MT4113 Assignment 3 Report

Abstract

This assignment aims to estimate three different methods to generate bootstrap confidence intervals, non-parametric and parametric bootstrap with percentile method, non-parametric bootstrap with Bias Corrected and Accelerated (BCa) with different samples, confidence levels, and resample times. The three methods perform almost equally for large samples and enough resample times (larger than 1,000). The performances are different when resources are limited. To estimate this coursework, it is not ideal because bootstrap confidence intervals' coverage rates are not close to confidence level. To solve this, further study is required time.

Introduction

To estimate the parameter of interest and make inferences about a sampling distribution, bootstrap in statistics is widely used. It is of great significance for a bootstrap confidence interval to have a good performance of the coverage of the true mean. This assignment aims to test performances of three different methods to generate bootstrap confidence intervals (CIs). They are non-parametric and parametric bootstrap with percentile method, non-parametric bootstrap with Bias Corrected and Accelerated (BCa). Results show that for a large sample, if the bootstrap resamples at least 1,000 times, there is no obvious difference between the three methods. For small samples, situations vary. In general, non-parametric bootstrap has a better performance. In addition, bootstrap for a poisson sample performs better than that for a normal sample. Firstly, this report will introduce methods used for bootstrap and testing results. Secondly, results will be displayed and discussed. Finally, a conclusion will be given.

Methods

Bootstrap Methods

Non-Parametric Bootstrap

For non-parametric bootstrap, it is not necessary to know the distribution of samples. The core concept is to resample the original samples with replacements. Here two ways to generate bootstrap confidence intervals are used.

Percentile method

It is easy to understand and use. Thus, this method is widely used. The way of taking the upper and lower limits of a CI is in lecture notes. There is no detailed discussion.

BCa

This method is more precise and unbiased comparing to percentile method. The way of generating a CI is complicated. In R, standard pnorm(), rnorm() and functions in *BCaHelperFunctions_v2* by Dr Len Thomas are used. The references are mainly lecture notes. There are no detailed explanations.

Parametric Bootstrap with Percentile Method

For parametric bootstrap, the user needs to know the distribution of the sample. Different from non-parametric bootstrap, to resample, functions for generating samples for a certain distribution are used in R (e.g. rnorm()). The way of taking a CI's upper and lower limits is the same percentile method.

Discussion about other methods

Parametric bootstrap with BCa

I did this at the beginning. However, I give up it. The performance of it is very bad, almost less than 1% generated CIs can cover true means. It reminds me whether BCa can be used for parametric bootstrap. To use BCa in non-parametric bootstrap, standard normal distribution is applied to calculate CI limits. However, for other distributions, they may do not have standard distributions. If I still use the standard normal distribution, it seems to be a mix of non-parametric and parametric bootstrap.

Finally, the R document in RStudio of boot.ci () and some research proves BCa is usually applied for non-parametric bootstrap [1].

Non-parametric Balanced Bootstrap

I planned to use this bootstrap. Research states the method can balance the bootstrap results [2][3]. However, the implementation of it is difficult. It has a few conditions to control in R. I simplified the conditions. The performance of this method becomes not ideal. Its CIs can cover a true mean but the CI is very big. After reading some materials, I think this method is easier to implement by other programming such as C. Thus, it is not finished.

The Testing Method

The main idea of the testing is to control conditions[1], run bootstrap functions for different times (10, 100, and 1,000) to compare the results. The aim is to see whether the performance is different when a condition changes. To test the performance of a bootstrap CI, three standards are referred:

1. The coverage rate of CIs containing the true mean should be close to (1- α).
2. For CIs which do not contain a true mean, rates that the true mean is smaller than the lower limit of the CI and the value is bigger than the upper limit should be equal.
3. For CIs which contains the true mean, the lengths of CIs should be as short as possible.

Graphs

Histograms are used. Estimated means by bootstrap are ordered from low to high and the true mean is added in a histogram to compare.

Text report

Functions print() and sink() are used to give directly visual results on the console and a text file respectively.

[1] Different sample distributions (poisson and normal), sample sizes (10 and 50), resample times (10, 100, and 1,000), and confidence levels (α = 0.05, 0.01, 0.02) are conditions.

Simple Data Analysis by R

A text file in csv format is generated. $\mathrm{write.table()}$ and $\mathrm{read.table()}$ in R are used to store and get results. *analysis.r* is the simple data analysis example. Eight simulations are generated with some sub-simulations. To analyse a certain condition, statistics of this variable are applied (e.g. to analyse the performance of different bootstrap methods for a normal sample with 50 observations in 95% confidence level, means, median, percentile values, maximums and minimums of CIs' lengths of the three methods are compared).

Results and Discussion

Overall, in 22,000 results, the coverage rate is not ideal, from 0 to 1. The mean of the coverage rate usually is 0.5. For a small resample times (10), the performance is worse than big resample times (1,000 or more). For a small sample size (10), the performance is worse than the big ones (50 or more). This implies that if a bootstrap method cannot resample enough times, the performance will be impacted heavily. In addition, a good data set can help the bootstrap method performance well. In this assignment, it is noticed that the closer the sample mean to the true mean, the better performance a bootstrap can obtain (the coverage rate is close to $(1-\alpha)$).

It seems that bootstrap for a poisson sample is better than that for a normal sample. This may owe to the characteristics of poisson distribution[2]. In detail, firstly, for a small sample size, the coverage rate of poisson sample is higher than that of normal sample. For large sample sizes, there is no obvious difference. Secondly, for a high confidence level (99%), the coverage rate of poisson samples is normal. That of normal samples is less than 0.1 with a poor performance. Thirdly, for a low resample times, poisson also performs better than normal samples.

The three methods perform differently under different situations.

For 99% confidence level, BCa is the most possible way to generate a true mean. In terms of confidence intervals, for a small resample times (10), BCa tends to have a higher CI coverage rate than other methods when the sample size is bigger than 30. It also tends to have the shortest CI lengths when resample times are small. Non-parametric percentile method performs the best when the sample size is smaller than 30. When the resample times are bigger than 1,000, there is no obvious performance difference of performance. This implies that if the bootstrap resample time is small, BCa is more useful for this confidence level when the sample size is big, while non-parametric percentile should be used when the sample size is small.

For 98% confidence level, in terms of bootstrap confidence intervals, there is no obvious performance difference with a low resample times. When the resample times are bigger than 100, non-parametric methods perform better (relatively high CI coverage rates, short CI lengths, and estimated means are close to the true mean).

---

[2] Because time limit, no future research about this.

For 95% confidence level, there is a trend that means generated by BCa are closest to true means. Regarding to confidence intervals, for a small resample times (10, 100), non-parametric and parametric percentile methods have higher CI coverage rate than BCa. Their estimated means by bootstrap are closer to the true mean. However, when the resample times are larger than 1,000, tails of CIs which do not contain true means become not balanced.

Conclusion

In conclusion, the coverage of my code is not ideal. The reason can be that the way I test is not reasonable enough. Or functions for bootstrap are wrong. Potential reasons may also exist and further research to solve this problem is needed. Based on current results, to estimate a population mean, BCa should be a relatively precise and unbiased way. In terms of generating a bootstrap confidence interval, in general, bootstrap for a poisson sample performs better than a normal sample. Performance of non-parametric and parametric percentile methods and non-parametric BCa have no obvious difference when the conditions are good (a big sample, resample times bigger than 1,000, and the sample mean is close to the true mean). When the sample size or resample times are small, non-parametric bootstrap is recommended. In detail, please see the last three paragraphs in Results and Discussion. In terms of the limitation of this study, data analysis is too simple and not strong. Rather than observation, detailed and complex statistical analysis should be applied if time limit, such as classifying failed or bad coverage CIs to see the potential reasons why this happen.

References

1. Haukoos J, Lewis R. Advanced Statistics: Bootstrapping Confidence Intervals for Statistics with "Difficult" Distributions. Academic Emergency Medicine 2005 Apr; 12(4): 360-365.

2. CTS_Chapter4_5-8_Students. [Internet]. [cited 2017 Nov 12]; [8 screens]. Available from:

   http://www.maths.qmul.ac.uk/~bb/CTS_Chapter4_5-8_Students.pdf .

3. Chernick M, LaBudde Robert. An Introduction to Bootstrap Methods with Applications to R. The United States of America: A JOHN WILEY & SONS, INC., PUBLICATION; 2011.