From classics to punk: almost fifthy years of Extreme Value Analysis

Stefano Rizzelli

University of Padova

Asiago - September 13, 2023

Overview

1. Inference based on empirical quantiles
2. Likelihood based Peaks-over-Threshold methods: maximum likelihood and Bayesian inference *
3. From Peaks-over-Threshold to Block Maxima **
4. Emprical Bayes inference and a real data illustration **

* joint work with Clément Dombry and Simone Padoan

** joint work with Simone Padoan

Tail inference: preliminaries

Drees (1998): «In many statistical applications it is necessary to make inferences about the tail of a distribution, where little or no data is available. For example, if dike is projected in order to protect a costal line, then usually it will be higher than any flood recorded yet. Hence the estimation of the minimal height that ensures that the probability of being flooded in a particular year is less than a given small value requires an extrapolation of the underlying distribution beyond the observed data.»

Tail inference: preliminaries

We restrict ourselves to data which are realisations of independent and identically distributed (i.i.d.) random variables.

Tail inference: preliminaries

We restrict ourselves to data which are realisations of independent and identically distributed (i.i.d.) random variables.

In order to make inference on the tail beyond the data, we need assumptions:

- on the distribution F of the observed random variable near its end-point
- or, equivalently, on its inverse function

$$F^{-1}(1-1/t) := \inf\{x : 1 - F(x) \geq 1/t\},$$

corresponding to the $1-1/t$-quantile of F, for large t.

A gentle start: the Pareto distribution

A random variable X follows the (standard) Pareto distribution if

$$\mathbb{P}(X > x) = x^{-1/\gamma}, \quad x > 1,$$

for some $\gamma > 0$.

A gentle start: the Pareto distribution

A random variable X follows the (standard) Pareto distribution if

$$\mathbb{P}(X > x) = x^{-1/\gamma}, \quad x > 1,$$

for some $\gamma > 0$.

The shape parameter γ, whose reciprocal is called tail index, fully characterises the distribution and, in particular, its tail features.

A gentle start: the Pareto distribution

Let $X_1, \dots, X_n$ be i.i.d. copies of X. Then, for $i = 1, \dots, n$

$$\log X_i \stackrel{d}{=} \gamma \log Y_i,$$

where $Y_1, \dots, Y_n$ are i.i.d. copies of Y, satisfying

$$\mathbb{P}(Y > y) = y^{-1}, \quad y > 1.$$

A gentle start: the Pareto distribution

Let $X_1, \ldots, X_n$ be i.i.d. copies of X. Then, for $i = 1, \ldots, n$

$$\log X_i \overset{d}{=} \gamma \log Y_i,$$

where $Y_1, \ldots, Y_n$ are i.i.d. copies of Y, satisfying

$$\mathbb{P}(Y > y) = y^{-1}, \quad y > 1.$$

As a result, the sample mean satisfies

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} \log X_i\right) = \gamma \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} \log Y_i\right) = \gamma.$$

A gentle start: the Pareto distribution

This is still true for log-spacings of the k-th largest order statistics among $X_{1,n} < \ldots < X_{n-k,n} < X_{n-k+1,n} < \ldots < X_{n,n}$, indeed

$$\log X_{n-k+1} - \log X_{n-k,n} \stackrel{d}{=} \gamma \log(Y_{n-k+i,n}/Y_{n-k,n}) \stackrel{d}{=} \gamma \log Y_{i,k}$$

for $i = 1, \ldots, k$. As a result,

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k}(\log X_{n-k+i,n} - \log X_{n-k,n})\right) &= \gamma\mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k}\log Y_{i,k}\right) \\
&= \gamma\mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k}\log Y_{i}\right) \\
&= \gamma.
\end{aligned}$$

More generally: positive tail index

Assume that, for some $\gamma > 0$ and a positive function a, as t $\to \infty$

$$\frac{F^{-1}(1 - 1/tx) - F^{-1}(1 - 1/t)}{a(t)} \to \frac{x^\gamma - 1}{\gamma}.$$

Then, as $t \to \infty$

$$\log F^{-1}(1 - 1/tx) - \log F^{-1}(1 - 1/t) \approx \gamma \log x.$$

More generally: positive tail index

Assume that, for some $\gamma > 0$ and a positive function a, as t $\to \infty$

$$\frac{F^{-1}(1-1/tx) - F^{-1}(1-1/t)}{a(t)} \to \frac{x^{\gamma}-1}{\gamma}.$$

Then, as $t \to \infty$

$$\log F^{-1}(1-1/tx) - \log F^{-1}(1-1/t) \approx \gamma \log x.$$

This fact has important consequences for the estimation of γ, still determining the degree of heavyness of the tail of F.

More generally: positive tail index

Indeed, since for $i = 1, \ldots, n$ we have

$$X_i \overset{d}{=} F^{-1}(1 - 1/Y_i),$$

then for a sequence $k \equiv k(n) \to \infty$, such that $k/n \to 0$ as $n \to \infty$,

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

$$\overset{d}{=} \frac{1}{k}\sum_{i=1}^{k}(\log F^{-1}(1 - 1/Y_{n-k+i,n}) - \log F^{-1}(1 - 1/Y_{n-k,n}))$$

$$\overset{d}{=} \frac{1}{k}\sum_{i=1}^{k}\left(\log F^{-1}\left(1 - \frac{1}{Y_{n-k,n}\frac{Y_{n-k+i,n}}{Y_{n-k,n}}}\right) - \log F^{-1}\left(1 - \frac{1}{Y_{n-k,n}}\right)\right)$$

$$\approx \gamma\frac{1}{k}\sum_{i=1}^{k}\log\frac{Y_{n-k+i,n}}{Y_{n-k,n}} \overset{d}{=} \gamma\frac{1}{k}\sum_{i=1}^{k}Y_i.$$

The basic estimator: Hill estimator

This justifies the construction of the Hill estimator (Hill, 1975)

$$\hat{\gamma}_H = \frac{1}{k}\sum_{i=1}^{k}(\log X_{n-k+i,n} - \log X_{n-k,n}).$$

The basic estimator: Hill estimator

This justifies the construction of the Hill estimator (Hill, 1975)

$$\hat{\gamma}_H = \frac{1}{k}\sum_{i=1}^{k}(\log X_{n-k+i,n} - \log X_{n-k,n}).$$

The proportion of data k/n used for estimation is known as effective sample fraction.

The basic estimator: Hill estimator

This justifies the construction of the Hill estimator (Hill, 1975)

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=1}^{k} (\log X_{n-k+i,n} - \log X_{n-k,n}).$$

The proportion of data k/n used for estimation is known as effective sample fraction.

When F is not a member of the Pareto family, the estimator is biased, i.e. $\mathbb{E}(\hat{\gamma}_H) - \gamma \neq 0$.

$\rightarrow$ More refined estimators have been proposed for bias reduction, see, e.g., Gomes and Pestana (2007).

More generally: nonnegative tail index

Assume that for some $\gamma \in \mathbb{R}$ and a positive function a

$$\frac{F^{-1}(1-1/tx) - F^{-1}(1-1/t)}{a(t)} \to \frac{x^\gamma - 1}{\gamma}, \quad t \to \infty.$$

Then, as $t \to \infty$

$$\frac{F^{-1}(1-1/tx) - F^{-1}(1-1/t)}{F^{-1}(1-1/ty) - F^{-1}(1-1/t)} \approx \frac{x^\gamma - 1}{y^\gamma - 1}.$$

As a result of this and convergence $Y_{n-sk,n}/Y_{n-k,n} \xrightarrow{\mathbb{P}} s^{-1}$...

More generally: nonnegative tail index

$$
\begin{aligned}
\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} &= \frac{X_{n-k,n} - X_{n-4k,n}}{X_{n-2k,n} - X_{n-4k,n}} - 1 \\
&= \frac{F^{-1}\left(1 - \frac{1}{Y_{n-4k,n}\frac{Y_{n-k,n}}{Y_{n-4k,n}}}\right) - F^{-1}\left(1 - \frac{1}{Y_{n-4k,n}}\right)}{F^{-1}\left(1 - \frac{1}{Y_{n-4k,n}\frac{Y_{n-2k,n}}{Y_{n-4k,n}}}\right) - F^{-1}\left(1 - \frac{1}{Y_{n-4k,n}}\right)} - 1 \\
&\approx \frac{\left(\frac{Y_{n-k,n}}{Y_{n-4k,n}}\right)^{\gamma} - 1}{\left(\frac{Y_{n-2k,n}}{Y_{n-4k,n}}\right)^{\gamma} - 1} - 1 \\
&\xrightarrow{\mathbb{P}} \frac{4^{\gamma} - 1}{2^{\gamma} - 1} - 1 = 2^{\gamma}.
\end{aligned}
$$

Another simple estimator: Pickands estimator

This justifies the construction of Pickands estimator (Pickands, 1975)

$$\hat{\gamma}_P = \frac{1}{\log 2} \log \left(\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \right)$$

Another simple estimator: Pickands estimator

This justifies the construction of Pickands estimator (Pickands, 1975)

$$\hat{\gamma}_P = \frac{1}{\log 2} \log \left(\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \right)$$

Again, the estimator is biased and more refined versions of it have been proposed for bias reduction, see, e.g., Drees (1996).

A general view: tail functionals

As discussed in Drees (1998), many estimators of γ are constructed via functionals, say T, of the *tail empirical process*

$$\mathcal{Q}_n(s) = X_{n-\lfloor sk \rfloor, n}, \quad s \in (0, 1].$$

A general view: tail functionals

As discussed in Drees (1998), many estimators of γ are constructed via functionals, say T, of the *tail empirical process*

$$\mathcal{Q}_n(s) = X_{n-\lfloor sk \rfloor, n}, \quad s \in (0,1].$$

Intuitively: the tail empirical process describes the tail of empirical distribution function

$$\mathcal{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \le x)$$

beyond its quantile of level $1 - k/n$.

A general view: tail functionals

If T is continuous, invariant to linear transformations and satisfies $T(\mathcal{Q}_\gamma) = \gamma$, with $\mathcal{Q}_\gamma(s) = (s^{-\gamma} - 1)/\gamma$, then

$$
\begin{aligned}
\hat{\gamma} = T(\mathcal{Q}_n) &= T\left(F^{-1}(1 - 1/Y_{n-\lfloor \cdot k,n \rfloor})\right) \\
&= T\left(\frac{F^{-1}(1 - 1/Y_{n-\lfloor \cdot k,n \rfloor}) - F^{-1}(1 - 1/Y_{n-\lfloor k,n \rfloor})}{a(Y_{n-k,n})}\right) \\
&= T\left(\frac{F^{-1}\left(1 - \frac{1}{Y_{n-k,n}\frac{Y_{n-\lfloor \cdot k \rfloor,n}}{Y_{n-k,n}}}\right) - F^{-1}\left(1 - \frac{1}{Y_{n-k,n}}\right)}{a(Y_{n-k,n})}\right) \\
&\approx T(\mathcal{Q}_\gamma) \\
&= \gamma.
\end{aligned}
$$

A general view: tail functionals

Pickands estimator fulfills these properties with

$$T_P(z) = \frac{1}{\log 2} \log \left(\frac{z(1/4) - z(1/2)}{z(1/2) - z(1)} \right).$$

A general view: tail functionals

Pickands estimator fulfills these properties with

$$T_P(z) = \frac{1}{\log 2} \log \left(\frac{z(1/4) - z(1/2)}{z(1/2) - z(1)} \right).$$

Another example: Probability Weighted Moment estimator $\hat{\gamma}_{PWM}$ (see Drees, 1998), obtained with

$$T_{PWM}(z) = \frac{\int_{[0,1]} z(s) w_1(s) \mathrm{d}s}{\int_{[0,1]} z(s) w_2(s) \mathrm{d}s}$$

for weight functions satisfying $\int_0^1 w_j(s) \mathrm{d}s = 0$, $j = 1, 2$.

From empirical to model-based inference

Recall that if, as $t \to \infty$,

$$\frac{F^{-1}(1 - 1/tx) - F^{-1}(1 - 1/t)}{a(t)} \to \mathcal{Q}_\gamma(x),$$

then

$$\mathbb{P}\left(\frac{X - F^{-1}(1-1/t)}{a(t)} > x \,\middle|\, X > F^{-1}(1 - 1/t) \right) \to (1 + \gamma x)^{-1/\gamma}$$
$$=: 1 - H_\gamma(x)$$

From empirical to model-based inference

Recall that if, as $t \to \infty$,

$$\frac{F^{-1}(1 - 1/tx) - F^{-1}(1 - 1/t)}{a(t)} \to \mathcal{Q}_\gamma(x),$$

then

$$\mathbb{P}\left(\frac{X - F^{-1}(1-1/t)}{a(t)} > x \,\middle|\, X > F^{-1}(1 - 1/t) \right) \to (1 + \gamma x)^{-1/\gamma} =: 1 - H_\gamma(x)$$

I.e., rescaled exceedances of a large threshold $F^{-1}(1 - 1/t)$ approximately follow a *generalized Pareto (GP) distribution*.

From empirical to model-based inference

In practice, F is unkown and so is $F^{-1}(1-1/t)$, thus a typical inferential routine is:

1. Setting $t = n/k$, use the k-th largest order statistic $X_{n-k,n}$ as a threshold.
2. Compute exceedances

$$Z_i := X_{n-k+i,n} - X_{n-k,n}, \quad i = 1, \ldots, k.$$

3. Fit a GP distribution to exceedances

$$\{H_\gamma(\cdot \sigma), \quad \gamma \in \mathbb{R}, \sigma > 0\}$$

including a scale parameter to account for the unkown $a(n/k)$.

Fitting a GP to data: maximum likelihood

Note that for any z such that $1 + \frac{\gamma}{\sigma}x > 0$

$$H(z/\sigma) = \int_0^z \left(1 + \frac{\gamma}{\sigma}x\right)^{-1/\gamma-1} \sigma^{-1}\mathrm{d}x =: \int_0^z h_{\gamma,\sigma}(x)\mathrm{d}x.$$

Fitting a GP to data: maximum likelihood

Note that for any z such that $1 + \frac{\gamma}{\sigma}x > 0$

$$H(z/\sigma) = \int_0^z \left(1 + \frac{\gamma}{\sigma}x\right)^{-1/\gamma - 1} \sigma^{-1} \mathrm{d}x =: \int_0^z h_{\gamma,\sigma}(x)\mathrm{d}x.$$

Treating the Z_i's as approximately:

- independent
- distributed according to a GP

the associated *likelihood function* at parameters $\theta = (\gamma, \sigma)$ is

$$\mathcal{L}(\theta) \equiv \mathcal{L}(Z_1, \ldots, Z_k; \theta) = \prod_{i=1}^{k} h_{\gamma,\sigma}(Z_i).$$

Fitting a GP to data: maximum likelihood

To estimate θ and infer tail probabilities, large quantiles, etc. one can compute the maximum likelihood estimator (MLE)

$$\hat{\theta} \in \arg\max \mathcal{L}(\theta)$$

Fitting a GP to data: maximum likelihood

To estimate θ and infer tail probabilities, large quantiles, etc. one can compute the maximum likelihood estimator (MLE)

$$\hat{\theta} \in \arg\max \mathcal{L}(\theta)$$

Caveat: a maximum over the full range of possible values for (γ, σ) does not exist. We focus on the region

$$\Theta := \{(\gamma, \sigma) : \gamma > -1/2, \sigma > 0\}$$

since the MLE behaves irregularly if $\gamma \leq -1/2$. Moreover...

Dombry, Padoan and R. (2023): U.r.c., with probability tending to 1, $\arg\max_{\theta \in \Theta} \mathcal{L}(\theta)$ consists of a singe value.

From parameter to quantile estimation

Recall that if, as $t \to \infty$,

$$\frac{F^{-1}(1 - 1/tx) - F^{-1}(1 - 1/t)}{a(t)} \to \mathcal{Q}_\gamma(x),$$

then, for a small $p \equiv p(n)$ such that $np \to c \in (0, 1)$

$$\begin{aligned} F^{-1}(1-p) &= F^{-1}\left(1 - \frac{1}{\frac{n}{k}\frac{k}{np}}\right) \\ &\approx F^{-1}\left(1 - \frac{1}{\frac{n}{k}}\right) + a\left(\frac{n}{k}\right)\mathcal{Q}_\gamma\left(\frac{np}{k}\right) \end{aligned}$$

From parameter to quantile estimation

Recall that if, as $t \to \infty$,

$$\frac{F^{-1}(1 - 1/tx) - F^{-1}(1 - 1/t)}{a(t)} \to \mathcal{Q}_\gamma(x),$$

then, for a small $p \equiv p(n)$ such that $np \to c \in (0, 1)$

$$\begin{aligned} F^{-1}(1-p) &= F^{-1}\left(1 - \frac{1}{\frac{n}{k}\frac{k}{np}}\right) \\ &\approx F^{-1}\left(1 - \frac{1}{\frac{n}{k}}\right) + a\left(\frac{n}{k}\right)\mathcal{Q}_\gamma\left(\frac{np}{k}\right) \end{aligned}$$

An estimator of the extreme quantile $F^{-1}(1-p)$ is thus

$$X_{n-k,n} + \hat{\sigma}\mathcal{Q}_{\hat{\gamma}}\left(\frac{np}{k}\right)$$

From frequentist to Bayesian inference

The Bayesian approach tackles the problem from a different angle:

- The data are seen as having a joint probability density *conditionally* on a give value θ

  $\rightarrow$ at realisations $z_{1:k} := (z_1, \ldots, z_k)$, it is *analytically* equal to $\mathcal{L}(z_{1:k}; \theta)$.

- In turn, θ is assigned a *prior distribution* Π, representing the researcher's belief on more plausible values of θ.

- Finally, inference is based on the *posterior distribution*, obtained via *Bayes rule*:

$$\Pi(B | Z_{1:k} = z_{1:k}) = \frac{\int_B \mathcal{L}(z_{1:k}; \theta) \mathrm{d}\Pi(\theta)}{\int_\Theta \mathcal{L}(z_{1:k}; \theta) \mathrm{d}\Pi(\theta)}.$$

From frequentist to Bayesian inference

- **Point estimation**: e.g., via posterior means
  - $\hat{\gamma} = \int_{\Theta} \gamma \mathrm{d}\Pi(\theta | Z_{1:k} = z_{1:k})$
  - $\hat{\sigma} = \int_{\Theta} \sigma \mathrm{d}\Pi(\theta | Z_{1:k} = z_{1:k})$

From frequentist to Bayesian inference

- **Point estimation**: e.g., via posterior means
  - $\hat{\gamma} = \int_{\Theta} \gamma \mathrm{d}\Pi(\theta | Z_{1:k} = z_{1:k})$
  - $\hat{\sigma} = \int_{\Theta} \sigma \mathrm{d}\Pi(\theta | Z_{1:k} = z_{1:k})$

- **In practice**: computations via MCMC

$$\hat{\gamma} = \sum_{i=1}^{R} \gamma_i w(\theta_i), \quad \hat{\sigma} = \sum_{i=1}^{R} \sigma_i w(\theta_i)$$

where $\theta_i = (\gamma_i, \sigma_i)$, $i = 1, \ldots, R$, are approximately drawn from the posterior distribution and w is a weight function, e.g.

$$w(\theta) \equiv 1/R.$$

From frequentist to Bayesian inference

- **Beyond point estimation**: it is possible to derive regions C_n containing θ with posterior probability $1-\alpha$, i.e.

$$\Pi(C_n | Z_{1:k} = z_{1:k}) = 1 - \alpha$$

where $1-\alpha$ is the desired *credibility level*.

From frequentist to Bayesian inference

- **Beyond point estimation**: it is possible to derive regions C_n containing θ with posterior probability $1-\alpha$, i.e.

$$\Pi(C_n | Z_{1:k} = z_{1:k}) = 1 - \alpha$$

where $1-\alpha$ is the desired *credibility level*.

- It is possible to derive also *credible intervals* for single parameters.

In practice: computations based on MCMC, e.g.

$$C_n = (\gamma_{\lfloor R\,\alpha/2 \rfloor, R}, \gamma_{\lfloor R(1-\alpha/2 \rfloor, R})$$

where $\gamma_{1,R} < \gamma_{2,R} < \ldots < \gamma_{R,R}$ are ordered values of a sample (approximately) from the posterior distribution.

From frequentist to Bayesian inference

Since for a given p and realisation $X_{n-k,n} = x_{n-k,n}$ the map

$$\mathcal{T} : \theta = (\gamma, \sigma) \mapsto x_{n-k,n} + \sigma \mathcal{Q}_{\gamma}\left(\frac{np}{k}\right)$$

is continuous, the posterior distribution of θ induces

$$\Psi(\,\cdot\,|Z_{1:k} = z_{1:k}) := \Pi(\mathcal{T}^{-1}(\cdot)|Z_{1:k} = z_{1:k}),$$

a posterior distribution for the extreme quantile.

From frequentist to Bayesian inference

Since for a given p and realisation $X_{n-k,n} = x_{n-k,n}$ the map

$$\mathcal{T} : \theta = (\gamma, \sigma) \mapsto x_{n-k,n} + \sigma \mathcal{Q}_\gamma \left(\frac{np}{k} \right)$$

is continuous, the posterior distribution of θ induces

$$\Psi(\cdot | Z_{1:k} = z_{1:k}) := \Pi(\mathcal{T}^{-1}(\cdot) | Z_{1:k} = z_{1:k}),$$

a posterior distribution for the extreme quantile.

Summaries for point or interval estimation can be extracted from the posterior along the previous lines.

From frequentist to Bayesian inference

Since for a given p and realisation $X_{n-k,n} = x_{n-k,n}$ the map

$$\mathcal{T} : \theta = (\gamma, \sigma) \mapsto x_{n-k,n} + \sigma \mathcal{Q}_\gamma \left(\frac{np}{k} \right)$$

is continuous, the posterior distribution of θ induces

$$\Psi(\cdot | Z_{1:k} = z_{1:k}) := \Pi(\mathcal{T}^{-1}(\cdot) | Z_{1:k} = z_{1:k}),$$

a posterior distribution for the extreme quantile.

Summaries for point or interval estimation can be extracted from the posterior along the previous lines.

From an approximate sample $(\theta_i)_{i=1}^R$ from $\Pi(\cdot | Z_{1:k} = z_{1:k})$ one can get a sample $(q_i = \mathcal{T}(\theta_i))_{i=1}^R$ from $\Psi(\cdot | Z_{1:k} = z_{1:k})$.

Bayesian inference: prediction

In this framework, a natural estimator for the c.d.f. of a future exceedance $X - F^{-1}\left(1 - \frac{1}{n/k}\right) \left| X > F^{-1}\left(1 - \frac{1}{n/k}\right)\right.$ is

$$\hat{H}(z) := \int_{\Theta} H_{\theta}(z) \mathrm{d}\Pi(\theta | Z_{1:k})$$

which defines the *posterior predictive distribution* function, with

$$H_{\theta}(\,\cdot\,) = H_{\gamma}(\,\cdot\,/\sigma).$$

Bayesian inference: prediction

In this framework, a natural estimator for the c.d.f. of a future exceedance $X - F^{-1}\left(1 - \frac{1}{n/k}\right) \middle| X > F^{-1}\left(1 - \frac{1}{n/k}\right)$ is

$$\hat{H}(z) := \int_{\Theta} H_{\theta}(z) \mathrm{d}\Pi(\theta | Z_{1:k})$$

which defines the *posterior predictive distribution* function, with

$$H_{\theta}(\,\cdot\,) = H_{\gamma}(\,\cdot\,/\sigma).$$

As a result, a natural estimator of the c.d.f. of a future peak $X \middle| X > F^{-1}\left(1 - \frac{1}{n/k}\right)$ over $F^{-1}\left(1 - \frac{1}{n/k}\right)$ is given by

$$\hat{H}^{*}(\,\cdot\,) = \hat{H}(\,\cdot\, - X_{n-k,n})$$

Bayesian inference: prediction

- **More than point prediction:** e.g., it is possible to construct a *predictive interval* (x_l^*, x_u^*) for a future peak such that

$$\hat{H}^*(x_u^*) - \hat{H}^*(x_l^*) = 1 - \alpha$$

for any desired probability level $1 - \alpha$.

Bayesian inference: prediction

- **More than point prediction:** e.g., it is possible to construct a *predictive interval* (x_l^*, x_u^*) for a future peak such that

$$\hat{H}^*(x_u^*) - \hat{H}^*(x_l^*) = 1 - \alpha$$

for any desired probability level $1 - \alpha$.

- **In practice**: with realisations $Z_{1:k} = z_{1:k}$ and $X_{n-k,n} = x_{n-k,n}$
  1. for each approximate draw from the posterior distribution θ_i, sample independently a value z_i^* from H_{θ_i};
  2. compute $x_i^* = z_i + x_{n-k,n}$, $i = 1, \ldots, R$;
  3. set $x_l^* = x^*_{\lfloor R\alpha/2 \rfloor, R}$ and $x_u^* = x^*_{\lfloor R(1-\alpha/2) \rfloor, R}$.

From peaks to block maxima

Let $M_m := \bigvee_{i=1}^{m} X_i$ (max over a block of size m).

From peaks to block maxima

Let $M_m := \bigvee_{i=1}^{m} X_i$ (max over a block of size m).

Recall that if, as $t \to \infty$,

$$\frac{F^{-1}(1 - 1/tx) - F^{-1}(1 - 1/t)}{a(t)} \to \frac{x^\gamma - 1}{\gamma},$$

then $\exists a_m > 0$, $\exists b_m \in \mathbb{R}$ such that normalised maximum's law converges to a *generalised extreme value (GEV) distribution*:

$$\mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) \to \exp\left\{-(1 + \gamma x)^{-1/\gamma}\right\} =: G_\gamma(x)$$

as $m \to \infty$, for all x such that $1 + \gamma x > 0$.

Block maxima approach

Common practice:

1. divide obs. $X_1, X_2, \ldots$ into blocks of size m;
2. compute maxima on each block, obtaining a sample of size k $M_{m,1}, \ldots, M_{m,k}$;
3. assume $F_{M_m}(x) \approx G_\gamma(a_m^{-1}(x - b_m))$;
4. fit a GEV model $G_\theta(x) = G_\gamma(\sigma x + \mu)$, where

$$\theta = (\gamma, \mu, \sigma)$$

   includes (σ, μ), to account for the unknown (a_m, b_m).

Block maxima approach

Common practice:

1. divide obs. $X_1, X_2, \dots$ into blocks of size m;
2. compute maxima on each block, obtaining a sample of size k $M_{m,1}, \dots, M_{m,k}$;
3. assume $F_{M_m}(x) \approx G_\gamma(a_m^{-1}(x - b_m))$;
4. fit a GEV model $G_\theta(x) = G_\gamma(\sigma x + \mu)$, where

$$\theta = (\gamma, \mu, \sigma)$$

includes (σ, μ), to account for the unknown (a_m, b_m).

Inference can be based, e.g. on maximisation of the likelihood

$$\mathcal{L}(\theta) \equiv \mathcal{L}(M_{m,1}, \dots, M_{m,k}; \theta) = \prod_{i=1}^{k} g_\theta(M_{m,i})$$

with g_θ the density of G_θ, or on Bayesian procedures.

Return levels

The *T-return level* is intuitively defined as the value that occurs or is exceeded (on average) every T-periods.

Formally, it is the $1 - 1/T$-quantile of the periodic maximum

$$F_{M_m}^{\leftarrow}(1 - 1/T) \approx b_m + a_m G_\gamma^{-1}(1 - 1/T) =: \mathcal{R}(a_m, b_m, \gamma).$$

Estimation:

- MLE: plug maximum likelihood estimates in the formula, i.e. compute $\mathcal{R}(\hat{\theta})$ with $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\gamma}) \in \arg\max \mathcal{L}(\theta)$.
- Bayes: obtain a posterior distribution of $\mathcal{R}(\theta)$ from that of θ.

Bayesian analysis of extremes: a subtle issue

Coles and Powell (1996): «Extreme value problems are characterized by a scarcity of data and the requirement of modelling where the data are most sparse. This presents a dilemma when considering a Bayesian approach to inference: the value of additional prior information is likely to be substantial, but the plausibility of formulating such prior knowledge for extremal behaviour is questionable.»

Bayesian analysis of extremes: a subtle issue

Coles and Powell (1996): «Extreme value problems are characterized by a scarcity of data and the requirement of modelling where the data are most sparse. This presents a dilemma when considering a Bayesian approach to inference: the value of additional prior information is likely to be substantial, but the plausibility of formulating such prior knowledge for extremal behaviour is questionable.»

In particular: difficult to specifiy genuine prior distributions which

- appropriety account for the block-size (threshold) dependent nature of location and scale parameters (scale parameter);
- put the basis for efficient posterior computations.

Empirical Bayes analysis of maxima

- Prior on γ: use standard densities π_{sh} on $\mathbb{R}$.
- Difficulties in specifying priors on (σ, μ).

Padoan and R. (2023) propose data-dependent priors

$$\hat{\pi}_{\text{sc}}(\sigma) = \pi_{\text{sc}}\left(\frac{\sigma}{\hat{a}_m}\right)\frac{1}{\hat{a}_m}, \quad \hat{\pi}_{\text{loc}}(\mu) = \pi_{\text{loc}}\left(\frac{\mu - \hat{b}_m}{\hat{a}_m}\right)\frac{1}{\hat{a}_m}.$$

Empirical Bayes analysis of maxima

- Prior on γ: use standard densities π_{sh} on $\mathbb{R}$.
- Difficulties in specifying priors on (σ, μ).

Padoan and R. (2023) propose data-dependent priors

$$\hat{\pi}_{\mathrm{sc}}(\sigma) = \pi_{\mathrm{sc}}\left(\frac{\sigma}{\hat{a}_m}\right)\frac{1}{\hat{a}_m}, \quad \hat{\pi}_{\mathrm{loc}}(\mu) = \pi_{\mathrm{loc}}\left(\frac{\mu - \hat{b}_m}{\hat{a}_m}\right)\frac{1}{\hat{a}_m}.$$

This yields the empirical Bayes posterior distribution

$$\Pi(B|M_{m,1:k}) = \frac{\int_B \mathcal{L}(\theta)\pi(\theta)\mathrm{d}\theta}{\int_\Theta \mathcal{L}(\theta)\pi(\theta)\mathrm{d}\theta}$$

with $\pi(\theta) = \pi_{\mathrm{sh}}(\gamma)\hat{\pi}_{\mathrm{loc}}(\mu)\hat{\pi}_{\mathrm{sc}}(\sigma)$ on $\Theta = (-1,\infty) \times \mathbb{R} \times (0,\infty)$.

Prediction of future maximum levels

I.i.d. context: distribution of a future maximum over $m^* > m$ obs.

$$F_{M_{m^*}}(x) = \{F_{M_m}(x)\}^{m^*/m} \approx \left\{G_\gamma(a_m^{-1}(x - b_m))\right\}^{m^*/m}.$$

E.g., $m = 122$ and $m^*/m = 3$, i.e. M_{m^*} is annual maximum.

Prediction of future maximum levels

I.i.d. context: distribution of a future maximum over $m^* > m$ obs.

$$F_{M_{m^*}}(x) = \{F_{M_m}(x)\}^{m^*/m} \approx \left\{G_\gamma(a_m^{-1}(x - b_m))\right\}^{m^*/m}.$$

E.g., $m = 122$ and $m^*/m = 3$, i.e. M_{m^*} is annual maximum.

By max-stability $\left\{G_\gamma(a_m^{-1}(x - b_m))\right\}^{m^*/m} = G_{\theta^*}(x)$, where

$$\theta^* = \left(\gamma, b_m + a_m \frac{(m^*/m)^\gamma - 1}{\gamma}, a_m(m/m^*)^\gamma\right).$$

Prediction

A natural way to define a predictive distribution function for such maximum is then

$$\widehat{G}_{m^*}(x) = \int_{\Theta} G_{\theta^*}(x)\Pi(\mathrm{d}\theta|M_{m,1:k}).$$

Prediction

A natural way to define a predictive distribution function for such maximum is then

$$\widehat{G}_{m^*}(x) = \int_\Theta G_{\theta^*}(x)\Pi(\mathrm{d}\theta | M_{m,1:k}).$$

From it we can derive:

- Predictive intervals: $(\widehat{G}_{m^*}^{-1}(\alpha/2), \widehat{G}_{m^*}^{-1}(1-\alpha/2))$.

Prediction

A natural way to define a predictive distribution function for such maximum is then

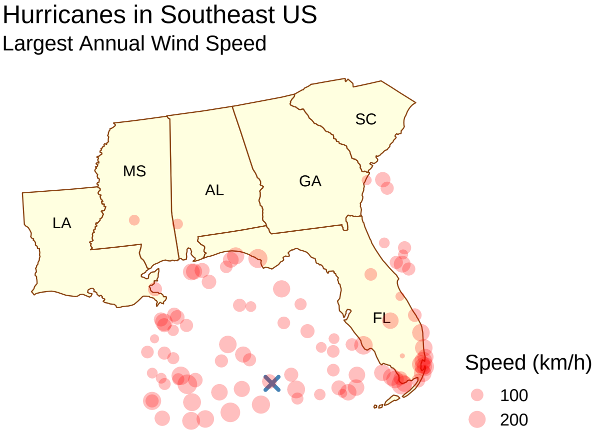
$$\widehat{G}_{m^*}(x) = \int_{\Theta} G_{\theta^*}(x)\Pi(\mathrm{d}\theta|M_{m,1:k}).$$

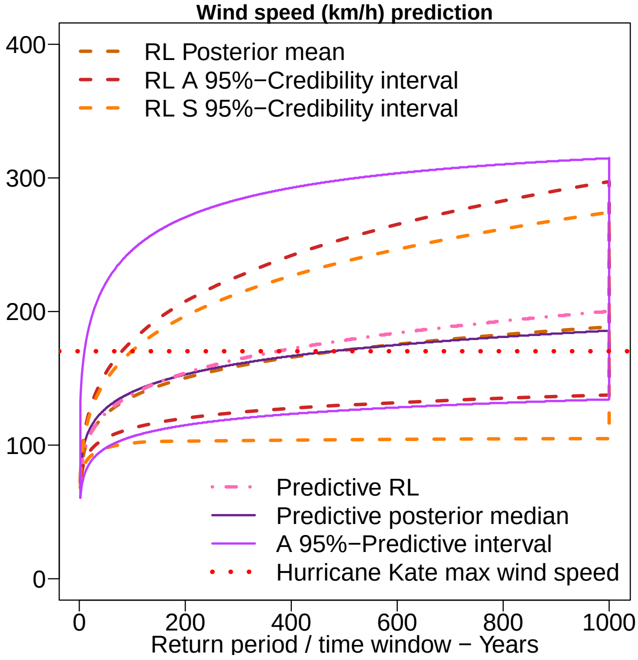
From it we can derive:

- Predictive intervals: $(\widehat{G}_{m^*}^{-1}(\alpha/2), \widehat{G}_{m^*}^{-1}(1-\alpha/2))$.
- Predictive return level: $\widehat{G}_{m^*}^{-1}(1-1/T)$, for $T > 1$.

Predicting hurricanes in Southeastern US

We analyse a sequence of daily wind speed maxima from 1976 to 2021, selecting training block size $m = 122 \implies k = 120$ maxima.





Wind speed (km/h) prediction
RL Posterior mean
RL A 95%−Credibility interval
RL S 95%−Credibility interval
Predictive RL
Predictive posterior median
A 95%−Predictive interval
Hurricane Kate max wind speed
0
100
200
300
400
0
200
400
600
800
1000
Return period / time window − Years

Thank you!

References

- Dombry, C., Padoan, S. A. and Rizzelli, S. (2023). Asymptotic theory of Bayesian inference and prediction: from ordinary to conditional peaks over a threshold methods. *Submitted*.
- Drees, H. (1996). Refined Pickands estimators with bias correction. *Commun. Statist. Th. Meth.* **25**, 837–851.
- Drees, H. (1998). *Estimating the Extreme Value Index*. Habilitation thesis, Univ. Cologne.
- Gomes, M. I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *J. Amer. Statist. Assoc.* **102**, 280–292.
- Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**,1163–1174.
- Padoan, S. A. and Rizzelli, S. (2023). Empirical Bayes inference for the block maxima method. *Bernoulli*, to appear.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131.