MT 4113: Computing in Statistics, Assignment 3

Set: 20 October 2017

Due: 22 November 2017

Your assignment

Design and implement a simulation study in R to examine the performance of the bootstrap for obtaining 95% confidence intervals on the mean of a univariate sample of data. The aspect of performance you should focus on is confidence interval coverage – i.e., the proportion of times the confidence interval contains true mean. Of course, for a 95% confidence interval, this "should" be 0.95 – however, as we have mentioned in class, the various bootstrap methods do not always work as advertised, depending on factors such as sample size of data, data distribution, etc.

As a minimum, you should evaluate the following methods

- nonparametric bootstrap percentile method
- nonparametric bootstrap BCa method
- at least one other method of your choice

Please do not use any pre-packaged bootstrap routines available in R libraries. On the other hand, you are welcome to use the functions I have provided on the module's Moodle site for calculating the BCa factors $\hat{z}_0$ and a.

It's up to you to choose what distribution(s) to sample from, what sample size(s) of data to simulate each time, how many bootstrap resamples to generate each time, and how many simulations to perform. You may want to focus on simulation scenarios you feel will best bring out the difference between the bootstrap methods, or highlight their relative strengths and weaknesses.

What to hand in

Before 11:59am on 22nd November, please upload the following to MMS, archived as a single zip (or other archive) file:

- A *short* (≤ 5 pages + any figures and/or tables) report (Word document, pdf, etc), that explains what you did and what you found. This report should be understandable by anyone who has taken MT4113 - in other words you don't need to explain in great depth what each type of bootstrap is, but you do need to say what you did, what you found and what conclusions you draw from your results. You might want to use the following section headings: Abstract, Introduction, Methods, Results, Discussion and Conclusions. If you tried some things and they did not work out, feel free to include information about this in your report.
- A planning document (text, Word document, pdf, etc) that gives evidence you spent time designing the simulation program before you started coding. This could be, for example, an outline of the functions that will be written with rough pseudocode inside. You're welcome to comment on your design choices.
- One or more *plain text files* (e.g., .r or .txt extensions) containing code to run the simulation. I would like to be able to source the code from this file (or these files) directly into `R` and get *exactly* the same results that you did.

Rules of engagement are the same as for Assignment 1:

- If you hand work in late, it will be subject to the late work penalty described here [http://www.st-andrews.ac.uk/maths/current/ug/information/latepenalties/] and summarised as *A late piece of work is penalised with an initial penalty of 15% of the maximum available mark, and then a further 5% per 8-hour period, of part thereof.* Do not wait until the morning of the due date to try uploading your work to MMS – if the upload fails at that point, your work will be discounted as late.

- Because this assignment counts towards your final grade, it is important that you do not collaborate with others in completing the work. *You should be comfortable with the following statement, which you should include as a comment at the beginning of your code files*

```
# I confirm that the attached is my own work, except where clearly indicated in the text.
```

- If you got stuck and needed to ask your peers, please use comments to tell me in your code file where you got help from others. If you are really stuck, email me (Len). For more information, see the *Academic misconduct* section of the university web site [http://www.st-andrews.ac.uk/students/rules/academicpractice/].

Marking scheme

There are 50 marks on offer for the assignment:

- 7 for producing code that correctly calculates confidence intervals using the nonparametric bootstrap percentile and BCa methods
- 8 for the correctness of the rest of your code (correct code is code that functions the way I believe you intended it to)
- 12 for the scope of your work (how insightful the simulation study you designed should be in highlighting the main features of each bootstrap method used; how ambitious it is)
- 14 for nicely written, well documented, compact, modular code (even if it doesn't work!)
- 4 for the level of good planning evident in your planning document
- 5 for the quality (clarity, completeness, conciceness) of the statistical report

If you have any doubts about what you are required to do, please post a question on the module forum.