Asymptotic properties of the Hill estimator for error contaminated data

Mihyun Kim

Joint work with Piotr Kokoszka

Department of Statistics
Colorado State University

EVA, July 4 2019

Outline

The Hill estimator

- Suppose that $X_1, \ldots, X_n$ are i.i.d. nonnegative random variables with common distribution F_X, which has **regularly varying** tail probabilities ($\bar{F}_X \in RV_{-\alpha}$):

$$\bar{F}_X = 1 - F_X = P(X > \cdot) \in RV_{-\alpha}, \ \alpha > 0.$$

- α^{-1} is estimated by **the Hill estimator**

$$H_{k,n} := \frac{1}{k} \sum_{i=1}^{k-1} \log \frac{X_{(i)}}{X_{(k)}},$$

with the convention that $X_{(1)}$ is the largest order statistic.

Motivation

- In many applications, data are contaminated by noise, measurement errors or roundoff errors.

- An example of internet traffic anomalies

Fig 1: A map showing 14 two-directional links of the backbone internet network in the United States known as Internet2.

Motivation - Internet Traffic Anomalies

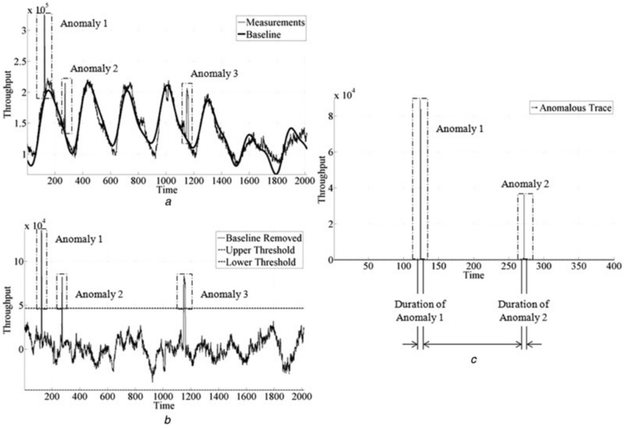


Fig 2: Anomaly extraction process implemented by Bandara *et al.* (2014)

- Due to a huge amount of data to be processed, the algorithm computes an anomaly arrival time only with the **precision of five minutes**.

The Hill estimator for error contaminated data

- We assume that we observe $Y_i = X_i + \varepsilon_i,\ 1 \leqslant i \leqslant n$, where $\bar{F}_X \in RV_{-\alpha}$, $\{\varepsilon_i\}$ are i.i.d. **random errors** following F_ε, and independent of the $\{X_i\}$.

- For example, in the case of the internet traffic data, $Y_i = X_i + \varepsilon_i$ where
  - the Y_i are observations,
  - the X_i are "true" interarrival time,
  - the errors ε_i are distributed on $[-1, 1]$. (We use 5 minutes as a unit lag.)

Questions

- The Hill estimator for the **observations** $Y_i = X_i + \varepsilon_i$ is defined as

$$\widehat{H}_{k,n} := \frac{1}{k} \sum_{i=1}^{k-1} \log \frac{Y_{(i)}}{Y_{(k)}}.$$

In our context, $\widehat{H}_{k,n}$ is the estimator that can be actually used.

- The question is, what would be suitable **assumptions on the measurement errors** ε_i to make $\widehat{H}_{k,n}$ **consistent** or **asymptotically normal**?

Consistency

- The **consistency** of the Hill estimator has been studied for i.i.d. and dependent heavy tailed data, see e.g. Mason (1982, The Annals of Probability), Hsing (1991, The Annals of Statistics), Davis and Resnick (1996, The Annals of Applied Probability), Resnick and Stărică (1998, The Annals of Applied Probability).

- Suppose that the X_i are i.i.d. with $\bar{F}_X \in RV_{-\alpha}$. Then

$$H_{k,n} := \frac{1}{k}\sum_{i=1}^{k-1} \log \frac{X_{(i)}}{X_{(k)}} \xrightarrow{P} \frac{1}{\alpha},$$

as

$$n \to \infty,\ k \to \infty,\ \frac{k}{n} \to 0. \tag{1}$$

Consistency

Theorem 1

Suppose that the X_i are i.i.d. random variables with $\bar{F}_X \in RV_{-\alpha}$, and ε_i are i.i.d. $\sim \bar{F}_\varepsilon$, which satisfies

$$P(|\varepsilon| > x) = o(P(X > x)), \text{ as } x \to \infty,$$

and independent of the $\{X_i\}$. Then **any estimator** of α computed from the $Y_i = X_i + \varepsilon_i$ is consistent as (1), if its counterpart computed from the unobservable X_i is consistent.

• **The measurement error ε has a lighter tail than X**; $\bar{F}_Y \in RV_{-\alpha}$ as well.

• This is no longer trivial if the X_i are dependent. (New theorems are established)

Asymptotic normality

- To obtain the **asymptotic normality** of the Hill estimator centered by the exponent α^{-1}, **second–order regular variation** needs to be assumed, see e.g. Hauesler and Teugels (1985, The Annals of Statistics), Resnick and Stărică (1997, Advances in Applied Probability), Resnick and Stărică (1997, Stochastic models).

- $2RV(-\alpha, \rho)$; there exists a positive function $g \in RV_\rho$ such that $g(t) \to 0$, as $t \to \infty$, and for $\alpha > 0$, $\rho \leqslant 0$, $K \neq 0$.

$$\lim_{t\to\infty} \frac{1}{g(t)} \left(\frac{\bar{F}_X(tx)}{\bar{F}_X(t)} - x^{-\alpha} \right) = H(x) := Kx^{-\alpha} \frac{x^\rho - 1}{\rho}, \quad x > 0.$$

- Suppose that the X_i are i.i.d. with $\bar{F}_X \in 2RV(-\alpha, \rho)$. Then

$$\sqrt{k}\left(H_{k,n} - \frac{1}{\alpha} \right) \Rightarrow N(0, 1/\alpha^2),$$

as

$$\sqrt{k} g(b(n/k)) \to 0.$$

Asymptotic normality - 2RV

• To obtain the centering α^{-1}, we consider conditions on F_X; **2RV** case and **Pareto** case.

Theorem 2 - 2RV

Suppose that $\bar{F}_X \in 2RV(\alpha, \rho)$, and ε_i are i.i.d. $\sim \bar{F}_\varepsilon$, which satisfies

$$P(|\varepsilon| > x) = o(x^{-\beta}), \text{ as } x \to \infty,$$

for some $\beta > \alpha - \rho$. In addition, the sequence $k = k(n)$ satisfies $\sqrt{k}g(b(n/k)) \to 0$ if $\rho > -1$ and $\sqrt{k}/b(n/k) \to 0$ if $\rho \leqslant -1$. Then, for $\alpha \geqslant 1$

$$\sqrt{k}\left(\widehat{H}_{k,n} - \frac{1}{\alpha}\right) \Rightarrow N(0, 1/\alpha^2).$$

• **The measurement error ε has a lighter tail than some power function. We need an additional restriction on the k.**

Asymptotic normality - Pareto

Theorem 2 - Pareto

Suppose that $\bar{F}_X(x) = x^{-\alpha}$, $x \geqslant 1$, and ε_i are i.i.d. $\sim \bar{F}_\varepsilon$, which satisfies

$$P(|\varepsilon| > x) = o(x^{-\beta}), \text{ as } x \to \infty,$$

for some $\beta > \alpha + 1$. In addition, the sequence $k = k(n)$ satisfies $\sqrt{k}/b(n/k) \to 0$. Then, for $\alpha \geqslant 1$

$$\sqrt{k}\left(\widehat{H}_{k,n} - \frac{1}{\alpha}\right) \Rightarrow N(0, 1/\alpha^2).$$

Finite sample behaviors

- Investigate **the impact of errors on the Hill estimator in finite samples**.

- We generate observations $Y_i = X_i + \varepsilon_i, \; i = 1, 2, \ldots, N, \; N = 500, 2000.$

- We use two models for the X_i, both having true tail index $\alpha = 2$.

Model 1 [Pareto] The X_i are i.i.d. random variables, which follow a Pareto distribution with $\alpha = 2$, $P(X_i > x) = x^{-2}, \quad x \geqslant 1$.

Model 2 [2RV] The X_i are i.i.d. random variables, which follow the Hall/Weiss class with $\alpha = 2$ and $\rho = -5$,
$P(X_i > x) = x^{-2}(1 + x^{-5})/2, \quad x \geqslant 1$.

Finite sample behaviors

- We consider four different distributions for the errors ε_i. For each of them, $P(|\varepsilon| > x) = o(x^{-\beta})$, for some $7 < \beta < 8$.
  - a **normal** distribution with mean 0 and standard deviation σ
  - a scaled t-distribution with 8 degrees of freedom (**scaled** t_8)
  - a generalized Pareto distribution (**GPD**),

    $$P(|\varepsilon| > z) = (1 + \xi(z - \mu)/\sigma)^{-1/\xi},$$

    with location $\mu = 0$, shape $\xi = 1/8$, and scale σ_{GPD}.
  - a **uniform** distribution on the interval $[-a, a]$, $a > 0$.
- For each model/error pair, we have 1000 replications.

Finite sample behaviors

- The asymptotic level $1 - p$ confidence interval for α^{-1} implied by the Theorem 2 is

$$\left(\frac{1}{\hat{\alpha}} - z_{p/2} \frac{1}{\hat{\alpha}\sqrt{k}}, \ \frac{1}{\hat{\alpha}} + z_{p/2} \frac{1}{\hat{\alpha}\sqrt{k}} \right),$$

where $\hat{\alpha}^{-1} = \widehat{H}_{k,n}$, and z_q is the upper quantile of the standard normal distribution defined by $\Phi(z_q) = 1 - q$.

- We investigate **the impact of these errors on the empirical coverage probability of the interval**.

Finite sample behaviors

- We examined four methods based on different underlying ideas of **selecting a data–driven cut–off k**.
  - **Hall** : It uses a **bootstrap procedure** to find the k which **minimizes the AMSE**, introduced by Hall (1990, Journal of Multivariate Analysis).
  - **MAD** : It is based on **minimizing a penalty function** of the distance between the observed quantile and the fitted Pareto type tail. The **mean absolute distance** is used for the penalty and it is introduced by Danielsson *et al.* (2016).
  - **KS** : The underlying idea is the same, but the **supremum of the absolute distance** is used for the penalty. It is introduced by Danielsson *et al.* (2016).
  - **Eye** : It is an Eye–Ball technique trying to **find a stable portion of the Hill plot** and obtain the k at which a considerable drop in the variance occurs, as k increases. Danielsson *et al.* (2016).

Result - Pareto

Method	Error Type	Error SD/Model SD Ratio 0	0.01	0.02	0.05	0.1	0.2	0.3
Hall	Normal	88.9	87.6	88.4	88.9	83.8	77.5	71.2
	scaled t_8	88.7	88.0	88.6	88.9	83.2	77.6	68.9
	GPD	89.4	88.9	88.8	88.7	83.7	76.9	72.7
	Uniform	89.1	88.2	88.3	87.9	80.1	73.1	61.3
MAD	Normal	97.0	97.4	96.8	97.6	96.8	97.4	96.2
	scaled t_8	97.1	97.2	97.4	97.8	97.2	97.2	97.2
KS	Normal	83.4	82.2	84.0	81.2	77.2	75.0	67.6
	scaled t_8	83.6	83.6	83.5	84.2	81.7	77.4	71.9
Eye	Normal	95.3	95.1	94.8	95.2	94.8	93.2	90.5
	scaled t_8	95.3	95.4	95.5	95.3	93.5	92.7	88.2

Table 1: Proportion (in percent) of the approximate 95% confidence intervals including $1/\alpha$, for $n = \mathbf{500}$ and the **Pareto** model. The target coverage is 95 %.

Result - 2RV

Method	Error Type	Error SD/Model SD Ratio 0	0.01	0.02	0.05	0.1	0.2	0.3
Hall	Normal	75.3	75.6	75.0	12.9	8.8	37.2	34.7
	scaled t_8	75.8	75.3	74.4	29.0	0.8	29.1	37.4
	GPD	75.8	76.2	72.5	28.3	1.9	17.6	38.3
	Uniform	75.6	75.2	74.0	33.8	26.4	35.0	30.3
MAD	Normal	18.7	18.2	18.5	16.3	7.5	36.6	70.8
	scaled t_8	18.7	18.9	18.8	17.0	8.5	21.0	51.7
KS	Normal	66.6	66.6	67.0	66.4	56.7	52.5	53.0
	scaled t_8	66.6	66.9	66.8	67.0	66.3	53.9	53.0
Eye	Normal	93.6	93.9	93.4	93.7	92.6	88.7	77.8
	scaled t_8	93.6	93.9	94.6	93.9	91.9	91.1	83.3

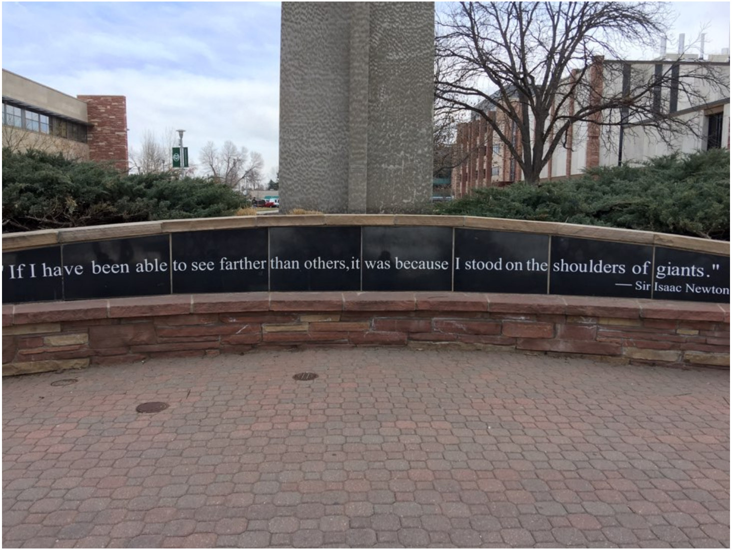
Table 2: Proportion (in percent) of the approximate 95% confidence intervals including $1/\alpha$, for $\mathbf{n} = \mathbf{500}$ and the **2RV** model. The target coverage is 95 percent.

Conclusions

- We derive broadly applicable **conditions on errors** under which the Hill estimator is **consistent** or **asymptotically normal**.

- From the simulation, **the impact of the errors is robust** for small ratios.

- There is no clear evidence that the coverage probability depends on the error distributions.


"If I have been able to see farther than others, it was because I stood on the shoulders of giants."
— Sir Isaac Newton

Thank you!