Estimation de quantiles extrêmes pour des lois à queues lourdes et légères

PAR

Jonathan EL METHNI

en collaboration avec

Laurent GARDES & Armelle GUILLOU & Stéphane GIRARD

Séminaire de Strasbourg 13 Mars 2012
1 Introduction to the extreme value theory
   - Motivations
   - Domains of attraction

2 Model

3 Estimators
   - Definition
   - Asymptotic properties

4 Illustration on simulations and on a real data set
   - Simulations
   - A real data set

5 Concluding remarks
Motivations

Statistical Framework

Let $X_1, \ldots, X_n$ be a sample of independent and identically distributed random variables driven from $X$ with cumulative distribution function $F$, and let $X_{1,n} \leq \cdots \leq X_{n,n}$ denote the order statistics associated to this sample.

- We want to estimate the extreme quantile $x_{p_n}$ of order $p_n$ associated to the random variable $X \in \mathbb{R}$ defined by
  $$x_{p_n} = \overline{F}^{-1}(p_n) = \inf\{x, \overline{F}(x) \leq p_n\},$$
  with $p_n \to 0$ when $n \to \infty$. The function $\overline{F}^{-1}$ is the generalized inverse of the non-increasing function $\overline{F} = 1 - F$.
- **Difficulty**: If $np_n \to 0$ then $\mathbb{P}(x_{p_n} > X_{n,n}) \to 1$. 

Fisher-Tippett-Gnedenko theorem (1943)

Under some conditions of regularity on the cumulative distribution function $F$, there exists a real parameter $\gamma$ and two sequences $(a_n)_{n \geq 1} > 0$ and $(b_n)_{n \geq 1} \in \mathbb{R}$ such that for all $x \in \mathbb{R}$,

$$\lim_{n \to \infty} \mathbb{P} \left( \frac{X_{n,n} - b_n}{a_n} \leq x \right) = \mathcal{H}_\gamma(x),$$

with

$$\mathcal{H}_\gamma(x) = \begin{cases} \exp\left(-(1 + \gamma x)_+^{-1/\gamma}\right) & \text{if } \gamma \neq 0, \\ \exp(-e^{-x}) & \text{if } \gamma = 0, \end{cases}$$

where $y_+ = \max(0, y)$. 

Principal results on extreme value theory
Three domains of attraction

- $\mathcal{H}_\gamma$ is called the cumulative distribution function of the extreme value distribution.
- If $F$ verifies the Fisher-Tippett-Gnedenko theorem, we say that $F$ belongs to the domain of attraction of $\mathcal{H}_\gamma$.

<table>
<thead>
<tr>
<th>Fréchet ($\gamma &gt; 0$)</th>
<th>Gumbel ($\gamma = 0$)</th>
<th>Weibull ($\gamma &lt; 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto</td>
<td>Normal</td>
<td>Uniform</td>
</tr>
<tr>
<td>Student</td>
<td>Exponential</td>
<td>Beta</td>
</tr>
<tr>
<td>Burr</td>
<td>Lognormal</td>
<td></td>
</tr>
<tr>
<td>Fréchet</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weibull</td>
<td></td>
</tr>
</tbody>
</table>
Three domains of attraction

- $\gamma$ is called the extreme value index.

Extreme value distribution with $\gamma = -1$, $\gamma = 0$ and $\gamma = 1$
Fréchet maximum domain of attraction

Heavy-tailed distributions

All cumulative functions which belong to the Fréchet maximum domain of attraction denoted by $\mathcal{D}(\text{Fréchet})$ can be rewritten as

$$\overline{F}(x) = x^{-1/\gamma} \ell(x),$$

where $\gamma > 0$ and $\ell(x)$ is a slowly varying function at infinity i.e. $\forall \lambda \geq 1$,

$$\lim_{x \to \infty} \frac{\ell(\lambda x)}{\ell(x)} \to 1.$$ 

- $\overline{F}(x)$ is said to be regularly varying at infinity with index $-1/\gamma$.
- This property is denoted by $\overline{F} \in \mathcal{R}_{-1/\gamma}$.
- $F$ is called a Pareto-type distribution.
To make inference on the distribution tail, most approaches consist in using the $k_n$ upper order statistics.

Since the tail information is only contained in the extreme upper part of the sample.

$(k_n)$ is an intermediate sequence of integers i.e. such that

$$(H1) \quad \lim_{n \to \infty} k_n = \infty \quad \text{and} \quad \lim_{n \to \infty} k_n/n = 0.$$ 

**Definition**

Let us consider $(k_n)$ an intermediate sequence of integers such that $(H1)$ holds the Hill estimator is defined in terms of log-spacings by

$$H_n(k_n) = \frac{1}{k_n - 1} \sum_{i=1}^{k_n-1} i \left( \log(X_{n-i+1,n}) - \log(X_{n-i,n}) \right) \xrightarrow{P} \gamma > 0.$$
Light-tailed distributions

There is no simple representation for distributions which belong to \( D(Gumbel) \). We focus on an interesting sub-family called Weibull tail-distributions

\[
\bar{F}(x) = \exp(-H(x)),
\]

where \( H^{-}(t) = \inf\{x, H(x) \geq t\} \in \mathcal{R}_\alpha \).

- The tail of such distributions is driven by the shape parameter \( \alpha > 0 \) called the Weibull tail-coefficient.
- Recents estimators of \( \alpha \) are based on the log-spacings between the \( k_n \) upper order statistics.
- All these estimators are thus similar to the Hill statistic.
Model established by L. Gardes, S. Girard & A. Guillou (2011)

### First order condition \((A_1(\tau, \theta))\)

Let us consider the family of survival distribution functions defined as:

\[
(A_1(\tau, \theta)) \quad \bar{F}(x) = \exp(-K_\tau^{-}(\log H(x))) \text{ for } x \geq x^* > 0 \text{ and } \tau \in [0, 1]
\]

with

- \(K_\tau(y) = \int_1^y u^{\tau-1} du \text{ where } y \in \mathbb{R},\)
- \(H\) an increasing function such that \(H^{-} \in \mathcal{R}_\theta\) where \(\theta > 0.\)

Let us consider the three cases:

- \(\tau = 0\) : Under \((A_1(0, \theta))\), \(\bar{F}(x) = \exp(-H(x))\) is the survival function of a Weibull-tail distribution.

- \(\tau = 1\) : \((A_1(1, \theta))\) entails \(\bar{F}(x) = e/H(x) = x^{-1/\theta} \tilde{\ell}(x)\) is the survival function of a Pareto-type distribution, where \(\tilde{\ell}\) is a slowly varying function.

- \(\tau \in (0, 1)\) : Corresponds to distribution tails lighter than Pareto tails but heavier than Weibull tails.
Fréchet / Gumbel

Proposition

- $F$ verifies $(A_1(0, \theta))$ if and only if $F$ is a Weibull-tail distribution function with Weibull tail-coefficient $\theta$.
- If $F$ verifies $(A_1(\tau, \theta))$, $\tau \in [0, 1)$ and if $H$ is twice differentiable then $F$ belongs to the Gumbel maximum domain of attraction.
- $F$ verifies $(A_1(1, \theta))$ if and only if $F$ is in the Fréchet maximum domain of attraction with tail-index $\theta$.

- The tail heaviness of $\bar{F}$ is mainly driven by $\tau \in [0, 1]$ and secondarily by $\theta > 0$.
- Thus the larger is $\tau$, the heavier is the tail.
- The parameter $\tau$ allows us to represent a large panel of distribution tails ranging from Weibull-type to Pareto-type.
Definition

Denoting by \((k_n)\) an intermediate sequence of integers, the following estimator of \(\theta\) is considered:

\[
\hat{\theta}_{n,\tau}(k_n) = \frac{H_n(k_n)}{\mu_{\tau}(\log(n/k_n))},
\]

with, for all \(t > 0\),

\[
\mu_{\tau}(t) = \int_0^\infty (K_{\tau}(x + t) - K_{\tau}(t)) e^{-x} \, dx.
\]
Estimators of $\theta$ and $x_{p_n}$

**Definition**

Denoting by $(k_n)$ an intermediate sequence of integers, the following estimator of $\theta$ is considered:

$$\hat{\theta}_{n,\tau}(k_n) = \frac{H_n(k_n)}{\mu_{\tau}(\log(n/k_n))},$$

with, for all $t > 0$,

$$\mu_{\tau}(t) = \int_0^\infty (K_{\tau}(x + t) - K_{\tau}(t)) e^{-x} \, dx.$$

**Definition**

An estimator of the extreme quantile $x_{p_n} = F^{\leftarrow}(p_n)$ with $p_n \to 0$ is derived

$$\hat{x}_{p_n,\hat{\theta}_{n,\tau}(k_n)} = X_{n-k_n+1,n} \exp \left( \hat{\theta}_{n,\tau}(k_n) [K_{\tau}(\log(1/p_n)) - K_{\tau}(\log(n/k_n))] \right).$$
Asymptotic properties

Second order condition \((A_2(\rho))\)

To establish the asymptotic normality of the estimators, a second-order condition on \(\ell\) is necessary

\((A_2(\rho))\) There exist \(\rho < 0\) and \(b(x) \to 0\) such that uniformly locally on \(\lambda \geq \lambda_0 > 0\)

\[
\log \left( \frac{\ell(\lambda x)}{\ell(x)} \right) \sim b(x)K_\rho(\lambda), \text{ when } x \to \infty,
\]

with \(|b|\) asymptotically decreasing. It can be shown that necessarily \(|b| \in \mathcal{R}_\rho\).

- The second order parameter \(\rho < 0\) tunes the rate of convergence of \(\ell(\lambda x)/\ell(x)\) to 1.
- Condition \((A_2(\rho))\) is the cornerstone in all the proofs of asymptotic normality for extreme value estimators.
Asymptotic properties

Theorem (a) : Asymptotic normality of $\hat{\theta}_{n,\tau}(k_n)$

Suppose that $(A_1(\tau, \theta))$ and $(A_2(\rho))$ hold. Let $(k_n)$ be an intermediate sequence such that (H1) holds and

$$\sqrt{k_n} b(\exp K_\tau(\log(n/k_n))) \to \lambda.$$

Then, introducing $d_{\tau,\rho} = 1$ if $\tau \in [0, 1)$ and $d_{1,\rho} = 1/(1 - \rho)$, we have

$$\sqrt{k_n} \left( \hat{\theta}_{n,\tau}(k_n) - \theta - d_{\tau,\rho} b(\exp K_\tau(\log(n/k_n))) \right) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$
Asymptotic properties

**Theorem (a) : Asymptotic normality of $\hat{\theta}_{n,\tau}(k_n)$**

Suppose that $(A_1(\tau, \theta))$ and $(A_2(\rho))$ hold. Let $(k_n)$ be an intermediate sequence such that (H1) holds and

$$\sqrt{k_n} \, b(\exp K_{\tau}(\log(n/k_n))) \to \lambda.$$ 

Then, introducing $d_{\tau,\rho} = 1$ if $\tau \in [0, 1)$ and $d_{1,\rho} = 1/(1 - \rho)$, we have

$$\sqrt{k_n} \left( \hat{\theta}_{n,\tau}(k_n) - \theta - d_{\tau,\rho} b(\exp K_{\tau}(\log(n/k_n))) \right) \overset{d}{\to} \mathcal{N}(0, \theta^2).$$

**Theorem (b) : Asymptotic normality of $\hat{x}_{p_n,\hat{\theta}_{n,\tau}(k_n)}$**

Suppose the assumptions of Theorem (a) hold with $\lambda = 0$. If, moreover,

$$(\log(n/k_n))^{1-\tau} (K_{\tau}(\log(1/p_n)) - K_{\tau}(\log(n/k_n))) \to \infty$$

then,

$$\frac{\sqrt{k_n}}{K_{\tau}(\log(1/p_n)) - K_{\tau}(\log(n/k_n))} \left( \frac{\hat{x}_{p_n,\hat{\theta}_{n,\tau}(k_n)}}{x_{p_n}} - 1 \right) \overset{d}{\to} \mathcal{N}(0, \theta^2).$$
Contribution

- Model \((A_1(\tau, \theta))\) provides a new tool for the analysis of tail estimators based on log-spacings.
- Estimators of \(\theta\) and \(x_{pn}\) both depend on the unknown parameter \(\tau\), making them useless in practical situations.
Contribution

- Model \((A_1(\tau, \theta))\) provides a new tool for the analysis of tail estimators based on log-spacings.
- Estimators of \(\theta\) and \(x_{pn}\) both depend on the unknown parameter \(\tau\), making them useless in practical situations.

Objectives and Contribution

1. Estimate \(\tau\) independently from \(\theta\).
Contribution

Model $\mathbf{A}_1(\tau, \theta)$ provides a new tool for the analysis of tail estimators based on log-spacings.

Estimators of $\theta$ and $x_{p_n}$ both depend on the unknown parameter $\tau$, making them useless in practical situations.

Objectives and Contribution

1. Estimate $\tau$ independently from $\theta$.
2. Replace $\tau$ by $\hat{\tau}_n$ in $\hat{\theta}_{n,\tau}$. 
Contribution

- Model \((A_1(\tau, \theta))\) provides a new tool for the analysis of tail estimators based on log-spacings.
- Estimators of \(\theta\) and \(x_{p_n}\) both depend on the unknown parameter \(\tau\), making them useless in practical situations.

Objectives and Contribution

1. Estimate \(\tau\) independently from \(\theta\).
2. Replace \(\tau\) by \(\hat{\tau}_n\) in \(\hat{\theta}_{n,\tau}\).
3. Replace \(\tau\) by \(\hat{\tau}_n\) and \(\hat{\theta}_{n,\tau}\) by \(\hat{\theta}_{n,\hat{\tau}_n}\) in \(\hat{x}_{p_n,\hat{\theta}_{n,\tau}}\).
Let \((k_n)\) and \((k'_n)\) with \(k'_n > k_n\) be two intermediate sequences of integers such that \(\left\{\begin{array}{l}
\hat{\theta}_{n,\tau}(k_n) \xrightarrow{P} \theta \\
\hat{\theta}_{n,\tau}(k'_n) \xrightarrow{P} \theta
\end{array}\right.\) we have

\[
\frac{\hat{\theta}_{n,\tau}(k_n)}{\hat{\theta}_{n,\tau}(k'_n)} = \frac{H_n(k_n) \mu_\tau(\log(n/k'_n))}{H_n(k'_n) \mu_\tau(\log(n/k_n))} \xrightarrow{P} 1.
\]

Then,

\[
\frac{H_n(k_n)}{H_n(k'_n)} \xrightarrow{P} \frac{\mu_\tau(\log(n/k_n))}{\mu_\tau(\log(n/k'_n))} =: \psi(\tau; \log(n/k_n), \log(n/k'_n)),
\]

where

\[
\psi(x; t, t') = \frac{\mu_x(t)}{\mu_x(t')} \quad \text{is a bijection from } \mathbb{R} \text{ to } (-\infty, \exp(t - t')).
\]
The following estimator of $\tau$ is considered

$$\hat{\tau}_n = \begin{cases} \psi^{-1} \left( \frac{H_n(k_n)}{H_n(k'_n)}; \log(n/k_n), \log(n/k'_n) \right) & \text{if } \frac{H_n(k_n)}{H_n(k'_n)} < \frac{k'_n}{k_n} \\ u & \text{if } \frac{H_n(k_n)}{H_n(k'_n)} \geq \frac{k'_n}{k_n} \end{cases}$$

where $u$ is the realization of a standard uniform distribution.
Estimators

**Definition**

The following estimator of $\tau$ is considered

$$
\hat{\tau}_n = \begin{cases} 
\psi^{-1} \left( \frac{H_n(k_n)}{H_n(k'_n)}, \log(n/k_n), \log(n/k'_n) \right) & \text{if } \frac{H_n(k_n)}{H_n(k'_n)} < \frac{k'_n}{k_n} \\
u & \text{if } \frac{H_n(k_n)}{H_n(k'_n)} \geq \frac{k'_n}{k_n}
\end{cases}
$$

where $u$ is the realization of a standard uniform distribution.

Plugging $\hat{\tau}_n$ in $\hat{\theta}_{n,\tau}$ we obtain a new estimator of $\theta$

$$
\hat{\theta}_{n,\hat{\tau}_n} = \frac{H_n(k_n)}{\mu_{\hat{\tau}_n}(\log(n/k_n))}.
$$
**Definition**

The following estimator of $\tau$ is considered

$$
\hat{\tau}_n = \begin{cases} 
\psi^{-1} \left( \frac{H_n(k_n)}{H_n(k'_n)}; \log(n/k_n), \log(n/k'_n) \right) & \text{if } \frac{H_n(k_n)}{H_n(k'_n)} < \frac{k'_n}{k_n} \\
u & \text{if } \frac{H_n(k_n)}{H_n(k'_n)} \geq \frac{k'_n}{k_n}
\end{cases}
$$

where $u$ is the realization of a standard uniform distribution.

Plugging $\hat{\tau}_n$ in $\hat{\theta}_{n,\tau}$ we obtain a new estimator of $\theta$

$$
\hat{\theta}_{n,\tau} = \frac{H_n(k_n)}{\mu_{\hat{\tau}_n}(\log(n/k_n))}.
$$

Replacing $\tau$ by $\hat{\tau}_n$ and $\hat{\theta}_{n,\tau}$ by $\hat{\theta}_{n,\hat{\tau}_n}$ yields a new estimator of extreme quantiles

$$
\hat{x}_{p_n,\hat{\tau}_n,\hat{\theta}_{n,\hat{\tau}_n}} = X_{n-k_n+1,n} \exp \left( \hat{\theta}_{n,\hat{\tau}_n} [K_{\hat{\tau}_n}(\log(1/p_n)) - K_{\hat{\tau}_n}(\log(n/k_n))] \right).
$$
Asymptotic properties

**Theorem 1 : Asymptotic normality of \( \hat{\tau}_n \)**

Suppose that \( (A_1(\tau, \theta)) \) and \( (A_2(\rho)) \) hold. If \( (k_n) \) and \( (k'_n) \) are two intermediate sequences of integers such that \( (H1) \) holds and

\[
\frac{k_n}{k'_n} \rightarrow 0, \quad \sqrt{k_n} b(\exp K_\tau(\log n/k'_n)) \rightarrow 0, \\
\log(n/k'_n) (\log_2(n/k_n) - \log_2(n/k'_n)) \rightarrow \infty, \\
\sqrt{k_n} (\log_2(n/k_n) - \log_2(n/k'_n)) \rightarrow \infty,
\]

then

\[
\sqrt{k_n} (\log_2(n/k_n) - \log_2(n/k'_n)) (\hat{\tau}_n - \tau) \overset{d}{\rightarrow} \mathcal{N}(0, 1).
\]

where \( \log_2 = \log(\log) \).
Asymptotic properties

Theorem 2 : Asymptotic normality of \( \hat{\theta}_n, \hat{\tau}_n \)

Suppose the assumptions of Theorem 1 hold. If, moreover,

\[
\frac{\log_2(n/k_n) - \log_2(n/k'_n)}{\log_2(n/k_n)} \to 0,
\]

\[
\sqrt{k_n} \frac{\log_2(n/k_n) - \log_2(n/k'_n)}{\log_2(n/k_n)} \to \infty,
\]

then

\[
\frac{\sqrt{k_n} (\log_2(n/k_n) - \log_2(n/k'_n))}{\log_2(n/k_n)} \left( \hat{\theta}_n, \hat{\tau}_n - \theta \right) \xrightarrow{d} \mathcal{N}(0, \theta^2).
\]
Asymptotic properties

Theorem 2: Asymptotic normality of $\hat{\theta}_n, \hat{\tau}_n$

Suppose the assumptions of Theorem 1 hold. If, moreover,

$$\frac{\log_2(n/k_n) - \log_2(n/k'_n)}{\log_2(n/k_n)} \rightarrow 0,$$

$$\sqrt{k_n} \left( \frac{\log_2(n/k_n) - \log_2(n/k'_n)}{\log_2(n/k_n)} \right) \rightarrow \infty,$$

then

$$\frac{\sqrt{k_n} (\log_2(n/k_n) - \log_2(n/k'_n))}{\log_2(n/k_n)} \left( \hat{\theta}_n, \hat{\tau}_n - \theta \right) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

The estimation of $\tau$ has a cost in terms of rates of convergence.

**Condition** \( \frac{\log_2(n/k_n) - \log_2(n/k'_n)}{\log_2(n/k_n)} \rightarrow 0, \)

implies that $\hat{\theta}_n, \hat{\tau}_n$ converges slower than $\hat{\theta}_n, \tau$. 
Asymptotic properties

Choice for \((k_n)\) and \((k'_n)\) satisfying Theorem 2

- If \(\tau = 0\) and \(\rho > -1\), it is not possible to choose sequences \((k_n)\) and \((k'_n)\) satisfying the above assumptions.
- If \(\tau \in (0, 1]\) or if \(\tau = 0\) and \(\rho < -1\), a possible choice for the two intermediate sequences is
  \[
  \begin{align*}
  \log(k_n) &= aK_\tau(\log(n)) \\
  \log(k'_n) &= a'K_\tau(\log(n))
  \end{align*}
  \]
  with the following restrictions on \((a, a') \in \mathbb{R}^2\)
  \[
  \begin{cases}
  0 < a < a' < 2\rho/(2\rho - 1) & \text{if } \tau = 1 \\
  0 < a < a' < -2\rho & \text{if } 0 < \tau < 1 \\
  2 < a < a' < -2\rho & \text{if } \tau = 0
  \end{cases}
  \]
- Finally, in the case where \(\tau = 0\) and \(\rho = -1\), the existence of sequences \((k_n)\) and \((k'_n)\) depends on the underlying distribution.
**Theorem 3**: Asymptotic normality of $\hat{x}_{p_n, \hat{\theta}_n, \hat{\tau}_n}$

Suppose the assumptions of Theorem 2 hold. If, moreover,

\[
\left(\log(n/k_n)\right)^{1-\tau} (K_\tau(\log(1/p_n)) - K_\tau(\log(n/k_n))) \to \infty,
\]

\[
\sqrt{k_n(\log_2(n/k_n) - \log_2(n/k_n'))/(\log_2(1/p_n))} \to \infty,
\]

\[
\log_2(n/k_n)[K_\tau(\log(1/p_n)) - K_\tau(\log(n/k_n))] \int_{\log(n/k_n)}^{\log(1/p_n)} \log(u)u^{\tau-1} du \to 0,
\]

then

\[
\frac{\sqrt{k_n(\log_2(n/k_n) - \log_2(n/k_n'))}}{\int_{\log(n/k_n)}^{\log(1/p_n)} \log(u)u^{\tau-1} du} \left( \frac{\hat{x}_{p_n, \hat{\theta}_n, \hat{\tau}_n}}{x_{p_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \theta^2).
\]

- A possible choice for the order $p_n$ of the extreme quantile satisfying Theorem 3 is given by

\[
\log_2(1/p_n) = [\log_2(n)]^{\beta} \quad \text{where} \quad \beta > 1.
\]
Simulations: behaviour of the extreme quantile estimator $\hat{\chi}_{p_n, \hat{\theta}_n, \hat{\tau}_n}$

- We generate $N = 100$ samples $(X_{n,i})_{i=1,...,N}$ of size $n = 500$.
- On each sample $(X_{n,i})$, the estimator $\hat{\chi}_{p_n, \hat{\theta}_n, \hat{\tau}_n}$ is computed for $k'_n = 3, \ldots, 500$ and $k_n = \lfloor ck'_n \rfloor$ with $c = 0.1$.
- The value $c = 0.1$ has been chosen on the basis of intensive Monte-Carlo simulations.
- In what follows we show simulation results for quantiles corresponding to $p_n = 10^{-3}$.
- The empirical Mean-Squared Error $\text{MSE}$ is plotted as a function of $k'_n$.
- Comparison with the Moment estimator of Dekkers et al. (1989) and the Peaks Over Threshold method.
The Moment estimator of Dekkers et al. (1989)

\[ \hat{\xi}_{pn} = X_{n-k_n,n} + X_{n-k_n,n} M_n^{(1)} (1 - \hat{\gamma} + M_n^{(1)}) \left( \frac{k_n}{npn} \right)^{\hat{\gamma}} - 1 \]

with

\[ \hat{\gamma} = M_n^{(1)} + 1 - \frac{1}{2} \left( 1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1} \]

where

\[ M_n^{(j)} = \frac{1}{k_n} \sum_{i=0}^{k_n-1} (\log(X_{n-i,n}) - \log(X_{n-k_n,n}))^j \]

- The Moment estimator of Dekkers et al. (1989) is designed to work in any domain of attraction.
The Peaks Over Threshold modelling using the Generalized Pareto distribution.

\[ \hat{x}_{pn} = X_{n-k_n+1,n} + \frac{\hat{\sigma}}{\hat{\gamma}} \left[ \left( \frac{k_n}{np_n} \right)^{\hat{\gamma}} - 1 \right] \]

where \( \gamma \) and \( \sigma \) are estimated using the Method of Moments in Hosking et al. (1987)

\[ \hat{\gamma} = \frac{1}{2} \left( 1 - \frac{\bar{x}^2}{s^2} \right) \quad \text{and} \quad \hat{\sigma} = \frac{1}{2} \bar{x} \left( \frac{\bar{x}^2}{s^2} + 1 \right) \]

with \( \bar{x} \) and \( s^2 \) respectively the sample mean and variance of the excesses.

* The Peaks Over Threshold Method (POT) works in any domain of attraction.
where the survival function of the Logweibull distribution is

$$
\bar{F}(x) = \exp \left( - \left( \frac{\log(x)}{\sqrt{2}} + 1 \right)^2 \right), \quad x > 0.
$$
Weibull tail-distributions

\[ \hat{\chi}_{p_n, \hat{\theta}_n, \hat{\tau}_n} \quad \text{Dekkers et al. (1989) POT} \]
Weibull tail-distributions

Normal distribution

Weibull distribution
log-Weibull tail-distributions

Lognormal distribution

Logweibull distribution
Pareto-type distributions

Pareto distribution

Student distribution
A real data set

- The performance of our estimators is illustrated though the analysis of extreme events on the Nidd river data set.
- The data set consists of 154 exceedances of the levels $65m^3s^{-1}$ by the river Nidd (Yorkshire, England) during the period 1934–1969.
- In what follows we show results corresponding to $k'_n = 3, \ldots, 154$ and $k_n = \lfloor ck'_n \rfloor$ with $c = 0.1$.
- There is no general agreement on a maximum domain of attraction for this data set.
- The estimation of $\tau$ is of great interest in order to know the domain of attraction of the distribution.
Estimation of $\tau$

The estimator of $\tau$ becomes stable for $k_n' \geq 80$ with $\hat{\tau}_n \simeq 1$
The estimator of $\theta$ becomes stable for $k'_n \geq 80$ with $\hat{\theta}_n, \bar{\tau}_n \approx 0.3$
We obtain \( \hat{\tau}_n \simeq 1 \) and \( \hat{\theta}_n, \hat{\tau}_n \simeq 0.3 \).

These results indicate that the data may be assumed to come from a distribution in the Fréchet maximum domain of attraction.

The Proposition indicates that in the case of heavy tailed-distribution \( \theta = \gamma \) thus, we obtain as an estimation of the tail index \( \gamma = 0.3 \).

In Diebolt et al. (2005) a Fréchet maximum domain of attraction is assumed and heavy tailed-distributions are considered as a possible model for such data.

Our results are in accordance with the ones obtained by Bayesian methods in Diebolt et al. (2005) where the tail index is also estimated at 0.3.
The N-year return level

- The standard quantity of interest in environmental studies is the \( N \)-year return level.
- Defined as the level which is exceeded on average once in \( N \) years.
- We focus on the estimation of the 50- and 100- year return levels.
- Comparison with the Moment estimator of Dekkers et al. (1989) and the Peaks Over Threshold method.
- We plot the associated \( N \)-year return level as a function of \( k'_n \) for \( N = 50 \) and \( N = 100 \).
Choosing $k_n' \geq 50$ we obtain an estimation of the 50-year return level which belongs to the interval $[340 m^3 s^{-1}, 375 m^3 s^{-1}]$.
Choosing $k'_n \geq 50$ we obtain an estimation of the 100-year return level which belongs to the interval $[400\, m^3\, s^{-1}, 470\, m^3\, s^{-1}]$.
Concluding remarks

Conclusions and Further Work

- The choice of the parameters $k_n$ and $k'_n$ in practice.
Concluding remarks

Conclusions and Further Work

- The choice of the parameters $k_n$ and $k'_n$ in practice.

- Adapt our results to the case $\tau > 1$ and investigate the possible link with super-heavy tails (Fraga Alves et al. (2009)).
Concluding remarks

Conclusions and Further Work

- The choice of the parameters $k_n$ and $k'_n$ in practice.
- Adapt our results to the case $\tau > 1$ and investigate the possible link with super-heavy tails (Fraga Alves et al. (2009)).
- Extend this work to random variable $Y = \varphi(X)$ where $X$ has a parent distribution satisfying $(A_1(\tau, \theta))$. 
Concluding remarks

Conclusions and Further Work

- The choice of the parameters $k_n$ and $k'_n$ in practice.
- Adapt our results to the case $\tau > 1$ and investigate the possible link with super-heavy tails (Fraga Alves et al. (2009)).
- Extend this work to random variable $Y = \varphi(X)$ where $X$ has a parent distribution satisfying $(A_1(\tau, \theta))$.
- For instance, choosing $\varphi(x) = x^* - 1/x$ would allow to consider distributions in the Weibull maximum domain of attraction (with finite endpoint $x^*$).


Main references


Merci de votre attention.