Distributed Inference for Extreme Value Index

Liujun Chen
Fudan University

joint work with Deyuan Li and Chen Zhou

On Extreme Value Analysis 2021

Distributed Inference

- Data privacy issue.
    - Bank may not share their operational loss.
    - Insurance firms cannot share insurance claims.

Distributed Inference

- Data privacy issue.
  - Bank may not share their operational loss.
  - Insurance firms cannot share insurance claims.
- Computational issue.
  - Size of the dataset is beyond a computer's memory.

Distributed Inference

- Divide and Conquer algorithm:
  - estimates on each machine
  - transmits the results to the center machine
  - takes "average" in the central machine

Distributed Inference

- Divide and Conquer algorithm:
  - estimates on each machine
  - transmits the results to the center machine
  - takes "average" in the central machine
- Oracle Property
  Whether the aggregated estimator achieves the same statistical efficiency as the Oracle estimator (the imaginary estimator using all observations).

Model Setting

- Consider a distribution function $F \in D(G_\gamma)$ with $\gamma > 0$ (heavy tailed distribution).
- This is equivalent to that $U := \{1/(1-F)\}^{\leftarrow}$ is a regular varying function:
$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma.$$
- A key question in extreme value analysis is to estimate the extreme value index.

Model Setting

- Assume that the i.i.d. observations $X_1, \ldots, X_N$ are stored in m machines with n observations in each machine, i.e. $N = nm$.
- Assume that $m = m(N) \to \infty, n = n(N) \to \infty$ as $N \to \infty$.
- Practically, we cannot apply statistical procedures to the oracle sample, i.e, the hypothetically combined dataset $\{X_1, \ldots, X_N\}$.

Oracle Hill estimator

If we can use the oracle sample, the (oracle) Hill estimator is defined as

$$\hat{\gamma}_H := \frac{1}{l} \sum_{i=1}^{l} \log M^{(i)} - \log M^{(l+1)}$$

where $l = l(N) \to \infty, l/N \to 0$ as $N \to \infty$.
Here, $M^{(1)} \geq \cdots \geq M^{(N)}$ are the order statistics of the oracle sample.

Distributed Hill estimator

Following a divide and conquer algorithm:

- Apply the Hill estimator at each machine:

$$\hat{\gamma}_{j,H} = \frac{1}{k}\sum_{i=1}^{k} \log M_j^{(i)} - \log M_j^{(k+1)}.$$

  Here, $M_j^{(1)} \geq \cdots \geq M_j^{(n)}$ are the order statistics of the observations in machine j.

- Take the average of the Hill estimates from all machines

$$\hat{\gamma}_{DH} := \frac{1}{m}\sum_{j=1}^{m} \hat{\gamma}_{j,H}.$$

Main Conditions

(A) $m = m(N) \to \infty, n = n(N) \to \infty$ and $n/\log m \to \infty$ as $N \to \infty$.

(B) (Second Order Condition.) There exist an eventually positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ and a real number $\rho \leq 0$ such that

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\gamma}}{A(t)} = x^{\gamma} \frac{x^{\rho} - 1}{\rho},$$

for all $x \geq 0$.

Asymptotics: when k is fixed

Theorem Suppose that $F \in D(G_\gamma)$ with $\gamma > 0$ and Conditions A and B hold. Assume $k \geq 1$ is a fixed integer. If $\sqrt{km}A(n/k) = O(1)$ as $N \to \infty$,

$$\sqrt{km}\,\{\hat{\gamma}_{DH} - \gamma - A(n/k)g(k,n,\rho)\} \xrightarrow{d} N(0,\gamma^2),$$

where

$$g(k,n,\rho) = \frac{1}{\rho}\left(\frac{n}{k}\right)^{-\rho}\frac{\Gamma(n+1)\Gamma(k-\rho+1)}{\Gamma(n-\rho+1)\Gamma(k+1)}.$$

Oracle Property: when k is fixed

- Assume that $\sqrt{km}A\left(\frac{N}{km}\right) = \sqrt{km}A\left(\frac{n}{k}\right) \to \lambda \in \mathbb{R}$, the Oracle Hill estimator possesses the asymptotic normality

$$\sqrt{km}\left(\hat{\gamma}_H - \gamma\right) \xrightarrow{d} N(\lambda/(1-\rho), \gamma^2).$$

Oracle Property: when k is fixed

- Assume that $\sqrt{km}A\left(\frac{N}{km}\right) = \sqrt{km}A\left(\frac{n}{k}\right) \to \lambda \in \mathbb{R}$, the Oracle Hill estimator possesses the asymptotic normality

$$\sqrt{km}\left(\hat{\gamma}_H - \gamma\right) \xrightarrow{d} N(\lambda/(1-\rho), \gamma^2).$$

- Under the same condition, we have that

$$\sqrt{km}\left(\hat{\gamma}_{DH} - \gamma\right) \xrightarrow{d} N\left(\lambda \frac{k^\rho}{1-\rho} \frac{\Gamma(k-\rho+1)}{\Gamma(k+1)}, \gamma^2\right).$$

Oracle Property: when k is fixed

- Assume that $\sqrt{km}A\left(\frac{N}{km}\right) = \sqrt{km}A\left(\frac{n}{k}\right) \to \lambda \in \mathbb{R}$, the Oracle Hill estimator possesses the asymptotic normality

$$\sqrt{km}\left(\hat{\gamma}_H - \gamma\right) \xrightarrow{d} N(\lambda/(1-\rho), \gamma^2).$$

- Under the same condition, we have that

$$\sqrt{km}\left(\hat{\gamma}_{DH} - \gamma\right) \xrightarrow{d} N\left(\lambda \frac{k^\rho}{1-\rho} \frac{\Gamma(k-\rho+1)}{\Gamma(k+1)}, \gamma^2\right).$$

- The oracle property holds only when $\rho = 0$ or $\lambda = 0$.

Asymtotics: when k is an intermediate sequence

Theorem Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and Conditions A and B hold. Assume $k = k(N) \to \infty, k/n \to 0$ as $N \to \infty$. If $\sqrt{km}A\left(n/k\right) = O(1)$, then

$$\sqrt{km}\left\{\hat{\gamma}_{DH} - \gamma - A(n/k)g(k, n, \rho)\right\} \xrightarrow{d} N(0, \gamma^2).$$

Oracle property: when k is an intermediate sequence

- Assume that $\sqrt{km}A\left(\frac{N}{km}\right) = \sqrt{km}A\left(\frac{n}{k}\right) \to \lambda \in \mathbb{R}$, the Oracle Hill estimator possesses the asymptotic normality

$$\sqrt{km}\,(\hat{\gamma}_H - \gamma) \xrightarrow{d} N(\lambda/(1-\rho), \gamma^2).$$

Oracle property: when k is an intermediate sequence

- Assume that $\sqrt{km}A\left(\frac{N}{km}\right) = \sqrt{km}A\left(\frac{n}{k}\right) \to \lambda \in \mathbb{R}$, the Oracle Hill estimator possesses the asymptotic normality

$$\sqrt{km}\left(\hat{\gamma}_H - \gamma\right) \xrightarrow{d} N(\lambda/(1-\rho), \gamma^2).$$

- Under the same condition, $g(k, n, \rho) \to 1$, we have that

$$\sqrt{km}\left(\hat{\gamma}_{DH} - \gamma\right) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

Oracle property: when k is an intermediate sequence

- Assume that $\sqrt{km}A\left(\frac{N}{km}\right) = \sqrt{km}A\left(\frac{n}{k}\right) \to \lambda \in \mathbb{R}$, the Oracle Hill estimator possesses the asymptotic normality

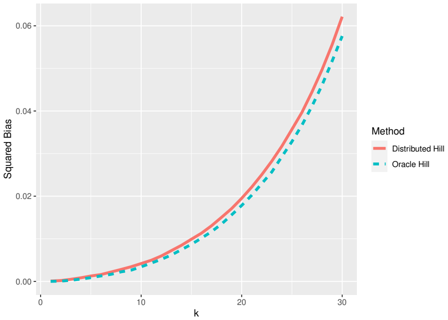
$$\sqrt{km}\left(\hat{\gamma}_H - \gamma\right) \xrightarrow{d} N(\lambda/(1-\rho), \gamma^2).$$

- Under the same condition, $g(k, n, \rho) \to 1$, we have that

$$\sqrt{km}\left(\hat{\gamma}_{DH} - \gamma\right) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

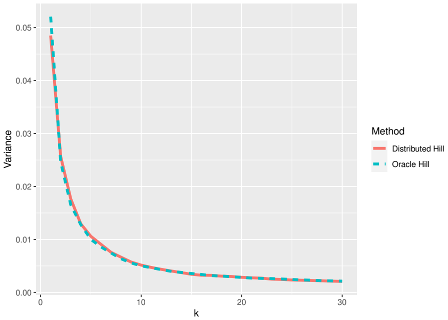
- The oracle property always holds in this case.

Simulation: squared bias for different level of k



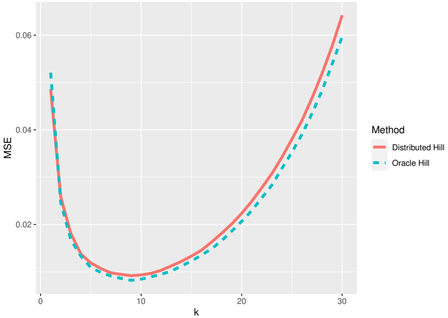
- Unit Fréchet distribution ($\rho = -1$), $N = 1000, m = 20, n = 50$
- Oracle Hill estimator uses $km = 20k$ exceedance ratios

Simulation: variance for different level of k



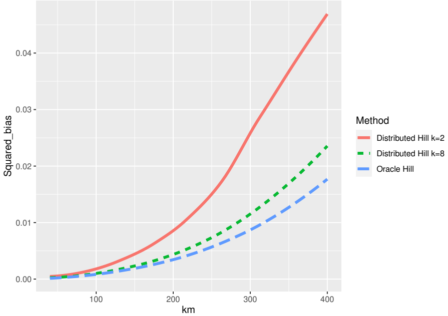
- Unit Fréchet distribution ($\rho = -1$), $N = 1000$, $m = 20$, $n = 50$
- Oracle Hill estimator uses $km = 20k$ exceedance ratios

Simulation: MSE for different level of k



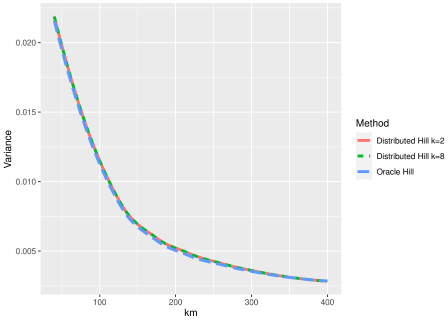
- Unit Fréchet distribution $(\rho = -1)$, $N = 1000$, $m = 20$, $n = 50$
- Oracle Hill estimator uses $km = 20k$ exceedance ratios

Simulation: squared bias for different level of m



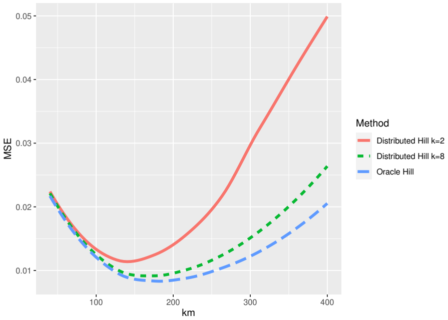
- Unit Fréchet distribution $(\rho = -1)$, $N = 1000$
- We fix k at two levels: $k = 2$ and $k = 8$.
- The x-axis is the effective number of exceedance ratios km.

Simulation: variance for different level of m



- Unit Fréchet distribution ($\rho = -1$), $N = 1000$
- We fix k at two levels: $k = 2$ and $k = 8$.
- The x-axis is the effective number of exceedance ratios km.

Simulation: MSE for different level of m



- Unit Fréchet distribution ($\rho = -1$), $N = 1000$
- We fix k at two levels: $k = 2$ and $k = 8$.
- The x-axis is the effective number of exceedance ratios km.

The choice of k

- The distributed Hill estimator is sensitive to the choice of k in each machine.
- This choice leads to a bias-variance tradeoff.
- Such a problem is more pronounced in the context of distributed inference.

The choice of k

In extreme value statistics literatures, there are two types of solutions:

- Finding the optimal level that balances the asymptotic bias and variance.
- Bias correction.

Bias Correction Methodology

Recall that, for the distributed Hill estimator,

$$\hat{\gamma}_{DH,k} - \gamma = \frac{N_\gamma}{\sqrt{km}} \frac{A(n/k)}{1-\rho} g(k, n, \rho) + \frac{1}{\sqrt{km}} o_P(1)$$

- The bias is an explicit function $\frac{A(n/k)}{1-\rho} g(k, n, \rho)$
- We shall estimate the bias, subtract it from the original estimator, and create the asymptotically unbiased distributed estimator.

Estimation for ρ

- We use a high intermediate sequence k_ρ for estimating the second order parameter ρ.
- Assume that as $N \to \infty$, $k_\rho \to \infty, k_\rho/n \to 0$, and

$$\begin{aligned} \sqrt{k_\rho m} A(n/k_\rho) &\to \infty, \\ \sqrt{k_\rho m} A^2(n/k_\rho) &\to \lambda_1, \\ \sqrt{k_\rho m} A(n/k_\rho) B(n/k_\rho) &\to \lambda_2, \end{aligned}$$

where B is the third order scale function.

Estimation for ρ

Consider the following statistics computed based on the observations in each machine j,

$$R_{j,k}^{(\alpha)} = \frac{1}{k}\sum_{i=1}^{k}\left\{\log M_j^{(i)} - \log M_j^{(k+1)}\right\}^{\alpha}, \quad \alpha = 1, 2, 3.$$

- We request that each machine sends the values $R_{j,k}^{(\alpha)}, \alpha = 1, 2, 3$ to the central machine.
- On the central machine, we take the average of the $R_{j,k}^{(\alpha)}$ to obtain

$$R_k^{(\alpha)} = \frac{1}{m}\sum_{j=1}^{m} R_{j,k}^{(\alpha)}$$

Estimation for ρ

We define the estimator for the second order parameter ρ as

$$\hat{\rho}_{k,\tau} = -3 \left| \frac{T_{k,\tau} - 1}{T_{k,\tau} - 3} \right|,$$

where

$$T_{k,\tau} := \frac{\left(R_k^{(1)}\right)^{\tau} - \left(R_k^{(2)}/2\right)^{\tau/2}}{\left(R_k^{(2)}/2\right)^{\tau/2} - \left(R_k^{(3)}/6\right)^{\tau/3}}.$$

Asymptotically unbiased distributed estimator for γ

- We can choose a high level of k in the eventual asymptotically unbiased distributed estimator.
- In our context, we choose k_n such that,

$$
\begin{aligned}
k_n/k_\rho &\to 0, \\
\sqrt{k_n m}A(n/k_n) &\to \infty, \\
\sqrt{k_n m}A^2(n/k_n) &\to 0, \\
\sqrt{k_n m}A(n/k_n)B(n/k_n) &\to 0.
\end{aligned}
$$

Asymptotically unbiased distributed estimator for γ

The asymptotically unbiased distributed estimator for γ is defined as

$$\tilde{\gamma}_{k_n,k_\rho,\tau} := R_{k_n}^{(1)} - \frac{R_{k_n}^{(2)} - 2\left(R_{k_n}^{(1)}\right)^2}{2R_{k_n}^{(1)}\hat{\rho}_{k_\rho,\tau}\left(1-\hat{\rho}_{k_\rho,\tau}\right)^{-1}}.$$

Note that, the estimator $\tilde{\gamma}_{k_n,k_\rho,\tau}$ adheres to a DC algorithm since each machine only sends five values $\left\{R_{j,k_n}^{(1)}, R_{j,k_n}^{(2)}, R_{j,k_n}^{(3)}, R_{j,k_\rho}^{(1)}, R_{j,k_\rho}^{(2)}\right\}$ to the central machine.

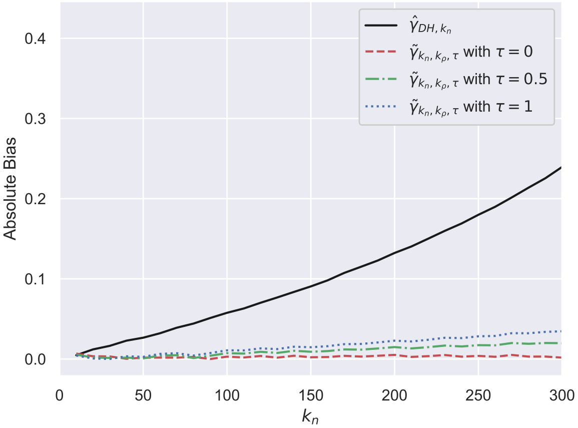
Asymptotics

Theorem

Under some mild conditions,

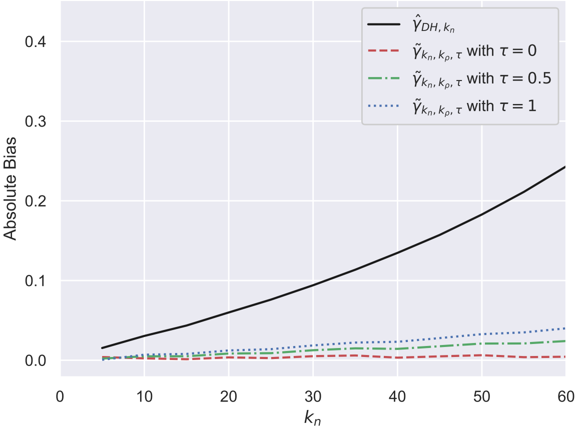
$$\sqrt{k_n m}\left(\tilde{\gamma}_{k_n,k_\rho,\tau} - \gamma\right) \xrightarrow{d} N\left(0, \gamma^2\left\{1 + (\rho^{-1} - 1)\right\}^2\right).$$

Simulation: Absolute bias for different levels of k_n



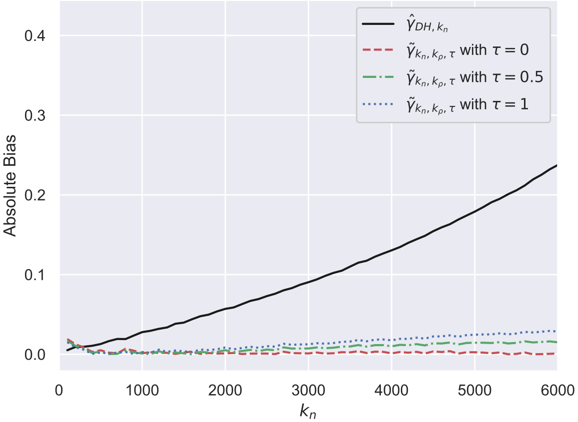
- Unit Fréchet distribution ($\rho = -1$)
- $N = 10000$, $m = 20$

Simulation: Absolute bias for different levels of k_n



- Unit Fréchet distribution ($\rho = -1$)
- $N = 10000$, $m = 100$

Simulation: Absolute bias for different levels of k_n



- Unit Fréchet distribution $(\rho = -1)$
- $N = 10000, m = 1$

Thank You!