Extreme Value Theory Without the Largest Values

Jingjing Zou\textsuperscript{1}
Joint work with Richard A. Davis\textsuperscript{1}
and Gennady Samorodnitsky\textsuperscript{2}

1. Department of Statistics, Columbia University; 2. School of Operations
Research and Information Engineering, Cornell University
Motivation: Heavy-tailed Data

Heavy-tails in data are often modeled by Pareto-type distributions: for large $x$

$$P(X > x) \sim \frac{C}{x^\alpha},$$

where $\alpha > 0$ is the tail index
Hill Estimator

- Independent $X_1, X_2, \ldots, X_n \sim F(x)$
- Tail index $\alpha$
- Order statistics $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$
- Hill estimator for $\gamma = 1/\alpha$

$$H_n(k) = \frac{1}{k} \sum_{i=1}^{k} \log X_{(n-i+1)} - \log X_{(n-k)}$$
Figure: Hill plot of i.i.d. Pareto ($\alpha = 0.5$) variables ($n = 1000$)
Example: Google+ Data

- A snapshot of the social network taken on Oct, 2012
- 76,438,791 nodes
- 1,442,504,499 edges

**Figure:** Hill Plots of In-degrees
Hill Plot

Figure: With 100 largest observations removed
Re-parametrization of Hill Estimator

- \( n \): original sample size
- \( k_n \): arbitrary sequence such that \( k_n \to \infty \) and \( k_n/n \to 0 \)
- \( \delta k_n \): observations removed and NOT observed
- \( \theta k_n \): number of top observations included in estimation
- Hill estimator without the extremes

\[
H_n(\delta, \theta) = \frac{1}{[\theta k_n]} \sum_{i=1}^{[\theta k_n]} \log X(n-[\delta k_n]-i+1) - \log X(n-[\delta k_n]-[\theta k_n])
\]
Functional Convergence of Hill Estimator without the Extremes

- $\sqrt{n}(H_n(\delta, \theta) - E(H_n))$ converges to a Gaussian random field
- Different values of $\delta$ and $\alpha$ are distinguishable through behaviors of sample paths of $H_n$

Figure: Pareto ($\alpha = 0.5$) variables ($n = 1000$, $k_n = 100$ and $\delta k_n = 100$)
Fitting mean curves to Pareto with top 100 values removed

\[ n = 1000, \ k_n = 100, \ \alpha = 0.5, \ \delta = 1 \]

\[ \begin{align*}
\delta &= 0.8 \\
\delta &= 1 \\
\delta &= 1.2 \\
\delta &= 1.4
\end{align*} \]

\[ \begin{align*}
0.0 & \quad 0.5 & \quad 1.0 & \quad 1.5 & \quad 2.0 \\
0.0 & \quad 0.2 & \quad 0.4 & \quad 0.6 & \quad 0.8 & \quad 1.0
\end{align*} \]

**Figure:** Fix \( \alpha = 0.5 \), change \( \delta \)
Fitting mean curves to Pareto with top 100 values removed

\( n = 1000, \ k_n = 100, \ \alpha = 0.5, \ \delta = 1 \)

Figure: Fix \( \delta = 1 \), change \( \alpha \)
Functional Convergence of Hill Estimator without the Extremes

Theoretical conditions for the convergence

- $F$ regularly varying
- Second-order regular variation condition
- Bias term in the mean of the Gaussian process if not Pareto
Functional Convergence

As \( n \to \infty \),

\[
\sqrt{k_n} \left( H_n(\cdot, \cdot) - \frac{g(\cdot, \cdot)}{\alpha} \right) - b_\rho(\cdot, \cdot) \xrightarrow{d} \frac{1}{\alpha} G(\cdot, \cdot)
\]

in \( D([0, \infty) \times (0, \infty)) \), where

\[
g(\delta, \theta) = \begin{cases} 
1, & \delta = 0, \\
1 - \delta \log \left( \frac{\theta}{\delta} + 1 \right), & \delta > 0,
\end{cases}
\]

\[
b_\rho(\delta, \theta) = \begin{cases} 
\frac{\lambda}{1 - \rho \theta^\rho}, & \delta = 0, \\
\frac{1}{1 + (\theta/\delta)^\rho - (\theta/\delta + 1)^\rho} \frac{\lambda}{(\delta + \theta)^\rho}, & \delta > 0,
\end{cases}
\]

\( G \) is a continuous Gaussian random field with mean zero.
Estimation Procedure

▶ Estimate parameters based on the asymptotic joint distribution of \( \{H_n\} \)

▶ Solve for maximum likelihood estimators for
  
  ▶ Number of removed extremes \( \delta k_n \)
  
  ▶ Tail index \( \alpha \)

▶ Beirlant et al. (2016)\(^1\) modeled truncation with threshold parameter \( T \) and estimated parameters based on Pareto likelihood

---

Estimation Results

- Cauchy distribution
- $\alpha = 1, \ n = 2000, \ k_n = 200, \ \text{removal} \ \delta k_n = 200$
- Averaged estimation results of 200 independent simulations

![Graph 1](image1.png)

- Estimates for $\alpha$ and $\delta k_n$

![Graph 2](image2.png)
Earthquake Data

- Earthquake fatalities by the U.S. Geological Survey (1900 - 2014) \(^2\)
- \(n = 125\) earthquakes with 1,000 or more deaths
- First apply the estimation procedures to the original data
- Then to the data with additional removal of 10 top observations
- Estimations should reflect the removal

Earthquake Data

**Figure: removal = 0**

**Figure: removal = 10**

**Figure: Estimates of number of extremes removed**
Earthquake Data

Figure: removal = 0

![Graph showing estimates of the tail index $\alpha$ for removal = 0, with two lines representing Proposed and Beirlant methods.]

Figure: removal = 10

![Graph showing estimates of the tail index $\alpha$ for removal = 10, with two lines representing Proposed and Beirlant methods.]

Figure: Estimates of the tail index $\alpha$
Earthquake Data

Figure: removal = 0

Figure: removal = 10

Figure: Hill estimators vs. fitted mean curves (with different number of observations included in estimation)
Earthquake Data

**Figure:** removal = 0

**Figure:** removal = 10

**Figure:** Hill estimators vs. fitted mean curves (with different number of observations included in estimation)
Earthquake Data

**Figure:** removal = 0

**Figure:** removal = 10

**Figure:** Hill estimators vs. fitted mean curves (with different number of observations included in estimation)
**Google+ Data**

**Figure:** removal = 0

**Figure:** removal = 400

**Figure:** Estimates of number of extremes removed
Google+ Data

Figure: removal = 0

Figure: removal = 400

Figure: Estimates of tail index $\alpha$
Google+ Data

Figure: removal = 0

Figure: removal = 400

Figure: Hill estimators vs. fitted mean curves (with different number of observations included in estimation)
Google+ Data

**Figure:** removal = 0

**Figure:** removal = 400

**Figure:** Hill estimators vs. fitted mean curves (with different number of observations included in estimation)
Google+ Data

Figure: removal = 0

Figure: removal = 400

Figure: Hill estimators vs. fitted mean curves (with different number of observations included in estimation)