Univariate Extreme Value Theory: Concepts and Applications

Inspired from Prof. Raphaël Huser (KAUST)

by Dr. Rishikesh Yadav (Postdoctoral Research Fellow)

HEC Montréal, McGill University, Canada

Table of contents

References

- Coles (2001), An Introduction to Statistical Modeling of Extreme Values, Springer
- Beirlant, Goegebeur, Teugels and Segers (2004), Statistics of Extremes: Theory and Applications, Wiley
- de Haan and Ferreira (2006), Extreme Value Theory: An Introduction, Springer
- Resnick (1987), Extreme Values, Regular Variation, and Point Processes, Springer

Motivation

Why Extreme Value Theory (EVT)?

- **Focus on Extremes:** EVT specifically addresses the tail of distributions, where extreme events lie, unlike classical statistics which focuses on the entire distribution.
- **Risk Assessment:** EVT provides tools for quantifying the probability and impact of rare events, essential for effective risk management.
- **Tail Behavior:** EVT accurately models the tails of distributions, capturing the behavior of extreme values better than classical approaches which may assume normality.
- **Extrapolation:** EVT enables extrapolation beyond the observed data to predict the likelihood of unobserved extreme events, whereas classical methods may struggle with such predictions.

Motivating Example: North sea Flood, 1953 (Netherlands)

- On the night of January 31 to February 1, 1953, a combination of a high spring tide and a severe European windstorm caused a storm surge in the North Sea.
- The water level locally exceeded 5.6 meters above mean sea level.
- The flood and waves overwhelmed sea defenses and led to extensive flooding.
- 1,835 people died in the Netherlands, 307 in the UK, and 28 in Belgium.

Delta Works II

- The Dutch government established the Delta Commission, which conceived and deployed an ambitious flood defense system called the Delta Works, designed to protect the estuaries of the rivers Rhine, Meuse, and Scheldt.
- The works were completed in 1997.
- The Delta Works were declared one of the Seven Wonders of the Modern World by the American Society of Civil Engineers.
- The design balanced cost and safety.
- The acceptable risk was set according to region. For instance, in North and South Holland, the flood defenses were built to withstand a failure once every 10,000 years.
  - $\Rightarrow$ Risk assessment was based on data observed over a much shorter period.
  - $\Rightarrow$ There was a need to extrapolate beyond the observed maximum.
  - $\Rightarrow$ This is precisely what the Statistics of Extremes is designed for.

Delta Works II

(Some) Application Areas

- **Environment:** rainfall, temperatures, snowfall (avalanches), river levels, sea levels, wind speed, pollution, etc.
- **Seismology:** large earthquakes and tsunamis
- **Finance:** financial time series and (re-)insurance
- **Material science:** strength of materials and structures
- **Athletics:** record times

Block Maxima Approach

Recall Central Limit Theorem

- Suppose that $Y, Y_1, Y_2, \ldots$ is a sequence of i.i.d. non-degenerate random variables defined with distribution $F(y)$.
- The i.i.d. assumption may be slightly relaxed...
- We define the partial sums as:

$$S_0 = 0, \quad S_n = Y_1 + \cdots + Y_n, \quad n \geq 1,$$

- And the arithmetic (or sample) means:

$$\overline{Y}_n = \frac{S_n}{n}, \quad n \geq 1.$$

- If $\sigma^2 = \text{var}(Y) < \infty$, then as $n \to \infty$:

$$Z_n = \left(n^{-1/2}\sigma\right)^{-1} (\overline{Y}_n - \mu) \xrightarrow{D} Z \sim \mathcal{N}(0,1),$$

where $\mu = E(Y)$ and $\mathcal{N}(0,1)$ is the standard normal distribution.

Maxima I

Would like to have a result similar to the CLT, but for extremes (minima/maxima)...

- We start by assuming that $Y, Y_1, Y_2, \ldots$ is a sequence of i.i.d. (non-degenerate) random variables with distribution $F(y)$.
- The i.i.d. assumption may be generalized for non-stationary extremes and accounting for temporal dependence.
- We define the partial (or sample) maximum as

$$M_n = \max(Y_1, \ldots, Y_n).$$

- The following theory also applies to minima because

$$\min(Y_1, \ldots, Y_n) = -\max(-Y_1, \ldots, -Y_n).$$

- Our general discussion is for maxima, and we make this transformation without comment when we model minima.

Maxima II

- The distribution function of the sample maximum M_n is

$$\begin{aligned}\Pr(M_n \leq y) &= \Pr(Y_1 \leq y, \ldots, Y_n \leq y) \\ &= \Pr(Y_1 \leq y) \times \cdots \times \Pr(Y_n \leq y) = F^n(y).\end{aligned}$$

- F is typically unknown, so we need to approximate F^n by some limit distribution.
- What distributions can arise? As $n \to \infty$, one has

$$F^n(y) \to \begin{cases} 0, & \text{if } F(y) < 1, \\ 1, & \text{if } F(y) = 1, \end{cases}$$

  so $M_n \xrightarrow{D} y_F$, where $y_F = \sup\{y : F(y) < 1\}$ is the upper support point of F. Hence, the limit distribution is degenerate!
- **Intuitively:** As M_n is always increasing with n, we should renormalize to get a non-degenerate limiting distribution!

Maxima III

We seek to mimic the Central Limit Theorem...

- We will investigate limiting distributions for renormalized maxima

$$M_n^* = a_n^{-1}(M_n - b_n),$$

  for some suitable sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$.
- Question: What kind of limiting distributions can arise for M_n^*, as $n \to \infty$?
- The distribution of M_n^* is

$$\begin{aligned}
\Pr(M_n^* \leq y) &= \Pr\left\{a_n^{-1}(M_n - b_n) \leq y\right\} \\
&= \Pr(M_n \leq a_n y + b_n) \\
&= F^n(a_n y + b_n) \\
&= F^n(u_n) \\
&= \Pr(M_n \leq u_n),
\end{aligned}$$

  where u_n is a suitable threshold sequence.

Back to Affine Transformations

Let's assume that a threshold sequence u_n exists.

- We will focus on sequences of the type $u_n = a_n y + b_n$ (affine transformation).
- This choice is simple, mimics the CLT, and allows us to get non-degenerate limits for a wide family of distributions.
- Questions:
  - How to choose a_n and b_n? Is the choice unique?
  - What distributions arise as limits for $M_n^* = a_n^{-1}(M_n - b_n)$ (properly renormalized)?
  - If there are several limit distributions, how to characterize their **"max-domains of attraction"**?
  - What is the speed of convergence to the limit?
- Similarly to the CLT, the notion of max-stability is crucial to start answering these questions.

Extremal Types Theorem – Fisher and Tippett, 1928

- Let $Y, Y_1, Y_2, \ldots$ be a sequence of i.i.d. random variables. If there exist sequences of constants $a_n > 0$ and b_n such that, as $n \to \infty$, $\Pr(M_n^* \leq y) = \Pr\left\{\frac{M_n - b_n}{a_n} \leq y\right\} \to G(y)$ for some non-degenerate distribution G, then G has the same type as one of the following distributions:
    - **I — Gumbel**: $G(y) = \exp\{-\exp(-y)\}$, $-\infty < y < \infty$;
    - **II — Fréchet**: $G(y) = \begin{cases} 0, & y \leq 0, \\ \exp(-y^{-\alpha}), & y > 0, \alpha > 0; \end{cases}$
    - **III — Weibull**: $G(y) = \begin{cases} \exp\{-(-y)^{\alpha}\}, & y < 0, \alpha > 0, \\ 1, & y \geq 0. \end{cases}$
- Conversely, each of these G's may appear as a limit for the distribution of $(M_n - b_n)/a_n$, and does so when G itself is the distribution of Y.

Generalized Extreme-Value (GEV) Distribution

This family encompasses all three of the previous extreme-value limit families:

$$G(y) = \exp\left\{ - \left[1 + \xi \left(\frac{y-\mu}{\sigma} \right) \right]_+^{-1/\xi} \right\},$$

defined on $\{y : 1 + \xi(y-\mu)/\sigma > 0\}$. Here, $a_+ = \max(0, a)$.

- μ is a location parameter, while $\sigma > 0$ is a scale parameter.
- ξ is a shape parameter determining the rate of tail decay, with
  - $\xi > 0$ giving the heavy-tailed (Fréchet) case,
  - $\xi = 0$ (interpreted as $\xi \to 0$) giving the light-tailed (Gumbel) case,
  - $\xi < 0$ giving the short-tailed (reversed Weibull) case.
- If $Y \sim \text{GEV}(\mu, \sigma, \xi)$, one has $E(Y^r) < \infty \iff \xi r < 1$.
- By the Extremal Types Theorem (ETT), the GEV distribution is the only univariate max-stable distribution.

Examples: Limit Laws for Maxima

- **Uniform distribution**: Maxima from the Uniform distribution $F(y) = \frac{y}{\theta}$, $y \in (0, \theta)$, $\theta > 0$, are "attracted" to the reversed Weibull distribution.
- **Gaussian distribution**: Maxima from the Gaussian distribution $F(y) = \Phi(y)$, $y \in \mathbb{R}$, are "attracted" to the Gumbel distribution.
- **Exponential distribution**: Maxima from the Exponential distribution $F(y) = 1 - \exp(-\theta y)$, $y > 0$, $\theta > 0$, are "attracted" to the Gumbel distribution.
- **Pareto distribution**: Maxima from the Pareto distribution $F(y) = 1 - y^{-\alpha}$, $y > 1$, $\alpha > 0$, are "attracted" to the Fréchet distribution.
- **Cauchy distribution**: Maxima from the Cauchy distribution $F(y) = \frac{1}{2} + \frac{1}{\pi} \arctan(y)$, $y \in \mathbb{R}$, are "attracted" to the Fréchet distribution.

Statistical Inference – Basic Idea

Suppose we have a time series of daily values $Y_1, Y_2, \ldots$ supposed to be independent and identically distributed from some distribution F.

- Extract maxima $M_n = \max(Y_1, \ldots, Y_n)$ of blocks of the original series:
  - For environmental time series, typically $n = 365$ for annual maxima, $n = 30$ for monthly maxima.
  - In finance, typically $n = 250$ for annual maxima, $n = 20$ for monthly data.
- Assume the resulting series of maxima $Z_1, \ldots, Z_N$, where Z_j is the maximum of the jth block of n consecutive observations, is distributed according to the GEV distribution $G(z)$ for some parameters μ, σ, ξ.
- Estimate the unknown parameters and use the fitted GEV model to make extrapolations.

Quantiles and return levels

- Let $0 < p < 1$ denote the "excess probability" and define

$$z_p = G^{-1}(1-p) = \mu + \frac{\sigma}{\xi}\left[\{-\log(1-p)\}^{-\xi} - 1\right],$$

  or equivalently: $1 - G(z_p) = p$.
- The level z_p is called the return level associated with the return period $1/p$.
- If the GEV(μ, σ, ξ) distribution G is used to model annual maxima, the M-year return level (with return period M) is simply

$$z_{1/M} = Q(M) = G^{-1}(1 - 1/M).$$

- Interpretation: "The M-year return level is exceeded (on average) once every M years."
- Provides a way to extrapolate to higher quantiles than the observed maximum (i.e., very small p) based on theoretically justified arguments.
- Depend on unknown parameters $\Rightarrow$ Need to be estimated from the data!

Likelihood Estimation – GEV Distribution

- **Log likelihood function for GEV distribution:**

  $$\ell(\mu,\sigma,\xi) = \sum_{i=1}^{N} \left[-\log\sigma - \left(1+\frac{1}{\xi}\right)\log\left\{1+\xi\frac{z_i-\mu}{\sigma}\right\} - \left\{1+\xi\frac{z_i-\mu}{\sigma}\right\}^{-1/\xi} \right].$$

  Note: This equals $-\infty$ if any $1+\xi\frac{z_i-\mu}{\sigma}<0$.
- **Numerical maximization of log likelihood.**
- **Calculation of standard errors from inverse of observed information matrix (also obtained numerically).**
- **Diagnostic checks:**
  - Probability plots
  - Quantile plots
  - Return level plots
- **Comparison of competing models:**
  - Nested models through deviance (likelihood ratio statistic)
  - Non-nested models by minimizing the Akaike information criterion (AIC):

    $$\mathrm{AIC} = -2\ell(\hat{\theta}) + 2p,$$

    (or similar criteria)
- **Calculation of confidence intervals for return levels using profile log likelihood.**

Regularity of MLE

- **Usual regularity conditions for validity of conventional likelihood inference:**
  - Require that the support of the density does not depend on the parameter values.
  - **Not the case for the GEV!**
- The score statistic $U(\theta) = \frac{\partial \ell(\theta)}{\partial \theta}$ must have finite first and second moments that satisfy:
$$\mathrm{var}\left(\frac{\partial \ell(\theta)}{\partial \theta}\right) = E\left\{-\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^T}\right\}.$$
- For the GEV, the rth moment of the score only exists if $r\xi > -1$ (exercise!).
- Smith (1985) established that the limiting behavior of the MLE depends on the value of the shape parameter ξ:
  - $\xi > -1/2$: MLE obeys standard theory;
  - $-1 < \xi \leq -1/2$: MLE is a solution to the score equation, but does not have the usual limiting distribution;
  - $\xi \leq -1$: MLE is not even a solution to the score equation.
- In most environmental problems, we find $\hat{\xi} \approx 0$ so can use MLE.
- Penalized likelihood methods (see exercise) or Bayesian inference can be used to force $|\xi| < 1/2$.

Peak-Over-Threshold (POT) Approach

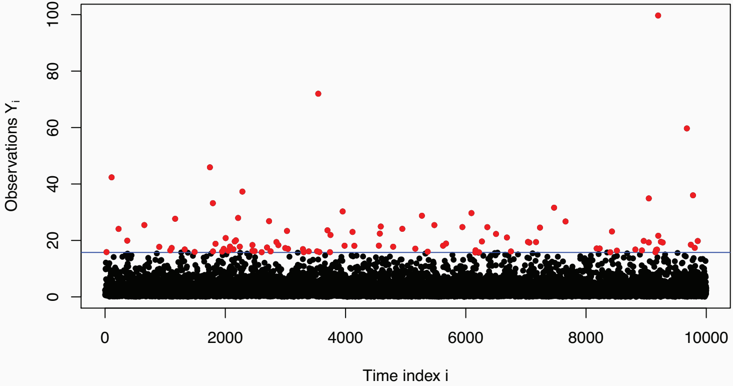
Issue With the Block Maxima Approach

- The annual maxima method can be inefficient if more data are available (as we keep only one observation per year).
- One solution: Take smaller blocks... But:
  - Loose interpretation if blocks do not correspond to natural cycles...
  - What does it mean to consider blocks of 53 days of observations, say?
  - Weekly or monthly blocks?
    - Sometimes too small for asymptotic model to be valid.
- Alternatives?
  - Peaks-Over-Thresholds (POT) method
  - r-largest order statistics method (not covered here)
- Both are special cases of a point process representation (not covered here), under which we use a Poisson process to approximate the occurrence of those values that exceed a (high) threshold.

Comments

- The fixed number of annual maxima has been replaced by a random number of exceedances over the threshold.
  - We now have more observations in the tail of the distribution.
- Dependence in the underlying series means that exceedances occur in clusters, which we may need to model (but we will not cover this aspect here and assumed independence throughout).
- For now, we suppose that the underlying series comprises independent identically distributed (i.i.d.) observations, whose maxima have a non-degenerate limit after renormalization.

Generalized Pareto (GP) Distribution



Pickands–Balkema–de Haan Theorem

- In extreme value theory generalized Pareto (GP) distribution plays a key role in the peaks-over-threshold (POT) approach and is the only possible limit for the marginal distribution of appropriately rescaled threshold exceedances.
- Provides a justification for fitting the generalized Pareto (GP) distribution to high threshold exceedances:

$$Y - u \mid Y > u \overset{d}{\sim} 1 - \left(1 + \xi \frac{y}{a(u)}\right)^{-1/\xi}, \quad \text{for large } u,$$

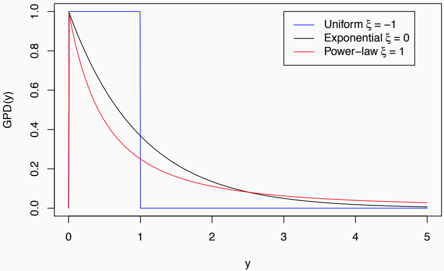
where ξ is the shape parameter and $a(u)$ is a scaling function dependent on the threshold u.

Generalized Pareto (GP) distribution)

- The GP(τ, ξ) is defined as:

$$H(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\tau}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\tau}\right), & \xi = 0 \end{cases}$$

- $\tau > 0$ is the scale parameter, $\xi \in \mathbb{R}$ is the shape parameter.
- **Special cases:**
    - $\xi > 0$: Power-law. For $\xi = 1$ standard Pareto distribution
    - $\xi = 0$: Exponential
    - $\xi < 0$: Upper bounded. For $\xi = -1$: Uniform distribution



Generalized Pareto (GP) Distribution

- **Uniform distribution**: For the $\text{Unif}(0, 1)$ distribution, with $F(y) = y$, $y \in [0, 1]$ properly renormalized exceedances over high thresholds converge in distribution to a GP distribution with $\xi = -1$.
- **Exponential distribution:** For the $\text{Exp}(\lambda)$ distribution, with $F(y) = 1 - \exp(-\lambda y)$, $y > 0, \lambda > 0$, properly renormalized exceedances over high thresholds converge in distribution to a GP distribution with $\xi = 0$.
- **Pareto distribution** For the $\text{Pareto}(\alpha)$ distribution, with $F(y) = 1 - y^{-\alpha}$, $y > 1, \alpha > 0$, properly renormalized exceedances over high thresholds converge in distribution to a GP distribution with $\xi = 1/\alpha > 0$.
- Under the same conditions on ξ as for the GEV, the GP distribution has finite moments and is regular for likelihood inference.

Likelihood estimation – POT approach (GP distribution)

- Choose a high threshold u, ensuring that the exceedances above u are well approximated by a GP distribution.
- The GP distribution likelihood is

$$L(\tau, \xi) = \prod_{i=1}^{N_u} \frac{1}{\tau} \left(1 + \xi \frac{y_i - u}{\tau}\right)_+^{-1/\xi - 1},$$

  where $y_1, \ldots, y_{N_u}$ is an enumeration of points exceeding the threshold u.
- ξ is the same as for maxima, but $\tau = \sigma + \xi(u - \mu)$ (τ depends on u)!
- Only two parameters (τ, ξ)? If not, what is the 3rd parameter?
  - Probability of exceedance $\zeta_u = \Pr(Y > u)$.
- Numerical maximization of GP distribution (log) likelihood is needed to obtain MLEs $\hat{\tau}, \hat{\xi}$.
- The MLE for ζ_u is simply obtained as $\hat{\zeta}_u = \frac{N_u}{n}$ (as $N_u \sim \mathrm{Bin}(n, \zeta_u)$).

Quantile and Return Levels

- **Return Levels for GP distribution:**
  - Return level z_N for an N-year period when data is observed on a daily scale.
  - For a given exceedance probability $p = \frac{1}{365N}$:

$$z_N = \begin{cases} u + \frac{\tau}{\xi}\left(\left(\frac{1}{365N}\right)^{-\xi} - 1\right), & \xi \neq 0, \\ u - \tau \log\left(\frac{1}{365N}\right), & \xi = 0 \end{cases}$$

    where u is the threshold, τ is the scale parameter, and ξ is the shape parameter.

- **Interpretation:**
  - The N-year return level z_N represents the value expected to be exceeded on average once every N years.
  - For example, a 100-year return level is the value expected to be exceeded once every 100 years.

Recap

Block Maxima Approach

- **Generalized Extreme Value (GEV) Distribution**
  - **Definition:**
    - Distribution function:

$$G(z;\mu,\sigma,\xi)=\begin{cases}\exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, & \text{if } \xi\neq 0,\\ \exp\left\{-\exp\left(-\frac{z-\mu}{\sigma}\right)\right\}, & \text{if } \xi=0.\end{cases}$$

  - **Special Cases:**
    - $\xi = 0$: Gumbel distribution
    - $\xi > 0$: Fréchet distribution
    - $\xi < 0$: Weibull distribution
  - **Log-Likelihood Function:** For N observations $z_1, \ldots, z_N$:

$$\ell(\mu,\sigma,\xi)=-\frac{\log\sigma}{N^{-1}}-\sum_{i=1}^{N}\left\{\left(1+\frac{1}{\xi}\right)\log\left(1+\xi\frac{z_i-\mu}{\sigma}\right)+\left(1+\xi\frac{z_i-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\}$$

  - **Parameter Estimation:**
    - Parameters (μ,σ,ξ) estimated via Maximum Likelihood Estimation (MLE).
  - **Model Diagnostics:**
    - Use probability plots and return level plots to assess model fit.

POT Approach

- **Generalized Pareto (GP) Distribution**
  - **Definition:**
    - Distribution function:

$$H(y;\tau,\xi) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\tau}\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\tau}\right), & \text{if } \xi = 0. \end{cases}$$

    - $\tau > 0$ is the scale parameter, $\xi \in \mathbb{R}$ is the shape parameter.
  - **Special Cases:**
    - $\xi = 0$: Exponential distribution
    - $\xi > 0$: Pareto distribution
    - $\xi < 0$: Bounded distribution
  - **Log-Likelihood Function:** For N_u exceedances $y_1, \ldots, y_{N_u}$ over threshold u:

$$\ell(\tau,\xi) = -\frac{\log \tau}{N_u^{-1}} - \sum_{i=1}^{N_u} \left\{ \left(1 + \frac{1}{\xi}\right) \log\left(1 + \xi \frac{y_i - u}{\tau}\right) + \left(1 + \xi \frac{y_i - u}{\tau}\right)^{-\frac{1}{\xi}} \right\}.$$

  - **Parameter Estimation:**
    - Parameters (τ, ξ) estimated via MLE.
  - **Model Diagnostics:**
    - Probability plots, return level plots, and evaluate the threshold choice.

Practical Examples Using R for IID Data

Extreme Value Analysis in R for IID Data

- See the code EVT.Rmd for the detailed study of extreme of precipitation data from western England using both block maxima approach and peaks over threshold approach
- Also see the code EVT-synthetic.Rmd for the application to synthetic data to verify the the max-domain of attractions for few of the standard distributions

Exercises

- Fit the block maxima approach to monthly maxima data. What differences you observed?
- Fit the threshold approach by changing the thresholds levels. What did you notice?

Non-Stationary Extremes

Regression Models for Extremes

- Classical regression models are useful to describe non-stationarity in the mean.
- Usually fitted by least squares (thanks to their optimality properties):
  - Identical to MLE under normality;
  - Best Linear Unbiased Estimator (BLUE) under broad assumptions.
- But:
  - What if noise is not normal?
    - LSE is not the MLE, so not fully efficient as $n \to \infty$.
  - What if noise has heavy tails?
    - LSE is sensitive to outliers, and increasingly inefficient as the tail becomes heavier.
  - What if mean is not finite?
    - LSE does not work! (to estimate an infinite mean...)
  - How to use the fitted curve to compute return levels?

Regression Models for Maxima

- By analogy with classical linear regression models, we can deal with non-stationary effects by including covariates into the usual extreme-value parameters.
- A suitable model for block maxima $Z_1, \ldots, Z_N$ may therefore be expressed as

$$Z_i \overset{\text{ind}}{\sim} \text{GEV}(\mu_i, \sigma_i, \xi_i),$$
$$\begin{cases} \mu_i = \mu(x_i; \beta_\mu) \\ \sigma_i = \sigma(x_i; \beta_\sigma) > 0 \\ \xi_i = \xi(x_i; \beta_\xi) \end{cases}$$

  where $\mu(x_i; \beta_\mu)$, $\sigma(x_i; \beta_\sigma) > 0$, and $\xi(x_i; \beta_\xi)$ are functions of a vector of covariates x_i and associated vectors of parameters β_μ, β_σ, β_ξ measuring the effect of covariates.
- For example, possible parametric models that depend on time t could be

$$\mu_i = \beta_{\mu 0} + \beta_{\mu 1}(t_i - t_0)$$
$$\sigma_i = \exp\{\beta_{\sigma 0} + \beta_{\sigma 1}(t_i - t_0)\}$$
$$\xi_i = \begin{cases} \beta_{\xi 0}, & t_i < t^\star \\ \beta_{\xi 1}, & t_i \geq t^\star \end{cases}$$

  where t_0 is a reference point in time and $t^\star$ is a change point.

Modeling Global Temporal Trends and Seasonality

- Global temporal trends can be modeled as a low-order polynomial:

$$\mu_i = \beta_{\mu 0} + \sum_{q=1}^{Q} \beta_{\mu q}(t_i - t_0)^q, \quad \sigma_i = \dots, \quad \xi_i = \dots,$$

  where Q is the polynomial order (usually $Q = 0$ or $Q = 1$).
- Seasonality (i.e., periodic effects) can be addressed by including harmonic terms. For example, in modeling monthly temperature maxima, one might use:

$$\mu_i = \beta_{\mu 0} + \sum_{k=1}^{K} \left(\beta_{\mu,2k-1} \sin\left(\frac{2k\pi t_i}{12}\right) + \beta_{\mu,2k} \cos\left(\frac{2k\pi t_i}{12}\right) \right),$$

  where t_i indicates the month (1–12).
    - The divisor 12 ensures yearly periodicity for monthly data, and it can be modified according to the periodicity of the data.
    - K determines the number of harmonic terms and is usually kept small for parsimony.
- Of course, one may combine seasonal terms, trends, and other covariates.

Modeling Considerations for Extremes

Parsimony is important, especially when dealing with extremes.

- Always compare the number of parameters to estimate with the "effective" number of data available (i.e., N maxima).
- The shape parameter ξ is especially difficult to estimate (i.e., with high uncertainty), so it is common to keep it constant:

$$\xi_i = \xi(x_i; \beta_\xi) = \beta_{\xi 0}, \quad i = 1, \ldots, N.$$

- Often in practice, $\sigma \propto \mu$, so it might make sense to fit a model where:

$$\sigma(x_i; \beta_\sigma) = c \times \mu(x_i; \beta_\mu), \quad i = 1, \ldots, N,$$

and c is a constant to estimate.

Estimation and Diagnostics

- Estimation may be performed by maximum likelihood inference. Let $g(z;\mu,\sigma,\xi)$ denote the GEV density. The log likelihood may be written as:

$$
\begin{aligned}
\ell(\beta_\mu,\beta_\sigma,\beta_\xi) &= \sum_{i=1}^{N} \log\left\{g(z_i;\mu_i,\sigma_i,\xi_i)\right\} \\
&= -\sum_{i=1}^{N} \log\left\{\sigma(x_i;\beta_\sigma)\right\} \\
&\quad - \sum_{i=1}^{N} \left(1+\frac{1}{\xi(x_i;\beta_\xi)}\right) \log\left(1+\xi(x_i;\beta_\xi)\frac{z_i-\mu(x_i;\beta_\mu)}{\sigma(x_i;\beta_\sigma)}\right) \\
&\quad + \sum_{i=1}^{N} \left(1+\xi(x_i;\beta_\xi)\frac{z_i-\mu(x_i;\beta_\mu)}{\sigma(x_i;\beta_\sigma)}\right)^{-1/\xi(x_i;\beta_\xi)} .
\end{aligned}
$$

- Numerical maximization with respect to the parameters is needed.
- Good starting values are essential for successful optimization, especially when the number of parameters is fairly large.

Non-Stationary Models for Threshold Exceedances

- Similar techniques are applicable for the POT models, but threshold selection can be tricky.
- Quantile regression may be useful for the choice of non-stationary thresholds.
- Let us study an example to illustrate the advantages, drawbacks, and difficulties of non-stationary POT approaches.

Non-Stationary Threshold Choice

- Linear regression:
  - Model: $u_i = \mathbf{x}_i^T\beta + \epsilon_i$, $\epsilon_i \overset{\text{iid}}{\sim} (0, \sigma^2)$, where $\mathbf{x}_i$ is a vector of covariates for the ith observation and β is a vector of unknown parameters.
  - Fit by least squares: $\hat{\beta}_2 = \min \sum_{i=1}^n (u_i - \mathbf{x}_i\beta)^2$.
  - $\mathbf{x}_i^T\hat{\beta}_2$ is an estimate of the conditional mean of Y_i.
- L_1 regression: Replace L_2 by L_1 loss function:
  - Fit by least absolute deviations: $\hat{\beta}_1 = \min \sum_{i=1}^n |u_i - \mathbf{x}_i\beta|$.
  - $\mathbf{x}_i^T\hat{\beta}_1$ is an estimate of the conditional median of Y_i.
- Quantile regression: Use "pinball" loss function $L(\cdot; p_u)$:
  - Estimated parameters: $\hat{\beta}_{p_u} = \min \sum_{i=1}^n L(u_i - \mathbf{x}_i\beta; p_u)$, where

$$L(y;\tau) = \begin{cases} p_u y, & y \geq 0; \\ -(1-p_u)y, & y < 0. \end{cases}$$

  - $\mathbf{x}_i^T\hat{\beta}_{p_u}$ is an estimate of the conditional p_u-quantile of u_i.

And so infernce is carried out in two steps;

- **Step 1:** quantile regression to obtain time-varying threshold $u(t)$.
- **Step 2:** GP distribution model to exceedances $Y_i - u(t) \mid Y_i > u(t)$.

Non-Stationary POT Model - Comments

- We may let the GP distribution parameters τ_u, ξ depend on covariates.
- Inference in two steps implies that uncertainty in $u(t)$ is neglected.
- How to perform model comparisons? Cannot use AIC/BIC when $u(t)$ is different.
- Fit may quite strongly depend on the choice of $u(t) \Rightarrow$ Bad!
- Care needed with model interpretation: with GP(τ_u, ξ), change of threshold $u \to v > u$ changes scale parameter

$$\tau_u \to \tau_v = \tau_u + \xi(v - u).$$

- So for example, if t denotes time, and x denotes another covariate, the model

$$\tau_u = s_1(t), \quad \xi = s_2(x).$$

  at threshold u will become

$$\tau_v = s_1(t) + s_2(x)(v - u), \quad \xi = s_2(x).$$

  at threshold $v > u$. So interpretation depends on threshold: avoid it.

Practical Examples Using R for non-IID Data

Extreme Value Analysis in R for non-IID Data

- See the code EVT-NonStationary.Rmd for the detailed study of non-stationary extreme of precipitation data from western England using both non-stationary block maxima approach and non-stationary peaks over threshold approach

Exercises

- Try other models and compare them?

There is Still Much More...

- Univariate extremes
    - point process approach
    - r-largest order statistics approach
    - extremes of temporally dependent extremes
- Multivariate extremes
- Spatio-temporal modeling of extremes