International Mathematical Forum, Vol. 11, 2016, no. 17, 845 - 851
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/imf.2016.67106

The Accuracy of Two-sided Confidence Interval Algorithm: An Alternative of Double Bootstrap Approach

Z. Nurul Hila

Pusat Pengajian Informatik & Matematik Gunaan
Universiti Malaysia Terengganu, Malaysia

L. Muhamad Safiih

Pusat Pengajian Informatik & Matematik Gunaan
Universiti Malaysia Terengganu, Malaysia

Abstract

The basis idea of double bootstrap is using iteration procedure in the set sample of first bootstrap. The iteration seems to consume a period of time in completing all steps of double bootstrap algorithm and could delaying interested estimation such as accuracy estimation. To overcome this problem, an alternative of double bootstrap is proposed in this research. Instead of using iteration, generally, the procedure is to repeat the whole algorithm of single bootstrap in resulting a sample of double bootstrap. In terms of accuracy, this research constructs two double bootstrap intervals and estimates the coverage rate and its length. The procedure of double bootstrap and its accuracy estimation apply in simulation study. The result shows double bootstrap intervals have the same coverage with nominal rate and gives shorter length when comparing with single bootstrap intervals. Thus, the proposed double bootstrap has given an accuracy estimation and not complex to be used in simulation study.

Keywords: Double bootstrap intervals, Coverage rate, Length of intervals

1 Introduction

The discussion on double bootstrap has become an interesting research either in parametric or nonparametric bootstrap. The aim for this discussion is to find a better result on solving the uncertainty estimation and provide more accuracy estimation when involving the used of either small or finite sample data. In terms of accuracy, the double bootstrap is widely used to estimate the one-sided and two-sided confidence interval, see for example [1], [2], [3] and [4]. These previous researches successfully increased the accuracy of interval estimation, where implementing the double bootstrap statistically proven to decrease the error from a factor of order n^{-2} to order n^{-1}. It was proven either in theoretical and simulation study, see for example [5], [6], and [7].

Even though the previous discussion on double bootstrap gave a high impact, it seems that the procedure of implementing this method required a difficult and excessive calculation. The conventional double bootstrap procedure is used the iteration method where generally the second bootstrap replication is made at each replication sets of first, more likely, single bootstrap [4], [7] and [8]. This procedure is quite easy to perform if the single bootstrap replication number is small but the major problem happen if the replication is a greater value. When it happens, the procedure requires a long term period to perform the algorithm and takes time to estimate the accuracy estimation, such as two-sided confidence interval. To overcome this problem, an alternative algorithm of double bootstrap proposes in this research, where the aim is to show an easier procedure of double bootstrap method and to prove the accuracy of confidence interval using the constructed algorithm of double bootstrap intervals.

This paper set out with some sections. Section 1 is for the brief introduction of an alternative double bootstrap and its application on confidence interval is discussed in Section 2. By considering an example of Exponentially Weighted Moving Average (EWMA) model, the corresponding algorithms of accuracy estimation are detailed in Section 3. By using Monte Carlo simulation study, a complete statistical evaluations are made and its detailed in Section 4. The concluding and some discussions are compiled in Section 5.

2 The Double Bootstrap Confidence Intervals

Consider an observation on a sample X from a population with cdf F and obtain $100(1-2\alpha)\%$ two-sided confidence interval estimators for a parameter $\theta = h(X)$ and its estimator $\hat{\theta}$. For the bootstrap, two samples bootstrap are obtained from replicating procedure of standard error. First procedure is to replicate standard error of X to obtain a single bootstrap sample, X^* , and second procedure is to replicate the standard error of X^* to obtain a double bootstrap sample denote as X^{**}. Both of bootstrap sample are used to estimated the corresponding estimator, $\hat{\theta}^*$ and $\hat{\theta}^{**} = h(X^{**} \mid X^* \cap X)$.

In this research, two bootstrap-base confidence intervals are proposed which are standard and bootstrap intervals. The confidence interval for double bootstrap Studentized is given:

$$I_{t^{BB}} = \left\lfloor \hat{\theta}^{**} - t^{**}_{1-\alpha}\hat{\sigma}(\hat{\theta}^{**}), \hat{\theta}^{**} - t^{**}_{\alpha}\hat{\sigma}(\hat{\theta}^{**}) \right\rfloor \quad (1)$$

where $\hat{\sigma}(\hat{\theta}^{**}) = \left[\sum(\hat{\theta}^{**} - \overline{\theta}^{**})^2 / n(n-1)\right]^{0.5}$. Meanwhile, the confidence interval for double bootstrap Bias Corrected and accelerated (BCa) is given by:

$$I_{BCa^{BB}} = \left\lfloor \hat{\theta}^{**}_{\alpha}, \hat{\theta}^{**}_{1-\alpha} \right\rfloor \quad (2)$$

Consider the Studentised double bootstrap interval given in Equation (1), then the probability of coverage can be given by:

$$P\left\lfloor \hat{\theta}^{**} - t^{**}_{1-\alpha}\hat{\sigma}(\hat{\theta}^{**}) \le \hat{\theta}^{**} \le \hat{\theta}^{**} - t^{**}_{\alpha}\hat{\sigma}(\hat{\theta}^{**}) \right\rfloor = \alpha \quad (3)$$

3 Algorithms on Estimating the Accuracy of Double Bootstrap Intervals

The construction of intervals from previous section requires *BB* sets of double bootstrap replication from the first bootstrapping sample of standard error. Suppose the replication, *BB* = 1500, then from the previous discussion section, the require intervals are I_t^{1500} and I_{BCa}^{1500} for two-sided double bootstrap confidence interval. Consider 90% confidence interval, then, for example Studentised, the value of $\hat{\theta}^{**}$ less than or equal to upper limit of $t^{**}_{0.25}$. The accuracy of these two-sided double bootstrap intervals are likely to be increase if the length of Studentised and BCa interval are estimated. The basis of difference between upper and lower limit is to obtain a shorter length of an interval, and statistically gives the accuracy of estimation.

To obtain useful numerical illustration of bootstrapping procedure, an example of interest model, i.e. EWMA and its residual are considered in this research. The aim of using the example model is to apply and construct the two-sided bootstrap interval of EWMA model. The approach of bootstrap method onto the EWMA model is used in two algorithms shown below.

Algorithm 1: Probability of coverage rates

Step 1 :Generate a sample of n_1 size randomly from, for example, normal distribution. Estimate the EWMA model, $W_i = \lambda x_i + (1-\lambda)W_{i-1}$, where $i = 1,2,...,n_1$

Step 2 :Estimate the residual model of EWMA, $e_i = x_i - W_{i-1}$

Step 3 :Set the single bootstrap replication number, *B*. Sampling the estimation data of e_i randomly with replacement.

Step 4 :Calculate the x_i^b, where $b = 1,2,...,B$ by rearranging the residual equation. Average the row matrix of every x_i^b and denote the sample of single bootstrap is x_i^*.

Step 5 :Estimate EWMA model, W_i^* and its residual, e_i^*

Step 6 :Set the double bootstrap replication number, *BB* and sample the estimation of e_i^* randomly with replacement.

Step 7 :Follow Step 4 and denote the double bootstrap sample is e_i^{**}

Step 8 :Set the confidence interval rate. Calculate the θ^{**} and $\hat{\sigma}(\theta^{**})$

Step 9 :Estimate the upper and lower limit for two-sided BCa and Studentised interval for double bootstrap. Calculate the coverage rate for both intervals

Step 10 :Repeat this algorithm by setting *N* times repetition

Algorithm 2: Length of two-sided double bootstrap intervals

Step 1 :Generate a n_2 sample size randomly and estimate the EWMA model and its residual.

Step 2 :Follow Step 3 – Step 7 of Algorithm 1.

Step 3 :Set the confidence level and α, for example 90% and $\alpha = 0.05$. Estimate the upper and lower limit of BCa interval. Find the length of this interval.

Step 4 :Repeat Step 4 for Studentised interval.

4 Monte Carlo Simulation

The Algorithm 1 and Algorithm 2 are considered to be implementing in simulation study. Consider a sample size n_1 = 30 and *N* = 10 000 repetition for Algorithm 1. Meanwhile, for Algorithm 2, three sample sizes are used, n_2 = 30, 60, 200. Consider a Gaussian distribution, $N(\mu, \sigma^2)$ generate randomly for both algorithms. In case of accuracy performance of two-sided single and double bootstrap intervals, consider confidence interval for mean is estimated by using two nominal coverages, i.e. 90% and 95%. The performance of accuracy estimation due to the use of Algorithm 1 and Algorithm 2 are shown in Table 1 and Table 3 respectively.

Table 1. Empirical Coverage Rates of Bootstrap Intervals

Interval	**90%**		**95%**	
	Single	**Double**	**Single**	**Double**
EWMA-BCa	0.90	0.90	0.95	0.95
EWMA-Studentised	0.64	0.88	0.65	0.93

Result obtained from *N*=10 000 replication with *B* = 1000 and *BB* = 1500 are single and double bootstrap replication respectively.

Table 1 shows the empirical coverage rates of single and double bootstrap intervals. The single and double BCa bootstrap intervals have empirical coverage rate equal to nominal coverage rates, i.e. 0.90 and 0.95. For the coverage rate of single and double Studentised bootstrap confidence interval are 0.64 and 0.88 for nominal coverage 90%. Meanwhile, for 95% coverage, the estimation of Studentised intervals are 0.65 and 0.90 for single and double bootstrap respectively. The double bootstrap estimation appears to be approximately approach to the both nominal rates, whereas single bootstrap is statistically different from nominal rates of 0.90 and 0.95. This result is similar to [9] study, where the coverage rate value of double bootstrap is more approaching to nominal rate and in terms of performance, i.e. for comparison, its found that bootstrap interval method gives a better performance compare to Studentised. For a better conclusive result of estimation in Table 1, a calculation of under-coverage percentage is made and can be further refer to Table 2. An interesting result showed that the double bootstrap interval only given 2.05 (90%) and 2.10 (95%) whereas a higher percentage given using single bootstrap interval, i.e. 25.25 (90%) and 30.20 (95%).

Table 2. Under-coverage Percentage of EWMA-Studentised Intervals

Nominal Coverage	**Single**[a]	**Double**[b]
90%	25.52	2.05
95%	30.20	2.10

[a,b]Two-sided single bootstrap interval.

Table 3 shows the numerical accuracy of single and double bootstrap intervals, where the length is estimated. In both case of confidence level, the length either single or double bootstrap, are decreased as n_2 increase. For example, 95% confidence level for BCa bootstrap intervals, the estimation values are decreased from 0.05097 → 0.04130 → 0.03185, and 0.04008 → 0.03285 → 0.02526. Both of BCa and Studentised are decreased the length estimation as n_2 increase from 30 → 60 → 200. As mentioned by [10], theoretical simulation of performance can be describe as size sample increased, the estimation of accuracy will going to be decreased. Moreover, the interesting part of this result is that, BCa interval method is given shorten length either in single or double bootstrap. This pattern of result was proven by several studies, for example [11], [12] and [13].

Table 3. Empirical Length of Bootstrap Intervals Estimation

Interval	**Single**			**Double**		
	30[b]	**60**	**200**	**30**	**60**	**200**
(a) 90%[a]						
EWMA-BCa	0.042	0.034	0.026	0.033	0.027	0.021
EWMA-Studentised	0.574	0.371	0.218	0.532	0.352	0.205

Table 3. (Continued): Empirical Length of Bootstrap Intervals Estimation

(b) 95%[a]						
EWMA-BCa	0.050	0.041	0.031	0.040	0.032	0.025
EWMA- Studentised	0.660	0.432	0.260	0.641	0.421	0.245

[a]Confidence level choosen in Algorithm 2. [b]The set of sample size, n_2.

However, referring to Table 3, double bootstrap is seen to improve the estimation of single bootstrap, which in describing the decreasing factor of order that has mentioned by [5]. The lengths of Studentised and BCa are reported to be shorter than single bootstrap intervals. For example, $n_2 = 60$ and 90% confidence level for Studentised, the double bootstrap is given 0.027 whereas the single bootstrap is 0.034. In this case, it might due to the used of smaller standard deviation in procedure of estimating double bootstrap confidence intervals. The smaller standard deviation is resulted from replication procedure of residual of single bootstrap, i.e. refers to Algorithm 2, Step 2. The replication decreased the residual onto some small new values which statistically lead to small standard deviation estimation.

5 Conclusions

This research proposed an alternative double bootstrap algorithm using the sampling of residual with replacement. This alternative algorithm implemented onto constructing two-sided double bootstrap confidence interval, and two algorithms of accuracy estimation of interval were conducted. Using Monte Carlo simulation study, Algorithm 1 increased the coverage rate of double bootstrap intervals and decreased the under-coverage percentage of interval. Furthermore, for Algorithm 2, the length of Studentised and BCa intervals were decreased as n_2 increase, and pointed that the lengths were shorter compare to all single bootstrap intervals.

References

[1] F.W. Scholz, On Exactness of the Parametric Double Bootstrap, *Statistica Sinica*, **4** (1994), 477-492.

[2] R. Davidson and J. G. MacKinnon, Improving the Reliability of Bootstrap Test With the Fast Double Bootstrap. *Journal of Econometric*, (2006), 1-31.

[3] F. T. Burbrink and R. A. Pyron, The Taming of the Skew: Estimating Proper Confidence Intervals for Divergence Dates, *Systematic Biology*, **57** (2008), 317–328. http://dx.doi.org/10.1080/10635150802040605

[4] I. C. Nedelea, and J. M. Fannin, Technical Efficiency of Critical Access Hospitals: An Application of the Two-stage Approach with Double Bootstrap, *Health Care Management Science*, **16** (2013), 27–36. http://dx.doi.org/10.1007/s10729-012-9209-8

[5] R. Beran, Prepivoting to Reduce Level Error of Confidence Sets, *Biometrika*, **74** (1987), 457–468. http://dx.doi.org/10.1093/biomet/74.3.457

[6] P. Hall and J. Horowitz, A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions, *The Annals of Statistics*, **41** (2013), 1892-1921. http://dx.doi.org/10.1214/13-aos1137

[7] J. Chang and P. Hall, Double-Bootstrap Methods That Use A Single Double-Bootstrap Simulation, *Biometrika*, **102** (2015), 203-214. http://dx.doi.org/10.1093/biomet/asu060

[8] P. Hall and M.A. Martin, On Bootstrapping Resampling and Iteration, *Biometrika*, **75** (1988), 661–671. http://dx.doi.org/10.1093/biomet/75.4.661

[9] J. C. Nankervis, Computational Algorithms for Double Bootstrap Confidence Intervals, *Computational Statistics and Data Analysis*, **49** (2005), 461–475. http://dx.doi.org/10.1016/j.csda.2004.05.023

[10] L. M. Safiih and Z. N. Hila, Performance of Bootstrap MCEWMA: Study Case of Sukuk Musyarakah Data, *Proceeding of The 21st National Symposium on Mathematical Sciences*, November 6 – 8, 2013, Penang, Malaysia, 1022–1027. http://dx.doi.org/10.1063/1.4887731

[11] B. Chen, Y. R. Gel, N. Balakrishna, and B. Abraham, Computationally Efficient Bootstrap Prediction Intervals for Returns and Volatilities in ARCH and GARCH Process, *Journal of Forecasting*, **30** (2011), 51–71. http://dx.doi.org/10.1002/for.1197

[12] L.I. Tong, C.W. Chang, S.E. Jin and R. Saminathan, Quantifying Uncertainty of Emission Estimates in National Greenhouse Gas Inventories Using Bootstrap Confidence Intervals, *Atmospheric Enviroment*, **56** (2012), 80–87. http://dx.doi.org/10.1016/j.atmosenv.2012.03.063

[13] L. Pascual, J. Romo and E. Ruiz, Bootstrap Prediction for Returns and Volatilities in GARCH Models, *Computational Statistics and Data Analysis*, **50** (2006), 2293–2313. http://dx.doi.org/10.1016/j.csda.2004.12.008

Received: November 15, 2015; Published: August 12, 2016