ON THE IMPACT OF APPROXIMATION ERRORS ON EXTREME QUANTILE ESTIMATION WITH APPLICATIONS TO FUNCTIONAL DATA ANALYSIS

Based on collaboration with Pauliina Ilmonen, Lauri Viitasaari, Valentin Garino and Benny Avelin

https://doi.org/10.48550/arXiv.2307.03581

Jaakko Pere

7th May, 2025

Dep. of Mathematics and Statistics, University of Helsinki

Agenda of the presentation

Univariate Extreme Value Theory

Impact of Approximation Errors

Extreme Quantile Estimation for L^p-Norms

Table of Contents

What is Extreme Value Theory?

Extreme value theory is concerned about inference of rare events.

What is Extreme Value Theory?

Extreme value theory is concerned about inference of rare events.

- Extreme quantile estimation
- Tail probability estimation
- Estimation of the endpoint of a given distribution

Maximum Domain of Attraction

Definition

Let $Y_1, \ldots, Y_n$ be i.i.d. observations of a random variable Y. If there exist sequences $a_n > 0$ and $b_n \in \mathbb{R}$, and a random variable G with a nondegenerate distribution such that

$$\frac{\max(Y_1, \ldots, Y_n) - b_n}{a_n} \xrightarrow{\mathcal{D}} G, \quad n \to \infty,$$

we say that Y belongs to the maximum domain of attraction of G, and denote $Y \in \mathrm{MDA}(G)$.

Extreme Value Index

Theorem (Fisher and Tippett 1928; Gnedenko 1943)

Up to location and scale, the distribution of $G = G_\gamma$ is characterized by the parameter γ, called the extreme value index. That is, the distribution of G_γ is of the type

$$F_{G_\gamma}(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right), \quad 1+\gamma x > 0 & \text{if} \quad \gamma \neq 0, \\ \exp\left(-e^{-x}\right), \quad x \in \mathbb{R} & \text{if} \quad \gamma = 0. \end{cases}$$

Extreme Value Index

Theorem (Fisher and Tippett 1928; Gnedenko 1943)

Up to location and scale, the distribution of $G = G_\gamma$ is characterized by the parameter γ, called the extreme value index. That is, the distribution of G_γ is of the type

$$F_{G_\gamma}(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right), \quad 1+\gamma x > 0 & \text{if} \quad \gamma \neq 0, \\ \exp\left(-e^{-x}\right), \quad x \in \mathbb{R} & \text{if} \quad \gamma = 0. \end{cases}$$

In the case $\gamma > 0$ the type of G_γ is Fréchet,

$$\Phi_\gamma(x) = \begin{cases} 0, & x \leq 0 \\ \exp\left(-x^{-1/\gamma}\right), & x > 0. \end{cases}$$

Tail Quantile Function

Define tail quantile function corresponding to distribution F by

$$U(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right), \quad t > 1,$$

where we denote left-continuous inverse of a nondecreasing function by $f^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : f(x) \geq y\}$.

Tail Quantile Function

Define tail quantile function corresponding to distribution F by

$$U(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right), \quad t > 1,$$

where we denote left-continuous inverse of a nondecreasing function by $f^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : f(x) \geq y\}$.

That is, $U(1/p)$ is the $(1-p)$-quantile.

Definition (Regular variation)

A Lebesgue measurable function $f : \mathbb{R}^+ \to \mathbb{R}$ that is eventually positive is regularly varying with index $\alpha \in \mathbb{R}$ if for all $x > 0$,

$$\lim_{t \to \infty} \frac{f(tx)}{f(t)} = x^\alpha.$$

Then we denote $f \in RV_\alpha$. Furthermore, we say that a function f is slowly varying if $f \in RV_0$.

Definition (Regular variation)

A Lebesgue measurable function $f : \mathbb{R}^+ \to \mathbb{R}$ that is eventually positive is regularly varying with index $\alpha \in \mathbb{R}$ if for all $x > 0$,

$$\lim_{t \to \infty} \frac{f(tx)}{f(t)} = x^\alpha .$$

Then we denote $f \in RV_\alpha$. Furthermore, we say that a function f is slowly varying if $f \in RV_0$.

Intuition:

$$f \in \mathrm{RV}_\alpha \iff f(x) = L(x) x^\alpha, \quad L \in \mathrm{RV}_0 .$$

We also have

$$\lim_{x \to \infty} x^{-\varepsilon} L(x) = 0, \quad \forall \varepsilon > 0.$$

Construction of an Extreme Quantile Estimator

Theorem ((Gnedenko 1943; de Haan 1970))

Let $\gamma > 0$. *We have*

$$Y \in \mathrm{MDA}\,(G_\gamma) \iff 1 - F \in RV_{-1/\gamma} \iff U \in RV_\gamma.$$

Construction of an Extreme Quantile Estimator

Theorem ((Gnedenko 1943; de Haan 1970))

Let $\gamma > 0$. We have

$$Y \in \mathrm{MDA}\left(G_\gamma\right) \iff 1 - F \in RV_{-1/\gamma} \iff U \in RV_\gamma .$$

Choose $t = n/k$ and $x = k/(np)$ to get the approximation

$$U\left(\frac{1}{p}\right) \approx U\left(\frac{n}{k}\right)\left(\frac{k}{np}\right)^{\gamma} .$$

Extreme Quantile Estimation

Suppose $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ is an i.i.d. sample of $Y \in \mathrm{MDA}\,(G_\gamma)$, $\gamma > 0$. Denote order statistics corresponding to the sample $\boldsymbol{Y}$ by $\boldsymbol{Y}_{1,n} \leq \cdots \leq \boldsymbol{Y}_{n,n}$. Then an estimator for the extreme $(1-p)$-quantile $x_p = U\,(1/p)$ can be given as

$$\hat{x}_p\,(\boldsymbol{Y}) = \boldsymbol{Y}_{n-k,n} \left(\frac{k}{np}\right)^{\hat{\gamma}(\boldsymbol{Y})},$$

where $\hat{\gamma}$ is an estimator for the extreme value index γ.

The Hill Estimator (Hill 1975; Mason 1982)

Suppose $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ is an i.i.d. sample of $Y \in \mathrm{MDA}\,(G_\gamma)$, $\gamma > 0$. The Hill estimator is defined as

$$\hat{\gamma}_H\,(\boldsymbol{Y}) = \frac{1}{k}\sum_{i=0}^{k-1} \ln\left(\frac{\boldsymbol{Y}_{n-i,n}}{\boldsymbol{Y}_{n-k,n}}\right).$$

The Hill Estimator (Hill 1975; Mason 1982)

Suppose $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ is an i.i.d. sample of $Y \in \mathrm{MDA}\,(G_\gamma)$, $\gamma > 0$. The Hill estimator is defined as

$$\hat{\gamma}_H\,(\boldsymbol{Y}) = \frac{1}{k}\sum_{i=0}^{k-1} \ln\left(\frac{\boldsymbol{Y}_{n-i,n}}{\boldsymbol{Y}_{n-k,n}}\right).$$

If additionally as $n \to \infty$, $k = k_n \to \infty$, $k/n \to 0$, then

$$\hat{\gamma}_H\,(\boldsymbol{Y}) \xrightarrow{\mathbb{P}} \gamma, \quad n \to \infty.$$

Table of Contents

The General Framework

- What if instead of the sample $\boldsymbol{Y}$, only approximations $\hat{\boldsymbol{Y}} = (\hat{Y}_1, \dots, \hat{Y}_n)$ are available?

The General Framework

- What if instead of the sample $\boldsymbol{Y}$, only approximations $\hat{\boldsymbol{Y}} = (\hat{Y}_1, \ldots, \hat{Y}_n)$ are available?
- How the approximation error affects the asymptotics?

The General Framework

- What if instead of the sample $\boldsymbol{Y}$, only approximations $\hat{\boldsymbol{Y}} = (\hat{Y}_1, \ldots, \hat{Y}_n)$ are available?
- How the approximation error affects the asymptotics?
- $\Rightarrow$ Useful approach in multivariate and infinite dimensional settings:
  - Let $X \in \mathbb{S}$ be a random object, where, e.g., $\mathbb{S} = \mathbb{R}^d$ or $\mathbb{S} = L^p([0,1]^d)$.
  - Let $g : \mathbb{S} \to \mathbb{R}$ be some suitable map.
  - Apply extreme value theory to $g(X)$.

Approximated L^p-Norms

- Let $X \in L^p\left([0,1]^d\right)$, and let $X_1, \ldots, X_n$ be i.i.d. copies of X.
- We wish to estimate extreme value index and extreme quantiles corresponding to $\|X\|_p \in \mathrm{MDA}\left(G_\gamma\right)$, $\gamma > 0$.

Approximated L^p-Norms

- Let $X \in L^p\left([0,1]^d\right)$, and let $X_1, \ldots, X_n$ be i.i.d. copies of X.
- We wish to estimate extreme value index and extreme quantiles corresponding to $\|X\|_p \in \mathrm{MDA}\left(G_\gamma\right)$, $\gamma > 0$.
- In practice we never observe $X_1, \ldots, X_n$.

Approximated L^p-Norms

- Let $X \in L^p\left([0,1]^d\right)$, and let $X_1, \ldots, X_n$ be i.i.d. copies of X.
- We wish to estimate extreme value index and extreme quantiles corresponding to $\|X\|_p \in \mathrm{MDA}\left(G_\gamma\right)$, $\gamma > 0$.
- In practice we never observe $X_1, \ldots, X_n$.
- Approximate norms with Riemann sums or Monte Carlo integration.
- Use approximated norms $\hat{Y}_i$ in the estimation.

Approximated L^p-Norms

- Let $X \in L^p\left([0,1]^d\right)$, and let $X_1, \ldots, X_n$ be i.i.d. copies of X.
- We wish to estimate extreme value index and extreme quantiles corresponding to $\|X\|_p \in \mathrm{MDA}\left(G_\gamma\right)$, $\gamma > 0$.
- In practice we never observe $X_1, \ldots, X_n$.
- Approximate norms with Riemann sums or Monte Carlo integration.
- Use approximated norms $\hat{Y}_i$ in the estimation.
- As the estimator of the extreme value index we choose the Hill estimator

$$\hat{\gamma}(\hat{\boldsymbol{Y}}) = \frac{1}{k} \sum_{i=0}^{k-1} \ln\left(\frac{\hat{\boldsymbol{Y}}_{n-i,n}}{\hat{\boldsymbol{Y}}_{n-k,n}}\right).$$

Draft of the Main Result

Let $\gamma > 0$. Let $Y_1, \ldots, Y_n$ be i.i.d. copies of $Y \in \mathrm{MDA}\,(G_\gamma)$ and $\hat{\boldsymbol{Y}} = (\hat{Y}_1, \ldots, \hat{Y}_n)$ the corresponding approximations. Denote errors by $E_i = \left|\hat{Y}_i - Y_i\right|$. If

$$\sqrt{k}\,\frac{\boldsymbol{E}_{n,n}}{U_Y\,(n/k)} \xrightarrow{\mathbb{P}} 0, \quad n \to \infty,$$

then

$$\sqrt{k}\left(\hat{\gamma}(\hat{\boldsymbol{Y}}) - \gamma\right) \quad \text{and} \quad \frac{\sqrt{k}}{\ln\,(k/(np))}\left(\frac{\hat{x}_p(\hat{\boldsymbol{Y}})}{U(1/p)} - 1\right)$$

are asymptotically normally distributed under the standard assumptions (second-order condition, rate for $p = p_n$, $k = k_n \to \infty$, $k/n \to 0$, as $n \to \infty$).

Table of Contents

References I

de Haan, Laurens (1970). "On Regular Variation and Its Application to Weak Convergence of Sample Extremes". PhD thesis. Universiteit van Amsterdam.

Fisher, Ronald Aylmer and Leonard Henry Caleb Tippett (1928). "Limiting forms of the frequency distribution of the largest or smallest member of a sample". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 24.2, pp. 180–190. DOI: `https://doi.org/10.1017/S0305004100015681`.

Gnedenko, Boris (1943). "Sur La Distribution Limite Du Terme Maximum D'Une Série Aléatoire". In: *Annals of Mathematics* 44.3, pp. 423–453. DOI: `https://doi.org/10.2307/1968974`.

Hill, Bruce Marvin (1975). "A Simple General Approach to Inference About the Tail of a Distribution". In: *The Annals of Statistics* 3.5, pp. 1163–1174. DOI: `https://doi.org/10.1214/aos/1176343247`.

References II

Mason, David M. (1982). “Laws of Large Numbers for Sums of Extreme Values”. In: *The Annals of Probability* 10.3, pp. 754–764. DOI: https://doi.org/10.1214/aop/1176993783.