Risk Analytics

Extreme Value Theory
Peaks-over-Threshold Method

Frédéric Aviolat & Juraj Bodik

2025

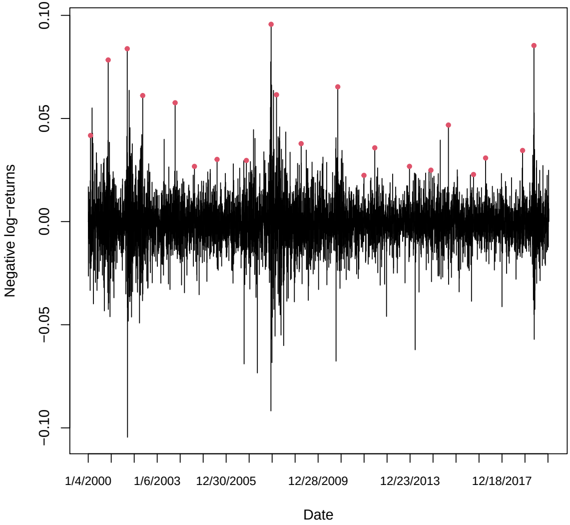
Previous Approach

- We saw how to analyse extreme values by their maxima.
- But this replaced a whole block of n values by a single one.
  - Lots of data loss, since n has to be "large enough"
  - Ignores behaviour inside a block (iid)

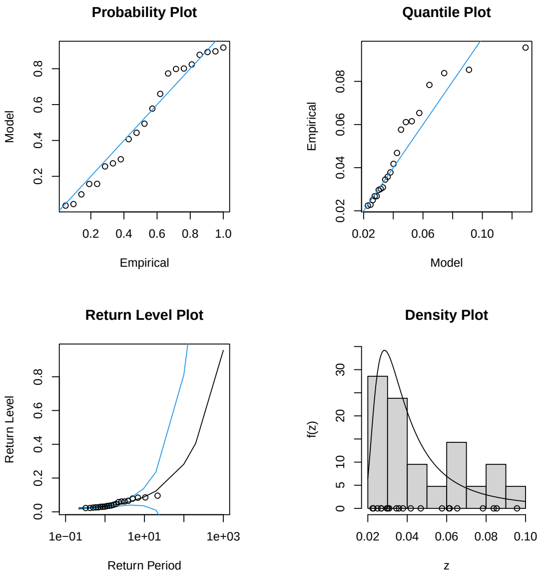
Threshold Methods

- Idea: instead of looking only at the highest, use all the values which are "high enough":
- The modelled values are $\{X_i : X_i > u\}$ for a high threshold u.

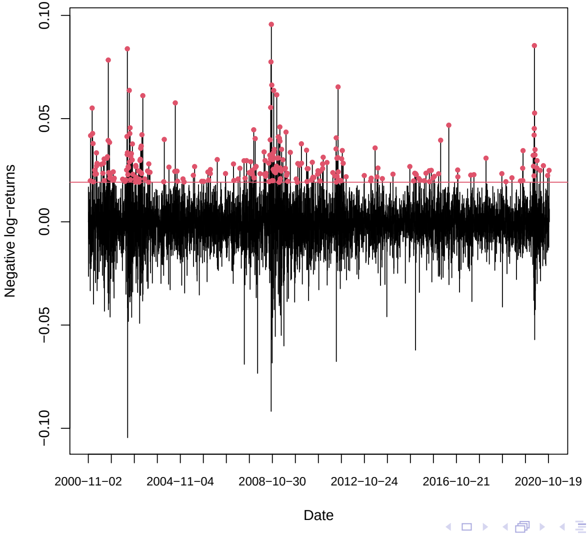
Nestlé negative log-returns (block maxima approach)



Nestlé negative log-returns (block maxima approach II)



Nestlé negative log-returns (Peaks-Over-Threshold approach)



Generalized Pareto Distribution

- The GPD is a two parameter distribution with df

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)_+^{-1/\xi} & \xi \neq 0\,, \\ 1 - \exp(-x/\beta) & \xi = 0\,, \end{cases}$$

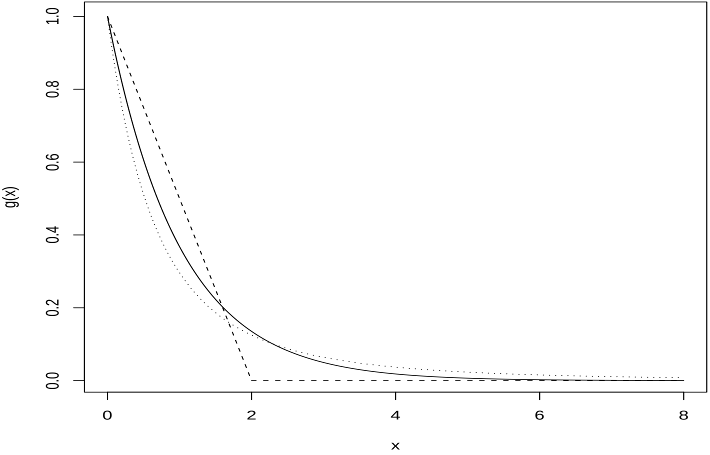
  where $\beta > 0$ and $a_+ = \max(a, 0)$.
  The support of $G_{\xi,\beta}$ is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

- Particular cases:

$$\left\{ \begin{array}{c} \xi > 0 \\ \xi = 0 \\ \xi < 0 \end{array} \right\} \text{ is equivalent to } \left\{ \begin{array}{l} \text{Pareto} \\ \text{Exponential} \\ \text{Pareto type II} \end{array} \right\}.$$

- **Moments.** For $\xi > 0$, the distribution is heavy tailed. $\mathbb{E}\left(X^k\right)$ does not exist for $k \geq 1/\xi$.

GPD Densities



Solid line corresponds to $\xi = 0$ (exponential); dotted line is $\xi = 0.5$ (Pareto); dashed line is $\xi = -0.5$ (Pareto type II). $\beta = 1$.

Peaks-over-threshold method

- **The excess distribution:** Given that a loss exceeds a high threshold, by how much can the threshold be exceeded?
- Let u be the high threshold and define the excess distribution above the threshold u to have the df

$$F_u(x) = \mathbb{P}\left(X - u \le x \mid X > u\right) = \frac{F(x+u) - F(u)}{1 - F(u)},$$

  for $0 \le x < x_F - u$ where $x_F \le \infty$ is the right endpoint of F.
- Extreme value theory suggests the GPD is a natural approximation for this distribution.

Examples

1. **Exponential.** $F(x) = 1 - e^{-\lambda x}$, $\lambda > 0$, $x \geq 0$. We find that

   $$F_u(x) = F(x)\,, \quad x \geq 0\,.$$

   The "lack–of–memory" property.

2. **GPD.** $F(x) = G_{\xi,\beta}(x)$. This time

   $$F_u(x) = G_{\xi,\beta+\xi u}(x)\,,$$

   where $0 \leq x < \infty$ if $\xi \geq 0$ and $0 \leq x < -\beta/\xi - u$ if $\xi < 0$. The excess distribution of a GPD remains a GPD with the same shape parameter; only the scaling changes.

What about the general case?

Asymptotics of Excess Distribution

Theorem. (Pickands–Balkema–de Haan (1974/75)) We can find a function $\beta(u)$ such that

$$\lim_{u \to x_F} \sup_{0 \le x < x_F - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0 \,,$$

if and only if $F \in \text{MDA}\,(H_\xi)$, $\xi \in \mathbb{R}$.

Essentially all the common continuous distributions used in risk management or insurance mathematics are in $\text{MDA}\,(H_\xi)$ for some value of ξ.

Using Pickands–Balkema–de Haan on data

- "For a wide class of distributions, the distribution of the excesses over high thresholds can be approximated by the GPD."
- This result suggests we choose u high and assume the limit result is more or less exact

$$F_u(x) \approx G_{\xi,\beta}(x)\,,$$

  for some ξ and β.
- To estimate these parameters we fit the GPD to the excess amounts over the threshold u, i.e. to $\{X_i - u \,|\, X_i > u\}$.
- Standard properties of maximum likelihood estimators apply if $\xi > -0.5$
- To implement the POT method we must choose a suitable threshold u, using data–analytic tools (e.g. mean excess plot) to help.

When does $F \in \mathrm{MDA}\,(H_\xi)$ hold?

- Fréchet case: $F \in \mathrm{MDA}\,(H_{\xi>0})$
  If the tail of the df F decays like a power function then $F \in \mathrm{MDA}\,(H_{\xi>0})$. Heavy-tailed distributions such as Pareto, Burr, log-gamma, Cauchy and t–distributions as well as various mixture models. Not all moments are finite.
- Gumbel Case: $F \in \mathrm{MDA}\,(H_0)$
  Essentially it contains distributions whose tails decay roughly exponentially and we call these distributions light–tailed. All moments exist for distributions in the Gumbel class.
  Examples are the normal, log-normal, exponential and gamma.
- Weibull Case: $F \in \mathrm{MDA}\,(H_{\xi<0})$
  If the tail of the df F is bounded above then $F \in \mathrm{MDA}\,(H_{\xi<0})$. Examples are the uniform, beta.

The Mean Excess Plot

The mean excess function of a random variable X is

$$e(u) = \mathbb{E}(X - u \mid X > u).$$

It is the mean of the excess distribution function $F_u(x)$ above the threshold u, expressed as a function of u.

GPD Model:
Excess losses over threshold u are exactly GPD with $\xi < 1$, i.e., $X - u \mid X > u \sim \text{GPD}(\xi, \beta)$. It is easily shown that for any higher threshold $v \geq u$

$$e(v) = \mathbb{E}(X - v \mid X > v) = \frac{\beta + \xi(v - u)}{1 - \xi},$$

so the mean excess function is linear in v.

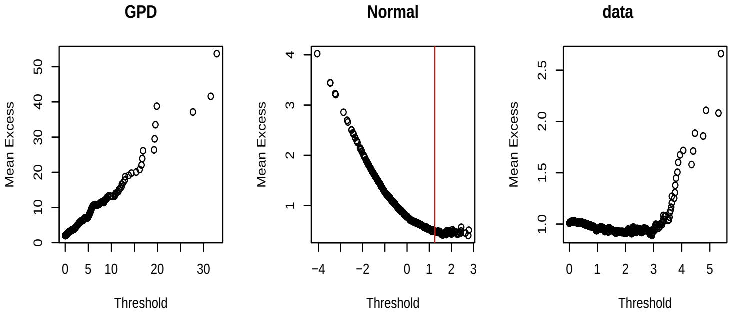
Choosing a threshold in practice

- Given some data, what is an appropriate threshold?
- The sample mean excess plot estimates $e(u)$ in the region where we have data:

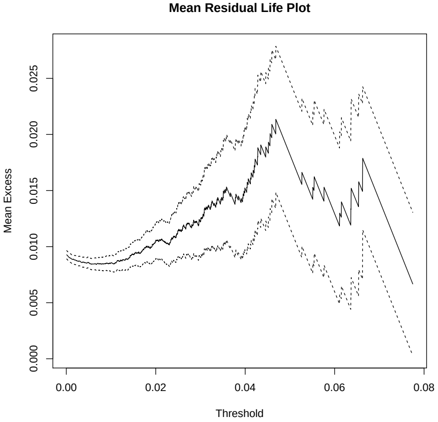
$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)_+}{\sum_{i=1}^{n} \mathbb{1}_{(X_i > u)}},$$

  We seek a threshold u, above which the plot is roughly linear.
- If we can find such a threshold, the result of theorem Pickands-Balkema-De Haan could be applied above that threshold.
- The plot is erratic for large u, when the averaging is over very few excesses. It is often better to omit these from the plot.
- **Bias-variance tradeoff:**
  - threshold too low $\Rightarrow$ bias because of the model asymptotics being invalid;
  - threshold too high $\Rightarrow$ variance is large due to few data points.

Examples of Mean Excess Plots

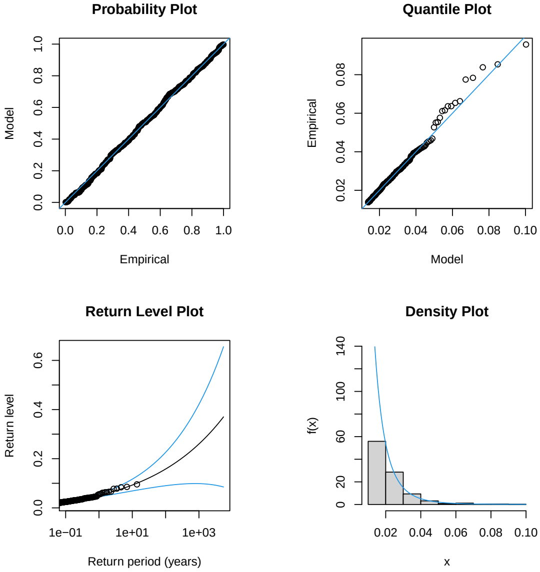


Nestlé negative log-returns: MRL-plot



Threshold could be put at 0.01? 0.04 looks maybe a bit high.

Nestlé negative log-returns: MRL-plot



Modelling Tails of Distributions

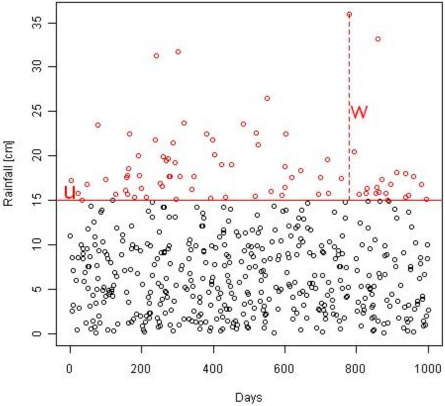
- Under our assumption that $F_u = G_{\xi,\beta}$ for some u, ξ, β we have, for $x \geq u$,

$$
\begin{aligned}
\bar{F}(x) &= \mathbb{P}(X > u)\mathbb{P}(X > x \mid X > u) \\
&= \bar{F}(u)\mathbb{P}(X - u > x - u \mid X > u) \\
&= \bar{F}(u)\bar{F}_u(x - u) \\
&= \bar{F}(u)\left(1 + \xi\frac{x - u}{\beta}\right)^{-1/\xi}
\end{aligned}
$$

  which if we know $\bar{F}(u)$, gives us a formula for estimating tail probabilities.

- This formula may be used to derive formulas for risk measures like VaR and expected shortfall (cf. Risk Measures Module).

Peaks-Over-Threshold approach in practice (I)



Peaks-Over-Threshold approach in practice (II)

Suppose $X_1, \dots, X_n$ a sample of iid data coming from the distribution F. Considering exceedances above some high threshold u, we have that

- The number of exceedances N_u above the high threshold u follows a Poisson distribution with parameter λ,
  and, independently,
- the sizes $W_i = X_i - u$, $i = 1, \dots, N_u$ of these N_u exceedances follow a GPD with parameters ξ and β.

In practice, we estimate the three parameters λ, ξ, β by maximizing the likelihood. Inference can be done separately for λ (from the Poisson likelihood) and for ξ, β (from the GPD likelihood).
The MLE of λ is $\hat{\lambda} = N_u/n$.

Dependence and clustering (I)

- Independence of widely separated extremes seems reasonable in most applications.
- But they almost always display short-range dependence in which clusters of extremes occur together (flow level maxima often occur during the strongest storm of the year).
- In these cases, it seems unrealistic to assume independence within each year.
- In the threshold method, the usual solution is to fit the point process model to the maxima of the clusters.

Dependence and clustering (II)

- An important practical problem is the identification of clusters from data.
- In this course both the choice of a suitable high threshold and of a method to identify independent clusters are based on a runs approach.
- Suppose that a series $X_1, \ldots, X_n$ has short range dependence, so extremes occur in clusters of mean size $1/\theta$, where $0 < \theta \leq 1$; θ is called the "extremal index".
- $\theta = 1$ corresponds to asymptotically independent clusters of size 1, whereas $\theta \approx 0$ means extremes tend to cluster.

Declustering method

For a fixed threshold u, we identify different groups of exceedances over u as independent clusters only if there are at least v consecutive observations under u between them. An estimate number of independent clusters for a sample of n observations is

$$C_n(u, v) = \sum_{j=1}^{n-v} Z_j(1 - Z_{j+1}) \cdots (1 - Z_{j+v}), \tag{1}$$

where $Z_j = \mathbb{1}_{(X_j > u)}$. The estimate of extremal index is then $\hat{\theta} = C_n(u, v)/N_u$, where N_u is the number of exceedances over u. Two important remaining points are the choices of u and v. In applications, we use sensible choices based on the specific situation.

Declustering method example

We consider the first-order process ARMAX(ρ) with parameter $0 \leq \rho \leq 1$ so that

$$Y_i = \max(\rho Y_{i-1}, \epsilon_i),$$

where $\{\epsilon_i\}$ are iid from F_ϵ with $F_\epsilon(x) = \exp\{-(1-\rho)/x\}$, for $x > 0$. It can be shown that $\theta = 1 - \rho$.

cf. in R, see doc of function `decluster` *in package* `extRemes`.

Comments

- POT modelling is a bit more flexible than block maxima, because the threshold can be moved continuously.
- If there is seasonality, a moving threshold will be required, and this requires separate modelling.
- In practice the benefit of including more high values in a threshold approach is limited: often one sees that the standard errors are about the same as when using block maxima.
- The real potential benefit of threshold methods is the possibility of more detailed modelling of temporal evolution of extremes, by looking at clusters of extremes, which are missed using the block maximum approach.