ECOLE POLYTECHNIQUE DE BRUXELLES

Computational statistics - Homework 3

Non-parametric Bootstrap

Matteo Silvestri

Professor Maarten Jansen

2024

Table des matières

1

Problem

The MATLAB routine, which can be found in the Github repository, simulates the construction of non-parametric bootstrap confidence intervals for the median of a positive random variable that was drawn from a lognormal distribution (the lognormal DGP is not supposed to be known in the simulation, as the bootstrap procedure is non-parametric). The routine compares the coverages of the basic and percentile bootstrap confidence intervals.

Discuss the experiment answering the following questions :

(a) Explain why the percentile confidence interval is valid here.

(b) From the simulations,which confidence interval is the better choice, and why?

(c) Increase the samplesize, e.g. typing samplesize=200. Before launching the routine, what can you conclude? (You may try other samplesizes as well)

2

Solution

Let's first comment the main results of the MATLAB routine. We used 1000 samples of size 20 generated by

$$X = \exp(\mu + Z)$$

with $\mu = 3$ and $Z \sim N_{20}(0, I)$, so the true median is $\text{med}_{\text{true}} = e^{\mu} = 20.085$. The objective of this routine is compute confidence intervals such that med_{true} belongs to these CIs with 0.95 probability and we represent the results in the figures 2.1,2.2. We used basic/percentile bootstrap computing B=10000 re-samples in order to get statistics and we obtain the following coverages(our goal is for them to be 0.95).

Bootstrap	Coverage
basic	0.808
percentile	0.935

Now let's answer the questions.

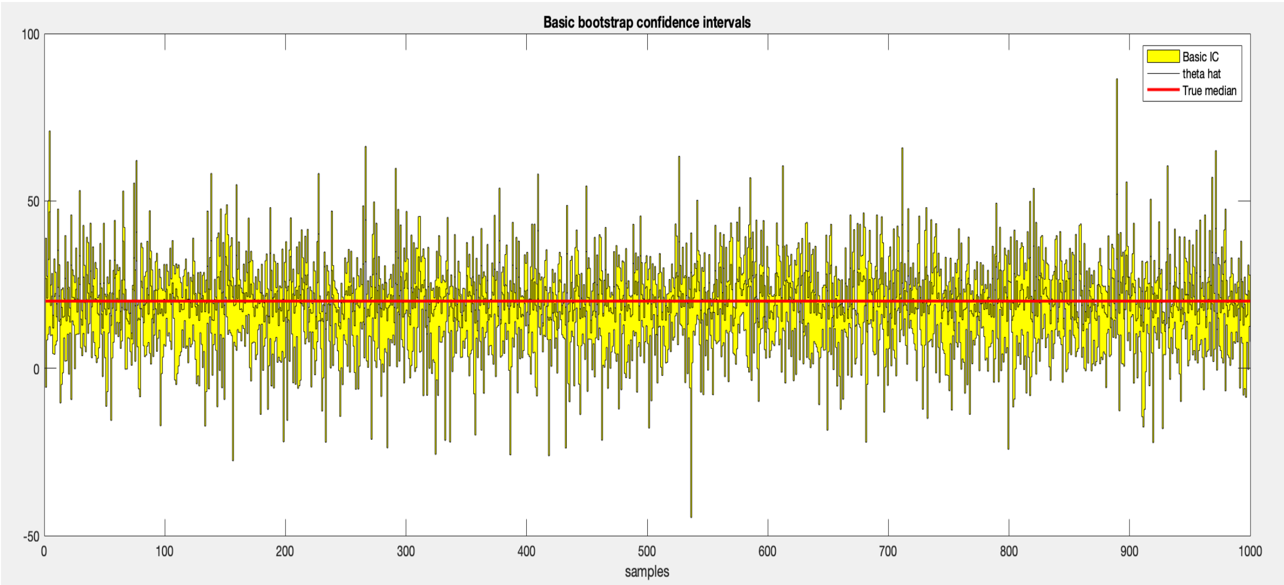
(a) As we saw in class, to use percentile bootstrap CI, two assumptions must be fulfilled : the first, is that one must work with the bootstrap estimator, i.e. the empirical analogue of the original parameter θ; the second, is that θ has to be the median. These assumptions are clearly satisfied in our case, since we are searching for the median of the distribution and $\hat{\theta}$ is the empirical median of the sample X. Then percentile CIs are valid in this case.

(b) We can say that, in this example, percentile bootstrap confidence intervals work better than the basic boostrap ones. Firstly, this is confirmed by the above table that shows that the coverage of basic CIs is 0.8 while we expect 0.95 coverage. Secondly, basic CIs cover an amplitude range of 130 while percentile CIs cover an amplitude range of 100; this should not happen since we are using the same data for both. Lastly, we note that basic CIs provides intervals even with negative values, which is not very good since we are working with positive samples and can therefore be assumed that the median is also positive. So we can conclude that in this case basic bootstrap is not a very good construction, whereas percentile bootstrap is.

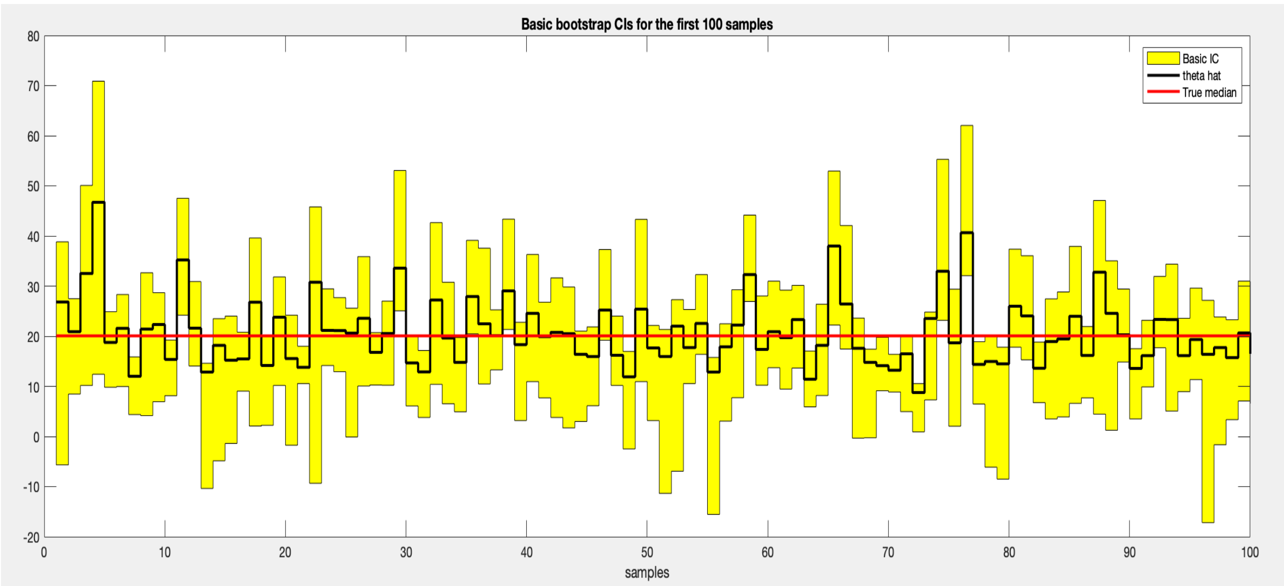
(c) Basic and percentile bootstrap CIs are bootstrap approximation of the true confidence interval. So, they are limited by the fact that they are based on bootstrap and bootstrap accuracy is limited by the sample size. In particular, if we increase the sample size then the accuracy of the bootstrap estimator should improve. Thus, we should see a positive effect on the performances of the methods as the sample size increases. Below are the "sample size"- coverages.

Bootstrap	20-coverage	200-coverage	500-coverage	1000-coverage	2000-coverage
basic	0.808	0.899	0.908	0.928	0.936
percentile	0.935	0.947	0.948	0.947	0.946

The results confirm the prediction.

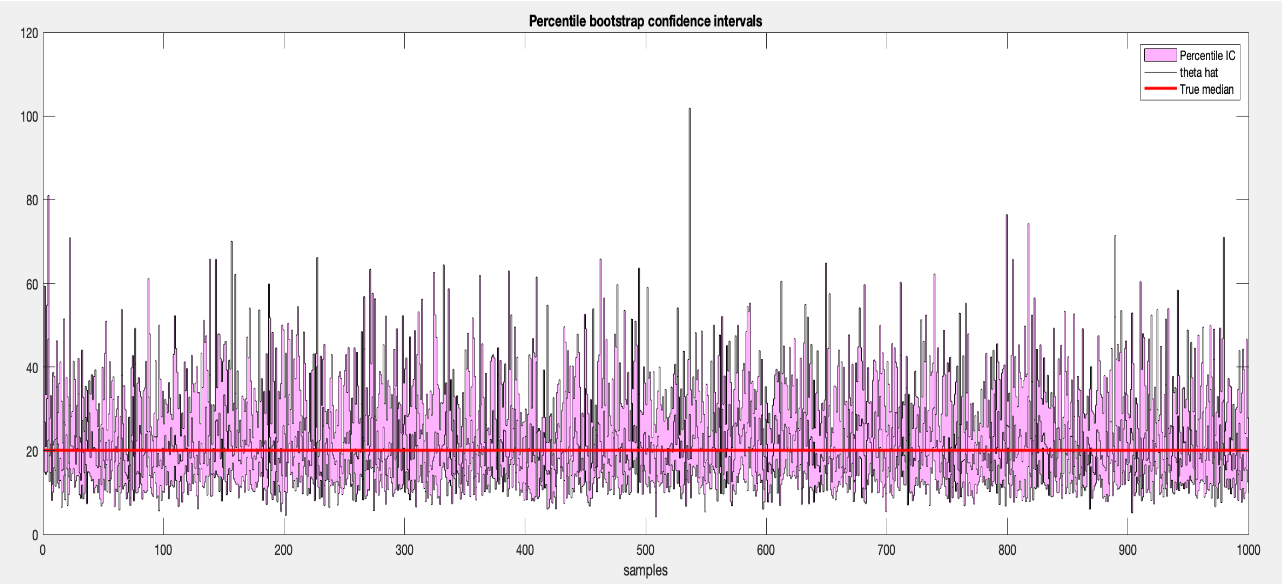


(a) 1000 samples

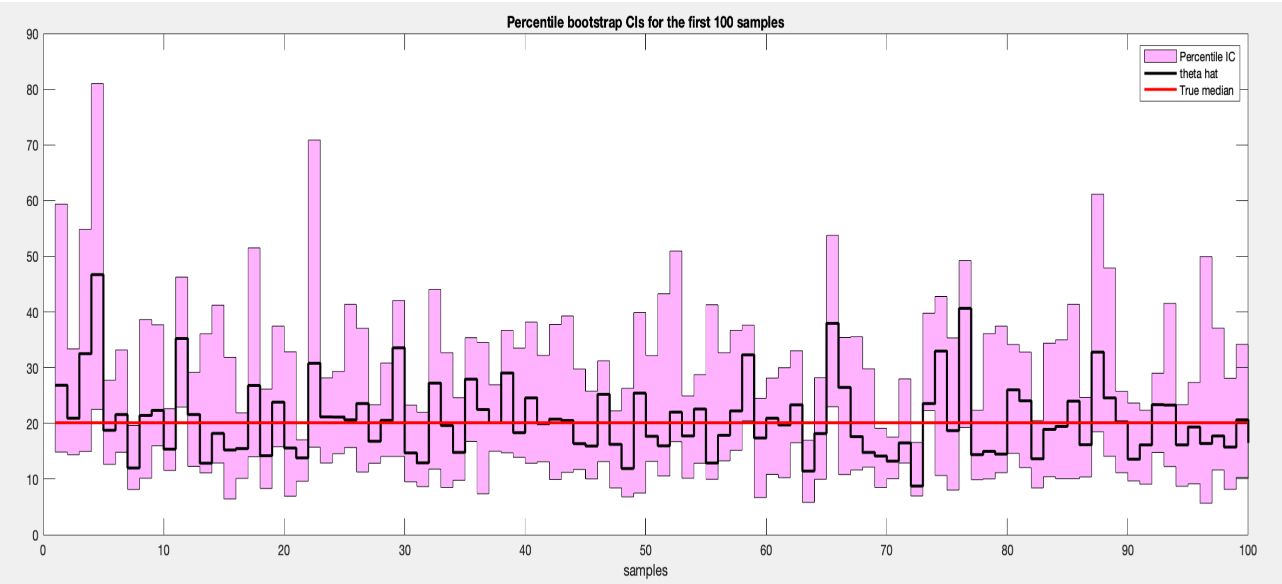


(b) First 100 samples

FIGURE 2.1 – Representations of basic-CIs, bootstrap estimator and true parameter



(a) 1000 samples



(b) First 100 samples

FIGURE 2.2 – Representations of percentile-CIs, bootstrap estimator and true parameter