Proceedings of the 1988 Winter Simulation Conference
M. Abrams, P. Haigh, and J. Comfort (eds.)

Bootstrap in time series models

Arup Bose
Department of Statistics
Purdue University
West Lafayette, IN 47907

ABSTRACT

The bootstrap is a resampling method of estimating distributional properties of estimators. We discuss how this method can be applied to time series models and indicate directions of theoretical and applied research in this area.

INTRODUCTION

The bootstrap was introduced by Efron (1979) mainly to estimate the statistical accuracy of estimators. A brief description of this procedure follows.

Let $X_1, \ldots, X_n$ be i.i.d. observations from a distribution F, which is completely or partially unknown. Let $\theta_n = \hat{\theta}(X_1, \ldots, X_n)$ be a symmetric estimator of a functional $\theta(F)$.

To have an idea of the accuracy of θ_n, one often considers $V(\theta_n) = \sigma^2(F, n, \theta_n) = \sigma^2(F)$ and one way to estimate it is to use $\sigma^2(F_n)$ where F_n is the empirical distribution function.

Example If $\theta(F) = \int x \, dF(x) = E_F(X_1)$ then $\theta_n = \int x \, dF_n(x) = n^{-1} \sum_{i=1}^{n} X_i = \bar{X}_n$ and $V(\theta_n) = n^{-1} \mu_2(F)$ where $\mu_2(F)$ is the population variance and is estimated by $\mu_2(F_n) = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$.

However, in most cases an explicit formula for $\sigma^2(\theta_n)$ is unavailable and the following algorithm (bootstrap) is used

(i) Estimate F by F_n.

(ii) Draw an i.i.d. sample $X_1^*, \ldots, X_n^*$ from F_n. Based on this bootstrap sample calculate $\hat{\theta}^* = \hat{\theta}(X_1^*, \ldots, X_n^*)$.

(iii) Repeat step (ii) a large number B of times and obtain $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$. Calculate $\hat{\sigma}^2 = \frac{1}{B-1} \sum_{i=1}^{B} (\hat{\theta}_b^* - \hat{\theta}_{\cdot}^*)^2$ where $\hat{\theta}_{\cdot}^* = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b^*$ and $\hat{\sigma}^2$ is an estimate of $\sigma^2(\theta_n)$.

The basis of doing this is that if $B \to \infty$ then $\hat{\sigma}^2 \to \sigma^2(\theta_n)$.

The three points worth noting are (a) the i.i.d. structure is crucial for the algorithm to work, (b) the method can be used for parametric or semiparametric models by replacing F_n by any other appropriate estimate at step (i), (c) any property of the estimate θ_n can be estimated in this manner. Indeed, the entire probability distribution of θ_n can be estimated by forming a histogram based on the bootstrap values. This is specially useful in confidence interval problems.

There is a large amount of theoretical and empirical studies on bootstrap in various types of problems. Refer to Bose (1988b) for a fairly extensive list of references till the year 1987. For modifications of the bootstrap procedure, refer to Efron (1987).

How can one possibly apply the bootstrap idea in the absence of an i.i.d. structure? In some situations an approximate i.i.d. structure can be recovered. One such situation is the class of autoregressive type models in time series.

In the next section we discuss how this method can be applied to such models. The classical normal approximation is available in most cases. We compare this with the bootstrap approximation. For many nonlinear models, even normal approximation is hard or impossible to justify. The bootstrap procedure has added importance in those situations. The need for both simulation and theoretical studies seem great in such models. The effect bootstrap has, in general on dependent models is far from understood. Further research in this area seems very pertinent.

BOOTSTRAP IN DEPENDENT MODELS

Freedman (1984) was the first to attempt bootstrap in dependent models. In a certain linear dynamic model he showed that the bootstrap distribution of the parameter estimates give the same asymptotic result as does the asymptotic normal theory. Almost no theoretical results for dependent models have appeared since then.

In the following subsections we deal with linear autoregressions, nonlinear autoregressions and other models. The stress is on questions which are waiting to be satisfactorily settled.

(a) Linear autoregression. One of the simplest dependent models is the stationary autoregressive process of order 1 (AR (1)).

Let (X_t) be an observable process generated by $X_t = \theta X_{t-1} + \varepsilon_t$, $t = 1, 2, \ldots$ where (ε_t) is a sequence of i.i.d. variables with $E\, \varepsilon_t = 0$ and $E\, \varepsilon_t^2 = 1$. We further assume that the unknown θ satisfies $|\theta| < 1$.

An efficient estimator of θ is the least squares estimator (l.s.e.) obtained by minimizing $\sum_{t=1}^{n} (X_t - \theta X_{t-1})^2$, yielding $\theta_n = (\Sigma X_{t-1}^2)^{-1} \Sigma X_t X_{t-1}$. Rubin (1950) showed that $\theta_n \to \theta$ a.s. no matter what the true value of θ is.

Clearly, the sampling distribution of θ_n is intractable. An approximation is obtained by showing that $n^{1/2} (\theta_n - \theta) \xrightarrow{\mathcal{L}} N(0, 1 - \theta^2)$. Hence an approximate 100 $(1 - \alpha)$% confidence interval for θ is given by $\theta_n \pm z(1 - \alpha/2)\, n^{-1/2} (1 - \theta_n^2)^{1/2}$. The accuracy of this interval is limited by the following theorem.

Theorem 1. If the distribution of $(\varepsilon_1, \varepsilon_1^2)$ satisfies Cramer's condition and $E\, \varepsilon_1^8 < \infty$ then $\sup_x | P(n^{1/2} (\theta_n - \theta) \leq x) - P(N(0, 1 - \theta^2) \leq x) | = 0(n^{-1/2})$. See Bose (1988a) for a proof. The Cramer's condition is satisfied, if, for example ε_1 has a density or has an absolutely continuous component.

In addition to this limitation, the confidence interval is always symmetric, whereas, the distribution of θ_n may be far from being so.

The bootstrap distribution, in general is not symmetric and in fact corrects for the skewness. See Efron (1979). Thus the bootstrap emerges as a competitor.

Note that from the model, $\varepsilon_t = X_t - \theta X_{t-1}$. Hence the ε_t's can be recovered in an approximate manner by defining $\tilde{\varepsilon}_t = X_t - \theta_n X_{t-1}$ and to improve this, let $\hat{\varepsilon}_t = \tilde{\varepsilon}_t - \bar{\varepsilon}_n$ where $\bar{\varepsilon}_n = n^{-1} \sum_{t=1}^{n} \tilde{\varepsilon}_t$ so that the mean of $\hat{\varepsilon}_t$'s is zero. F_n is the distribution function which puts mass $1/n$ at each $\hat{\varepsilon}_t$, $t = 1, \ldots, n$. The bootstrap distribution can now be obtained. Generate $\varepsilon_1^*, \ldots, \varepsilon_n^*$ i.i.d. F_n. Generate X_t^* by $X_t^* = \theta_n X_{t-1}^* + \varepsilon_t^*$, $X_0^* = X_0$. Having obtained $(X_1^*, \ldots, X_n^*)$ pretend that θ_n is unknown and obtain its l.s.e. by $\theta_n^* = (\Sigma X_t^* X_{t-1}^*)(\Sigma X_{t-1}^{*\,2})^{-1/2}$.

The (conditional) distribution of θ_n^* given $X_0, X_1, \ldots, X_n$ can be used to approximate the distribution of θ_n.

Theorem 2. Assume the conditions of

Theorem 1. For almost every sequence $X_0, X_1, \ldots,$

$$\sup_x \left| P^*\left(\frac{n^{1/2}\,(\theta_n^* - \theta)}{(1-\theta_n)^2} \le x \right) - P\left(\frac{n^{1/2}\,(\theta_n - \theta)}{(1-\theta^2)^{1/2}} \le x \right) \right| = o\,(n^{-1/2})$$

For a proof see Bose (1988b). In practice, the bootstrap distribution P^* cannot be explicitly calculated but can be approximated to any desired degree of accuracy by a sufficiently large number of bootstrap samples.

Clearly then, the bootstrap outperforms the normal approximation in some sense. The method can be extended to situations with $E\,\varepsilon_t = \mu$, $E\,\varepsilon_t^2 = \sigma^2$ with μ and σ^2 unknown. See Bose (1988b) for details. Some simulation results are available in Chatterjee (1985) but the author has not been able to obtain a copy of his work at the time of writing this paper.

The case when $|\theta| \ge 1$ is trickier. Normal approximation is not valid. However, for a suitably normed θ_n, the bootstrap still gives the correct result asymptotically. See Basawa et al. (1987). However, no results about its accuracy is known. Simulation work might yield an insight in this problem.

(b) Nonlinear autoregressions. The scope of linear models being limited, researchers are looking increasingly at nonlinear models to explain complicated data. One class of models which has drawn some attention is $X_t = f(\theta, X_{t-1}) + \varepsilon_t$ where f is a nonlinear function. See Ozaki (1980), Jones (1978), Tjostheim (1986) for examples of such models.

As before, we can obtain the l.s.e. θ_n of θ by minimizing $\sum_{t=1}^{n} (X_t - f(\theta, X_{t-1}))^2$, which might need the use of an algorithm. Asymptotic properties of θ_n do not follow immediately. Klimko and Nelson (1978) and Tjostheim (1986) prove the consistency and asymptotic normality of θ_n in a general set up and these can be applied to the present situation. However, the function f needs to be smooth (at least three derivates) for their results to apply. Bose (1988c) obtains these properties with conditions similar to the above authors' but with less restrictive conditions on f.

Theorem 3. Assume the following conditions:

(a) $|f(\theta, x) - f(\varphi, x)| \le K(|\theta - \varphi|)\,J(x)$

(b) $\int_{-1}^{1}\int_{-1}^{1} \frac{K^2(|\theta - \varphi|)}{g(|\theta - \varphi|)}\,d\theta\,d\varphi \le C$ with $L(x) = \int_0^x u^{-1}\,d\,g^2(u) < \infty$

(c) $\sum_{t=1}^{n} E\,J^2(X_{t-1}) = 0\,(n^2 (\log n)^{-(1+\varepsilon)})$

(d) $n^{-1}\sum_{t=1}^{n} [f(\theta, x) - f(\theta_0, x)]^2 \longrightarrow I(\theta)$ a.s. where θ_0 is the unknown true value and $I(\theta) \ne 0$ if $\theta \ne \theta_0$

(e) $|f'(\theta, x) - f'(\varphi, x)| \le K(|\theta - \varphi|)\,J(x)$ and $|f'(\theta, x)| \le J(x)$ with $K(x) = 0\,(x^\alpha)$ as $x \longrightarrow 0$ for some $\alpha > 0$.

(f) $n^{-1}\sum_{t=1}^{n} f'^2(\theta_0, X_{t-1}) \longrightarrow J(\theta_0) > 0$ a.s.

Under these conditions, if the parameter space is $[-1, 1]$, $\theta_n \longrightarrow \theta_0$ a.s. and $n^{1/2}(\theta_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, J(\theta_0))$.

The above result can be generalized to some extent by relaxing some of the conditions.

However, no results are known regarding the rate of the above convergence. Simulation studies might help to show this rate for various choices of f and ε_t's. It can also help to judge the asymptotics when some of the above conditions are violated (e.g. (e) and (f)) which impose restrictions on the smoothness of f.

The bootstrap idea assumes more importance due to the sharpness of normal approximation being unknown. It can be implemented by defining $\tilde{\varepsilon}_t = X_t - f(\theta_n, X_{t-1})$ and proceeding as in the linear case. Simulation studies can help gauge the performance of the bootstrap. Theoretical aspects of the bootstrap also promises to be a difficult and interesting problem.

(c) Other models. As is clear, the bootstrap idea can be used whenever an approximate i.i.d. structure can be recovered, usually by replacing unknown values with their estimates. It appears that this is tied up with the notion of invertibility of models. See Bose (1987) and Hannan (1970). As a test case Bose (1987) shows that it works well in the moving average model, $X_t = \varepsilon_t + \alpha\, \varepsilon_{t-1}$ where resampling is a little trickier. As we have already discussed the bootstrap works in a AR (1) model with $|\theta| \geq 1$. So the invertibility does not seem crucial. This shows the inherent automatic nature and power of the bootstrap procedure.

CONCLUSION

In conclusion, bootstrap works in a wide variety of situations, even when the normal theory is no longer valid. When the normal theory is valid, bootstrap can outperform it. Finding a class of models where bootstrap works is an interesting question. Theoretical studies and simulation work are needed to gauge its performance and the factors affecting it. It will also be of interest to see how it performs where no known results of the properties of the estimator is available. Some of these problems are being looked at by the author currently.

REFERENCES

Basawa, I.V., Mallik, A.K., McCormick, W.P., Taylor, R.L. (1987). Bootstrapping nonstationary autoregressive processes. Technical Report # STA 64, Department of Statistics, University of Georgia.

Bose, Arup (1987). Bootstrap in moving average models. Technical Report # 87-55, Department of Statistics, Purdue University.

Bose, Arup (1988a). Higher order approximations for autocovariances from linear processes with applications. *Statistics*, *19*, *2*, 259-269.

Bose, Arup (1988b). Edgeworth correction by bootstrap in autoregressions, Technical Report # 87-37, Department of Statistics, Purdue University. To appear in *Ann. Statist.*, *6*, *4*, December.

Bose, Arup (1988c). Asymptotic estimation in non linear time series models. Under preparation.

Chatterjee, S. (1985). Bootstrapping ARMA models: Some simulations. Preprint, College of Business Administration, Northeastern University.

Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *Ann. Statist.*, 7, 1-26.

Efron, B. (1987). Better bootstrap confidence intervals. *Jour. Amer. Stat. Assoc.*, *82*, *397*, 171-185.

Freedman, D. (1984). On bootstrapping two stage least squares estimates in stationary linear models. *Ann. Statist.*, *12*, 827-842.

Hannan, E.J. (1970). *Multiple Time Series*. John Wiley and Sons.

Jones, D.A. (1978). Non-linear autoregressive processes. *Proc. Roy. Soc. London A*, *360*, 71-95.

Klimko, L.A., and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.*, *6*, 629-642.

Ozaki, T. (1980). Non-linear time series models for non-linear random vibrations. *J. Appl. Probab.*, *17*, 84-93.

Rubin, H. (1950). Consistency of maximum likelihood estimates in the explosive case. In *Statistical Inference in Dynamic Economic Models* (*T.C. Koopman, Ed.*), 356-364. Wiley, New York.

Tjostheim, D. (1986). Estimation in nonlinear time series models. *Stoch. Proc. Appl.*, *21*, 251-273.

BIOGRAPHY

DR. ARUP BOSE received his Bachelor's, Master's and Doctorate degrees in statistics from the Indian Statistical Institute, Calcutta in 1980, '81, and '87 respectively. Since 1987, he is a visiting Assistant Professor at the department of Statistics, Purdue University. He has more than a dozen published papers in leading statistical journals. His main research interest is in the area of asymptotic theory, with special reference to estimation of parameters in dependent models.

Arup Bose
Department of Statistics
Purdue University
West Lafayette, IN 47907
(317) 494-6043