Bootstrapping Time Series

Peter Bühlmann
Seminar für Statistik, ETH Zentrum, CH-8092 Zürich, Switzerland
buhlmann@stat.ethz.ch

1. Introduction

Extracting information from data is usually based on some statistics, formalized as a vector- or curve-valued estimator $\hat{\theta}$. With time series data, estimating the distribution of $\hat{\theta}$ is much more difficult than for independent observations. The bootstrap is then a very important tool. For simplicity, we consider only univariate data $X_1, \ldots, X_n$.

2. Bootstraps for time series

A straightforward approach is model based: the dependence structure is modelled explicitly and Efron's bootstrap is applied to the remaining residual part. Such an approach is, of course, heavily driven by the choice of the model and very sensitive to model-misspecification. We focus here on bootstraps which are nonparametric and, in a to be specified sense, model-free.

Block-Bootstrap. We briefly outline the method and refer to the seminal paper by Künsch (1989) for more details.

Proper application of the block-bootstrap scheme involves first an adaptation to the problem. Assume that the statistic $\hat{\theta}$ estimates a functional θ, depending on the p-dimensional marginal distribution of the time series. For example, the lag(1)-correlation $\mathrm{Corr}(X_0, X_1)$ in a stationary time series is a functional of the distribution of (X_t, X_{t+1}), corresponding to $p = 2$. Then, build vectors of consecutive observations

$$Y_t = (X_{t-p+1}, \ldots, X_t),\ t = p, \ldots, n. \tag{1}$$

Apply the block-bootstrap to these vectorized observations. Build overlapping blocks of consecutive vectors $(Y_p, \ldots, Y_{p+\ell-1}), (Y_{p+1}, \ldots, Y_{p+\ell}), \ldots, (Y_{n-\ell+1}, \ldots, Y_n)$, where $\ell \in \mathbb{N}$ is the block-length parameter. For simplicity, assume that $n - p + 1 = k\ell$ with $k \in \mathbb{N}$. Then, resample k blocks independently and recompute the statistic on such resampled blocks to obtain the block-bootstrapped estimator $\hat{\theta}^{*B}$.

The block-bootstrapped estimator $\hat{\theta}^{*B}$ is *not* constructed with a plug-in rule, i.e., we can not write $\hat{\theta}^{*B} = h_n(X_1^*, \ldots, X_n^*)$, where $\hat{\theta} = h_n(X_1, \ldots, X_n)$ and $X_1^*, \ldots, X_n^*$ a suitable bootstrap sample. At least two difficulties remain to be answered in a case by case manner.
(A) Redesigning the *computation* of $\hat{\theta}^{*B}$, which can become practically very inconvenient.
(B) Vectorization as in (1) is not always appropriate. For example, the MA-parameter in an MA(1) model or the spectral density of a stationary process depend on the entire distribution of the process, corresponding to $p = \infty$.

Whenever problems (A) and/or (B) become too awkward, an ad-hoc solution is to neglect the vectorization step in (1) and work with $p = 1$. Then, a substantial efficiency loss of the method has generally to be paid.

Sieve-Bootstraps. The strategy here is to get a reasonable *approximation* $\hat{P}_n$ for the true distribution P of the underlying process of the observed data. We base it on a (semi-) parametric family of models belonging to the general method of sieves. Unlike the fully model-based bootstrap approach, we allow here for finite-sample model-misspecification: the correct model is only required in some asymptotic sense. In contrast to the block-bootstrap, a sieve-bootstrapped estimator is based on the plug-in rule,

$$\hat{\theta}^{*S} = h_n(X_1^{*S}, \ldots, X_n^{*S}), \tag{2}$$

where $\hat{\theta} = h_n(X_1, \ldots, X_n)$ and $X_1^{*S}, \ldots, X_n^{*S}$ is a sieve-bootstrapped time series, resampled from the estimated distribution $\hat{P}_n$. The computation of the bootstrapped estimator is straightforward and practically very convenient. The difficulty is to get a good estimate $\hat{P}_n$ and in turn a good resample $X_1^{*S}, \ldots, X_n^{*S}$.

AR-sieve bootstrap for stationary linear time series. We refer to a linear, invertible time series if it allows an autoregressive representation of order infinity (AR(∞)),

$$X_t - \mathbb{E}[X_t] = \sum_{j=1}^{\infty} \phi_j (X_{t-j} - \mathbb{E}[X_{t-j}]) + \varepsilon_t, \; t \in \mathbb{Z}, \tag{3}$$

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. random variables with $\mathbb{E}[\varepsilon_t] = 0$ and additional regularity conditions for ε_t and $(\phi_j)_{j \in \mathrm{N}}$ have to be made.

The AR-sieve bootstrap then resamples from a fitted (typically high order) AR(p). The order p is usually data-driven and in theory, it should be thought of as an increasing sequence p_n tending to infinity with a slower order than sample size n. If the true underlying process is AR(∞), or 'close' to it, such a bootstrap scheme is highly recommendable and often far superior to the general block-bootstrap. The procedure uses crucially the linear AR(∞) assumption for the underlying true process.

VLMC-sieve bootstrap for general stationary categorical time series. Sieve approximation can also be used for general stationary categorical processes. The difficulty is the construction of a sieve being flexible enough to cover general processes, but allowing a parsimonious parameterization so that reasonable statistical accuracy can be maintained. We propose the sieve of so-called variable length Markov chains (VLMC) for approximating the $\mathcal{X}$-valued time series $(X_t)_{t \in \mathbb{Z}}$, with $\mathcal{X}$ a finite space.

An $\mathcal{X}$-valued, stationary VLMC $(X_t)_{t \in \mathbb{Z}}$ is characterized as a Markov chain of high order whose time-homogeneous transition probabilities are functions depending on a variable number ℓ of lagged values

$$\mathbb{P}[X_t = x_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \ldots] = \mathbb{P}[X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_{t-\ell} = x_{t-\ell}],$$

for all x_t, where $\ell = \ell(x_{t-1}, x_{t-2}, \ldots)$ is itself a function of the past. If $\ell(x_{t-1}, x_{t-2}, \ldots) \equiv k$ for all $x_{t-1}, x_{t-2}, \ldots$, we obtain the full Markov chain model of order k. For variable $\ell(.)$ with $\sup\{\ell(x_{t-1}, x_{t-2}, \ldots);\; x_{t-1}, x_{t-2}, \ldots\} = k$, we have an embedding (full) Markov chain of order k, but with an additional *structure of a variable length memory*, implying that some transition probabilities of the embedding Markov chain are lumped together.

Fitting a VLMC involves a version of the tree structured context algorithm (Rissanen, 1983) for estimating the variable length memory and the transition probabilities; the former is a highly complex model selection problem with an extremely large number of possible models. The method yields a consistent estimate $\hat{P}_{n;VLMC}$ for the distribution of suitably regular processes which are not necessarily a VLMC, see Bühlmann and Wyner (1999) and Ferrari (1999).

The construction of the VLMC-sieve bootstrap is as follows. Resample

$$X_1^{*VLMC-S}, \ldots, X_n^{*VLMC-S} \sim \hat{P}_{n;VLMC}. \tag{4}$$

We then proceed as in (2).

3. Comparison

Theoretical results about adaptivity. For asymptotic comparisons we consider estimation of the variance of the mean $\hat{\theta} = \overline{X}_n$. Despite the estimator's simplicity, the problem is representative since many other estimators behave asymptotically as a mean of influence functions. Under suitable regularity conditions, $n\mathrm{Var}(\overline{X}_n)$ converges to $\sum_{k=-\infty}^{\infty} R(k),\ R(k) = \mathrm{Cov}(X_0, X_k)$.

If $R(k)$ is not vanishing as $k \to \infty$, it is known that the block-bootstrap yields for any stationary, suitably regular process $(X_t)_{t\in\mathbb{Z}}$,

$$\mathbb{E}[\left(n\mathrm{Var}^{*B}(\overline{X}_n^{*B}) - n\mathrm{Var}(\overline{X}_n)\right)^2] \sim const.n^{-2/3}, \tag{5}$$

achieved with the optimal block-length $\ell \sim const.n^{-1/3}$, see Künsch (1989). The result in (5) shows that the optimal rate of the block-bootstrap accuracy does *not* depend on the decay of the covariances $R(k)$ $(k \to \infty)$ (or some mixing coefficients): it is in this sense *non-adaptive* to the underlying dependence structure. In the case where $R(k) = 0$ for all $k \geq K$ with some $0 < K < \infty$, the result in (5) could be improved under the unrealistic assumption of *knowing a priori* the value of K.

Adaptivity with respect to the decay of $R(k)$ $(k \to \infty)$ can be achieved with the AR-sieve bootstrap for the class of AR(∞) processes in (3), see Bühlmann (1997). The same is true for the VLMC-sieve bootstrap: if the covariances $R(k)$ decay exponential fast as $k \to \infty$, we obtain for any stationary, suitably regular categorical process,

$$n\mathrm{Var}^{*VLMC-S}(\overline{X}_n^{*VLMC-S}) - n\mathrm{Var}(\overline{X}_n) = O_P(\log(n)/\sqrt{n}), \tag{6}$$

which is an adaptation in a *fully data-driven* way to the case of exponential decaying covariances. Note that the process $(X_t)_{t\in\mathbb{Z}}$ is generally not a VLMC.

Numerical results. The differences between (5) and (6) can be well seen in finite-sample problems, and the gain of the VLMC-sieve bootstrap is often substantial. Figure 1 displays the quality of distribution estimation (which is beyond variance estimation) for the following situation: sample size $n = 128$, estimator $\hat{\theta} = \overline{X}_n$ with $X_t = 1_{[Y_t > 0]}$ $(t \in \mathbb{Z})$, $(Y_t)_{t\in\mathbb{Z}}$ an AR(1) with parameter 0.8 and t_6 innovations. The bootstrap procedures are block-bootstrap with block-length $\ell = 5 \approx n^{1/3}$, block-bootstrap with estimated block-length $\hat{\ell}$ from the data (see Bühlmann and Künsch, 1999) and VLMC sieve-bootstrap (which involves a tuning parameter as well). The number of simulation runs for the model is 100. Figure 1 is representative for other, more complicated situations with exponential decaying dependence structure: the VLMC sieve-bootstrap is superior to the block-bootstrap, the latter being sensible to the specification of the block-length tuning parameter.

4. Open problems and conclusions

Improving upon the non-adaptivity in theory (see (5)) and practical inefficiencies with the block-bootstrap for *general* stationary, *real-valued* time series is still an open problem. The VLMC-sieve bootstrap idea can be extended to real-valued time series by first applying a scalar quantization $q : \mathbb{R} \to \mathcal{X}$. Preliminary heuristics suggest in some cases an adaptivity result as in (6) which is also visible in finite-sample numerical examples. A strategy different from sieve-bootstrapping leading to adaptivity as in (6) would be interesting, too.

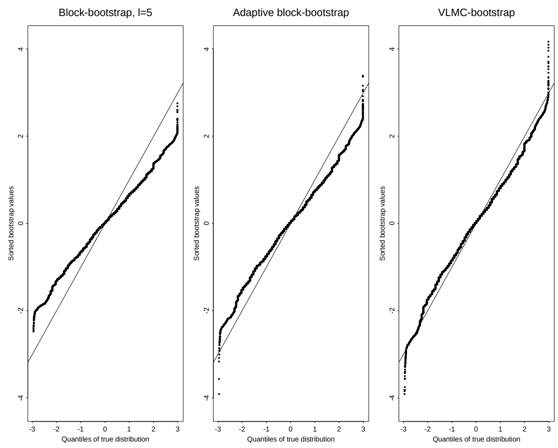

Figure 1. QQ-plots with reference line $y = x$ of three bootstraps for estimating the distribution of $(\overline{X}_n - \mathbf{E}[\overline{X}_n])/\sqrt{Var(\overline{X}_n)}$ (100 simulations, 500 bootstrap replicates).

Conclusions.

(A) The block-bootstrap is in some cases outperformed: by the AR- and VLMC-sieve bootstraps for stationary linear (as in (3)) and stationary, categorical time series, respectively.

(B) Sieve-bootstraps enjoy the user-friendly advantage of using the plug-in rule. In case of the VLMC-sieve, the difficulty is the complicated construction of the resample. The block-bootstrap resampling step is always easy to implement, but the computation of the block-bootstrapped estimator might have to be redesigned, due to its non-plug-in definition.

(C) Sieve bootstrap schemes seem more stable than the block-bootstrap with respect to choosing a tuning parameter. The former involve interpretable approximating models, whereas the latter uses a block-length (a type of bandwidth) as tuning element.

REFERENCES

Bühlmann, P. (1997). Sieve bootstrap for time series. Bernoulli **3**, 123-148.

Bühlmann, P. and Künsch, H.R. (1999). Block length selection in the bootstrap for time series. To appear in Computat. Statist. Data Anal.

Bühlmann, P. and Wyner, A.J. (1999). Variable length Markov chains. To appear in Ann. Statist. **27**.

Ferrari, F. (1999). Estimation of general stationary processes by variable length Markov chains. Preprint, ETH Zürich.

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

Rissanen, J. (1983). A universal data compression system. IEEE Trans. Inform. Theory **IT-29**, 656-664.

RÉSUMÉ

Nous considérons le block- et des versions de sieve-bootstrap pour des series temporelles stationnaires. Les dés-/avantages de chaques methodes sont discutés, particulièrement une qualité adaptive pour les sieve-bootstraps.