**MT4113 Assignment 3 Simulation Study**

# Abstract

There are various methods of bootstraps that can produce confidence intervals, each with their own benefits and downfalls that mean they often do not perform as expected, with the coverage not being the expected 95% in same cases. In this simulation study I will look at 4 methods of bootstrap confidence intervals, the standard non-parametric percentile bootstrap, the bias-corrected and accelerated bootstrap, the bootstrap-t and a smoothed bootstrap. I simulated numerous confidence intervals for each method for varying numbers of bootstrap resamples and varying sizes of resample sizes. I investigated three different distributions - the Normal, Poisson and Gamma distributions - and looked at the coverage for each simulation. Surprisingly, the BCa performs poorly in comparison to these other methods, where in most scenarios, the percentile appears to have the most consistent and highest coverage.

# Introduction

Bootstrap confidence intervals are an essential tool in the statistician's toolkit, able to give estimates of variance of a parameter of interest when an analytic solution is not possible. There are numerous methods to create bootstrap confidence intervals but in this study we only look at four.

The non-parametric percentile bootstrap confidence interval is the most standard of all the bootstrap methods due to its simplistic nature. We can create multiple replications of the data by simulating from the empirical distribution function (EDF), which is done by resampling with replacement from the data. Since we assume our resamples come from the same distribution as the original data, then we can use these resamples to obtain the distribution of the quantity of interest eg. μ. We can create an interval that contains the true value of μ 100(1-α)% of the time, which is our confidence interval. We order the samples and take the α/2(b+1)%th value to be our lower limit and the upper limit to be the (1-α/2)(b+1)%th value. There are downfalls to this method though, if the distribution of our quantity of interest is not symmetric then the coverage error is often substantial. (Carpenter & Bithell, 2000) There are also problems when the standard error of our estimator depends on its value, this is a problem for the Poisson distribution where the variance is equal to its mean. Additionally, for low n the confidence intervals of our estimator tend to be too small. (Thomas, 2017)

A smooth bootstrap can also be created from the percentile method that fares better for low n. The procedure is as for the percentile method but before we calculate the quantity of interest for each of our bootstrap samples we add noise to our samples, typically from a Normal(0, h) or Uniform(-h, h), where h is a constant that indicates the degree of smoothing. The greater h, the greater effect of the smooth. (Thomas & Buckland, 2017)

The Bias-Corrected and accelerated (BCa) method is similar to the percentile method, in that we resample from our data and create estimates of our estimator. However when it comes to taking the quantiles of our ordered statistics α1(B+1)th and α2(B+1)th values, the α's are different from the percentile method. The α's for the BCa are fairly difficult to calculate, we need to compute a bias correction factor ẑ0 , the estimate of median bias of our bootstrap estimates as well as â an acceleration parameter. This is the estimated rate of change of the standard error of our estimate with respect to its true value.

This method typically has smaller coverage error than the percentile, and performs better at small n, and when the standard error of our estimator depends on the true value. But it is not very intuitive and can be computationally demanding. In addition, the coverage error increases as α tends to zero. (Carpenter & Bithell, 2000)

The bootstrap-t is conceptually simpler than the BCa method but can be numerically unstable and in non-parametric bootstrapping can produce exceedingly long confidence intervals. If we obtain an estimate φ for our estimator θ from data set then an estimate of the standard deviation for φ, σ, then we can define T = (φ - θ)/σ. If T(α) is the 100αth percentile of T then a 100(1-α)% confidence can be obtained by (φ-σ T(α), φ-σ T(1-α)). This assumes we know the T percentiles, but we do not in this case. We can create bootstrap replications by estimating φ and σ from each bootstrap resample and creating a T statistic for each sample, T\*, and find T(α)\* through the Bα ordered value of our T\*s. We can find our t confidence interval through (φ-σ T(α)\*, φ-σ T(1-α)\*). (DiCiccio & Efron, 1996)

In my study I shall be looking at each of these methods and seeing how they compare under different situations by analysing their confidence interval coverage using a Normal(0,1), Poisson(10) and a Gamma(2,0.5).

# Methods

The bootstrap methods I implemented were the non-parametric percentile, BCa, t and smooth bootstraps to create my confidence interval simulations.

I simulated *B* samples of length *n* from the distribution that I wanted using my *data.gen* function for Normal, Poisson and Gamma distributions. Then calculated the mean or t-statistic (depending on which method I was using) for each of these bootstrap samples in my *b.ests.np* function. Following this I then calculated the correct α for the appropriate method, in order to use the quantile function on these bootstrap estimates to obtain a confidence interval for the mean of my samples. I did this with my *bootstrap.ci* function. I then simulated this many times in my *do.sim.np* function. Into this, I inputted a vector for the number of bootstrap samples I wanted to change whilst keeping the size of the samples constant. There was also a vector inputted into this function that was for the sizes of the bootstrap samples whilst keeping the number of samples constant in order to see the difference between methods regarding performance when *B* and *n* change independently. I carried out these simulations lots of times for each level of *B* and *n.* Performance was assessed using the proportion of times that the true mean of my chosen distribution was included in the simulated confidence interval over the simulations for each level of *B* and *n.* ggplots were then created in order to see how coverage changed for each method and distribution as these variables changed.

# Results

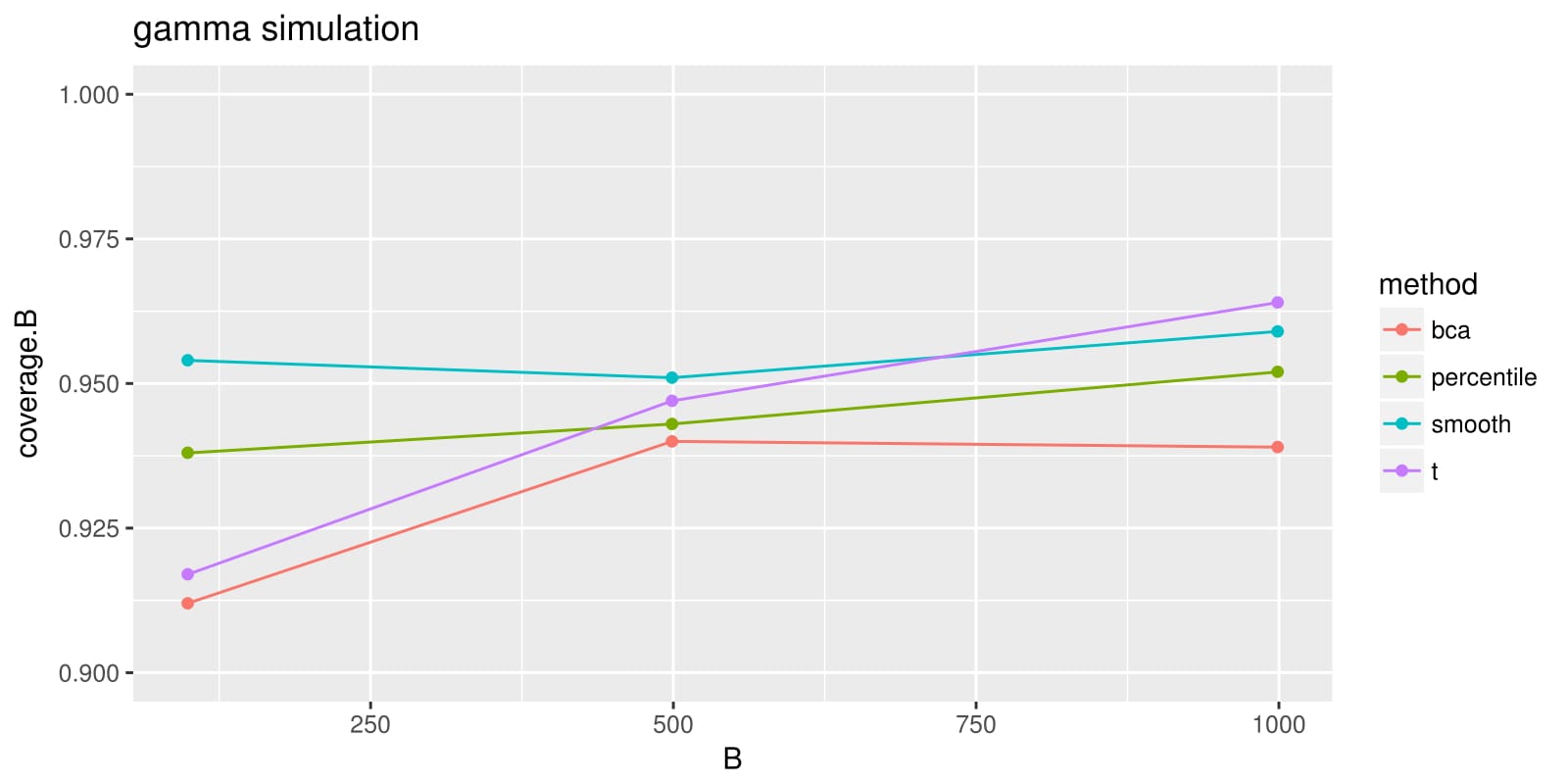


Figure 1. Gamma simulation for varying levels of B

# gamma.n-1.jpg

Figure 2. Gamma simulation for varying levels of n

# normal.n-1.jpg

Figure 3. Normal simulation for varying levels of n

# normal.B-1.jpg

Figure 4. Normal simulation for varying levels of B

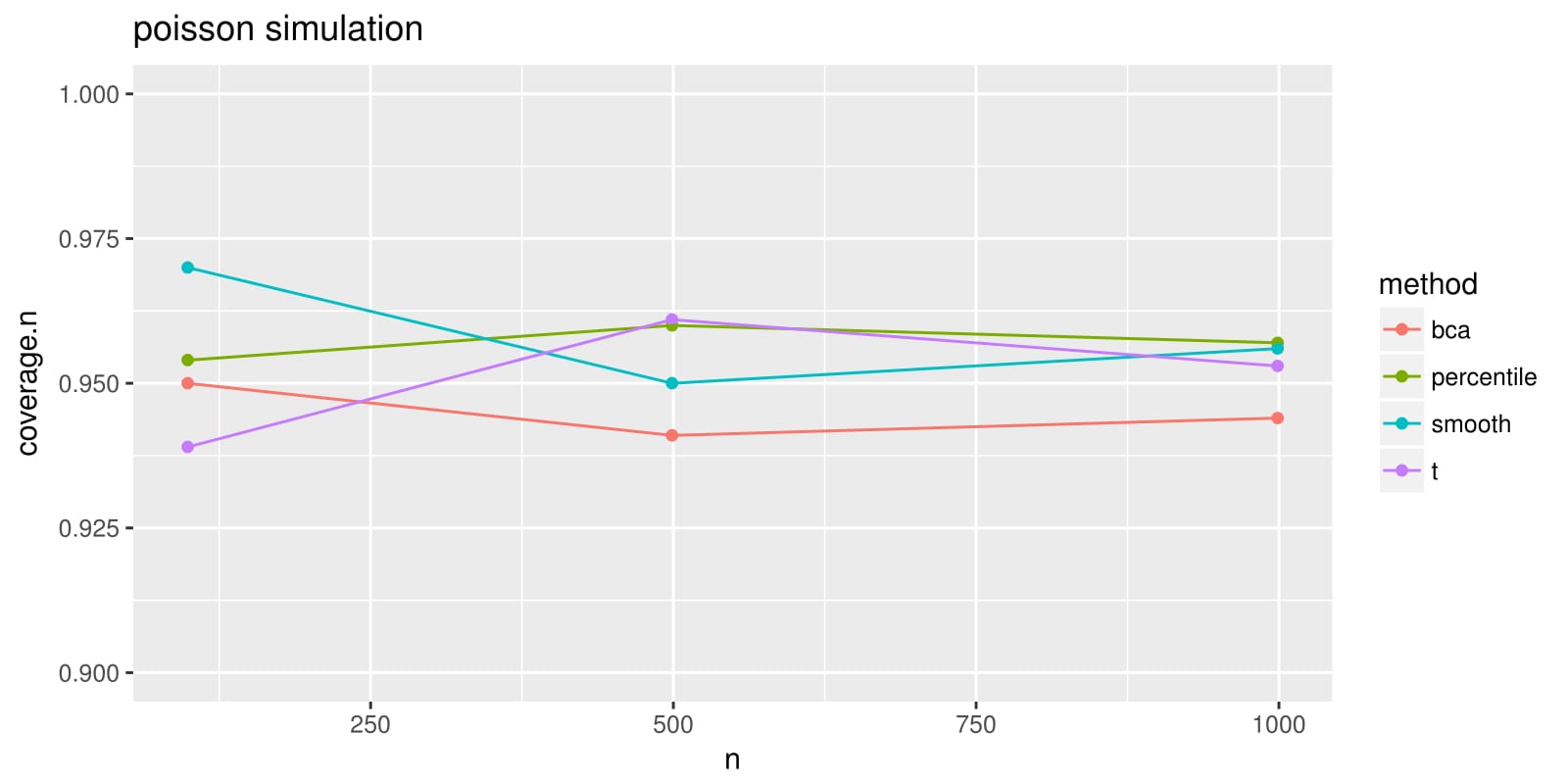


Figure 5. Poisson simulation for varying levels of n

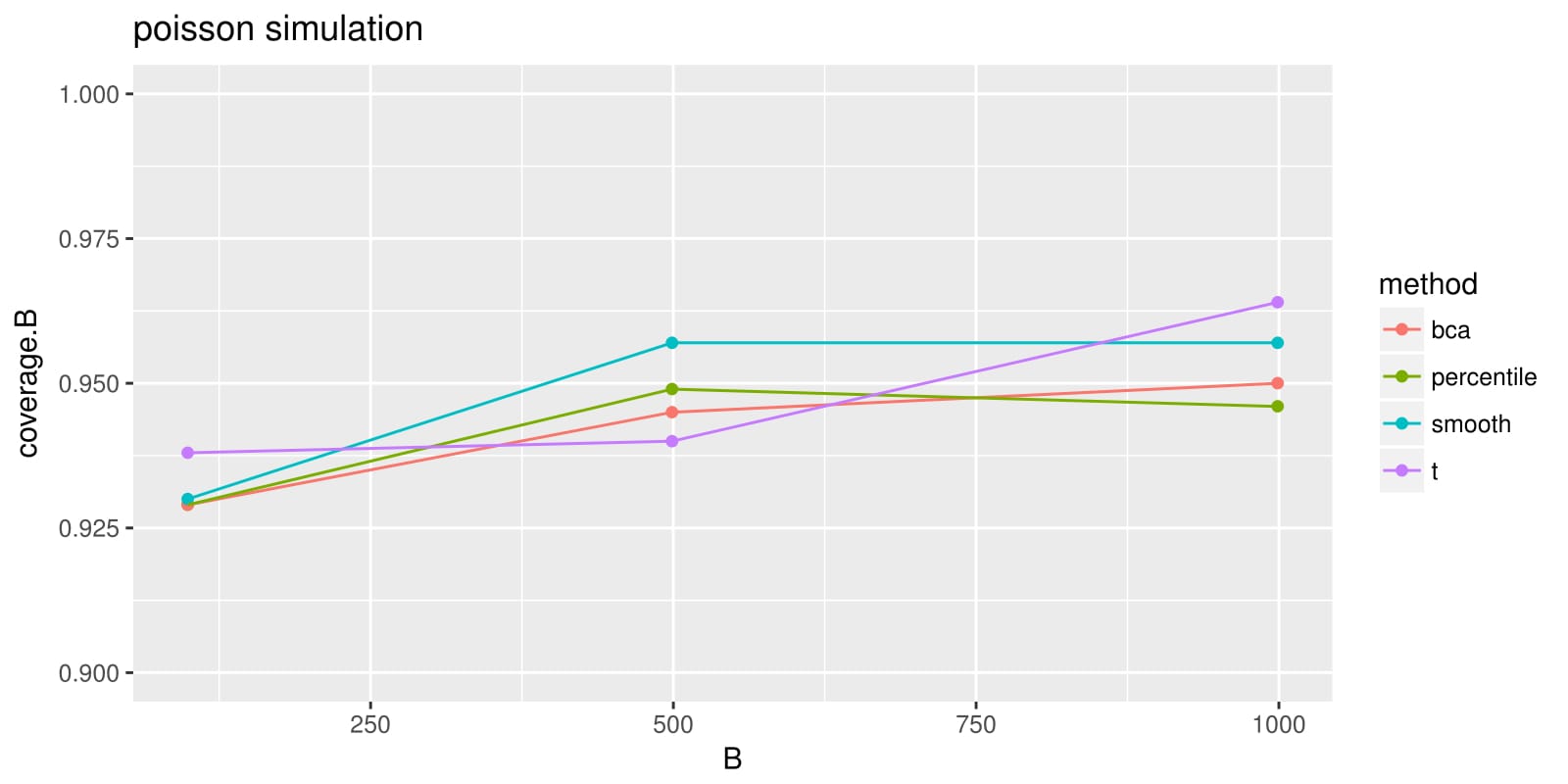
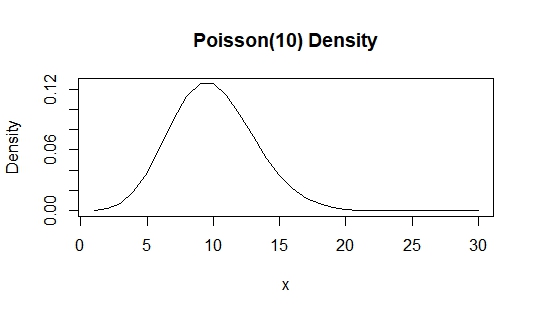
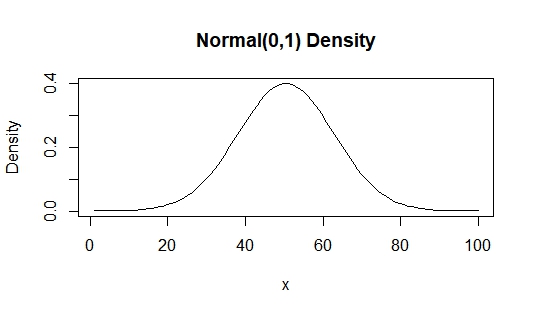


Figure 6. Poisson simulation for varying levels of B

These are the results of 1000 simulations, for B=99, B=499, B=999 and n=99, n=499 and n=999 for the percentile method, BCa, t and smooth bootstrap confidence intervals. I used a Normal(0,1), Poisson(10) and Gamma(2, 1/2). These distributions can be seen in Figure 7.





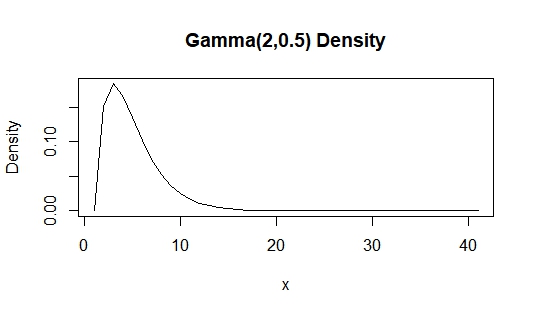


Figure 7. Density plots of my distributions

I also carried out further simulations to look at smaller values of B and n. I carried out 1000 simulations for B=49, 99 , 149 and 199 and n=49, 99, 149 and 199 using the same distributions and methods. I will look only at the normal distribution as I want to look at the effect of low sample size and number of samples differs across methods.

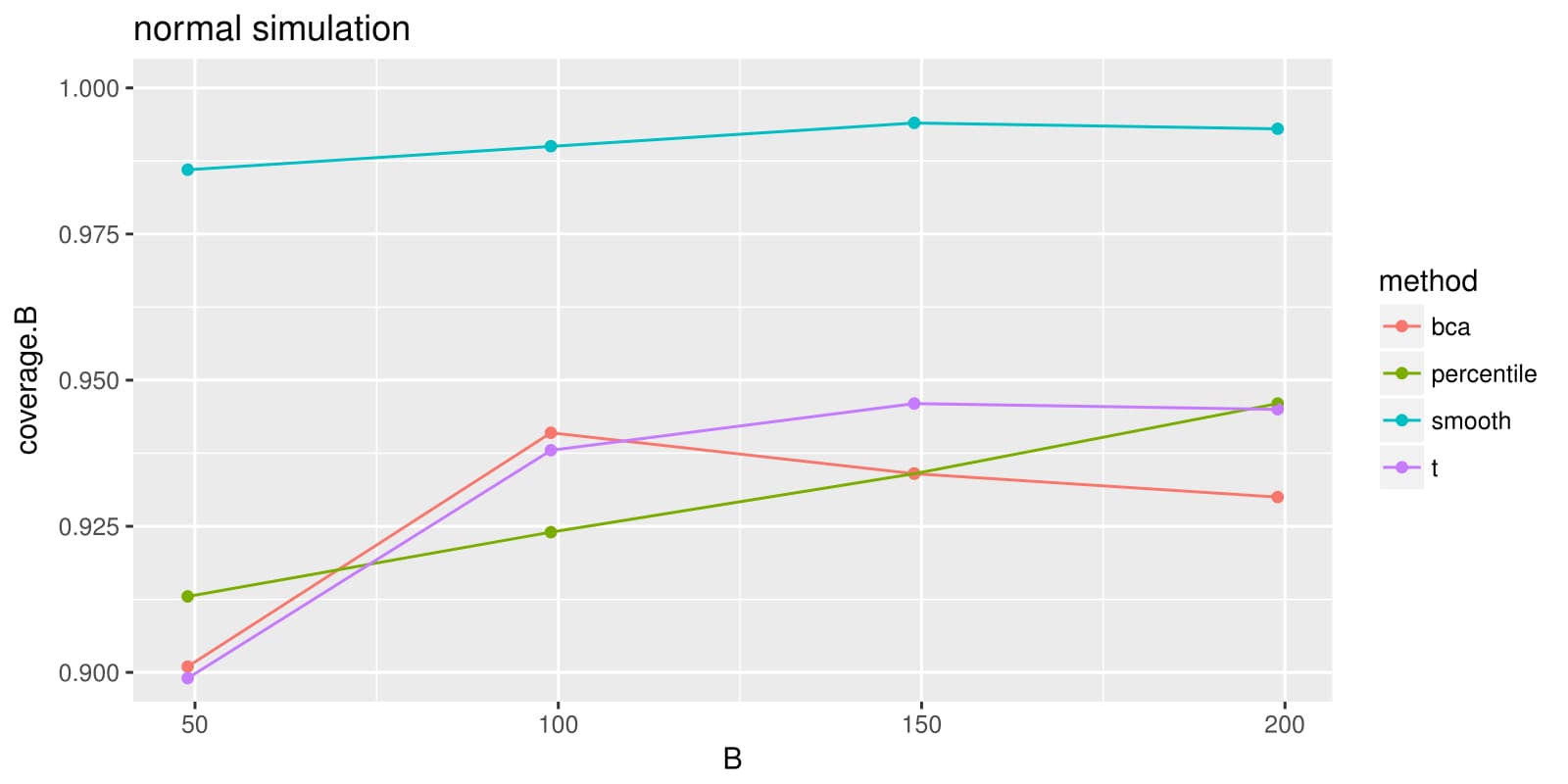


Figure 8. Normal simulation for small B

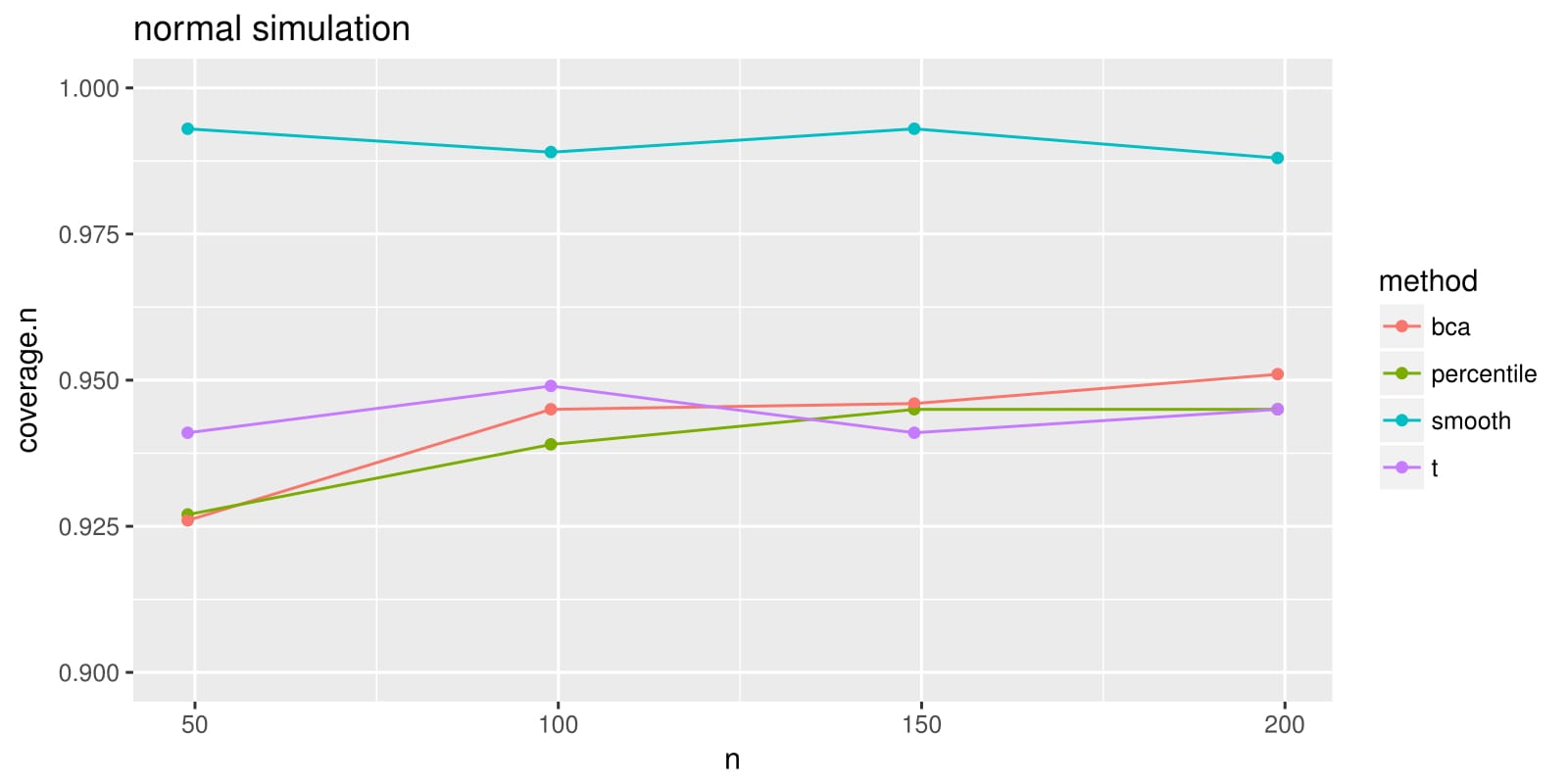


Figure 9. Normal simulations for small n

## Discussion

From Figures 1-6, it would appear that the BCa method does not perform as expected. I was expecting the BCa to perform the best out all the methods chosen but seems to consistently perform poorly in comparison. For the "simplest" distribution, the Normal, the BCa performs poorly when n is kept constant. As n changes and B is kept constant, the coverage improves a bit and is better than the bootstrap-t at least.

The percentile method often has better coverage in my simulations than the BCa which is very surprising for the Poisson and Gamma distributions as these distributions have features which the percentile method is meant to perform poorly on. The Poisson distribution's variance depends on the true value of its mean and so it doesn't make sense for it to outperform the BCa for this distribution, as the BCa does not suffer from this disadvantage. In fact for varying levels of n, the percentile has a coverage over 95%! It seems that all methods are better than the BCa for the Poisson.

The Gamma distribution I chose was very skewed and non-symmetric as can be seen in Figure 7. and the percentile method does not deal very well with non-symmetric distributions, but in Figures 1 and 2 the percentile seems to deal with this problem very well which was not expected and again, outperforms the BCa. In fact, both the t and the smooth do as well. All methods except the smooth perform poorly for when B is changing which was a bit odd, since for varying n the coverage of all methods is better. I thought the two would be similar.

From Figures 8 and 9, the BCa overall does appear the perform better at low values of n and B ( < 100) for the Normal distribution, the smooth still has exceedingly high coverage. The t also has better coverage at really low values of n, but then it appears to oscillate around this value.

The smooth bootstrap appears to perform very well for when B is changing, and at low B. It regularly has better performance than the percentile, however I chose h, the degree of smoothing to be the sd(data)/3 which may have been too much smoothing. For the normal distribution it performs exceedingly well.

The t often seems to perform well and then worse in other scenarios, it's coverage is a bit erratic and seems to display numerical instability.

# Conclusions

In conclusion, from my simulation study it would appear that the percentile is exceedingly robust to skewness, low sample size and distributions that have variance dependent on the mean which is I doubt is truly the case. The BCa is meant to perform better in all of these scenarios and so I think I must have implemented the BCa incorrectly or there's a very small chance the BCa isn't as good as advertised. I have looked at my code for hours and can't seem to pinpoint the problem. I expected the BCa to have the best coverage of all of the methods I chose, the t, even if a bit erratic, performs better. The smooth seems to improve the percentile coverage, especially when using the normal distribution.

Ideally I would like to run even longer simulations to look in more detail at how the coverage for each method behaves but this would have been computationally infeasible and far too time-consuming.

# Bibliography

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, *19*(9), 1141–64. http://doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F

DiCiccio, T. J., & Efron, B. (1996). Bootstrap Confidence Intervals. *Statistical Science*, *11*(3), 189–212. Retrieved from http://www.jstor.org/stable/2246110

Thomas, L. (2017). MT4113 Computer Intensive Methods Lecture Notes.

Thomas, L., & Buckland, S. (2017). MT4113 Background Reading for Lectures 6-8.