BOOTSTRAP FOR THE SAMPLE MEAN AND FOR U-STATISTICS OF WEAKLY DEPENDENT OBSERVATIONS

Olimjon Sharipov[1] and Martin Wendler[2]

[1]Institute of Mathematics and Information Technologies, Tashkent, Uzbekistan
[2]Bochum University, Bochum 44780, Germany
[1]*osharipov@yahoo.com*, [2]*Martin.Wendler@rub.de*

In many statistical applications the data does not come from an independent stochastic process. A standard assumption of weak dependence is given by the strong mixing condition:
Definition 1. Let $(X_n)_{n\in N}$ be a stationary process. Then the strong mixing coefficient is given by

$$\alpha(k)=\sup\left\{\left|P(A\cap B)-P(A)P(B)\right|: A\in F_1^n, B\in F_{n+k}^{\infty}, n\in N\right\}$$

where F_a^l is the $\sigma-$field generated by r. v. 's $X_a,...,X_l$., and $(X_n)_{n\in N}$ is called strongly mixing, if $\alpha(k)\to 0$ as $k\to\infty$.
For further information on strong mixing and a detailed description of other mixing conditions see Doukhan [4] and Bradley [2].
In many statistical applications, for example in the determination of confidence bands, one faces the task to compute the distribution of a statistic $T_n=T_n(X_1,...,X_n)$. Thus is usually rather difficult, as the distribution F of X_i is unknown, so one often has to use approximation by the normal distribution. Efron [5] proposed the bootstrap as an alternative. For i.i.d. data, the validity of the bootstrap was established by Bickel and Freedman [1], and by Singh [11]. Using Edgeworth expansion, one can often show that the bootstrap works better than normal approximation, see Hall [6] for details. Computation of the distribution of T_n becomes even more difficult when the observations are dependent, e.g., in the case of the sample mean $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, one gets for weakly dependent data under some technical assumptions $\sqrt{n}\left(\bar{X}_n-EX_1\right)\to N(0,\sigma^2)$ in distribution, where $\sigma^2=Var[X_1]+2\sum_{i=1}^{\infty}Cov[X_1,X_{i+1}]$.So one has not only the variance to estimate, but also the autocovariances of the process. The naïve bootstrap can fail under dependence, as Singh [11] mentioned. Therefore, block bootstrappings method are commonly used for nonparametric inference under dependence. There are different ways to resample blocks, for example the circular block bootstrap or the moving block bootstrap (for a detailed description of the different bootstrapping methods see Lahiri [7]. For the circular block bootstrap, Shao and Yu [10] have shown that under strong mixing the distribution of the block bootstrap version $\bar{X}_n^*$ of the sample mean converges almost surely to the same distribution as the sample mean $\bar{X}_n$. Peligrad [8] has proved asymptotic normality of $\bar{X}_n^*$ under another set of conditions, which does not necessarily imply the central limit theorem for $\bar{X}_n$. Radulovic [9] has established weak consistency under very weak conditions. We consider the nonoverlapping bootstrap, proposed

by Carlstein [3], for the sample mean and for U-statistics. Let $\left(X_n\right)_{n\in N}$ be a sequence of r.v.'s. Let $p \in N$ be the block length such that $p = p(n) = o(n),\ p \to \infty$ as $n \to \infty$. We introduce the following blocks of indices and r.v.'s:

$$I_i = \left(X_{(i-1)p+1},...,X_{ip}\right),$$

$$B_i = \left\{(i-1)p+1,...,ip\right\},\ i = 1,...,k$$

where $k = k(n) = \left[\frac{n}{p}\right]$ is the number of blocks. We consider a new sample $X_1^*,...,X_{kp}^*$, which is constructed by choosing randomly and independently blocks k times with

$$P\left(\left(X_1^*,...,X_p^*\right) = I_i\right) = \frac{1}{k}\ ,\ i = 1,2,...,k.$$

As a bootstrap version of the sample mean we consider

$$\overline{X}_{n,kp}^* = \frac{1}{kp}\sum_{i=1}^{kp} X_i^*.$$

With P*,E*, var* we denote the probability, expectation and variance conditionally on $\left(X_n\right)_{n\in N}$. Note that

$$E^*\overline{X}_{n,kp}^* = \frac{1}{kp}\sum_{i=1}^{kp} X_i =: \overline{X}_{n,kp}^*.$$

In what follows, we denote by $\bar{X}_n$ the sample mean of the observations $X_1,...,X_n$, by $N\left(0,\sigma^2\right)$ a Gaussian r.v. with mean zero and variance σ^2 and by $1_{\{\cdot\}}$ an indicator function. Here we will give results for the sample mean only. First we will give theorems for general stationary sequences which are analogues to the results of Peligrad [8], and Shao and Yu [10].

Theorem 1. Let $\left\{X_i, i \geq 1\right\}$ be a stationary sequence of r.v'.s such that $EX_1 = \mu$ and $VarX_1 < \infty$ .Assume that the following conditions hold

(1) $$Var\ n^{\frac{1}{2}}\left(\bar{X}_n - \mu\right) \to \sigma^2 > 0,$$

(2) $$n^{\frac{1}{2}}\left(\bar{X}_n - \mu\right) \to N\left(0,\sigma^2\right) \text{ in distribution,}$$

(3) $$p^{\frac{1}{2}}\left(\bar{X}_{n,kp} - \mu\right) \to 0 \quad \text{a. s. ,}$$

(4) $$\frac{1}{kp}\sum_{i=1}^{k}\left[\left(\sum_{j\in B_i}^{p}\left(X_i - \mu\right)\right)^2 - E\left(\sum_{j\in B_i}\left(X_j - \mu\right)\right)^2\right] \to 0 \quad \text{a. s.}$$

$$\frac{1}{kp}\sum_{i=1}^{k}\left(\sum_{j\in B_i}\left(X_j-\mu\right)\right)^2 1_{\left\{\left|\sum_{j\in B_i}\left(X_j-\mu\right)\right|^2>\varepsilon kp\right\}}\to 0 \qquad \text{a. s.}$$

for any $\varepsilon > 0$. Then the following takes place as $n\to\infty$

$$Var^*\left(\sqrt{kp}\,\bar{X}^*_{n,kp}\right)\to\sigma^2 \qquad \text{a. s.}$$

$$\sup_{x\in R}\left|P^*\left(\sqrt{kp}\left(\bar{X}^*_{n,kp}-\bar{X}_{n,kp}\right)\le x\right)-P\left(\sqrt{n}\left(\bar{X}_n-\mu\right)\le x\right)\right|\to 0 \quad \text{a.s.}$$

Theorem 2. Let $(X_n)_{n\in N}$ be a stationary sequence of r.v.'s. with $EX_1=\mu$, Var $X_1<\infty$. Assume that conditions (1), (2), (4) and for each fixed $x\in R$

$$\frac{1}{kp}\sum_{k=1}^{k}\left(1_{\left\{\frac{1}{\sqrt{p}}\sum_{j\in B_i}\left(X_j-\mu\right)\le x\right\}}-P\left(\frac{1}{\sqrt{p}}\sum_{i=1}^{p}\left(X_i-\mu\right)\le x\right)\right)\to 0 \quad \text{a.s.}$$

hold. Then the statement of Theorem 1 remains true.

Theorem 3. Let $(X_n)_{n\in N}$ be a stationary sequence of bounded almost surely r.v. 's with $EX_1=\mu$. Assume that (3) and following conditions hold

(5)
$$\frac{p^2}{n}\to 0 \qquad \text{as } n\to\infty,$$
$$\frac{1}{n}VarS_n\to\sigma^2 \qquad \text{as } n\to\infty,$$
$$\frac{1}{kp}\sum_{i=1}^{k}\left(\sum_{j\in B_i}\left(X_j-\mu\right)\right)^2\to\sigma^2 \qquad \text{a.s. as } n\to\infty.$$

Then almost surely as $n\to\infty$

(6) $$Var^*\left(\sqrt{kp}\,\bar{X}^*_{n,kp}\right)\to\sigma^2,$$

(7) $$\sqrt{kp}\left(\bar{X}^*_{n,kp}-\bar{X}_{n,kp}\right)\to N\left(0,\sigma^2\right).$$

We formulate theorems under assumptions on the strong mixing coefficients which are analogues to the results of Peligrad [8] and Shao, Yu [10].

Theorem 4. Let $(X_n)_{n\in N}$ be a stationary sequence of strong mixing r.v.'s with

$EX_1=\mu$ and $\left(E\left|X_1\right|^{2+\delta}\right)^{\frac{1}{2+\delta}}<\infty$ for some $0<\delta\le\infty$. Assume

$$\alpha(n)\le Cn^{-1} \text{ for some } C>0,\ r>\frac{2+\delta}{\delta},$$

$$p(n)\le Cn^{\varepsilon} \text{ for some } 0<\varepsilon<1 \text{ and}$$

(8) $$p(n) = p(2^l) \text{ for } 2^l < n \le 2^{l+1},\ l = 1,2,...$$

Then $\sigma^2 = EX_1^2 + 2\sum_{i=1}^{\infty} Cov(X_1, X_i) < \infty$ and in the case $\sigma^2 > 0$ the statement of Theorem 1 holds.

Theorem 5. Let $(X_n)_{n\in N}$ be a stationary sequence of almost surely bounded strongly mixing r.v. 's. Assume that (5), (8) and the following conditions hold

$$\sum_{n=1}^{\infty} \frac{p^2(n)\alpha(p(n))}{n} < \infty,$$

$$\sum_{n=1}^{\infty} \frac{p^3(n)}{n^2} < \infty.$$

Then (6), (7) hold.

We have established consistency of the bootstrap version of U-statistics of mixing observations, but results will be given in another paper.

References

1. Bickel, P.J. and Freedman, D.A. (1981) Some asymptotic theory for the bootstrap.Ann.Stat. , v.9, 1196-1217.
2. Bradley, R.C. (2007) Introduction to strong mixing conditions.Vol.1-3, Kendrick Press,Heber City, Utah.
3. Carlstein, E. (1986) The use of subseries values for estimating the variance of a general statistic from stationary sequence. Ann.Stat.v.14, 1171-1179.
4. Doukhan, P. (1994) Mixing.Springer,New York.
5. Efron, B. (1979) Bootstrap methods: another look at the jackknife. Ann.Stat., v.7, 1-26.
6. Hall, P. (1992) The bootstrap and Edgeworth expansions. Springer. New York.
7. Lahiri, S.N. (2003) Resampling methods for dependent data. Springer. New York.
8. Peligrad, M. (1998) On the blockwise bootstrap for empirical processes of stationary sequences. Ann.Prob., v.26, 877-901.
9. Radulovic, D. (1996) The bootstrap of the mean for strong mixing sequences under minimal conditions. Statist.Probab.Letters,28, 65-72.
10. Shao, Q.M. and Yu, H. (1993) Bootstrapping the sample meansfor stationary mixing sequences. Stochastic Process.Appl., 48, 175-190.
11. Singh, K. (1981) On the asymptotic accuracy of Efron's bootstrap. Ann.Stat., v.9, 1187-1195.