

# BOOTSTRAP METHOD AND THEIR APPLICATION USING R-PROGRAMMING

MirandaPajo<sup>1</sup>

<sup>1</sup>University of Vlora "Ismail Qemali", Faculty of Technical Science,  
Department of Mathematics, Square Pavarësia, Street Kosova,  
Skelë Vlora, Albania  
[pajo.miranda@gmail.com](mailto:pajo.miranda@gmail.com)

## Abstract

Bootstrap Method examines a data practical approach to illustrate the support applications and other Resampling methods. It can also provide a powerful approach of statistical data analysis, as they have more general applications than standard parametric methods.

Bootstrap is a method which provides accurate measurements in the evaluation of a choice. This technique allows the evaluation of a choice distribution. In the case of supposed independent observations, it can be implemented by creating a number of new choices of observed data, each of which taken randomly.

This article begins with a description of Bootstrap methods and shows how they are related to other resampling methods. In the meantime it examines a wide variety of application approaches.

The article deals with the generation of bootstrap copies by computational methods which enable us the calculation of variance, average and quantiles coefficient; the evaluation of standard error, and BCA confidence interval. It also examines the applications of these methods in a wide range of problems and the rising assumptions and hypotheses together with evaluation problems.

Bootstrap method uses statistical software such as R, S-Plus, SPSS to attain a lot of values in different choices. In this article we have used the statistical software R, which is a software that is 'free', based on the S language, providing the implementation of a wide range of statistical functions.

This paper presents experimental results of the population which are taken into consideration to study the variance coefficient using Bootstrap Method by emphasizing the implementation of programming language. It will also present the comparison of experiment conclusions which are drawn by using the standard statistical methods.

**Keywords:** *Bootstrap, Resampling, Coefficient of Variation, BCa Intervals, R-Programming.*

## 1. Introduction

The bootstrap method introduced in Efron (1979) is a very general resampling procedure for estimating the distributions of statistics based on independent observations. The bootstrap method is shown to be successful in many situations, which is being accepted as an alternative to the asymptotic methods. In fact, it is better than some other asymptotic methods, such as the traditional normal approximation and the Edgeworth expansion. However, there are some counterexamples that show the bootstrap produces wrong solutions, i.e., it provides some inconsistent estimators.

## 2. Bootstrap Method

Assumed the data  $X_1, X_2, \dots, X_n$ , are independent and identically distributed (*iid*) samples from  $k$ -dimensional population distribution  $F$  and the problem of estimating the distribution:

$$H_n(x) = P\{R_n \leq x\}$$

Where  $R_n = R_n(T_n, F)$  is a real-valued functional of  $F$  and  $T_n = T_n(X_1, X_2, \dots, X_n)$  a statistic of interest.

Let  $X_1^*, X_2^*, \dots, X_n^*$  be a “bootstrap” samples *iid* from  $F_n$ , the empirical distribution based on  $X_1, X_2, \dots, X_n$ ,  $T_n^* = T_n(X_1^*, X_2^*, \dots, X_n^*)$  and  $R_n^* = R_n(T_n^*, F_n)$ .  $F_n$  is constructed by placing at each observation  $X_i$  mass  $1/n$ . Thus,  $F_n$  may be represented as:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq x\}.$$

A bootstrap estimator of  $H_n$  is:

$$\bar{H}_n(x) = P\{R_n^* \leq x\}.$$

where for given  $X_1, X_2, \dots, X_n$ ,  $P$  is the conditional probability with respect to the random generator of bootstrap samples. Since the bootstrap samples are generated from  $F_n$ , this method is called the nonparametric bootstrap. Note that  $\bar{H}_n(x)$  will depend on  $F_n$  and hence  $\bar{H}_n(x)$  itself is a random variable. To be specific,  $\bar{H}_n(x)$  will change as the data  $\{x_1, \dots, x_n\}$ .

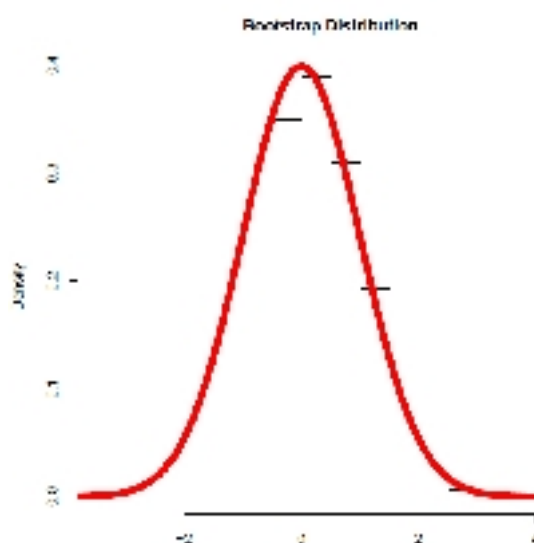


Fig.1 Bootstrap distribution Graphics

### Example 1.

Let us consider a sample containing two hundred values generated randomly from a standard normal population  $N(0, 1)$ . This is the original sample. In this example, the sampling distribution of the arithmetic mean is approximately normal with a mean roughly equal to 0 and a standard deviation approximately equal to  $1/200$ . Now, let us apply the nonparametric bootstrap method to infer the result. One thousand and five hundred resamples are drawn from the original sample, and the arithmetic mean is calculated for each resample. These calculations are performed by using `R` functions as follows.

**Step 1.** Randomly draw two hundred points from a standard normal population.

```
Gauss <- rnorm(200,0,1)
```

**Step 2.** Perform the nonparametric bootstrap study (1500 resamples).

```
Bootmean <- 1 : 1500
```

```
For ( i in 1 : 1500) bootmean[i] <- mean (sample (gauss, replace = T))
```

**Step 3.** Do the normalization and comparison with  $N(0,1)$ .

```
Bootdistribution <- sqrt(200)*(bootmean - mean (gauss))
```

```
Hist(bootdistribution, freq = FALSE, main = "Bootstrap Distribution", xlab="")
```

```
Points(x,y col=2)
```

### Theorem 1.

If  $x_1 < x_2$  and  $x_1$  and  $x_2$  are points of continuity of  $F$  then we have:

$$F(x_2) - F(x_1) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_2} - e^{-itx_1}}{it} f(t) dt.$$

where  $f(t)$  is the characteristic function.

## 3. Bootstrap standard error and bias

### a) Bootstrap estimate of standard error

Arguably, one of the most important practical applications of the bootstrap is in providing conceptually simple estimates of the standard error and bias for a statistic of interest. As usuals stands for the statistic of interest based on the observed sample. In output of the bootstrap method, we get bootstrap replicates  $(\bar{x}_1, \dots, \bar{x}_b)$ . Let us denote its standard error by  $\sigma_s$  standard error defined as  $\sqrt{\text{Var}(s(\bar{X}_1, \dots, \bar{X}_n))}$ . Since the form of the statistic  $s$  may be very complicated, the exact formulas for the corresponding bootstrap estimate of standard error (BESE) is approximated by the empirical standard deviation of  $(\bar{x}_1, \dots, \bar{x}_b)$ :

$$\hat{\sigma}_b = \sqrt{\frac{1}{b-1} \sum_{i=1}^b (\bar{x}_i - \bar{\bar{x}})^2} \approx \sigma_s$$

Where  $\bar{\bar{x}}$  is the mean of bootstrap replicates  $(\frac{1}{b} \sum_{i=1}^b \bar{x}_i)$ .

Let us note that the total number of resamples  $b$ , may be taken as large as we wish, since we are in complete control of the resampling process. For instance, it has been shown that for estimating BESE,  $b$  equal to about 500 typically gives already a satisfactory approximation.

### b) Bootstrap estimate of bias

Similarly, we have an analogous estimate of the bias of statistics  $s$ . We recall that the bias is the difference between the expectation of our statistics and the mathematical quantity  $S$  we want to estimate. For instance, the bias of the mean is defined as  $B_{\bar{X}_n} = E[\bar{X}_n] - E[X]$ . Whereas in a parametric context, the bias of  $\bar{X}_n$  is  $B_{\bar{X}_n} = E[\bar{X}_n] - \lambda$ , where  $\lambda$  can denote the parameter of an exponential distribution for example.

We estimate the bias of the statistics  $s$  by :

$$B_b = \frac{1}{b} \sum_{i=1}^b s_i - s(x_1, \dots, x_n) = B_b$$

We denote  $B_b$  as the bootstrap estimate of bias (BEB). When  $b = 1000$ , we generally have a good estimation of the true bias  $B_b$ .

## 4. Bootstrap confidence interval

Confidence sets are widely used in statistics because they provide bounds for our statistic of interest. A subset  $C_n$  of  $R$  is said to be a confidence set of level  $\alpha$  if  $P(S \in C_n) = 1 - \alpha$ .

In most cases, confidence sets are constructed by considering a "pivotal" function  $g(x_1, \dots, x_n, F)$ , whose distribution  $G_n$  is independent of unknown distribution  $F$ . But the problem is that the distribution  $G_n$  is not necessarily known. In classical theory, we replace the distribution  $G_n$  by its limit and so the derived confidence set is only valid for large  $n$ .

The bootstrap approach of this problem consists in using the bootstrap estimation  $G_{boot}$  of  $G_n$ , and then construct the confidence set.

### a) The Gaussian Bootstrap

The normal-theory interval assumes that the statistic  $s$  is normally distributed, which can be the case for statistics in sufficiently large samples. And then we use the bootstrap estimate of standard error to construct a  $100 - (1 - \alpha)\%$  confidence interval for  $S$  of the form:

$$[s - u_{1-\frac{\alpha}{2}} \sigma_b, s + u_{1-\frac{\alpha}{2}} \sigma_b]$$

Where  $\sigma_b$  is the bootstrap estimate of the standard error of  $s$ , and  $u_{1-\frac{\alpha}{2}}$  is the  $1 - \frac{\alpha}{2}$  quantile of the standard-normal distribution.

### b) The Bootstrap-t

The problem of the previous approach is that we use the normal approximation on the statistic  $s$ . That is to say we assume  $\frac{s - S}{\sigma_s}$  follows a normal distribution, where  $s$  is our statistic estimating the quantity  $S$  and  $\sigma_s$  denotes the true standard error of  $s$ . The principle of the bootstrap-t is to estimate the distribution of  $\frac{s - S}{\sigma_s}$  by its bootstrap distribution.

If we denote by  $G_{boot}$  the distribution computed with the bootstrap replicates, then the bootstrap-t confidence interval for  $S$  is:

$$[s - \sigma_s G_{boot}^{-1}\left(1 - \frac{\alpha}{2}\right), s + \sigma_s G_{boot}^{-1}\left(1 - \frac{\alpha}{2}\right)]$$

The main drawback is we need an estimate of the standard error  $\sigma_s$  of  $s$ . But we have to use difference bootstrap replicates for  $U_{boot}$  and  $\bar{s}$ .

**c) The Bootstrap Percentile**

An alternative approach, called the bootstrap percentile interval, is to use the empirical quantiles of bootstrap replicates  $(s_1, \dots, s_n)$  to form a confidence interval for  $s$ :

$$[s_j; s_k]$$

Where  $j = \lfloor n \frac{\alpha + \beta}{2} \rfloor$ ,  $k = \lfloor (1 - \frac{\alpha + \beta}{2}) n \rfloor$  and  $\bar{s}$  denotes the ordered bootstrap replicates.

**d) The bootstrap accelerated bias-corrected percentile (BCa)**

Although they do not assume normality, percentile confidence intervals often do not perform well.

So-called bias-corrected, accelerated (or BCa) percentile intervals are preferable. To find the BCa interval for  $s$ , we first compute  $z = \Phi^{-1}(G_{boot}(s))$ , where  $\Phi^{-1}$  stands for the standard-normal quantile function and  $G_{boot}$  is the empirical distribution function of  $s$  calculated with bootstrap replicates. If the bootstrap sampling distribution is symmetric, and if  $s$  is unbiased, then this proportion will be close to 0.5, and the "correction factor"  $z$  will be close to 0.

$$G_{boot}(x) = \frac{1}{b} \sum_{i=1}^b I_i(x)$$

Let  $s_{(-i)}$  represent the value of  $s$  produced when the  $i$ -th observation is deleted from the original sample  $(x_1, \dots, x_n)$ . There are  $n$  of these quantities  $(s_{(-1)} = s(x_2, \dots, x_n), s_{(-2)} = s(x_1, x_3, \dots, x_n), \dots)$ .

Let  $\bar{s}$  represent the average of the 1-deleted statistic sample  $(s_{(-1)}, \dots, s_{(-n)})$ ,  $\bar{s} = \frac{1}{n} \sum_{i=1}^n s_{(-i)}$ . Then we calculate:

$$a = \frac{\sum_{i=1}^n (s_{(-i)} - \bar{s})^2}{6(\sum_{i=1}^n (s_{(-i)} - \bar{s}))^2}$$

With the correction factors  $z$  and  $a$  in hand, we compute:

$$a_1 = \Phi\left(z + \frac{z - u_1 - \frac{a}{z}}{1 - a(z - u_1 - \frac{a}{z})}\right) \text{ dhe } a_2 = \Phi\left(z + \frac{z + u_1 - \frac{a}{z}}{1 - a(z - u_1 - \frac{a}{z})}\right)$$

Finally, the values  $a_1$  and  $a_2$  are used to locate the endpoints of the corrected percentile confidence interval:

$$[s_j; s_k]$$

where  $j = \lfloor na_1 \rfloor$  and  $k = \lfloor na_2 \rfloor$ .

When the correction factors  $a$  and  $z$  are both 0, we get  $a_1 = \Phi(-u_{1-\frac{\alpha}{2}}) = \Phi(u_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$  and  $a_2 = \Phi(u_{1-\frac{\alpha}{2}}) = \Phi(u_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$  which corresponds to the (uncorrected) percentile interval. So we explained that the correction factor  $z$  issued to correct the bias of the confidence interval, while  $a$  is dedicated to speed up the convergence of the confidence interval.

The whole idea behind this correction two correction factors is to find a relation between the statistic  $s$  and our quantity of interest  $S$  such as  $P(f(s) - f(S) \leq x) = \varphi(x)$  where the transformation function  $f$  is such that  $\varphi$  is a known distribution (typically the normal distribution).

But the previous equality does not necessarily hold. So Efron & Tibshirani (1993) suggests an equation of the form:

$$P\left(\frac{f(s) - f(S)}{1 + af(S)} + z \leq x\right) = \varphi(x)$$

To obtain sufficiently accurate 95-percent BCa confidence intervals, the number of bootstrap samples,  $b$ , should be at least 1000.

## 5. Bootstrap hypothesis test

A hypothesis test is a statistical rule based on a random sample which aims to decide if a hypothesis (a statement on a characteristic of data) can be rejected or accepted. Formally, we choose two hypothesis  $H_0$  and  $H_1$  (respectively the null and the alternative hypothesis). A hypothesis test give us a criterion to choose between  $H_0$  and  $H_1$ . We generally denote by  $\alpha$  the error probability to reject  $H_0$  when  $H_0$  is true.

As for the other applications of bootstrap method, two cases have to be distinguished: nonparametric and parametric bootstrap.

- ❖ For parametric bootstrap testing, we assumed data come from a known distribution  $F$  but with an unknown parameter  $\theta$ . Under the null hypothesis,  $\theta$  is contained in a particular subset of parameter say  $\Theta_0$ . (which could be just a singleton). So we just have to sample bootstrap replicates  $\theta_1^*, \dots, \theta_b^*$  under the null hypothesis. And then, approximate the probability that we reject  $H_0$  which is  $\frac{\sum_{i=1}^b I(\theta_i^* \in \Theta_1)}{b}$ .
- ❖ For non-parametric bootstrap testing, we generally test some characteristics on multi sample (e.g. the independence or the difference in means of two samples). This has consequences on the resampling method, because we need to resample under the null hypothesis.

Here are some very basic examples of bootstrap hypothesis tests. First, we use a non-parametric bootstrap method to compare two soporic drugs on patents by the extra number of sleeps compare to control. We want to test if the two samples ( $x$  and  $y$ ) have the mean i.e.  $H_0: \bar{x}_n = \bar{y}_n$  and  $H_1: \bar{x}_n \neq \bar{y}_n$ .

With R, we find a p-value of 99%, hence we utterly reject the null hypothesis. More complicated non-parametric bootstrap test can be done with a studentized "pivot".

Secondly, we illustrate the parametric bootstrap hypothesis test. We consider an exponential distributed sample, i.e. we assumed  $(X_i)_{i=1, \dots, n} \sim \mathcal{E}(\lambda)$ . The maximum likelihood estimator is  $\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i}$ . With R, we simulate a sample  $(x_1, \dots, x_{10})$ .

The maximum likelihood estimator is  $\hat{\lambda} = \frac{10}{\sum_{i=1}^{10} x_i} = 1.8339$ . We want to test the hypothesis that  $H_0: \lambda = 1.8$ . We sample 100 sample exponentially distributed with a parameter equal to 1:8 in order to compute bootstrap replicate. Then the p-value is 0:43 approximated by the ratio of bootstrap replicate greater than  $\hat{\lambda}$  and bootstrap number ( $b = 100$ ). Hence to reject the null hypothesis, we can have a better confidence than 44%.

## Shembull 2.

Take in consideration the population of the students in the University of Vlora "Ismail Qemali ", bachelor in Information Technology and Informatics, whose will study the arithmetic average of the 3-year study after they were graduated. For this data we will estimate the coefficient of variation, o generate a single bootstrap sample, the value corresponding to the upper 97.5, the value corresponding to the lower 2.5%, Efron confident limit, BCA confidence limits and a plot of the histogram of these values. The values are:

(8.26, 6.33, 10.0, 5.27, 5.35, 5.61, 6.12, 6.19, 5.2, 7.01, 8.74, 7.78, 7.02, 6, 6.5, 5.8, 5.12, 7.41, 6.52, 6.21, 8.28, 5.6, 5.38, 6.6, 8.74, 8.16, 5.33, 10.0, 6.27, 5.35, 5.51, 6.17, 6.15, 5.3, 7.01, 8.74, 7.78, 7.02, 6, 6.5, 5.8, 5.12, 7.41, 6.52, 6.21, 7.28, 5.6, 5.38, 6.6, 8.74, 8.26, 6.33, 10.0, 5.27, 5.35, 5.61, 6.12, 6.19, 5.2, 7.01, 8.74, 7.78, 7.02, 6, 6.5, 5.8, 5.12, 7.41, 6.52, 6.21, 7.28, 5.6, 5.38, 6.6, 8.74, 8.26, 6.33, 10.0, 5.27, 5.35, 5.61, 6.12, 6.19, 5.2, 7.01, 8.74, 7.78, 7.02, 6, 6.5, 5.8, 5.12, 7.41, 6.52, 6.21, 5.28, 5.6, 5.38, 6.6, 8.74).

1. So, let's compute the CV

$$CV(x) = 0.1909538$$

2. Compute the CV using a single bootstrap sample,

$$CV(\text{sample}(x, \text{replace}=T)) = 0.1692212$$

3. Note that the sum command is fairly general, for example

$$\text{sum}((x - \text{mean}(x))^2) = 158.7123$$

$$4. \text{mean}(\text{boot}) = 0.1889268$$

$$5. \text{Find the variance } \text{var}(\text{boot}) = 0.0001496165$$

$$6. \text{The value corresponding to the (say) upper 97.5 } \text{quantile}(\text{boot}, 0.975) = 0.2117789$$

$$7. \text{The value corresponding to the lower 2.5\% } \text{quantile}(\text{boot}, 0.025) = 0.1647055$$

$$8. \text{Estimate of the bias } <- \text{mean}(\text{boot}) - CV(x) = 0.1929808$$

9. The approximate 95% confidence interval is given by

$$CV(x) - \text{bias} - 1.96 * \text{sqrt}(\text{var}(\text{boot})) = 0.1690065$$

$$CV(x) - \text{bias} + 1.96 * \text{sqrt}(\text{var}(\text{boot})) = 0.2169551$$

10. Efron's confident limit has an upper and lower value of

$$\text{quantile}(\text{boot}, 0.975) = 0.2117789$$

$$\text{quantile}(\text{boot}, 0.025) = 0.1647055$$

11. BCa confident limit has an upper and lower value of

$$2 * CV(x) - \text{quantile}(\text{boot}, 0.025) = 0.2172021$$

$$2 * CV(x) - \text{quantile}(\text{boot}, 0.975) = 0.1701286$$

12. Histograma:

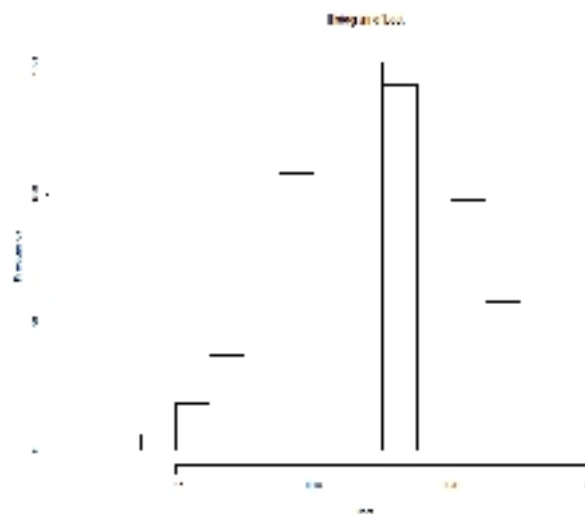


Fig. 2 The histogram

## Conclusion

In statistics, Bootstrap method is a method that gives us the right measurements in the evaluation of the choice. This technique allows the evaluation of distribution of a choice. A standard choice for the approximate distribution is the empirical distribution of the observed data. In the case of the observations can be assumed to be independent, it can be implemented by constructing a number of re-election to the observed data, each of which was obtained by random choice.

Bootstrap method is a method for statistic. It is mainly used in the calculation of confidence intervals, but also used to calculate the displacement of an evaluator and control hypotheses. The task is to use the scientific method to determine Bootstrap correctness of the statistical assessments, thus approximating the standard error and confidence interval limits.

The bootstrap method needs a good statistical software, where intensive computation can be done. R statistical software is a free software based on the S language, providing an extreme widerange of statistical functionalities R allows users to add additional functionality with packages.



## References

1. B. Efron, R. Tibshirani ( 1986). Bootstrap Methods for Standart Errors, Confidence Intervals, and Other Measures of Statistical Accuracy.
2. Shao, J. & Tu, D. (1996), The Jackknife and Bootstrap, Springer.
3. Abrevaya, J. and Huang, J. (2005). On the Bootstrap of the Maximum Score Estimator. *Econometrica*, 73 1175{1204.
4. Bickel, P. and Freedman, D. (1981). Some Asymptotic *Theory for the Bootstrap*. *Ann. Statis.*, 9, 1196{1217.
5. Sen, B. and Banerjee, M. and Eoodroofe, M. (2008). *Inconsistency of Bootstrap: the Grenander estimator*. *Submitted*.
6. Bodhisattva, S. (2008). A Study of Bootstrap and Likelihood Based Methods in Non-Standard Problems
7. Dutang, Ch. and Nguyen, N. and Nguyen, P. (2008). Bootstrap
8. Kim, P. and Pan, L. and Tony S. Eirjanto (2005). Bootstrapping Neural-Netöork Models of Exchange Rate
9. Resampling Methods—3rd Edition Program Code
10. LeBaron, B. (1997). An Evolutionary Bootstrap Approach to Neural Netöork Pruning and Generalization Philip M. Dixon ( 2001).
11. The Bootstrap Hesterberg, T. and Monaghan, Sh. and S. Moore, D. and Clipson, A. and Epstein, R. (2003). Bootstrap Method and Permutation Tests.
12. Kesar Singh and Ming Xie 'Bootstrap: A Statistical Method' Rutgers University
13. G. Jogesh Babu, Jackknife and Bootstrap, Center for Astrostatistics The Pennsylvania State University.
14. [http://en.wikipedia.org/wiki/Bootstrapping\\_\(statistics\)](http://en.wikipedia.org/wiki/Bootstrapping_(statistics))
15. <http://www.math.ntu.edu.tw/~hchen/teaching/LargeSample/notes/notebootstrap.pdf>
16. <http://www.biostat.jhsph.edu/~ririzarr/Teaching/649/section-08.pdf>

### Appendix: R-Program for finding the variance coefficient

```
# R has a number of nice features for easy calculation of bootstrap estimates and confidence
#intervals. To see how to use these features, consider the following 100 observations
# Suppose we wish to estimate the coefficient of variation. Let's do this with a bootstrap estimator.
# First, let's put the data into a vector, which we will call x,
x <-c(8.26, 6.33, 10.0, 5.27, 5.35, 5.61, 6.12, 6.19, 5.2,
7.01, 8.74, 7.78, 7.02, 6, 6.5, 5.8, 5.12, 7.41, 6.52, 6.21,
8.28, 5.6, 5.38, 6.6, 8.74,8.16, 5.33, 10.0, 6.27, 5.35, 5.51, 6.17, 6.15, 5.3,
7.01, 8.74, 7.78, 7.02, 6, 6.5, 5.8, 5.12, 7.41, 6.52, 6.21,
7.28, 5.6, 5.38, 6.6, 8.74,8.26, 6.33, 10.0, 5.27, 5.35, 5.61, 6.12, 6.19, 5.2,
7.01, 8.74, 7.78, 7.02, 6, 6.5, 5.8, 5.12, 7.41, 6.52, 6.21,
7.28, 5.6, 5.38, 6.6, 8.74,8.26, 6.33, 10.0, 5.27, 5.35, 5.61, 6.12, 6.19, 5.2,
7.01, 8.74, 7.78, 7.02, 6, 6.5, 5.8, 5.12, 7.41, 6.52, 6.21,
5.28, 5.6, 5.38, 6.6, 8.74)
# Now let's define a function in R, which we will call CV, to compute the coefficient of variation,
CV <- function(x) sqrt(var(x))/mean(x)
# So, let's compute the CV
CV(x)
# To generate a single bootstrap sample from this data vector, we use the command which generates a
#bootstrap sample of the data vector x by sampling with replacement.
sample(x,replace=T)
# Hence, to compute the CV using a single bootstrap sample, The particular value that R returns for
#you will be different as the sample is random.
CV(sample(x,replace=T))
# Note that the sum command is fairly general, for example
sum((x-mean(x))^ 2)
# So, lets now generate 1000 bootstrap samples. We first need to specify a vector of real
#values of length 1000, which we will call boot
boot <-numeric(1000)
for (i in 1:1000) boot[i] <- CV(sample(x,replace=T))
# The mean and variance of this collection of bootstrap samples are easily obtained using the mean
#and var commands (again, your values may differ),
mean(boot)
#Find the variance
var(boot)
# Likewise, the value corresponding to the (say) upper 97.5
quantile(boot,0.975)
# while the value corresponding to the lower 2.5% follows from
quantile(boot,0.025)
# Recall from the notes that the estimate of the bias is given by the difference between the mean of the
#bootstrap values and the initial estimate, and an bootstrap-corrected estimate of the CV is just the
#original estimate minus the bias,
bias <- mean(boot) - CV(x)
CV(x) -bias
# Assuming normality, the approximate 95% confidence interval is given by
CV(x) - bias - 1.96*sqrt(var(boot))
CV(x) - bias + 1.96*sqrt(var(boot))
# Efron's confident limit has an upper and lower value of
quantile(boot,0.975)
quantile(boot,0.025)
# BCa confident limit has an upper and lower value of
2*CV(x) - quantile(boot,0.025)
2*CV(x) - quantile(boot,0.975)
hist(boot)
```