootstrap method is not directly applicable. However, if the underlying stochastic structure is slowly changing with time, one may employ a local block-resampling procedure. We define such a procedure, and give an example of its applicability.

Résumé

Bloc re-échantillonnage local

Pour les séries chronologiques qui ne sont pas stationnaires, la méthode de bloc re-échantillonnage n'est pas directement applicable. Cependant, si la structure stochastique fondamentale change lentement, on peut utiliser une méthode de bloc re-échantillonnage local. Nous définissons une telle procédure et donnons un exemple de son applicabilité.

1 Introduction

Let $X_1, ..., X_n$ be an observed stretch of a real-valued time series $\{X_t, t \in \mathbf{Z}\}$. In many applications, for instance in connection with financial or meteorological time series, the sample size n may be quite large. Consequently, it may be unrealistic to assume that such a long time series is stationary; a more realistic model might be to assume a slowly-changing stochastic structure in the sense that the joint probability law of $(X_t, X_{t+1}, \ldots, X_{t+k})$ changes smoothly (and slowly) with t for any k. Such nonstationary models have been provided by the evolutionary spectra models of Priestley (1988), and the locally stationary models of Dahlhaus (1996, 1997).

Now because of the nonstationarity of $\{X_t\}$, the block bootstrap method of Künsch (1989) is not directly applicable; rather, a modification that takes into account the changing stochastic structure should be constructed. We therefore propose the Local Block Bootstrap (LBB) procedure to address this situation. The basic premise of the LBB is to only resample blocks that are close to each other, i.e., a block that starts at time t, can only be replaced with blocks whose starting point is close to t. The LBB algorithm can then be briefly described as follows: an LBB bootstrap pseudo-series is constructed by a concatenation of q blocks of size b (such that $qb \simeq n$), where the jth block of the resampled series is chosen randomly from a distribution (say, uniform) on all the size-b blocks of consecutive data whose time indices are 'close' to those in the original block. Rigorous definition of the Local Block–Bootstrap (LBB) algorithm is given in Section 2; Section 3 gives an application to a simple but illustrative example.

2 Local Block Bootstrap

Given the data $X_1, ..., X_n$, the LBB algorithm creates a bootstrap pseudo-series $X_1^*, ..., X_n^*$ as follows:

- Select an integer block size b, and a real number $B \in (0, 1]$ such that nB is an integer; both b and B are thought of as functions of n.

- For $m = 0, 1, \ldots, (\lceil n/b \rceil - 1)$, let[1] $X^*_{mb+j} := X_{I_m+j-1}$ for $j = 1, \ldots, b$, where $I_1, I_2, \ldots$ are independent, integer-valued random variables satisfying $P(I_m = k) = W_{n,m}(k)$.

The probability distribution $W_{n,m}(k)$ is the practitioner's choice. The easiest choice is the uniform probability over the integers in the interval $[J_{1,m}, J_{2,m}]$, where $J_{1,m} = \max\{1, mb - nB\}$ and $J_{2,m} = \min\{n - b + 1, mb + nB\}$. However, many other choices are possible; our only requirement is:

Condition (C_0) : Let $w(x)$ be a probability density over the interval $[-1, 1]$ that is continuous, symmetric around 0, and monotone non-increasing for $x \in [0, 1]$. Then, let $W_{n,m}(k) = c_m^{-1} w(\frac{k-mb}{nB}) \mathbf{1}_{[J_{1,m}, J_{2,m}]}(k)$, where $\mathbf{1}_{[J_{1,m}, J_{2,m}]}(k)$ is 1 or 0 according to whether $k \in [J_{1,m}, J_{2,m}]$ or not, and $c_m = \sum_{k=J_{1,m}}^{J_{2,m}} w(\frac{k-mb}{nB})$.

Note that the range $[k - nB, k + nB]$ indicates our 'local' neighborhood of the point k; in other words, the idea is that the time series $\{X_t\}$ can be considered as 'almost' stationary in a window of length $2nB$ so that the stretch $X_{k-nB}, \ldots, X_{k+nB}$ is approximately stationary for any k.

Recall that in the block bootstrap method of Künsch (1989) it is suggested to take the block size b to be big but small with respect to the sample size n. The LBB requirements are the following:

Condition (C_1) : The block size b should be big but small with respect to the local window size $2nB$ which determines the effective 'local' sample size, i.e., we require $b \to \infty$ but $b/(nB) \to 0$.

Condition (C_2) : The local window size $2nB$ must be big but small with respect to the sample size n, i.e., we require $nB \to \infty$ but $B \to 0$.

[1]Here $\lceil k \rceil$ denotes the smallest integer that is greater or equal to k. Thus, in the case where n is not an integer multiple of b, our LBB algorithm generates a bootstrap series whose length is slightly bigger than n, although only $X_1^*, ..., X_n^*$ are kept in the end.

Remark 2.1 Condition (C_2) is where the local (but not global) stationarity enters; if the sequence $\{X_t\}$ is globally stationary, then Condition (C_2) is not needed, i.e., one may take $B = 1$. Consequently, Künsch's (1989) block bootstrap is a special case of the LBB: just let $B = 1$, and w be a uniform density. In addition, if $\{X_t\}$ happens to be independent, then taking $b \to \infty$ is unnecessary; therefore, Shi's (1991) local bootstrap for heteroskedastic regression is another special case of the LBB: just take $b = 1$, $B \to 0$ and $nB \to \infty$.

3 Application

As is well-known, the applicability of bootstrap methods is usually checked on a case-by-case [2] basis. Thus, we focus on a particular interesting example. Let

$$X_t = \mu + v_t \epsilon_t, \quad \text{for all } t \in \mathbf{Z} \tag{1}$$

where $\{\epsilon_t,\ t \in \mathbf{Z}\}$ is a mean-zero, strong mixing and strictly stationary sequence satisfying

$$E|\epsilon_t|^{6+\delta} < \infty, \quad \text{and} \quad \sum_{k=1}^{\infty} k^2 \alpha_\epsilon^{\delta/(6+\delta)}(k) < \infty \quad \text{for some} \quad \delta > 0, \tag{2}$$

where $\alpha_\epsilon(k)$ indicates the strong mixing coefficient associated with $\{\epsilon_t,\ t \in \mathbf{Z}\}$.

Consequently, the first moment of the sequence $\{X_n\}$ is time-invariant and given by μ; nevertheless, the second moments are not time-invariant because of the heteroskedasticity induced by the unknown (but deterministic) factor v_t. Our goal is interval estimation of the unknown parameter μ based on the data $X_1, ..., X_n$. For this reason we require an approximation to the sampling distribution of the sample mean $\hat{\mu} = n^{-1}\sum_{t=1}^{n} X_t$ that will serve as our estimator of μ. We propose the LBB as a method for this approximation.

Let $X_1^*, ..., X_n^*$ denote an LBB pseudo-series constructed using the algorithm of Section 2, and let $\hat{\mu}^* = n^{-1}\sum_{t=1}^{n} X_t^*$. To investigate the asymptotic behavior of $\hat{\mu}$ and $\hat{\mu}^*$ it is helpful to use the set-up of Dahlhaus (1997) and map the range $\{1, \ldots, n\}$ onto the unit interval. For model (1), this entails the assumption:

$$v_t = V(t/n) \quad \text{where } V : [0,1] \to \mathbf{R} \text{ is a bounded, piecewise continuous function.} \tag{3}$$

Under (3), the data $X_1, ..., X_n$ constitute the nth row of a triangular array; however, since no confusion arises, we will not use the usual double-index notation.

The following theorem shows that the LBB is successful in giving a consistent approximation to the sampling distribution of the sample mean $\hat{\mu}$. Hence, asymptotically valid confidence intervals for μ can be based on the quantiles of the bootstrap distribution $P^*(\sqrt{n}(\hat{\mu}^* - E^*\hat{\mu}^*) \le x)$ that are computable, as opposed to the quantiles of the unknown true distribution $P(\sqrt{n}(\hat{\mu} - \mu) \le x)$; here,

[2] The only exception so far seems to be the i.i.d. bootstrap with smaller resample size that is generally consistent; see Politis et al. (1999, Ch. 2.3).

P^*, E^* and Var^* denote probability, expectation and variance under the LBB bootstrap mechanism that—as usual—is performed conditional on $X_1, ..., X_n$.

Theorem 3.1 *Assume Conditions C_0, C_1, C_2 as well as eq. (1), (2) and (3). Then, as $n \to \infty$,*

$$Var(\sqrt{n}\hat{\mu}) - Var^*(\sqrt{n}\hat{\mu}^*) \xrightarrow{P} 0 \tag{4}$$

and

$$\sup_x |P(\sqrt{n}(\hat{\mu} - \mu) \leq x) - P^*(\sqrt{n}(\hat{\mu}^* - E^*\hat{\mu}^*) \leq x)| \xrightarrow{P} 0. \tag{5}$$

Proof (sketch). The first step consists of showing that the two distributions, $P(\sqrt{n}(\hat{\mu} - \mu) \leq x)$ and $P^*(\sqrt{n}(\hat{\mu}^* - E^*\hat{\mu}^*) \leq x)$, are both asymptotically normal. For the former, the proof follows the results of Roussas et al. (1992); for the latter, the proof involves checking the conditions of a Lindeberg-Feller Central Limit Theorem in the bootstrap world for the sum of the independent but not identically distributed blocks. Since the two distributions in (5) are asymptotically normal with mean zero, it is apparent that verifying (4) is sufficient to also show (5). Checking (4) is straightforward but tedious; suffice it to note that both variances in (4) have the same asymptotic limit that is given by:

$$\int_0^1 V^2(s)ds \sum_{k=-\infty}^{\infty} Cov(\epsilon_0, \epsilon_k).$$

Remark 3.1 The LBB can handle more complicated models than model (1); in a follow-up paper with Arif Dowla (UCSD), heteroskedastic models that also possess a nonstationary first moment are analyzed, and the LBB is shown to be successful in capturing the distribution of the nonparametric trend estimator.

References

[1] Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes, *Stoch. Proc. Appl.*, 62, 139-168.

[2] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes, *Annals of Statistics*, 25, 1-37.

[3] Künsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations, *Annals of Statistics*, 17, 1217-1241.

[4] Politis, D.N., Romano, J.P., and Wolf, M. (1999). *Subsampling.* Springer, New York.

[5] Priestley, M.B. (1988). *Non-linear and non-stationary time series analysis*, Academic Press, London.

[6] Roussas, G.G., Tran, L.T. and Ioannides, D.A. (1992). Fixed design regression for time series: asymptotic normality, *J. Multivariate Analysis*, 40, 262-291.

[7] Shi, S.G. (1991). Local bootstrap, *Annals Inst. Statist. Math.*, 43, 667-676.