

Figure 1: PP Plot for Entire Pareto Set

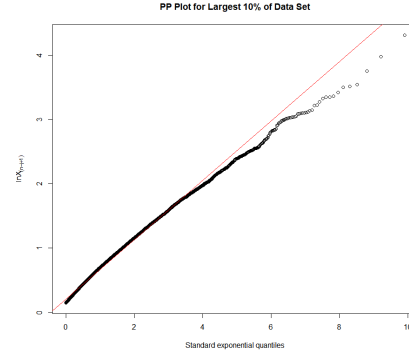


Figure 2: PP Plot for Largest 10% of Pareto Set

Above are the Zipf plots for 200000 randomly generated realizations from a Pareto Random Variable with parameter 3. We can see the typical behaviour of such a plot for the entire data set as well as the largest 10% of the realizations. The red line is a fitted line using linear regression.

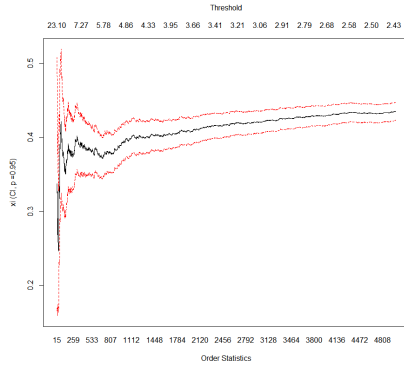


Figure 3: Hill Plot of Largest 5000 Order statistics

Table 1: The Hill estimates for Pareto with parameter 3. Standard errors are in parenthesis.

k	Right Tail		
	2000	3000	4000
X_i	0.4081(0.009)	0.4211(0.007)	0.4300(0.006)

Hence, we can say that $\xi \approx 0.42$, then tail density is $1 - F(x) \rightarrow \mathcal{L}(x)^{-2.38}$ as x get sufficiently large. This is close enough to the original Pareto Parameter.

Similiarly, if we have another 200000 realizations with pareto parameter 2, we have the following Hill Plot and Estimators.

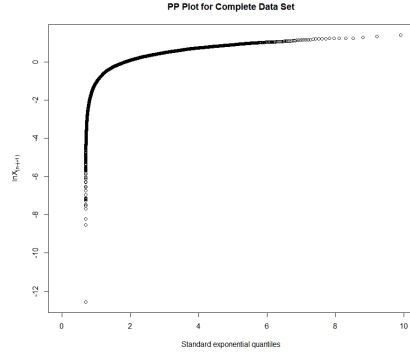


Figure 5: PP Plot for Entire Normal Set

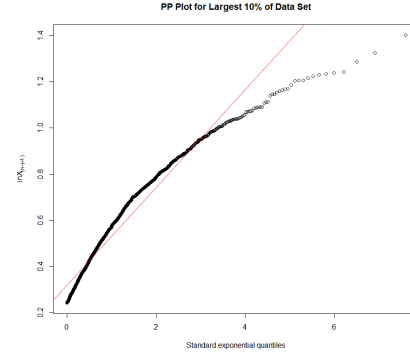


Figure 6: PP Plot for Largest 10% of Normal Set

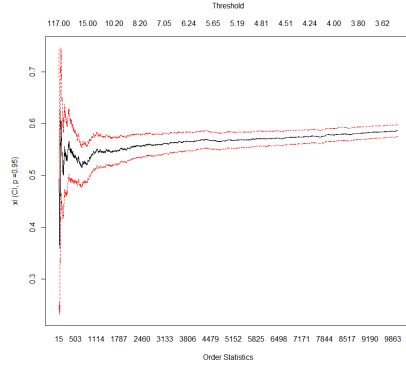


Figure 4: Hill Plot of Largest 5000 Order statistics

Hence, we can say that $\xi \approx 0.55$, then tail density is $1 - F(x) \rightarrow \mathcal{L}(x)^{-1.82}$ as x get sufficiently large. This is close enough to the Pareto Parameter 2.

If you take instead 200000 realizations from a standard normal distribution, we have the following Zipf plots

Table 2: The Hill estimates for Pareto with parameter 3. Standard errors are in parenthesis.

k	2000	Right Tail 3000	4000
X_i	0.5495(0.012)	0.5597(0.0102)	0.5659(0.0089)

In contrast to the Pareto, we see that the largest 10% of the data set shows non-linearity and bows away from the fitted least squares line, suggesting thinner tails.

Below is the Hill Plot for the Largest 5000 order statistics. We can see even in such a small subset, there is no region of stability, making it much more difficult to estimate ξ .

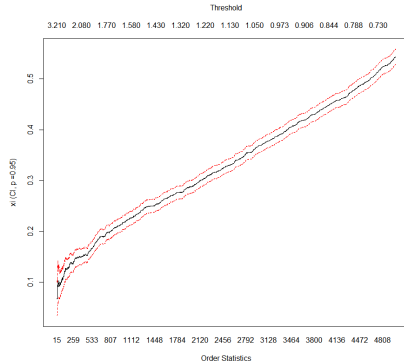


Table 3: The Hill estimates for Standard Normal. Standard errors are in parenthesis.

k	2000	Right Tail 3000	4000
Y_i	0.2882(0.006)	0.3687(0.006)	0.4478(0.007)

Figure 7: Hill Plot of Largest 5000 Order statistics

In general, there is no hard a fast rule for choosing k . We use heuristic rules to find regions where the Hill Plot becomes stable. That is, find a region in the hill plot where it seems that the variance subsides. This is referred to as the “Eye-Ball technique” [1].

Bibliography

- [1] Danielsson, Jon and Ergun, Lerby Murat and de Haan, Laurens and de Vries, Casper G. *Tail index estimation: Quantile driven threshold selection*. Available at SSRN 2717478, 2016.