COMPARISON OF CONVENTIONAL, BALANCED AND SUFFICIENT BOOTSTRAPPING APPROACHES VIA CONFIDENCE INTERVALS AND EFFICIENCY

Engin YILDIZTEPE*, Department of Statistics, Dokuz Eylül University, Turkey, engin.yildiztepe@deu.edu.tr
(https://orcid.org/0000-0002-7617-4934)

Abstract

There are various bootstrapping approaches depending on how bootstrap samples are selected. The conventional bootstrapping obtains random bootstrap samples by using all the units in the original sample. Balanced bootstrapping based on having individual observations with equal overall frequencies in all bootstrap samples and sufficient bootstrapping based on using only the distinct individual observations instead of all the units in the original sample are the two basic attempts proposed in this manner. This study compares the balanced, sufficient and conventional bootstrapping approaches in terms of efficiency, bootstrap confidence interval coverage accuracy, and average interval length. Although sufficient bootstrapping approach resulted in more efficient estimators and the narrower confidence intervals than the other two in all cases, none of the actual coverage level of confidence intervals was controlled within the desired limits. Conventional and balanced bootstrapping approaches have given quite similar results in terms of efficiency, coverage accuracy and average length.

Keywords: resampling, bootstrap, coverage accuracy, robust estimators of location

GELENEKSEL, DENGELİ VE YETERLİ BOOTSTRAP YAKLAŞIMLARININ GÜVEN ARALIKLARI VE ETKİNLİK AÇISINDAN KARŞILAŞTIRILMASI

Özet

Bootstrap örneklemlerinin nasıl oluşturulduğuna bağlı olarak değişen farklı bootstrap yaklaşımları mevcuttur. Geleneksel bootstrap, bootstrap örneklemlerini orjinal örneklemdeki tüm gözlem birimlerini kullanarak rasgele oluşturur. Dengeli bootstrap, orjinal örneklemdeki her bir gözlemin tüm bootstrap örneklemlerdeki toplam frekansının eşit olması ilkesine dayanır. Yeterli bootstrap yaklaşımında ise, bootstrap örneklemler oluşturulurken orjinal örneklemdeki tüm gözlemler yerine sadece farklı örneklem birimleri kullanılır. Bu çalışma, dengeli, yeterli ve geleneksel bootstrap yaklaşımlarını bootstrap güven aralıklarının kapsama oranı, ortalama genişliği ve etkinlik açılarından karşılaştırmaktadır. İncelenen tüm durumlar için, yeterli bootstrap yaklaşımı kullanılarak diğer iki yaklaşıma göre daha dar güven aralıkları ve daha etkin kestiriciler elde edilmiştir. Ancak, yeterli bootstrap yaklaşımı ile elde edilen güven aralıklarının kapsama oranları belirlenen limitler içerisinde yer almamıştır. Geleneksel ve dengeli bootstrap yaklaşımları, etkinlik, güven aralıklarının kapsama oranı ve ortalama genişliği kriterleri açısından oldukça benzer sonuçlar vermişlerdir.

Anahtar Kelimeler: yeniden örnekleme, bootstrap, kapsama oranı, dayanıklı konum ölçüleri

1. Introduction

It is known that the coverage accuracy of a t-based confidence interval for the population mean is (1-α) for normal distributed population, and approximately (1-α) for non-normally distributed populations. However, the coverage accuracy decreases when the population distribution is asymmetric [1]. In cases where the data is obtained from an asymmetric or heavy-tailed distribution, robust measures of location can be used. Since it is not possible to calculate the standard errors of many robust measures of location analytically, bootstrap methods can be utilised to compute confidence intervals and test hypotheses [2].

There are many studies on the confidence interval estimation of the location parameter for asymmetric population distributions comparing the coverage accuracy of various methods including the bootstrap (e.g., [3-6]). In all of these studies, the confidence interval methods were compared using bootstrap samples generated by the conventional bootstrapping approach. In conventional bootstrapping, the bootstrap samples are obtained by using all the units in the original sample.

There are various bootstrapping approaches that are claimed to be more efficient than the conventional one in the literature. In this study, we focus on the first order balanced and the sufficient bootstrapping procedures. The first-order balanced bootstrap was introduced by Davison et al. [7]. They claim that the balanced bootstrapping approach results in better estimation of bias in terms of efficiency. In a first-order balanced bootstrap, each of the original observations appears with equal frequency in the samples and in a second-order balanced bootstrap, each pair of observations occurs exactly the same number of times [8]. Do and Hall [9] investigated the theoretical aspects of the balanced resampling.

In 2011, Singh and Sedory [10] introduced the sufficient bootstrapping that uses only the distinct individual responses instead of all the units in the bootstrap resample. They consider the estimation of the sample mean, variance, standard deviation, coefficient of variation and proportion. Based on a simulation study, they conclude that the use of the proposed sufficient bootstrapping method may lead to better efficiency than conventional bootstrapping for certain situations.

In this study, we compare the conventional, the first order balanced and the sufficient bootstrapping methods based confidence interval estimations for some robust location measures. We compare those methods by using coverage accuracy and average interval length. We focus on three different bootstrap confidence interval methods: bootstrap percentile, bootstrap-t and bootstrap BCa. The details of these bootstrap confidence interval methods can be found in [8, 11].

The organization of the paper is as follows. Section 2 describes the bootstrapping approaches. Section 3 gives a brief information on the robust location estimators used in the study. Section 4 presents the conducted simulation study and its results. Section 5 gives concluding remarks. The R codes used in the simulation study are provided in Appendix and https://github.com/eyildiztepe/comp_bootstrap.

2. Bootstrapping Approaches

This section concisely describes the conventional, the first-order balanced, and the sufficient bootstrapping approaches that are used in the study.

2.1. Conventional Bootstrapping

In conventional bootstrap, bootstrap samples of size n are selected by random sampling with replacement from the original sample with the size n.

Let $X^{(b)*} = \{x_1^{(b)*}, x_2^{(b)*}, \dots, x_n^{(b)*}\}$ be the b-th bootstrap resample by sampling with replacement from a sample of $X = \{x_1, x_2, \dots, x_n\}$. The statistic $\hat{\theta}^{(b)*}$ is calculated for the corresponding bootstrap sample. This process is repeated for B times where θ denotes the targeted population parameter. The distribution of these bootstrap statistics $\{\hat{\theta}^{(1)*}, \hat{\theta}^{(2)*}, \dots, \hat{\theta}^{(B)*}\}$ forms the bootstrap sampling distribution that is used to estimate the standard error, bias, shape, dispersion and location of the sampling distribution of the corresponding sample statistic $\hat{\theta}$.

2.2. The First-Order Balanced Bootstrapping

In a balanced bootstrap method, every x_i in $X = \{x_1, x_2, \dots, x_n\}$ exactly occurs B times in all of the B bootstrap samples. It is said to have the first-order balance if each of the original observations appears exactly equal. A simple first-order balanced bootstrap algorithm proposed by [8] can be described as follows;

(1) Concatenate B copies of $x_1, x_2, \dots, x_n$ in a list L of length $B \times n$

(2) Randomly permute the elements of L.

(3) Take as bootstrap samples the successive sets of length n from L.

Then those resamples are used to form the bootstrap sampling distribution of the statistic $\hat{\theta}$.

2.3. The Sufficient Bootstrapping

Let $X^{s(b)*} = \{x_1^{s(b)*}, x_2^{s(b)*}, \dots, x_\nu^{s(b)*}\}$ be the bth sufficient bootstrap sample consisting ν distinct units in bth $X^{(b)*} = \{x_1^{(b)*}, x_2^{(b)*}, \dots, x_n^{(b)*}\}$ conventional bootstrap sample where $b = 1, 2, \dots, B$. The sufficient bootstrap sample estimate $\hat{\theta}^{s(b)*}$ is estimated from $X^{s(b)*} = \{x_1^{s(b)*}, x_2^{s(b)*}, \dots, x_\nu^{s(b)*}\}$. Then the distribution of $\{\hat{\theta}^{s(1)*}, \hat{\theta}^{s(2)*}, \dots, \hat{\theta}^{s(B)*}\}$ gives the sufficient bootstrap sampling distribution [10]. It is possible that some distinct units in a sufficient bootstrap sample may have the same value. But, it should be clear that those same values in a sufficient bootstrap sample must have come from different units.

3. Robust Estimators of Location

The standard errors of the trimmed mean, the one-step M-estimator, and the modified one-step M-estimator were calculated using the aforementioned bootstrap methods in order to estimate the confidence interval for the population counterpart of these estimators. Only the brief information for those robust estimators are presented here. For detailed information, see [2].

3.1. Trimmed Mean

The γ trimmed mean can be described as follows:

$$\mu_t = \frac{1}{1-2\gamma} \int_{x_\gamma}^{x_{1-\gamma}} x \, dF(x) \tag{1}$$

where x_γ and $x_{1-\gamma}$ are the γ and $1-\gamma$ quantiles ($0 \leq \gamma \leq 0.5$). Let $x_1, x_2, \dots, x_n$ be a random sample and let $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ be the observations in ascending order. Let $g = [\gamma n]$, where $[\gamma n]$ is the value of γn rounded down to the closest integer. The $\bar{x}_t$ is computed as follows [2];

$$\bar{x}_t = \frac{x_{(g+1)} + \dots + x_{(n-g)}}{n-2g} \tag{2}$$

3.2. One-step M-estimator

Estimation of Huber's M-measure of location uses an iterative estimation procedure such as the Newton-Raphson method. One iteration of this procedure produces the one-step M-estimator which has good asymptotic properties even with only a single iteration [12]. The $\hat{\mu}_m$ is

$$\hat{\mu}_m = \frac{1.28 MADN(i_2 - i_1) + \sum_{i=i_1+1}^{n-i_2} x_{(i)}}{n - i_1 - i_2} \tag{3}$$

where $MADN = MAD/0.6745$, and
$MAD = Med\{|x_1 - M|, |x_2 - M|, \dots, |x_n - M|\}$. MAD is the median absolute deviation statistic and M is the sample median. Here i_1 is the number of observations x_i such that $(x_i - M)/MADN < -1.28$ and i_2 is the number of observations x_i such that $(x_i - M)/MADN > 1.28$ [2].

3.3. Modified One-step M-estimator

Ignoring $1.28MADN(i_2 - i_1)$ in eq.(3) gives the modified one-step M-estimator. This estimator averages non-outlier values. However, to obtain a reasonably good efficiency under normality, utilized outlier detection rule is changed. The one-step M-estimator is

$$\hat{\mu}_{mom} = \frac{\sum_{i=i_1+1}^{n-i_2} x_{(i)}}{n - i_1 - i_2} \tag{4}$$

where i_1 is the number of observations for which $(x_i - M)/MADN < -2.24$, and i_2 is the number of observations for which $(x_i - M)/MADN > 2.24$ (Hampel identifier) [2].

4. Numerical Studies

This section includes the detailed simulation study and two real data examples that are used to compare three bootstrap methods for estimating confidence intervals for five location estimators: trimmed mean ($\gamma = 0.1, 0.2$), median, one-step M-estimator, modified one-step M-estimator) and sample mean. We compare the actual coverage level and the average length of the confidence intervals. In each cell of the tables that we present the coverage accuracies and average lengths, the first, second and third values represent the results for the conventional bootstrap, the sufficient bootstrap and the first-order balanced bootstrap, respectively. Coverage accuracies closer to the nominal values are highlighted. All computations were performed in R statistical programming language [13].

4.1. The Simulation Study

In this section, we generate our data from the standard normal and the *g*-and-*h* distributions for each of the sample sizes *n*=20 and *n*=50. For each of the 5000 simulations, we select *B*=1500 bootstrap resamples. To estimate the unknown standard error with an unknown analytical form, we conduct the second level bootstrap with *B*=500. Then we calculate the coverage accuracy by dividing the number of intervals which include the parameter's real value by the total number of simulations. We also consider the average length of the confidence intervals. For confidence intervals with coverage accuracy close to the nominal value, the method with a smaller average length is accepted as more appropriate. $(1 - \alpha) \pm (0.2\alpha)$ range are accepted as close to the nominal value and it corresponds to the limits of (0.94, 0.96) for the 95% ($\alpha = 0.5$) confidence level that we study [5, 6, 14].

The *g*-and-*h* distribution is equivalent to the standard normal distribution for *g*=*h*=0 which provides the opportunity to observe, by ranking, the distance from normal distribution, in which skewness can be controlled by *g*, and kurtosis by *h* parameters [15]. The skewness and kurtosis values for the *g*-and-*h* distribution obtained with *g*=0.5 and *h*=0.4 are 76.877 and 14091.46 respectively. Fig. 1 illustrates the shapes of the *g*-and-*h* distributions for selected values of the parameters.

Tables 1 and 2 present the results for the Normal Distribution and *g*-and-*h* distributions, respectively. For all of the cases considered, the sufficient bootstrapping approach fails keeping the coverage accuracy between the determined limits. All the average confidence interval lengths calculated for the samples of size 50 are narrower compared to the ones calculated for *n*=20 as expected. Coverage accuracy of the all confidence intervals for trimmed mean with trimming percentage 0.2 stays within the desired limits proving its success when one wishes to use it with a bootstrap confidence interval method. When sufficient bootstrapping is used, we get the bootstrap-t confidence intervals smaller in the average length compared to the average lengths of the bootstrap-BCa and the percentile bootstrap confidence intervals. However, it is just the opposite when the conventional and the first-order balanced approaches are used. The percentile bootstrap is the best among three bootstrap confidence interval methods considered for both conventional and balanced approaches under both the standard normal distribution and the *g*-and-*h*(0.5,0.4) distribution. None of the confidence intervals calculated for the population arithmetic mean preserve the nominal coverage accuracy for the *g*-and-*h* population.
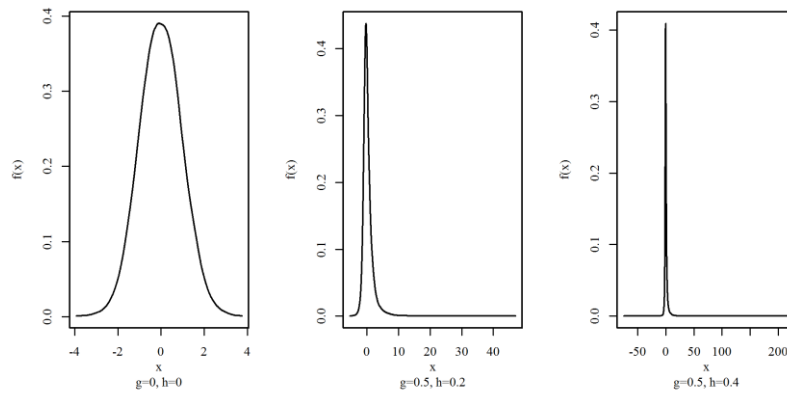
Figure 1. The pdf of various *g*-and-*h* distributions

Table 1. Coverage accuracies (Coverage) and average lengths (Av.L.) when the population distribution is Normal(0,1).

		n=20			n=50		
		bs-BCa	bs-t	bs-perc	bs-BCa	bs-t	bs-perc
$\bar{x}$	Coverage	0.9302	**0.9540**	0.9320	0.9366	**0.9460**	0.9378
		0.8512	0.6866	0.8502	0.8596	0.6586	0.8584
		0.9254	**0.9498**	0,9284	**0.9442**	**0.9512**	**0.9446**
	Av.L.	0.8505	0.9452	0.8455	0.5473	0.5693	0.5468
		0.6595	0.4633	0.6592	0.4209	0.2702	0.4209
		0.8491	0.9430	0.8441	0.5468	0.5689	0.5462
$\bar{x}_{tr=0.1}$	Coverage	**0.9456**	**0.9548**	**0.9468**	**0.9462**	**0.9504**	**0.9460**
		0.8434	0.6480	0.8478	0.8530	0.6340	0.8576
		0.9446	**0.9540**	**0.9450**	**0.9508**	**0.9540**	**0.9512**
	Av.L.	0.9113	0.9839	0.9090	0.5732	0.5889	0.5730
		0.6719	0.4368	0.6738	0.4317	0.2683	0.4321
		0.9077	0.9766	0.9056	0.5718	0.5872	0.5715
$\bar{x}_{tr=0.2}$	Coverage	**0.9420**	**0.9478**	**0.9450**	**0.9450**	**0.9514**	**0.9452**
		0.8390	0.6352	0.8476	0.8570	0.6362	0.8566
		0.9438	**0.9540**	**0.9458**	**0.9480**	**0.9550**	**0.9478**
	Av.L.	0.9445	1.0333	0.9433	0.5954	0.6181	0.5955
		0.6895	0.4430	0.6941	0.4471	0.2785	0.4478
		0.9387	1.0246	0.9376	0.5931	0.6156	0.5931
M	Coverage	**0.9406**	0.9100	**0.9400**	**0.9460**	0.9132	**0.9476**
		0.8408	0.5730	0.8370	0.8530	0.5950	0.8540
		0.9434	0.9128	**0.9420**	**0.9468**	0.9194	**0.9482**
	Av.L.	1.0718	1.2541	1.0528	0.6904	0.7789	0.6945
		0.7902	0.5204	0.7749	0.5212	0.3384	0.5206
		1.0680	1.2457	1.0468	0.6807	0.7660	0.6859
One step M-estimator	Coverage	0.9376	**0.9592**	**0.9400**	**0.9442**	**0.9494**	**0.9460**
		0.8442	0.6702	0.8456	0.8564	0.6388	0.8570
		0.9377	**0.9578**	0.9386	**0.9444**	**0.9526**	**0.9448**
	Av.L.	0.9146	1.0828	0.9089	0.5723	0.5968	0.5721
		0.6866	0.4668	0.6856	0.4352	0.2726	0.4352
		0.9145	1.0847	0.9089	0.5727	0.5975	0.5722
Modified One step M-estimator	Coverage	0.9304	0.9170	**0.9554**	**0.9434**	**0.9400**	**0.9576**
		0.8478	0.6352	0.8734	0.8606	0.6470	0.8736
		0.9388	0.9280	**0.9572**	**0.9478**	**0.9402**	**0.9586**
	Av.L.	1.0422	1.1032	1.0335	0.6342	0.6410	0.6350
		0.7861	0.5057	0.7875	0.4860	0.3035	0.4870
		1.0448	1.1018	1.0360	0.6339	0.6393	0.6351

Note: n: sample size; bs-BCa: Bias corrected accelerated bootstrap method; bs-t: Bootstrap-t method; bs-perc: Percentile bootstrap method.

Table 2. Coverage accuracies and average lengths when population distribution is *g*-and-*h* (*g*=0.5, *h*=0.4) distribution

		n=20			n=50		
		bs-BCa	bs-t	bs-perc	bs-BCa	bs-t	bs-perc
$\bar{x}$	Coverage	0.8052	0.8438	0.8108	0.8432	0.8650	0.8462
		0.7154	0.5196	0.7154	0.7370	0.5168	0.7334
		0.8046	0.8430	0.8150	0.8422	0.8654	0.8468
	Av.L.	2.8857	7.9970	2.3394	2.1626	4.7769	1.7444
		1.6770	4.4556	1.6827	1.2271	2.6187	1.2296
		2.9470	6.9919	2.3806	2.1806	4.1680	1.7614
$\bar{x}_{tr=0.1}$	Coverage	**0.9442**	**0.9546**	**0.9512**	**0.9430**	0.9366	**0.9486**
		0.8588	0.6012	0.8468	0.8648	0.5934	0.8568
		0.9388	**0.9556**	**0.9462**	**0.9500**	**0.9446**	**0.9544**
	Av.L.	1.9264	1.8220	1.8347	0.8948	0.8573	0.8909
		1.0732	0.6931	1.0950	0.6468	0.3670	0.6583
		1.9412	1.8206	1.8526	0.8948	0.8557	0.8907
$\bar{x}_{tr=0.2}$	Coverage	**0.9440**	**0.9470**	**0.9472**	**0.9460**	**0.9412**	**0.9462**
		0.8620	0.5776	0.8498	0.8636	0.6000	0.8588
		0.9468	**0.9470**	**0.9486**	**0.9498**	**0.9472**	**0.9524**
	Av.L.	1.3454	1.3187	1.3442	0.7451	0.7325	0.7458
		0.9195	0.5056	0.9444	0.5503	0.3219	0.5560
		1.3412	1.3107	1.3400	0.7427	0.7299	0.7435
M	Coverage	**0.9402**	0.9044	**0.9400**	**0.9458**	0.9080	**0.9470**
		0.8434	0.5472	0.8378	0.8534	0.5876	0.8540
		0.9442	0.9118	**0.9422**	**0.9466**	0.9162	**0.9482**
	Av.L.	1.2028	1.3131	1.2087	0.7221	0.7878	0.7347
		0.8586	0.5102	0.8576	0.5405	0.3368	0.5433
		1.2033	1.3109	1.2058	0.7107	0.7736	0.7242
One step M-estimator	Coverage	0.9306	**0.9516**	0.9354	**0.9400**	**0.9424**	**0.9454**
		0.8372	0.6364	0.8408	0.8472	0.6234	0.8526
		0.9306	**0.9522**	0.9344	**0.9400**	**0.9448**	**0.9438**
	Av.L.	1.3361	1.5403	1.3031	0.7804	0.8038	0.7734
		0.9437	0.6386	0.9373	0.5781	0.3624	0.5767
		1.3354	1.5415	1.3025	0.7793	0.8039	0.7728
Modified One step M-estimator	Coverage	0.9238	0.8984	0.9612	0.9322	0.9128	**0.9600**
		0.8448	0.5732	0.8848	0.8518	0.6196	0.8830
		0.9298	0.9068	0.9620	0.9376	0.9198	0.9622
	Av.L.	1.3354	1.3077	1.3140	0.7733	0.7643	0.7692
		0.9630	0.5617	0.9690	0.5822	0.3561	0.5833
		1.3290	1.2974	1.3036	0.7703	0.7625	0.7671

Note: n: sample size; bs-BCa: Bias corrected accelerated bootstrap method; bs-t: Bootstrap-t method; bs-perc: Percentile bootstrap method.

Table 3. Descriptive statistics of the real data sets

Data Set Name	N	Mean	Median	Std. Dev.	IQR	Skewness	Kurtosis
Total AA Degrees	2859	268.98	114.0	451.96	281	4.98	47.31
GRE Verbal	4637	468.81	470	103.23	150	0.13	2.54

4.2. Real Data Examples

In this section, we study performances of the considered confidence intervals on two different real data sets. Those data sets are available on https://github.com/eyildiztepe/datasets_cbs. One of the data sets has symmetric distribution while the other has an asymmetric shape (Fig. 2). Table 3 gives some descriptive statistics of these data sets. We treat those

sets as our populations and randomly select samples with the sizes of 20 and 50 from each sets. We repeat this selection for 5000 times with 1500 bootstrap resample for each simulation. 500 second bootstrap resamples are selected for estimating standard errors. Tables 4 and 5 present the results when the random sample is selected from the right skewed "Total AA Degrees" data set and from the symmetrical "GRE Verbal" data set, respectively.
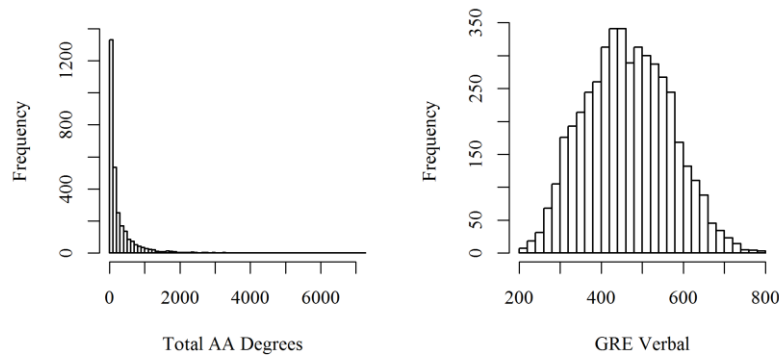
Figure 2. The histograms of the real data sets

Table 4. Coverage accuracies and average lengths for the "Total AA Degrees" real data set.

		n=20			n=50		
		bs-BCa	bs-t	bs-perc	bs-BCa	bs-t	bs-perc
$\bar{x}$	Coverage	0.8728	0.9270	0.8492	0.9164	**0.9401**	0.8988
		0.7542	0.6510	0.7450	0.8058	0.6554	0.7988
		0.8760	0.9272	0.8494	0.9160	0.9386	0.8974
	Av.L.	387.11	622.77	327.84	254.91	306.82	227.27
		242.59	326.99	242.30	168.19	153.60	168.30
		385.98	621.51	327.81	254.88	306.69	227.27
$\bar{x}_{tr=0.1}$	Coverage	**0.9424**	**0.9600**	**0.9442**	**0.9476**	**0.9532**	**0.9490**
		0.8378	0.6492	0.8406	0.8610	0.6420	0.8666
		0.9414	**0.9580**	**0.9440**	**0.9470**	**0.9520**	**0.9486**
	Av.L.	301.94	337.48	288.70	165.82	174.02	164.15
		187.31	136.76	194.37	119.49	75.36	123.70
		301.30	336.66	288.57	165.73	174.01	164.10
$\bar{x}_{tr=0.2}$	Coverage	**0.9442**	**0.9598**	**0.9442**	**0.9464**	**0.9512**	**0.9486**
		0.8446	0.6120	0.8422	0.8540	0.6390	0.8560
		0.9436	**0.9526**	**0.9444**	**0.9476**	**0.9504**	**0.9484**
	Av.L.	241.82	273.10	241.60	139.87	148.57	140.14
		162.61	104.15	174.13	101.82	63.73	105.35
		241.46	272.49	241.60	139.81	148.55	140.07
M	Coverage	**0.9456**	0.9286	**0.9434**	**0.9458**	0.9242	**0.9504**
		0.8492	0.5714	0.8432	0.8628	0.5996	0.8628
		0.9448	0.9292	**0.9442**	**0.9451**	0.9230	**0.9494**
	Av.L.	211.02	261.97	221.09	125.51	141.93	129.90
		151.45	96.15	156.49	93.48	58.20	95.09
		211.37	261.71	220.80	125.77	142.02	129.99
One step M-estimator	Coverage	0.9382	0.9674	0.9330	**0.9472**	0.9648	**0.9452**
		0.8404	0.7176	0.8412	0.8548	0.6994	0.8572
		0.9400	0.9688	0.9334	**0.9462**	0.9670	**0.9448**
	Av.L.	266.09	424.16	248.50	157.24	198.20	152.48
		186.14	162.93	181.38	116.00	86.37	114.46
		265.73	424.89	248.45	157.38	198.35	152.54
Modified One step M-estimator	Coverage	**0.9410**	0.9082	**0.9460**	**0.9428**	0.9264	**0.9532**
		0.8518	0.5346	0.8592	0.8668	0.5964	0.8684
		0.9412	0.9034	**0.9454**	**0.9434**	0.9286	**0.9518**
	Av.L.	258.20	281.28	238.15	151.05	165.53	144.25
		178.24	108.52	174.23	111.15	69.92	108.82
		258.28	279.87	238.12	150.88	165.20	144.22

Note: n: sample size; bs-BCa: Bias corrected accelerated bootstrap method; bs-t: Bootstrap-t method; bs-perc: Percentile bootstrap method.

Table 5. Coverage accuracies and average lengths for the "GRE Verbal" real data set

		n=20			n=50		
		bs-BCa	bs-t	bs-perc	bs-BCa	bs-t	bs-perc
$\bar{x}$	Coverage	0.9336	**0.9538**	0.9288	**0.9466**	0.9548	0.9436
		0.8520	0.7000	0.8538	0.8638	0.6694	0.8634
		0.9340	**0.9530**	0.9292	**0.9470**	0.9556	0.9468
	Av.L.	87.76	97.56	87.41	56.72	59.06	56.68
		68.30	47.46	68.27	43.62	27.85	43.61
		87.78	97.56	87.42	56.67	58.98	56.62
$\bar{x}_{tr=0.1}$	Coverage	**0.9422**	**0.9546**	**0.9416**	**0.9470**	0.9518	**0.9462**
		0.8414	0.6650	0.8508	0.8566	0.6594	0.8622
		0.9442	**0.9564**	**0.9432**	**0.9458**	0.9522	**0.9458**
	Av.L.	96.20	105.34	96.01	61.46	63.64	61.46
		71.47	47.02	71.62	46.32	29.00	46.34
		96.22	105.37	95.99	61.41	63.58	61.39
$\bar{x}_{tr=0.2}$	Coverage	**0.9406**	**0.9524**	**0.9414**	**0.9458**	0.9516	**0.9480**
		0.8326	0.6472	0.8452	0.8552	0.6540	0.8614
		0.9408	**0.9532**	**0.9418**	**0.9454**	0.9514	**0.9472**
	Av.L.	101.77	113.47	101.67	65.24	68.30	65.25
		74.58	48.83	75.06	49.03	30.77	49.09
		101.79	113.53	101.68	65.22	68.25	65.21
M	Coverage	**0.9432**	0.8906	**0.9498**	**0.9566**	0.9028	0.9622
		0.8562	0.5582	0.8576	0.8850	0.5820	0.8878
		0.9483	0.8924	**0.9480**	**0.9537**	0.9042	0.9628
	Av.L.	117.31	140.49	115.87	78.19	90.03	78.62
		86.86	58.81	85.85	59.07	39.73	58.91
		120.49	140.38	115.79	80.04	89.97	78.71
One step M-estimator	Coverage	**0.9408**	0.9622	0.9382	**0.9460**	0.9538	0.9450
		0.8554	0.6944	0.8542	0.8666	0.6652	0.8614
		0.9442	0.9639	0.9393	**0.9464**	0.9550	0.9460
	Av.L.	98.54	118.13	97.85	61.45	64.13	61.41
		74.16	50.84	74.10	46.85	29.33	46.85
		98.67	118.22	98.06	61.48	64.20	61.45
Modified One step M-estimator	Coverage	**0.9400**	0.9256	**0.9508**	**0.9500**	0.9456	0.9556
		0.8572	0.6746	0.8736	0.8750	0.6790	0.8776
		0.9384	0.9240	**0.9510**	**0.9476**	0.9462	0.9556
	Av.L.	113.41	121.41	112.25	67.46	67.33	67.64
		85.03	54.98	85.20	51.40	31.51	51.54
		113.48	121.11	112.23	67.38	67.29	67.60

Note: n: sample size; bs-BCa: Bias corrected accelerated bootstrap method; bs-t: Bootstrap-t method; bs-perc: Percentile bootstrap method.

For real data sets, conventional and balanced bootstrapping methods behave similarly. In terms of confidence interval type, bootstrap-BCa seems to be the best one.

4.3. Relative Efficiencies

Table 6 shows the relative efficiencies of the location estimators calculated by the conventional, sufficient and balanced bootstrapping approaches. The first value in each cell of Table 6 is the relative efficiency of the sufficient bootstrap estimate with respect to the conventional bootstrap estimate (e.g. the variance of the estimator from the conventional bootstrap is divided by the variance estimated by the sufficient bootstrap). The second value is for the relative efficiency of the first-order balanced bootstrap estimate with respect to the conventional bootstrap estimate. With the relative efficiencies greater than 1, the sufficient bootstrap estimators are more efficient than conventional bootstrap estimators under the considered scenarios. The first-order balanced bootstrap and conventional bootstrap estimators seem to behave equally in terms of efficiency.

Table 6. The relative efficiencies of the bootstrap estimates

Estimator	n	Distribution		Real Data	
		Normal(0,1)	g-and-h	Total AA Degrees	GRE Verbal
$\bar{x}$	20	1.6396	1.6383	1.6504	1.5508
		0.9904	0.9672	1.0575	1.0439
	50	1.8016	1.7578	1.7041	1.6476
		1.0291	0.9373	1.0256	0.9586
$\bar{x}_{tr=0.1}$	20	1.8244	2.8370	1.6345	1.6626
		0.9785	0.9411	1.0551	1.0391
	50	1.9155	2.1441	2.0174	1.7174
		1.0449	1.0150	0.9596	0.9575
$\bar{x}_{tr=0.2}$	20	1.8310	2.2398	2.0871	1.6983
		0.9720	0.9821	1.0749	1.0529
	50	1.9058	1.9864	1.7360	1.6342
		1.0531	1.0470	0.9722	0.9530
M	20	1.7109	1.8055	1.8822	1.4682
		0.9888	0.9743	1.0955	1.0244
	50	2.7719	2.6824	2.6799	1.4953
		1.0428	1.0499	1.1372	0.9632
One step M-estimator	20	1.7126	1.9645	2.8121	1.6184
		0.9751	0.9658	1.2143	1.0412
	50	1.8891	2.0034	1.9873	1.5353
		1.0530	1.0401	1.0309	0.9395
Modified One step M-estimator	20	1.7916	1.8567	3.6037	1.9142
		0.9656	0.9874	1.5260	1.0580
	50	1.9358	2.0353	2.2577	1.4475
		1.1020	1.1240	1.0729	0.9439

First row= $var^*_{con}(est)/var^*_{suf}(est)$, second row= $var^*_{con}(est)/var^*_{bal}(est)$ where the subscript shows the resampling approach under which estimator was computed.
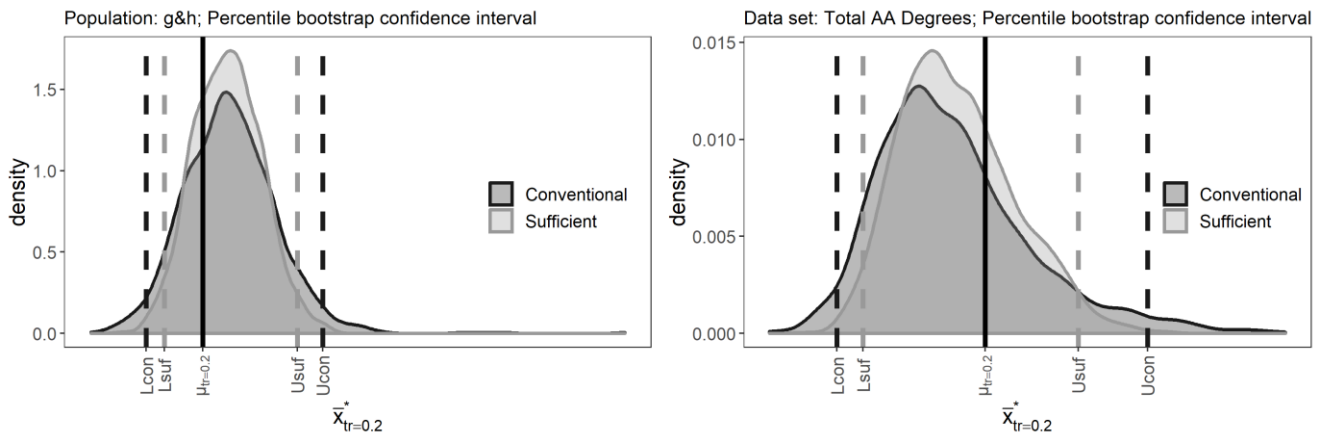

Figure 3. Density plots of bootstrap sampling distributions for $\bar{x}_{tr=0.2}$. The vertical solid line indicates the population parameter. The black dashed lines show confidence interval limits obtained with conventional bootstrapping and the grey dashed lines show confidence interval limits obtained with sufficient bootstrapping.

5. Conclusion

We compared the sufficient, balanced and conventional bootstrapping approaches in terms of efficiency for estimating three different bootstrap confidence interval methods using the coverage accuracy and the average interval length criteria. We considered both symmetric and skewed distributions.

The variances of the sampling distribution of all the estimators calculated by the sufficient bootstrapping approach is smaller than the variances estimated by the conventional bootstrapping approach. Figure 3 illustrates the sampling distributions of the 20% trimmed means calculated with both approaches, and the limits of the confidence interval obtained using the percentile bootstrap method.

As a result of smaller variance of an estimator obtained by the sufficient bootstrap method, the length of the confidence intervals obtained via sufficient bootstrapping is always narrower than the one obtained using conventional bootstrapping. The small leftward or rightward shifts of the sampling distribution on the *X* scale, which can result from randomness, might easily yield a confidence interval not containing the targeted parameter when the sufficient bootstrapping approach was used. For these reasons, the coverage accuracies of the confidence intervals calculated using the samples generated with the sufficient bootstrapping approach are lower values compared to conventional bootstrapping approach, and the actual coverage accuracies could not be preserved between the $(1 - \alpha) \pm (0.2\alpha)$ range for any of the confidence intervals obtained using sufficient bootstrapping. Having a low coverage accuracy always means having high actual significance levels in terms of one sample statistical inference for the sufficient bootstrap approach. All the actual significance levels were larger than the nominal level (0.05) which means the sufficient bootstrapping approach caused a liberal one sample inference about the population parameters included in the study. In general, conventional approach was successful in 98 and balanced approach was successful in 97 cases out of 144 different experimental situations investigated. The percentile bootstrap confidence intervals worked better than the other two when samples are generated from theoretical distributions whereas the bootstrap-BCa worked better when samples are generated from real data sets. 20% and 10% trimmed means were the best two robust estimators regarding the number of coverage accuracies that controlled within the desired limits. Using sufficient bootstrapping approach while generating bootstrap samples resulted in small standard errors for estimators but this did not create any advantage in terms of one sample inference and any type of bootstrap confidence interval's coverage accuracy in particular.

6. References

[1] Boos, D. D. and Hughes-Oliver, J. M., "How Large Does n Have to be for Z and t Intervals?", *American Statistician,* 54, 121–128, 2000.

[2] Wilcox, R. R., *Introduction to Robust Estimation and Hypothesis Testing: 4th Edition*, Academic Press, 2017.

[3] Zhou, X.-H. and Gao, S., "One-sided confidence intervals for means of positively skewed distributions", *American Statistician,* 54, 100–104, 2000.

[4] Zhou, X. H. and Dinh, P., "Nonparametric confidence intervals for the one-and two-sample problems", *Biostatistics,* 6, 187–200, 2005.

[5] Kuonen, D., "Studentized bootstrap confidence intervals based on M-estimates", *Journal of Applied Statistics,* 32, 443–460, 2005.

[6] Banik, S. and Kibria, B. M. G., "Comparison of some parametric and nonparametric type one sample confidence intervals for estimating the mean of a positively skewed distribution", *Communications in Statistics—Simulation and Computation,* 39, 361–389, 2010.

[7] Davison, A. C., Hinkley, D. V. and Schechtman, E., "Efficient bootstrap simulation", *Biometrika,* 73, 555–566, 1986.

[8] Davison, A. C. and Hinkley, D. V., *Bootstrap Methods and Their Application*, Cambridge University Press, 1997.

[9] Do, K.-A. and Hall, P., "Quasi-random resampling for the bootstrap",. *Statistics and Computing,* 1, 13–22, 1991.

[10] Singh, S. and Sedory, S. A., "Sufficient bootstrapping", *Computational Statistics and Data Analysis,* 55, 1629–1637, 2011.

[11] Efron, B. and Tibshirani, R. J., *An Introduction to the Bootstrap,* Chapman and Hall, 1993.

[12] Serfling, R., *Approximation Theorems of Mathematical Statistics*, Wiley, 1980.

[13] R Core Team, R: A language and environment for statistical computing, 2019.

[14] Ng, H. K. T., Filardo, G. and Zheng, G., "Confidence interval estimating procedures for standardized incidence rates", *Computational Statistics and Data Analysis,* 52, 3501–3516, 2008.

[15] Hoaglin, D. C. (eds. Hoaglin, D. C., Mosteller, F. and Tukey. J.), Summarizing Shape Numerically: The g-and-h Distributions. in *Exploring Data Tables, Trends, and Shapes,* 461–513, John Wiley and Sons, 2011.

Appendix - R codes used in the simulation study

```
#R codes used in the simulation study are provided for
#only 20% trimmed mean.
#Distribution: N(0,1), estimator: 20% trimmed mean
#initial values
alpha<-0.05; B<-1500 #bootstrap replications
k<-5000 #iteration number
n<-20 #sample size
tr=0.2 #trim value
theta<-0 # parameter value
#functions
trmean.f<-function(d,i){z<-mean(d[i], trim=0.2)
        return(z)}
coverage <- function(ci, value,kr=nrow(ci)) {
#function that takes a matrix of confidence intervals and
#the true value of parameter
#and returns the coverage level.
 coverage<-(kr-(sum((ci[,1]>value)|(ci[,2]<value))))/kr
 below<-(sum(ci[,2]<value))/kr
 over<-(sum(ci[,1]>value))/kr
 return(list("coverage" = coverage, "below" = below,
"over"= over))  }
# BCa bootstrap CI
boot.BCa <-function(data, statistic, statistic.vector, fun,
conf = .95) {
 data <- as.matrix(data);  n <- nrow(data)
```

```
 N <- 1:n;  alpha <- (1 + c(-conf, conf))/2
 zalpha <- qnorm(alpha)
 z0<-qnorm(sum(statistic.vector<statistic)/
length(statistic.vector))
 statistic.vector.jack <- numeric(n)
 for (i in 1:n) { J <- N[1:(n-1)]
  statistic.vector.jack[i] <- fun(data[-i, ], J)  }
 L <- mean(statistic.vector.jack) - statistic.vector.jack
 a <- sum(L^3)/(6 * sum(L^2)^1.5)
 adj.alpha <- pnorm(z0 + (z0+zalpha) / (1-a* (z0 +
zalpha)))
 limits <- quantile(statistic.vector, adj.alpha, type=6)
 return(list("estimated"=statistic, "BCa"=limits))   }
# (Bootstrap-t CI)
boot.t<-function(statistic, statistic.vector, se, conf = .95) {
 t.stats <- (statistic.vector - statistic) /se
 se0 <- sd(statistic.vector)
 Qt <- quantile(t.stats, c((1-conf)/2,(1+conf)/2),type = 6)
 names(Qt) <- rev(names(Qt))
 CI <- rev(statistic - Qt * se0)
 return(list("estimated"=statistic, "t"=CI)) }
# (Bootstrap percentile CI)
boot.perc <-function(statistic,statistic.vector, conf=.95){
 CI <- quantile(statistic.vector,c((1-conf)/2, (1+conf)/2),
type = 6);  names(CI)<-names(quantile(statistic.vector,
c((1-conf)/2, (1+conf)/2)))
 return(list("estimated"=statistic, "percentile"=CI)) }
ci.width <- function(cis) {  # function that takes a matrix
#of conf. int. and computes the average of ci width.
 z<-mean(cis[,2]-cis[,1]);  return(z) }
bootse<-
function(x,nboot=1000,est=median,SEED=TRUE,...){
#  Wilcox, R.R. (2017). Introduction to Robust Est. and
#Hypothesis Testing. 4th Ed.
#  Rallfun-vxx.txt can be downloaded from
#  http://dornsife.usc.edu/labs/rwilcox/software/
#   Compute bootstrap estimate of the standard error of
# the estimator est
 if(SEED)set.seed(2)  # set seed of random number
#generator so that results can be duplicated.
 data<-matrix(sample(x,size=length(x)*nboot,
replace=TRUE), nrow=nboot)
 bvec<-apply(data,1,est,...); bootse<-sqrt(var(bvec))
 bootse}
#simulation
con.theta <- numeric(B)
suf.theta <- numeric(B) ; bal.theta <- numeric(B)
con.theta.se<-numeric(B)
suf.theta.se<-numeric(B) ; bal.theta.se<-numeric(B)
con.bs.bca<-matrix(nrow=k,ncol=2)
suf.bs.bca<-matrix(nrow=k,ncol=2)
bal.bs.bca<-matrix(nrow=k,ncol=2)
con.bs.t<-matrix(nrow=k,ncol=2)
suf.bs.t<-matrix(nrow=k,ncol=2)
bal.bs.t<-matrix(nrow=k,ncol=2)
con.bs.perc<-matrix(nrow=k,ncol=2)
suf.bs.perc<-matrix(nrow=k,ncol=2)
bal.bs.perc<-matrix(nrow=k,ncol=2)

clist<-list() ; set.seed(100)
for (i in 1:k) { x<-rnorm(n)
 theta.hat <- mean(x,trim=tr)
#Randomly permute the indices for balanced bootstrap
 bal.ind<-sample(rep(1:n,B),B*n)
 for (b in 1:B) {
   ind.con <- sample(1:n, size = n, replace = TRUE)
   resample.x <- x[ind.con]
   con.theta[b] <- mean(resample.x,trim=tr)
   con.theta.se[b]<-bootse(resample.x,nboot=500,
est=mean,SEED=F, trim=tr)
#bootse computes bootstrap estimate of the standard
#error of the estimator est
   ind.suf <- unique(ind.con)#sufficient bootstrapping
   resample.x <- x[ind.suf]
   suf.theta[b] <- mean(resample.x, trim=tr)
   suf.theta.se[b]<-bootse(resample.x,nboot=500,
est=mean,SEED=F, trim=tr)
   #balanced bootstrapping
   ind.bal<-bal.ind[((b-1)*n+1):(b*n)]
   resample.x <- x[ind.bal]
   bal.theta[b] <- mean(resample.x,trim=tr)
   bal.theta.se[b]<-bootse(resample.x,nboot=500,
est=mean,SEED=F, trim=tr)   }
con.bs.bca[i,]<-boot.BCa(x, statistic = theta.hat,
statistic.vector = con.theta, fun = trmean.f, conf=1-
alpha)$BCa[c(1,2)]
suf.bs.bca[i,]<-boot.BCa(x, statistic = theta.hat,
statistic.vector = suf.theta, fun = trmean.f, conf=1-
alpha)$BCa[c(1,2)]
bal.bs.bca[i,]<-boot.BCa(x, statistic = theta.hat,
statistic.vector = bal.theta, fun = trmean.f, conf=1-
alpha)$BCa[c(1,2)]
con.bs.t[i,]<-boot.t(statistic = theta.hat, statistic.vector =
con.theta, se=con.theta.se, conf=1-alpha)$t[c(1,2)]
suf.bs.t[i,]<-boot.t(statistic = theta.hat, statistic.vector =
suf.theta, se=suf.theta.se, conf=1-alpha)$t[c(1,2)]
bal.bs.t[i,]<-boot.t(statistic = theta.hat, statistic.vector =
bal.theta, se=bal.theta.se, conf=1-alpha)$t[c(1,2)]
con.bs.perc[i,]<-boot.perc(statistic = theta.hat,
statistic.vector = con.theta, conf=1-
alpha)$percentile[c(1,2)]
suf.bs.perc[i,]<-boot.perc(statistic = theta.hat,
statistic.vector = suf.theta, conf=1-
alpha)$percentile[c(1,2)]
bal.bs.perc[i,]<-boot.perc(statistic = theta.hat,
statistic.vector = bal.theta, conf=1-
alpha)$percentile[c(1,2)]
print(paste(i,".completed",sep=""))   }
ci.list<-list(con.bs.bca=con.bs.bca, suf.bs.bca=suf.bs.bca,
bal.bs.bca=bal.bs.bca, con.bs.t=con.bs.t, suf.bs.t=suf.bs.t,
bal.bs.t=bal.bs.t, con.bs.perc=con.bs.perc,
suf.bs.perc=suf.bs.perc, bal.bs.perc=bal.bs.perc)
result <- lapply(ci.list, coverage, value=theta, kr=k)
ci.widths<-lapply(ci.list, ci.width)
result
ci.widths
```