BOCCONI UNIVERSITY
DEPARTMENT OF DECISION SCIENCES
PhD in Statistics and Computer Science

UNIVERSITY OF PADUA
DEPARTMENT OF STATISTICAL SCIENCES

Empirical Bayes inference for the Peaks Over a Threshold method

Pietro Scanzi   Simone Padoan   Nicola Sartori

1 Introduction

Extreme Value Theory (EVT) is a branch of statistical theory aimed at quantifying the stochastic behaviour of a process at unusually large (or small) levels. Since the main goal of EVT is the prediction of events that are expected to fall far beyond the observable data $x_1, \ldots, x_n$, the probabilistic models and statistical tools that it provides are asymptotically motivated. From a statistical point of view, such events are represented by the quantiles corresponding to an exceeding probability $p \leq 1/n$, of the unknown unconditional distribution F that has generated the data.

2 EVT tools

This section refers to [2] and [3].

Let $X_1, X_2, \ldots X_n \overset{iid}{\sim} F$ unknown over $\mathbb{R}$ and let its upper end-point $x^+ = \sup\left\{x \in \bar{\mathbb{R}} : F(x) < 1\right\}$. For $m \leq n$, we consider the **maximum of block size m:**

$$M_m := \max\{X_1, \ldots, X_m\} \qquad \text{with distribution}$$
$$F^m(x) := \mathbb{P}(M_m \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_m \leq x) \overset{iid}{=} (F(x))^m .$$

It's easy to derive that, as $n \to \infty$, which implies $m \to \infty$,

$$M_m \overset{p}{\to} x^+ \qquad \Longrightarrow \qquad F^m(x) \Rightarrow \delta_{x^+}(x) \ \ \forall\, x \in \left(-\infty, x^+\right] .$$

Now, we can fix a high threshold $u \in \mathbb{R}$ such that only k among the X_i's overcome it, i.e. $X_{(n-k)} \leq u < X_{(n-k+1)}$. We define, for every $j = 1, \ldots, k$, the **u-threshold exceedances:**

$$Y_j := \left(X_j - u\right) \mid \left(X_j > u\right) \qquad \text{with distribution}$$
$$F_u(y) := \mathbb{P}\left(Y_j \leq y\right) = \mathbb{P}\left(X_j - u \leq y \,|\, X_j > u\right) = \frac{F(u+y)}{1-F(u)}, \quad y > 0 .$$

Theorem: If there exist sequences of scale $(a_m) \in (0, +\infty)$ and location $(b_m) \in \mathbb{R}$ constants such that

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) = \lim_{m\to\infty} F^m\left(a_m x + b_m\right) = G_\gamma\left(x\right) \tag{1}$$

for a non-degenerate distribution function G_γ, then $F \in \mathcal{D}\left(G_\gamma\right)$ and G_γ is a member of the **Generalized Extreme Value (GEV)** family where

$$G_\gamma(x) = \exp\left\{-\left(1+\gamma\, x\right)^{-\frac{1}{\gamma}}\right\} \qquad \text{defined on} \qquad \{x \in \mathbb{R} : 1 + \gamma\, x > 0\} .$$

Moreover, for large u, the conditional distribution function F_u is approximated by

$$H(y) = 1 - \left(1 + \gamma \frac{y}{\tilde{\sigma}_u}\right)^{-\frac{1}{\gamma}} \qquad \text{defined on} \qquad \left\{y > 0 : 1 + \gamma \frac{y}{\tilde{\sigma}_u} > 0\right\} ,$$

with $\tilde{\sigma}_u = 1 + \gamma\, u$. H belongs to the **Generalized Pareto (GP)** family.

As a consequence,

- $F^m(x) \approx G_\gamma\left(\frac{x-b_m}{a_m}\right)$ and we say $M_m \mathrel{\dot\sim} GEV\left(\gamma, b_m, a_m\right)$ for large m;
- $Y_1, \ldots, Y_k \mathrel{\dot\sim} GP\left(\gamma, 1 + \gamma\, u\right)$ for large u;
- the parameters of the GEV and GP models can be estimated from the observed block maxima and threshold exceedances by maximum likelihood;
- when $\gamma^0 > -0.5$ we can rely on the usual likelihood asymptotics;

3 Censored Peaks Over a Threshold (CPOT) method

This section refers to [1].

We have $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F \in \mathcal{D}\left(G_\gamma\right)$ sample units where k of them exceed some high threshold $u \in \mathbb{R}$. Here, the **peaks over the threshold** are $X_{(n-k+1)}, \ldots, X_{(n)}$ and we set $s = n/k$. Accordingly, we can split units into k blocks of size s such that each block contains only one peak and $s - 1$ non-exceeding units. Under this blocking scheme $M_s = X_{(n-j+1)}$ for some $j \in \{1, \ldots, k\}$ and the k block maxima coincide with the k peaks over the threshold. Moreover,

$$s = \frac{n}{k} \to \infty \ \text{ as } \ n \to \infty \quad \text{just like} \quad m \to \infty \ \text{ as } \ n \to \infty$$

and the asymptotic approximation (1) still holds with s in place of m, leading to $X_{(n-k+1)}, \ldots, X_{(n)} \mathrel{\dot\sim} GEV\left(\gamma, b_s, a_s\right)$ for large s. It is also possible to derive a suitable approximation for the tail of F, i.e.

$$F(x) \approx \exp\left\{-\left(1+\gamma\frac{x-b_s}{a_s}\right)^{-\frac{1}{\gamma}}\right\}^{\frac{1}{s}} = \left[G_\gamma\left(\frac{x-b_s}{a_s}\right)\right]^{\frac{1}{s}}, \qquad \text{for large } x, s . \tag{2}$$

Exploting this approximation, we can obtain an expression for the **extreme quantile** of level $1 - p$ with $p \leq 1/n$, i.e.

$$x_p = F^{-1}(1-p) \approx b_s + a_s \frac{(s\,p)^{-\gamma} - 1}{\gamma} . \tag{3}$$

We have a sample $\boldsymbol{x} = (x_1, \ldots, x_n)^t$ of data and we define a high threshold $u = x_{(n-k)}$ for $\frac{k}{n} = \frac{1}{s} \in \{0.10, 0.05, 0.01\}$. Thanks to (2), we can assume for the peaks $X_{(n-k+1)}, \ldots, X_{(n)} \mathrel{\dot\sim} GEV^{\frac{1}{s}}\left(\gamma, b_s, a_s\right)$ for large s and in order not to waste information, we consider $x_{(1)}, \ldots, x_{(n-k)} = u$ as **left-censored** by u. Finally, we can define the **Censored POT** log-likelihood for the parameter $\boldsymbol{\theta} = (\gamma, b_s, a_s) \in \Theta = (-1, +\infty) \times \mathbb{R} \times (0, +\infty)$ as

$$l(\gamma, b_s, a_s\,; \boldsymbol{x}) = \sum_{i=1}^{n} \log \mathcal{L}\left(\gamma, b_s, a_s\,; x_i\right) , \tag{4}$$

where, for $i = 1, \ldots, n$,

$$\mathcal{L}\left(\gamma, b_s, a_s\,; x_i\right) = \begin{cases} \left[G_\gamma\left(\frac{u-b_s}{a_s}\right)\right]^{\frac{1}{s}} & \text{if } x_i \leq u , \\ \frac{d}{d\,x}\left\{\left[G_\gamma\left(\frac{x-b_s}{a_s}\right)\right]^{\frac{1}{s}}\right\}\Big|_{x=x_i} & \text{if } x_i > u . \end{cases}$$

The MLE $\hat{\boldsymbol{\theta}}_n = \left(\hat{\gamma}_n, \hat{b}_{s,n}, \hat{a}_{s,n}\right)$ can be found maximizing numerically (4).

4 Empirical Bayes CPOT

This section refers to [6].

The location b_s and scale a_s constants increase as $s \to \infty$ and so do the **true parameter values** $\boldsymbol{\theta}^0 = \left(\gamma^0, b_s^0, a_s^0\right)$ and their **MLEs** $\hat{\boldsymbol{\theta}}_n = \left(\hat{\gamma}_n, \hat{b}_{s,n}, \hat{a}_{s,n}\right)$. In a Bayesian context, an empirical Bayes approach is therefore necessary in order to avoid infinite and mathematically incorrect priors.

We assume for $\boldsymbol{\theta}$ a **data-dependent prior** density with independent components, i.e.

$$\pi\left(\gamma, b_s, a_s\right) = \pi\left(\gamma\right) \cdot \pi\left(b_s\right) \cdot \pi\left(a_s\right) , \qquad \text{where}$$
$$\pi\left(\gamma\right) = \frac{t_1\left(\gamma\right)}{1 - T_1\left(-1\right)} , \qquad \gamma \in (-1, +\infty) ;$$
$$\pi\left(b_s\right) \propto \frac{1}{\hat{a}_{s,n}^2} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{b_s - \hat{b}_{s,n}}{\hat{a}_{s,n}}\right)^2\right\} , \qquad b_s \in \mathbb{R} ;$$
$$\pi\left(a_s\right) \propto \frac{1}{\hat{a}_{s,n}^2} \exp\left\{-\frac{1}{\hat{a}_{s,n}} a_s\right\} , \qquad a_s \in (0, +\infty) .$$

Under some mild conditions on the data generating process (satisfied by the CPOT model) and on the prior distribution (satisfied by the aforementioned prior), the posterior distribution of $\boldsymbol{\theta}$ given the data $\boldsymbol{x}$ provides consistent estimation of the unknown true parameter $\boldsymbol{\theta}^0$ and is asymptotically Gaussian as $s \to \infty$.

5 An adaptive Metropolis-Hastings (AMH) algorithm

This section refers to [4] and [5].

We sample from $\pi\left(\boldsymbol{\theta} \,|\, \boldsymbol{x}\right)$ through an **adaptive Gaussian random walk Metropolis-Hastings algorithm**. This algorithm is refined in a way that it adapts the scaling parameter κ and the covariance matrix Σ of the **proposal distribution** $\mathcal{N}_3\left(\boldsymbol{\theta}^{(i)}, \kappa^{(i)}\,\Sigma^{(i)}\right)$ at each iteration $i + 1$ with the objective of reaching a fixed optimal **overall acceptance probability (OAP)** $\eta^* = 0.234$.

Algorithm 1: Adaptive Gaussian Random-Walk Metropolis-Hastings

Initialize: Set R, $\boldsymbol{\theta}^{(0)}$, $\kappa^{(0)}$ and $\Sigma^{(0)}$;
for $i = 1$ **to** R **do**
  draw proposal $\boldsymbol{\theta}^* \sim \mathcal{N}_3\left(\boldsymbol{\theta}^{(i)}, \kappa^{(i)}\,\Sigma^{(i)}\right)$;
  compute acceptance probability $\eta^{(i)} = \min\left(\frac{\pi(\boldsymbol{\theta}^*)\mathcal{L}(\boldsymbol{\theta}^*; \boldsymbol{x})}{\pi(\boldsymbol{\theta}^{(i)})\mathcal{L}(\boldsymbol{\theta}^{(i)}; \boldsymbol{x})}, 1\right)$;
  draw $U \sim \mathcal{U}(0,1)$;
  if $\eta^{(i)} > U$ **then**
    set $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^*$;
  else
    set $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)}$;
  update $\Sigma^{(i+1)} = \frac{1}{i-1}\sum_{s=1}^{i}\left(\boldsymbol{\theta}^{(s)} - \bar{\boldsymbol{\theta}}^{(i)}\right) \cdot \left(\boldsymbol{\theta}^{(s)} - \bar{\boldsymbol{\theta}}^{(i)}\right)^t + \frac{1}{i}\mathbb{I}_3$;
  update $\kappa^{(i+1)} = \exp\left\{\log\left(\kappa^{(i)}\right) + a\left(\eta^*\right)\frac{\eta^{(i)} - \eta^*}{\max\left\{200, \frac{i}{3}\right\}}\right\}$.

6 Simulation study

We construct a simulation study to test the frequentist accuracy of credible intervals based on the empirical Bayes CPOT method. We are interested in the **marginal posterior distribution** of γ, b_s, a_s and the extreme quantile x_p (3) for $p = 1/n$. We compute the **quantile**, **Gaussian approximation based** and **HPD** 95% credible intervals. To this aim, we draw $N = 1000$ independent random samples of increasing size n from 9 distributions pertaining to the max-domain of attraction of the GEV and we calculate the **coverage probabilities** over the N iterations. We set the posterior sample size $R = 50000$, the burn-in to 10000 and study 4 different simulation scenarios:

1. small extreme sample: $n = 800$, $k = 20 \ \Rightarrow\ s = 40$;
2. medium extreme sample: $n = 1800$, $k = 30 \ \Rightarrow\ s = 60$;
3. large extreme sample: $n = 5450$, $k = 50 \ \Rightarrow\ s = 109$;
4. "big-data" extreme sample: $n = 23400$, $k = 100 \ \Rightarrow\ s = 234$.

We obtain an overall accurate performance of the empirical Bayes CPOT method. Generally, the empirical coverages reach 95% at least once for every considered quantity, credible interval type, data generating process and sample size. Moreover, coverage probabilities tend to improve as s grows. Best results are provided by the quantile intervals, followed by the Gaussian and then by the HPD.

References

[1] Boris Beranger, Simone A Padoan, and Scott A Sisson. Estimation and uncertainty quantification for extreme quantile regions. *Extremes*, 24(2):349–375, 2021.

[2] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.

[3] Laurens De Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2006.

[4] Paul H Garthwaite, Yanan Fan, and Scott A Sisson. Adaptive optimal scaling of Metropolis–Hastings algorithms using the Robbins–Monro process. *Communications in Statistics-Theory and Methods*, 45(17):5098–5111, 2016.

[5] Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242, 2001.

[6] Simone A Padoan and Stefano Rizzelli. Empirical Bayes inference for the block maxima method. *arXiv preprint arXiv:2204.04981*, 2022.