Higher Order Bootstrapping

Quinlin Neuhaus
STAT 5001

April 26, 2026

1 Introduction

In today's statistics world, bootstrap methods have become a popular instrument for the approximation of sampling distributions, the construction of confidence intervals, and small sample correction. First introduced by Efron in 1979 as an evolution of the jackknife, ordinary bootstrapping requires mild assumptions and delivers first order accuracy ($O(n^{-1/2})$).

Higher order methods use studentization to achieve improved accuracy at the cost of stricter assumptions. When appropriate smoothness and moment conditions are met, the percentile-t bootstrap procedure reaches second order accuracy ($O(n^{-1})$). This was first shown by Singh in 1981 with Edgeworth expansions and expanded on by Hall in 1992.

My report showcases the theory behind the asymptotic behavior of ordinary and studentized bootstrap methods and demonstrates their differences empirically.

2 Bootstrap Methods

2.1 Ordinary Bootstrap

Let $X_1, \ldots, X_n$ be an independent and identically distributed sample from some unknown distribution F, and let $\hat{F}$ denote the empirical distribution constructed from said sample. A bootstrap sample $X_1^*, \ldots, X_n^*$ is drawn uniformly and with replacement from $\hat{F}$, and a statistic of interest $\theta = \theta(F)$ is estimated by $\hat{\theta} = \theta(\hat{F})$.

The bootstrap approximation to the sampling distribution of $\hat{\theta}$ is obtained by repeatedly computing

$$\hat{\theta}^* = \theta(X_1^*, \ldots, X_n^*).$$

A $100(1-\alpha)\%$ confidence interval constructed using the empirical quantiles of the bootstrap distribution takes the form

$$[q_{\alpha/2}, q_{1-\alpha/2}].$$

Under standard regularity conditions, including finite second moments and differentiability of $\theta()$, the bootstrap is consistent. However, the resulting confidence intervals exhibit coverage error of $O(n^{-1/2})$, reflecting the CLT nature of the approximation.

2.2 Percentile-t Bootstrap

Following the same problem setup as the ordinary bootstrap, the percentile-t bootstrap improves upon the ordinary method by studentizing the statistic. Specifically, one considers

$$T_n = \frac{\hat{\theta} - \theta}{\hat{\sigma}}.$$

The bootstrap analogue is given by

$$T_n^* = \frac{\hat{\theta}^* - \hat{\theta}}{SE(\hat{\theta}^*)}.$$

$100(1-\alpha)\%$ Confidence intervals are constructed using quantiles of the bootstrap distribution of T_n^*, yielding intervals of the form

$$\left[\hat{\theta} - q_{1-\alpha/2} * SE(\hat{\theta}),\ \hat{\theta} - q_{\alpha/2} * SE(\hat{\theta})\right].$$

This approach requires stronger assumptions, including the existence of fourth moments and twice differentiability of θ. Under these conditions, the percentile-t bootstrap achieves second order accuracy, with coverage error of $O(n^{-1})$.

3 Edgeworth Expansions and Higher Order Accuracy

To understand the improvement in accuracy, consider a statistic of the form

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{T_i - \mu}{\sigma}.$$

While the central limit theorem guarantees the convergence of T_n to a normal distribution, the Edgeworth expansion gives a better approximation of the following form:

$$P(T_n \leq t) = \Phi(t) + n^{-1/2}p_1(t)\phi(t) + n^{-1}p_2(t)\phi(t) + \cdots ,$$

where $\Phi(t)$ and $\phi(t)$ denote the standard normal CDF and density, respectively, and $p_k(t)$ are products of population cumulants and Hermite polynomials.

The bootstrap distribution admits an analogous expansion,

$$P^*(T_n^* \leq t) = \Phi(t) + n^{-1/2}\hat{p}_1(t)\phi(t) + n^{-1}\hat{p}_2(t)\phi(t) + \cdots ,$$

in which the unknown cumulants are replaced by their sample estimates.

The difference between the true and bootstrap expansions can therefore be written as

$$P(T_n \leq t) - P^*(T_n^* \leq t) = (\Phi(t) - \Phi(t)) + n^{-1/2}(p_1(t) - \hat{p}_1(t))\phi(t) + O(n^{-1}).$$

Since the difference between population and sample cumulants is of order $O(n^{-1/2})$ by the Law of Large Numbers

$$P(T_n \leq t) - P^*(T_n^* \leq t) = n^{-1/2}O(n^{-1/2})\phi(t) + O(n^{-1}).$$

Using $O()$ multiplication rules

$$P(T_n \leq t) - P^*(T_n^* \leq t) = O(n^{-1})\phi(t) + O(n^{-1}).$$

Because $\phi(t)$ does not grow with n

$$P(T_n \leq t) - P^*(T_n^* \leq t) = O(n^{-1}) + O(n^{-1}).$$

Using $O()$ addition rules

$$P(T_n \leq t) - P^*(T_n^* \leq t) = O(n^{-1}).$$

This cancellation of the first order term explains the improved accuracy of studentized bootstrap procedures. While countably infinite higher order terms exist after the second order one, they are inconsequential to the bootstrap's convergence since the lowest order error dominates the expansion.

4 Methods

The central component of higher order bootstrap methods is studentization, which requires estimation of the statistic's standard error. In some cases a known formula exists for said standard error, however, when a closed form expression for the standard error is unavailable, it may be approximated using an additional layer of resampling. This is the double bootstrap, in which a bootstrap procedure is applied within each bootstrap sample.

Although computationally intensive, the double bootstrap provides a practical means of implementing studentized procedures for complex statistics. It is worth noting that Moore's Law and the explosion of GPU computing have all but made concerns about computation cost and time negligent. The bootstrap algorithm, studentized or otherwise, is an embarrassingly parallel process. If compute resources are the only thing that stand in the way of a better performing method, this obstacle should not dissuade its use.

5 Simulation

5.1 Design

To evaluate empirical performance and demonstrate the effect of the violation of regularity conditions, the following simulation was conducted. Consider estimation of the population mean under three distributions: an Exponential(1) distribution, which has all finite moments, a Pareto(3) distribution, which has finite variance but infinite kurtosis, and a Pareto(1.5) distribution, which has finite mean but infinite variance. These populations allow the assessment of the role of moment conditions in bootstrap accuracy, and compare ordinary and studentized bootraps' accuracy and convergence.

For each of the three population distributions, samples of sizes ranging from 4 to 400 are generated. For each sample, 1000 bootstrap resamples are used to construct 95% confidence intervals for both methods, and the entire procedure is repeated 5000 times to estimate coverage probabilities.

5.2 Results

The first figure for each distribution shows empirical coverage probability across different sample sizes. The second compares log(Coverage Error) with log(Sample Size) and estimates the slope with a simple linear regression. This method is by no means an exact statistically sound way to measure the convergence, but rather serves as a simple way to look at its rate.

5.2.1 Exponential(1) Distribution

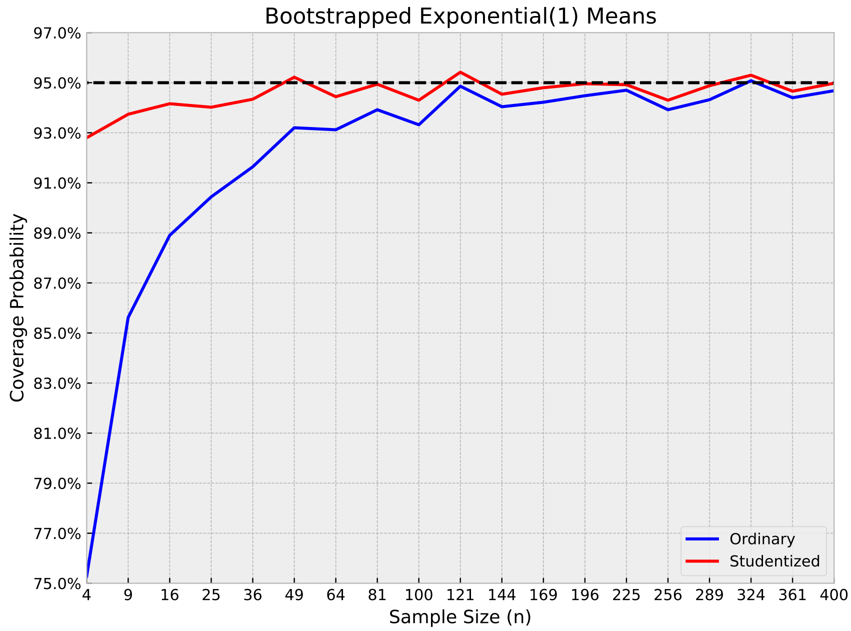

Figure 1: Coverage probabilities for Exponential(1)

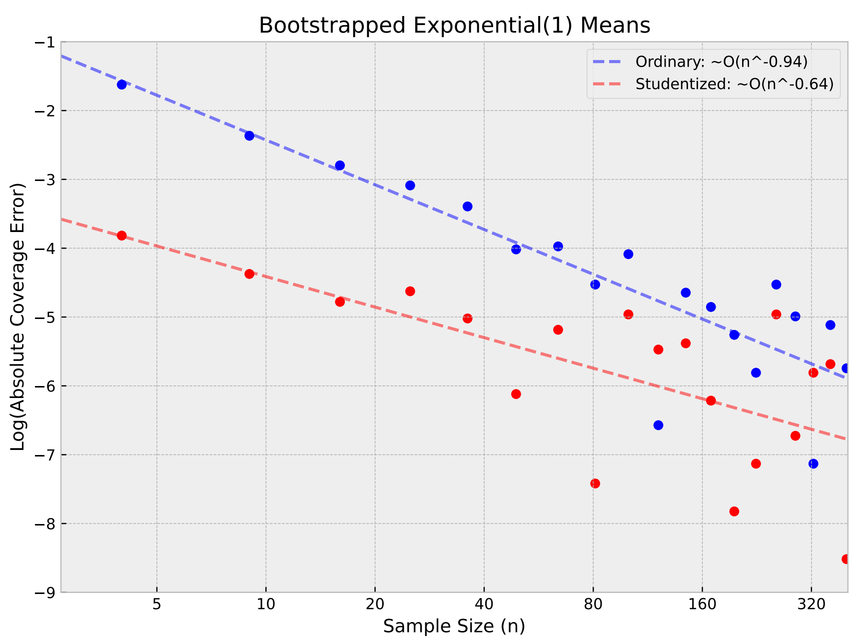

Figure 2: Convergence rate for Exponential(1)

5.2.2 Pareto(3) Distributions

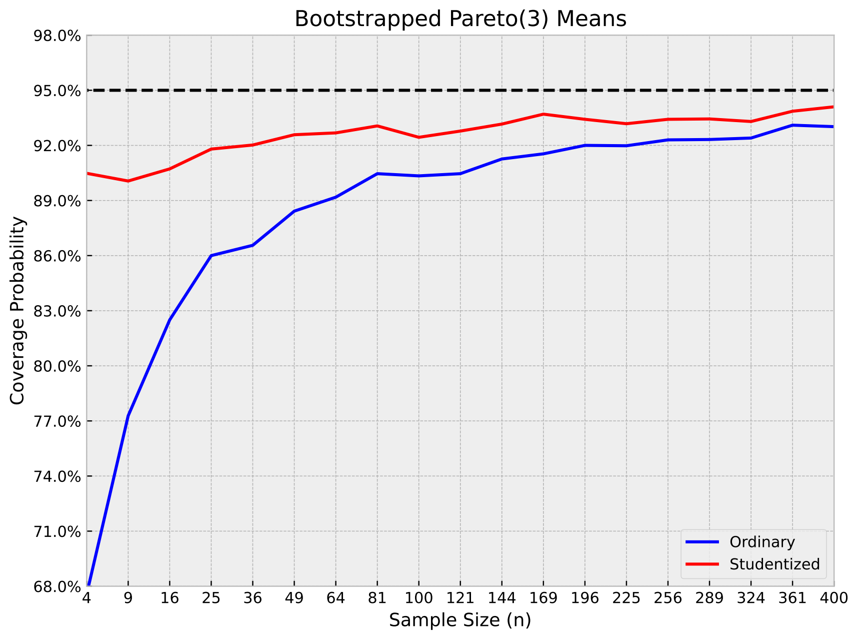

Figure 3: Coverage probabilities for Pareto(3)

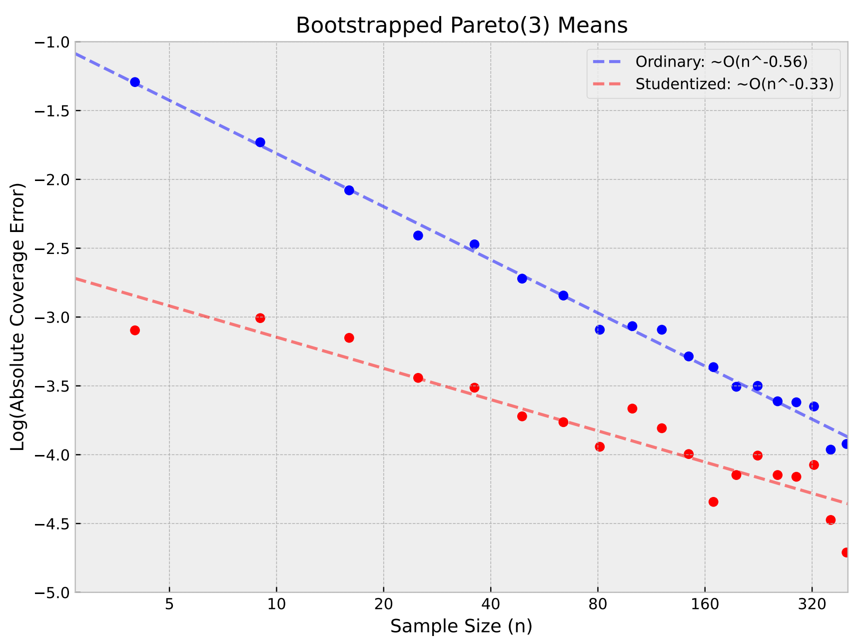

Figure 4: Convergence rate for Pareto(3)

5.2.3 Pareto(1.5) Distributions

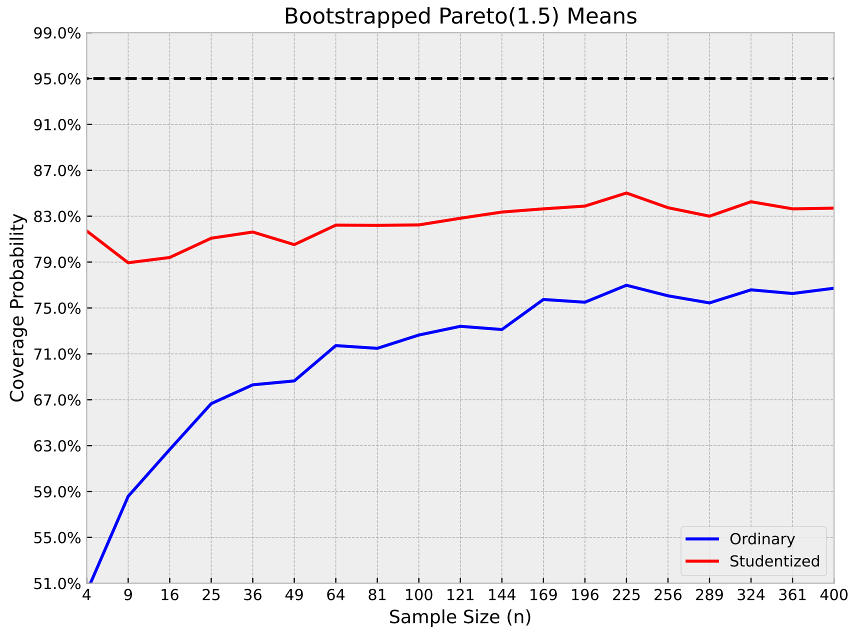

Figure 5: Coverage probabilities for Pareto(1.5)

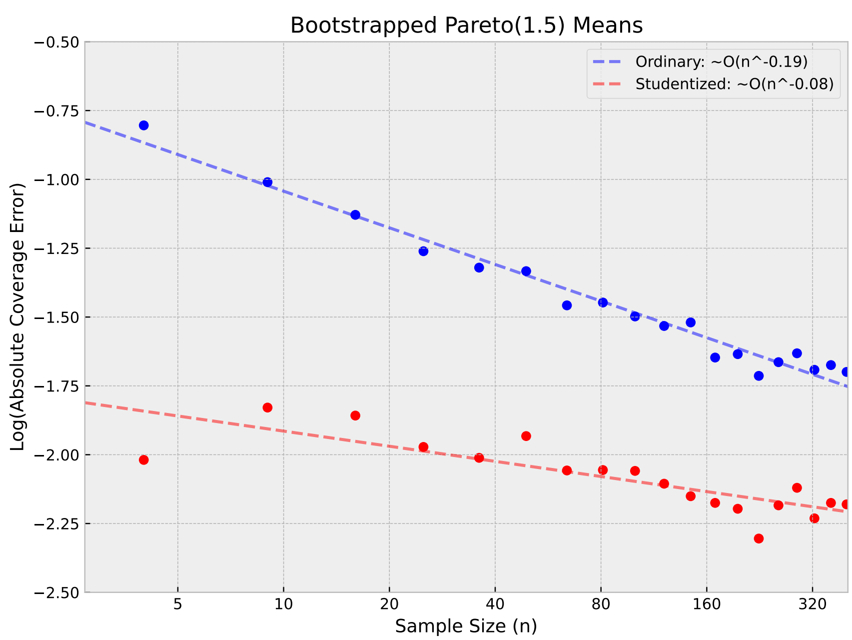

Figure 6: Convergence rate for Pareto(1.5)

5.3 Discussion

The results demonstrate that the effectiveness of bootstrap methods depends on satisfying their regularity conditions. In all three examples, the studentized bootstrap outperformed the ordinary bootstrap, with the greatest difference in performance at the Pareto(1.5) case. All three methods also increase coverage probability as the original sample size grows.

First, for the Exponential distribution, both methods achieve close to 95% coverage quickly, with the studentized bootstrap hardly beating the ordinary one at higher sample sizes. Figure 2 shows a surprising result, the ordinary bootstrap seems to converge faster than expected, and the studentized bootstrap converges slowly, although this could be because the studentized bootstrap starts at such a high coverage probability to begin with.

Next, for the Pareto(3) population, the studentized bootstrap marginally beats out the ordinary bootstrap in coverage probability even though this distribution violates its assumptions. Looking at figure 4 shows what is expected from this distribution, the ordinary bootstrap converges at the first order rate, and the studentized bootstrap fails to reach its convergence rate.

Finally, for the Pareto(1.5) case, the studentized bootstrap outperforms the ordinary one significantly. It is also shown that the violation of both methods' conditions leads to almost no convergence in coverage probability from figure 6.

6 Applications and Limitations

Higher order bootstrap methods are routinely useful in small samples, skewed distributions, or complex estimators where analytic variance formulas are unavailable. However, their general applicability can be constrained by the need for higher order moments and smoothness. Like countless other statistical methods, bootstrapping can give sharper error convergence at the trade off of stricter regularity conditions.

7 Conclusion

The ordinary bootstrap provides a broadly applicable method with first order accuracy, while the studentized bootstrap achieves second order accuracy from the cancellation of the first order term in the Edgeworth expansion. Empirical results suggest that studentized bootstrapping can outperform ordinary bootstrapping, potentially even despite a lack of satisfied assumptions. Overall, the choice of bootstrap method should be guided by both theoretical guarantees and the properties of the underlying data.

References

[1] Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife.

[2] Hall, P. (1992). The Bootstrap and Edgeworth Expansion.

[3] Hesterberg, T. (2014). What Teachers Should Know About the Bootstrap.

[4] Singh, K. (1981). On the Asymptotic Accuracy of Efron's Bootstrap.