Bootstrapping on Heavy Tail Data Meredith Clayton* & Rachel Thornton**

What Are Heavy Tails?

"The top 1% of a population owns 40% of the wealth."

"The top 2% of twitter users send 60% of the tweets."

"File sizes on a computer tend to be small with a few very large files thrown in the mix."



Even when the CLT applies to heavy tailed (HT) distributions, it can have to a relatively slow rate of convergence to the normal distribution.

A slow rate of convergence leads to imprecise results for small data sets.

Our Problem

We are looking for a data-analysis method that approximates quantiles for the following statistic based on HT data:

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$$

The new method should outperform the CLT in case of small and moderate sample sizes.

What is Bootstrapping?

Bootstrapping allows us to create simulated data samples by resampling original data. This "replicates" our distribution, which allows us to estimate quantiles in the data.



Accuracy of our Method

The charts below compare our new method for estimating quantiles with the studentized distribution approach based on a Pareto(1, 4.1) distribution and LogNormal(0, 1) with sample sizes of 25.





Our bootstrap method is more accurate at predicting quantiles for small, HT samples.

Acknowledgements

We would like to thank our mentor Mayya Zhilova, the School of Mathematics at Georgia Tech, and the NSF Award DMS-1851843.

*Stephen F. Austin State University, **University of Texas at Austin

