S.N. Lahiri

**Effects of block lengths on the validity of block resampling methods**

Received: 29 June 1998 / Revised version: 13 November 2000 / Published online: 14 June 2001 – © Springer-Verlag 2001

**Abstract.** Resampling methods for dependent variables typically involve selecting blocks of observations. In this paper, we investigate the effects of different block lengths on the consistency of approximations generated by block bootstrap and subsampling methods. A complete description is provided for the case of sample mean. It is shown that both methods are consistent if the block length grows at a rate slower than the sample size. When the growth rate of block lengths is comparable to the sample size, the resulting approximations are no longer consistent. In the latter case, we present a detailed account of the limit behavior of these resampling distribution approximations, considered as random elements in the space of all probability measures on the real line.

1. Introduction

Resampling methods have been introduced in the literature as ways of generating accurate approximations to unknown sampling distributions of different statistics without restrictive model assumptions. When the underlying random process generating the observations is correlated, the sampling distribution of a statistic typically depends on the joint distribution of the process-variables and can not be determined only from the knowledge of their one-dimensional marginal distribution(s). The bootstrap and the other resampling methods for such dependent processes thus differ from their classical versions for independent and identically distributed (iid) variables by resampling ‘blocks’ of observed variables as compared to resampling a single observation at a time. The intuitive reasoning behind this is that the dependence structure of the process is preserved within the blocks, and therefore, valid approximations can be obtained by resampling the blocks. However, the validity of approximations generated by a block resampling method crucially depends on several factors, such as, the length of the blocks used in the resampling procedure, dependence structure of the underlying process, resample size, etc. As a result, for each of these factors, it is important to know the range of values for which these block resampling methods provide consistent or valid approximations and outside
which these methods break down. In the case of iid random variables, a necessary and sufficient condition for the validity of Efron (1979)’s bootstrap for the sample mean is known (cf. Giné and Zinn (1989, 1990)). The situation is somewhat different in the dependent case because of the multiplicity of factors involved. Partial answers to the validity of some block resampling procedures under certain factors, like strength of dependence (weak dependence versus long-range dependence) and the existence of the second moment, have been obtained in the literature (cf. Lahiri (1993, 1995)). In this paper, we investigate the effects of block lengths on the validity (i.e., consistency) of approximations generated by some commonly used block resampling procedures.

We consider the simplest situation, namely that of approximating the sampling distribution of the sample mean by block resampling methods and under very mild regularity conditions, obtain a necessary and sufficient condition on the growth rate of the block length for consistency. The main results of the paper show that under some regularity conditions, approximations derived from block resampling methods are consistent if and only if the block length goes to infinity at a rate slower than the sample size $n$. The ‘if part’ has been noted by some authors under various regularity conditions. The main contribution of the paper is to analyze limit behavior of approximations when the block length say, $\ell$ goes to infinity as fast as the sample size, (thus providing an answer for the ‘only if part’). Here, two very distinct behaviors of the resampling approximations are noted. For the subcase $\ell/n \to \lambda \in (0, 1)$, assuming that a weak invariance principle holds for the sample partial sum process, it is shown that the distributional approximations generated by these block resampling methods have random limits on the space of all probability measures on the real line. The limiting random probability measures are defined in terms of the sample paths of the Brownian motion. In the other subcase where $\ell/n \to 1$, the resampling approximations are shown to converge in probability to the (nonrandom) probability measure degenerate at zero. Thus, whenever the block size $\ell$ goes to infinity as fast as $n$, the resampling approximations fail to be consistent.

On the other end, where the block length $\ell$ goes to infinity at a rate slower than the sample size $n$, we show that the resampling methods provide valid approximations to the distribution of the sample mean under very mild moment and mixing conditions. In this case, we also show that not only are the distributional approximations consistent, variance estimators generated by block resampling schemes are also consistent under very minimal regularity conditions. These results extend some recent results of Peligrad and Shao (1995, 1996) on variance estimation for stationary weakly dependent processes.

The effect of block lengths for more general estimators, that can be approximately expressed as smooth functions of multivariate sample means, (for example, the maximum likelihood-type estimators or M-estimators), can be derived from these results using linearization. Indeed, it is clear from the discussion in Sections 3 and 4 that inconsistency under the case $\ell/n \to \lambda \in (0, 1]$ and consistency under the case $\ell/n \to 0$ continue to hold for such general estimators.

The rest of the paper is organized as follows. A brief literature review concludes Section 1. Section 2 lays down some preliminary framework for the problem