SECOND ORDER CORRECTNESS OF THE BLOCKWISE BOOTSTRAP FOR STATIONARY OBSERVATIONS

F. Götze

Fakultät für Mathematik
University of Bielefeld
D-33501 Bielefeld 1, Germany

H.R. Künsch

Seminar für Statistik
ETH Zürich
CH-8092 Zürich, Switzerland

Abstract

We show that the blockwise bootstrap approximation for the distribution of a studentized statistic computed from dependent data is second order correct if we choose the variance estimators such that their biases are similar. For the proof we extend the results of Götze and Hipp (1983) on the validity of the formal Edgeworth expansion for a sum to the studentized mean. The paper concludes with a small simulation example.

Keywords. Edgeworth expansion, studentization, time series, dependent data, strong mixing, resampling.

AMS 1980 Subject classifications. Primary 62M10; secondary 62G15, 62E20.

Running head. Second order correctness.

1 Introduction

Efron's (1979) bootstrap is distribution free, but the assumption of independence is crucial since it ignores the time order of the observations. For serially dependent observations, Künsch (1989) has proposed a blockwise bootstrap which samples blocks of length l of consecutive observations with replacement. He has shown that this procedure estimates the asymptotic variance and the asymptotic normal distribution consistently if the statistic considered is the arithmetic mean or a function thereof. He assumes only suitable mixing and moment conditions for the observations and an increase of $l = l(n)$ to infinity, but at a slower speed than n. However for a simple statistic like a function of an arithmetic mean there are other methods to estimate the asymptotic variance which require fewer computations, e.g the blockwise jackknife or an estimate of the spectrum at zero for the estimated influence function (see also Künsch, 1989). Hence in these cases the additional effort required for the bootstrap is only justified if the resulting approximation to the distribution of the statistic is better than the one relying on asymptotic normality. With i.i.d. data Efron's bootstrap has indeed this property, see e.g. Singh (1981), Hall (1988). Here we show that this is also true for the blockwise bootstrap as announced some time ago in an abstract (Götze and Künsch, 1990). The essential reason for this result is that the blockwise bootstrap distribution has asymptotically the correct skewness (cf. Künsch, 1989, Formula (3.19)). This gain in accuracy is however covered by the error in the bootstrap variance which is in our case of the order $O(l^{-1})+O_p((l/n)^{1/2})$ and thus always larger than the order $O(n^{-1/2})$ of the skewness term. Thus for a better than normal approximation we have to make the true and the bootstrap variance equal, e.g. by studentizing the statistic. When implementing this idea it turns out that it is crucial how the variances are estimated, see Davison and Hall (1993). For our result the essential requirement is that the variance estimators for the original and the bootstrap sample have the same bias.

The most difficult part in proving our result is to show the validity of the Edgeworth expansion for the studentized statistic for dependent observations. Götze and Hipp (1983) have shown the validity of the Edgeworth expansion for the arithmetic mean under a set of conditions covering a broad range of situations. We are going to use the same conditions. The difficulty in passing from the arithmetic to the studentized mean comes from the fact that the estimated variance is the sum of l different arithmetic means, namely the sample covariances for lags $0, \ldots, l-1$, and l increases with n. Lahiri (1991) has shown that the blockwise bootstrap is second order correct if the asymptotic variance is known and we use it to standardize the mean. In the case where the observations are m-dependent with m known, his Theorem 2.2 can be used also for the studentized mean, but with a different way of studentization. In this way he avoids the problem with increasing l, but the result is of limited use since typically both the asymptotic variance and m are unknown.

Our result is somewhat surprising because dependence makes the estimation of the