Chapter 3: Application of extreme-value theory to extrapolation

Stéphane Girard

Inria Grenoble Rhône-Alpes
http://mistis.inrialpes.fr/people/girard/

November-december 2018

1. Application of the Extreme-Value Theorem

2. Application of Pickands theorem

3. Semi-parametric approach in MDA(Fréchet)

As a consequence of the first two chapters, we have an approximation of the survival function

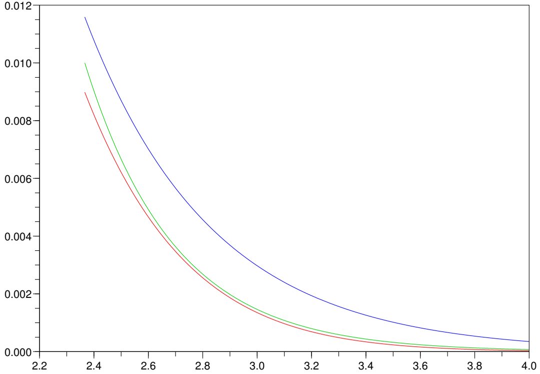
$$\begin{aligned}\bar{F}(x_n) &\simeq \frac{1}{n}\left[1+\gamma\left(\frac{x_n-b_n}{a_n}\right)\right]^{-1/\gamma} \quad \text{if} \quad \gamma \neq 0 \\ &\simeq \frac{1}{n}\exp\left(-\frac{x_n-b_n}{a_n}\right) \quad \text{if} \quad \gamma = 0\end{aligned}$$

when $x_n \to \infty$ as $n \to \infty$ and an approximation of its inverse

$$\begin{aligned}\bar{F}^{-1}(p_n) &\simeq b_n + \frac{a_n}{\gamma}\left[(np_n)^{-\gamma}-1\right] \quad \text{if} \quad \gamma \neq 0 \\ &\simeq b_n - a_n \log(np_n) \quad \text{if} \quad \gamma = 0.\end{aligned}$$

when $p_n \to 0$ as $n \to \infty$.

Illustration on a standard Gaussian distribution ($\gamma = 0$)



Comparison between $\bar{F}(x)$, $\frac{1}{n}\exp\left(-\frac{x-b_n}{a_n}\right)$ with $n = 10$ and $\frac{1}{n}\exp\left(-\frac{x-b_n}{a_n}\right)$ with $n = 100$

- Here, the theoretical values of a_n, b_n and γ (in case of the standard Gaussian distribution) have been used.
- In practice, a_n, b_n and γ are unknown (since F is unknown) and have to be estimated.

The cdf of interest is

$$H_{\gamma,a,b}(x) \stackrel{def}{=} H_\gamma\left(\frac{x-b}{a}\right) = \exp\left\{-\left[1+\gamma\left(\frac{x-b}{a}\right)\right]_+^{-1/\gamma}\right\}.$$

In the following, two estimators are considered:

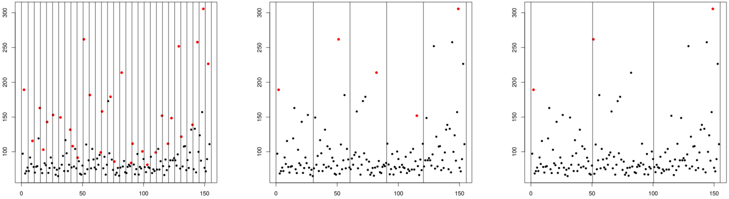
- Maximum Likelihood Estimators (MLE): not closed-form,
- Probability Weighted Moments (PWM).

Inference (with the EVD approach)

Let $\{Y_1, \ldots, Y_k\}$ be a sample of k iid random variables from the cdf $H_{\gamma,a,b}$. This starting point may be a problem since, usually, one does not observe directly a sample of maxima. They have to be extracted from an initial sample $\{X_1, \ldots, X_n\}$ from an arbitrary cdf F. The common practice is to divide the data into k blocks and to extract the maxima of each block (block maxima method). Two drawbacks:

- Loss of information (n data $\longrightarrow$ k data),
- The maxima are not exactly distributed from an EVD distribution.

Illustration on Nidd data



Block Maxima selection. Three block sizes *k* are considered. The associated block-maxima are depicted in red.

The negative log-likelihood is given by

$$\ell(a,b,\gamma) = k\log a + \left(1+\frac{1}{\gamma}\right)\sum_{i=1}^{k}\log\left(1+\gamma\frac{Y_i-b}{a}\right) + \sum_{i=1}^{k}\left(1+\gamma\frac{Y_i-b}{a}\right)^{-1/\gamma}$$

if (a,b,γ) is such that $1+\gamma(Y_i-b)/a>0$ for all $i=1,\dots,k$ and $\ell(a,b,\gamma)=+\infty$ otherwise.

- The MLE is not explicit, its computation requires a numerical optimization procedure which may fail on small datasets.
- Since the support of the GEV depends on its parameters, the asymptotic normality of the MLE does not hold in the general case but a least for $\gamma>-1/2$ [11]. In this case, the Fisher information matrix is provided in [10] which allows to build asymptotic confidence intervals. Recently, [4] proved the consistency under the weaker assumption $F\in \mathrm{MDA}(H_\gamma)$, $\gamma>-1$.

Illustration on Nidd data

```
> # Fit GEV to the data in nidd.annual, the annual maximum water
> # levels of the River Nidd, using the "BFGS" optimization method
> gevnidd <- gev(nidd.annual, method = "BFGS", control = list(maxit = 500))

$data
 [1]  65.08  65.60  75.06  76.22  78.55  81.27  86.93  87.76  88.89  90.28
[11]  91.80  91.80  92.82  95.47 100.40 111.54 111.74 115.52 131.82 138.72
[21] 148.63 149.30 151.79 153.04 158.01 162.99 172.92 179.12 181.59 189.04
[31] 213.70 226.48 251.96 261.82 305.75

$par.ests
       xi      sigma         mu
 0.321221  36.154177 103.118249

$varcov
            [,1]       [,2]       [,3]
[1,]  0.04758274 -0.4142656 -0.7770124
[2,] -0.41426557 43.6098796 35.7316149
[3,] -0.77701236 35.7316149 58.0116406
```

MLE estimation of the GEV parameters with R using the Block-Maxima approach on Nidd data. Annual blocks are considered so that $k = 35$ block-maxima are obtained. The estimated parameters are provided as well as the asymptotic covariance matrix.

In some context, one may assume that $\gamma = 0$ and fit a Gumbel distribution. The negative log-likelihood is obtained by letting $\gamma \to 0$ in the previous equation:

$$\ell(a, b, 0) = k \log a + \sum_{i=1}^{k} \left(\frac{Y_i - b}{a} \right) + \sum_{i=1}^{k} \exp \left[- \left(\frac{Y_i - b}{a} \right) \right].$$

- The MLE is still not explicit.
- Since the support of the Gumbel distribution does not depend on its parameters, the asymptotic normality automatically holds.

PWM

Recall that $\{Y_1, \ldots, Y_k\}$ is a set of k iid random variables with cdf $H_{\gamma,a,b}$. The Probability Weighted Moment of order $r \geq 0$ is defined by [8]:

$$\mu_r = \mathbb{E}\left[YH^r_{\gamma,a,b}(Y)\right].$$

Lemma 1

The PWM of order $r \geq 0$ exists provided that $\gamma < 1$ and is given by

$$\mu_r = \frac{1}{r+1}\left[b - \frac{a}{\gamma}\left\{1 - (r+1)^\gamma \Gamma(1-\gamma)\right\}\right],$$

where Γ is the Gamma function:

$$\Gamma(t) = \int_0^{+\infty} x^{t-1}\exp(-x)dx, \; t > 0.$$

Proof. Let $\gamma < 1$ and let $h_{\gamma,a,b}$ be the density associated with $H_{\gamma,a,b}$:

$$\mu_r = \int_{-\infty}^{+\infty} x H_{\gamma,a,b}^r(x) h_{\gamma,a,b}(x) dx = \int_0^1 H_{\gamma,a,b}^{-1}(y) y^r dy,$$

thanks to the change of variable

$$x = H_{\gamma,a,b}^{-1}(y) = b - \frac{a}{\gamma} - \frac{a}{\gamma}(-\log y)^{-\gamma}.$$

Replacing yields

$$\mu_r = \left(b - \frac{a}{\gamma}\right) \frac{1}{r+1} - \frac{a}{\gamma} \int_0^1 (-\log y)^{-\gamma} y^r dy.$$

Letting $z = -(r+1)\log y$ finally leads to

$$\mu_r = \left(b - \frac{a}{\gamma}\right) \frac{1}{r+1} - \frac{a}{\gamma}(r+1)^{\gamma-1} \int_0^{+\infty} z^{-\gamma} e^{-z} dz$$

and the conclusion follows. ■

To obtain a, b and γ, three PWMs are sufficient:

$$\begin{aligned} \mu_0 &= b - \frac{a}{\gamma}\{1 - \Gamma(1-\gamma)\} && (1) \\ 2\mu_1 - \mu_0 &= -\frac{a}{\gamma}(1-2^\gamma)\Gamma(1-\gamma) && (2) \\ \frac{3\mu_2 - \mu_0}{2\mu_1 - \mu_0} &= \frac{1-3^\gamma}{1-2^\gamma}. && (3) \end{aligned}$$

From (3), it is possible to obtain γ with a numerical procedure. Then, a can be deduced from (2) and b can be deduced from (1). It remains to estimate μ_0, μ_1 and μ_2 to obtain estimators for a, b and γ.

To estimate μ_r, the idea is as follows. First, the mathematical expectation is replaced by the empirical mean:

$$\mu_r \simeq \frac{1}{k}\sum_{i=1}^{k} Y_i H^r_{\gamma,a,b}(Y_i) = \frac{1}{k}\sum_{i=1}^{k} Y_{i,k} H^r_{\gamma,a,b}(Y_{i,k}).$$

Second, $H_{\gamma,a,b}$ is replaced by the empirical cdf:

$$\mu_r \simeq \frac{1}{k}\sum_{i=1}^{k} Y_{i,k}\hat{F}^r_k(Y_{i,k}) = \frac{1}{k}\sum_{i=1}^{k} Y_{i,k}\left(\frac{i-1}{k}\right)^r.$$

One thus obtain

$$\hat{\mu}_r = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i-1}{k}\right)^r Y_{i,k}.$$

One may also use the alternative estimator:

$$\tilde{\mu}_r = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{(i-1)\cdots(i-r)}{(k-1)\cdots(k-r)}\right) Y_{i,k}.$$

The joint asymptotic normality of $(\hat{a},\hat{b},\hat{\gamma})$ is established in [6] under the assumption $F \in \mathrm{MDA}(H_\gamma)$, $\gamma < 1/2$.

From Pickands theorem, we have, for all $y \geq 0$,

$$\bar{F}_u(y) = \frac{\bar{F}(u+y)}{\bar{F}(u)} \simeq \bar{G}_{\gamma,\sigma}(y),$$

where $\bar{G}_{\gamma,\sigma}$ is the survival function of the GPD. The change of variable $x = u + y$ yields for all $x \geq u$:

$$\bar{F}(x) \simeq \bar{F}(u)\bar{G}_{\gamma,\sigma}(x-u).$$

Finally, introducing $\alpha = \mathbb{P}(X > u) = \bar{F}(u)$, it follows

$$\bar{F}(x) \simeq \alpha \bar{G}_{\gamma,\sigma}(x - \bar{F}^{-1}(\alpha)).$$

We thus obtain an approximation of the small tail probabilities:

$$\begin{aligned} \bar{F}(x_n) &\simeq \alpha\left[1+\gamma\left(\frac{x_n-\bar{F}^{-1}(\alpha)}{\sigma}\right)\right]^{-1/\gamma} \quad \text{if } \gamma \neq 0 \\ &\simeq \alpha \exp\left(-\frac{x_n-\bar{F}^{-1}(\alpha)}{\sigma}\right) \quad \text{if } \gamma = 0 \end{aligned}$$

as well as an approximation of the extreme quantiles:

$$\begin{aligned} \bar{F}^{-1}(p_n) &\simeq \bar{F}^{-1}(\alpha)+\frac{\sigma}{\gamma}\left[\left(\frac{p_n}{\alpha}\right)^{-\gamma}-1\right] \quad \text{if } \gamma \neq 0 \\ &\simeq \bar{F}^{-1}(\alpha)-\sigma\log\left(\frac{p_n}{\alpha}\right) \quad \text{if } \gamma = 0. \end{aligned}$$

Comparison between EVD and GPD approaches

For both approaches, the approximations are the same. For instance, in the case of small tail probabilities and $\gamma \neq 0$:

$$\begin{aligned} \bar{F}(x_n) &\simeq \alpha \left[1 + \gamma \left(\frac{x_n - \bar{F}^{-1}(\alpha)}{\sigma} \right) \right]^{-1/\gamma} \quad \text{(GPD)} \\ \bar{F}(x_n) &\simeq \frac{1}{n} \left[1 + \gamma \left(\frac{x_n - a_n}{b_n} \right) \right]^{-1/\gamma} \quad \text{(EVD)} \end{aligned}$$

There are three parameters to be estimated:

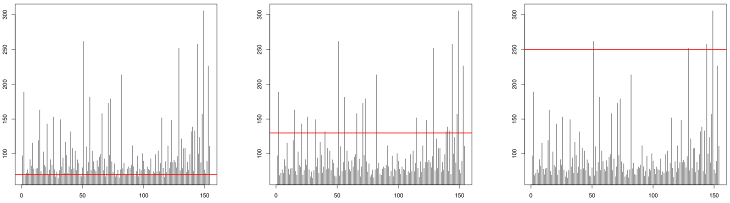
- the extreme value index γ (shape parameter),
- a scale parameter σ (GPD) or b_n (EVD),
- a position parameter $\bar{F}^{-1}(\alpha)$ (GPD) or a_n (EVD).

Inference (with the GPD approach)

Let $\{Y_1, \dots, Y_k\}$ be a sample of k iid random variables from the cdf $G_{\gamma,\sigma}$. Similarly to EVD case, this starting point may be a problem since, usually, one does not observe direclty a sample of excesses. They have to be extracted from an initial sample $\{X_1, \dots, X_n\}$ from an arbitrary cdf F. The common practice is to choose a number of k excesses and select $\{Y_1, \dots, Y_k\} := \{X_{n-k+1,n} - X_{n-k,n}, \dots, X_{n,n} - X_{n-k,n}\}$ (peaks over threshold method).

- The excesses are not exactly distributed from a GPD distribution,
- Dependence issues,
- $\alpha = k/n$, the position parameter $\bar{F}^{-1}(k/n)$ is estimated by the empirical quantile $X_{n-k,n}$. It only remains to estimate γ and σ.

Illustration on Nidd data



Excesses selection. Three numbers of excesses k are considered. The associated thresholds are depicted in red.

MLE

The negative log-likelihood is given by

$$\ell(\sigma,\gamma) = k\log\sigma + \left(1+\frac{1}{\gamma}\right)\sum_{i=1}^{k}\log\left(1+\gamma\frac{Y_i}{\sigma}\right)$$

if (σ,γ) is such that $1+\gamma Y_i/\sigma > 0$ for all $i=1,\dots,k$ and $\ell(\sigma,\gamma) = +\infty$ otherwise. The reparametrization $(\tau,\gamma) := (\gamma/\sigma,\gamma)$ yields

$$\tilde{\ell}(\tau,\gamma) = k\log(\gamma/\tau) + \left(1+\frac{1}{\gamma}\right)\sum_{i=1}^{k}\log\left(1+\tau Y_i\right) \tag{4}$$

and thus the equation

$$\frac{\partial\tilde{\ell}}{\partial\gamma} = \frac{k}{\gamma} - \frac{1}{\gamma^2}\sum_{i=1}^{k}\log\left(1+\tau Y_i\right) = 0$$

shows that $\hat{\gamma}$ can be expressed as a function of τ:

$$\hat{\gamma}(\tau) = \frac{1}{k}\sum_{i=1}^{k}\log\left(1+\tau Y_i\right).$$

Replacing in (4), the MLE can be computed thanks to a one-parameter numerical optimization of $\tilde{\ell}(\tau,\hat{\gamma}(\tau))$.

- Since the support of the GPD depends on its parameters, the asymptotic normality of the MLE does not hold in the general case but only for $F \in \text{MDA}(H_\gamma)$ with $\gamma > -1$ [13]. In this case, the Fisher information matrix is provided in [2] which allows to build asymptotic confidence intervals. [13] has also proved the non-consistency for $\gamma < -1$.
- In some context, one may assume $\gamma = 0$. In such a case, the GPD reduces to an exponential distribution and the associated estimator for extreme quantiles is referred to as the ET (exponential tail) estimator.

Illustration on Nidd data

```
> # Fit GPD to the data in nidd.thresh
> gpdnidd <- gpd(nidd.thresh, 100)

$data
 [1] 189.02 115.52 119.28 162.99 102.92 143.06 153.04 149.30 116.77 131.82
[11] 107.97 104.19 261.82 110.48 181.59 104.19 158.01 172.92 179.12 213.70
[21] 111.74 100.40 104.19 151.79 111.54 148.63 251.96 121.73 107.58 108.14
[31] 131.92 138.72 133.06 257.62 123.71 157.12 305.75 226.48 110.98

$threshold
[1] 100

$p.less.thresh
[1] 0.7467532

$n.exceed
[1] 39

$par.ests
          xi        beta
 0.003508321 50.608623759

$varcov
            [,1]       [,2]
[1,]  0.04562003  -2.303872
[2,] -2.30387212 182.476944
```

MLE estimation of the GPD parameters with R. The threshold $u = 100$ is considered so that $k = 39$ excesses are obtained. The estimated parameters are provided as well as the asymptotic covariance matrix.

PWM

Recall that $\{Y_1, \ldots, Y_k\}$ is a set of iid random variables with cdf $G_{\gamma,\sigma}$. In this case, the Probability Weighted Moment of order $s \geq 0$ is defined by [9]:

$$\nu_s = \mathbb{E}\left[Y \bar{G}^s_{\gamma,\sigma}(Y)\right].$$

Lemma 2

The PWM of order $s \geq 0$ exists provided that $\gamma < 1$ and is given by

$$\nu_s = \frac{\sigma}{(s+1)(s+1-\gamma)}.$$

Proof. By definition

$$\nu_s = \int_{1+\frac{\gamma x}{\sigma}>0} \frac{x}{\sigma}\left(1+\frac{\gamma x}{\sigma}\right)^{-(\gamma+s+1)/\gamma} dx = \sigma \int_{1+\gamma y>0} y\,(1+\gamma y)^{-(\gamma+s+1)/\gamma}\, dy$$

and an integration by parts concludes the proof. ∎

To obtain γ and σ two moments are enough:

$$\gamma = \frac{4\nu_1 - \nu_0}{2\nu_1 - \nu_0} \text{ and } \sigma = \frac{2\nu_1\nu_0}{\nu_0 - 2\nu_1},$$

where ν_0 and ν_1 can be easily estimated:

$$\hat{\nu}_s = \frac{1}{k}\sum_{i=1}^{k}\left(1 - \frac{i}{k}\right)^s Y_{i,k}.$$

Note that here, all estimators are explicit. The asymptotic normality of $(\hat{\gamma}, \hat{\sigma})$ is established in [1], for $F \in \text{MDA}(H_\gamma)$ with $\gamma < 1/2$.

We focus on the Fréchet Maximum Domain of Attraction for which a simple characterisation of the survival function is available:

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

where ℓ is a slowly-varying function and $\gamma > 0$. This can be interpreted as a semi-parametric model including

- a parametric part $x^{-1/\gamma}$ which only depends on a parameter $\gamma > 0$,
- a non-parametric part $\ell(x)$, the only information being

$$\lim_{u\to\infty} \frac{\ell(tu)}{\ell(u)} = 1,$$

for all $t > 0$.

Application to extrapolation

For all $t > 0$,

$$\lim_{u\to\infty} \frac{\bar{F}(tu)}{\bar{F}(u)} = t^{-1/\gamma} \left(\lim_{u\to\infty} \frac{\ell(tu)}{\ell(u)} \right) = t^{-1/\gamma},$$

and we thus have the approximation:

$$\bar{F}(tu) \simeq \bar{F}(u) t^{-1/\gamma}.$$

Letting $x = tu$ and $\alpha = \bar{F}(u)$, approximations can be derived for small tail probabilities and extreme quantiles:

$$\begin{aligned} \bar{F}(x) &\simeq \alpha \left(\frac{x}{\bar{F}^{-1}(\alpha)} \right)^{-1/\gamma} \\ \bar{F}^{-1}(p) &\simeq \bar{F}^{-1}(\alpha) \left(\frac{p}{\alpha} \right)^{-\gamma}, \end{aligned}$$

for all $x > u$ or equivalently $p \leq \alpha$.

Comparison with EVD and GPD approaches

For all three approaches, the approximations are the same. For instance, in the case of small tail probabilities and $\gamma > 0$:

$$\bar{F}(x_n) \simeq \alpha \left[1 + \gamma \left(\frac{x_n - \bar{F}^{-1}(\alpha)}{\sigma}\right)\right]^{-1/\gamma} \quad \text{(GPD)}$$

$$\bar{F}(x_n) \simeq \frac{1}{n} \left[1 + \gamma \left(\frac{x_n - a_n}{b_n}\right)\right]^{-1/\gamma} \quad \text{(EVD)}$$

$$\bar{F}(x_n) \simeq \alpha \left(\frac{x_n}{\bar{F}^{-1}(\alpha)}\right)^{-1/\gamma} \quad \text{(semi-parametric, Fréchet)}$$

The third approach is a particular case of the GPD method with $\sigma = \gamma \bar{F}^{-1}(\alpha)$. Thus, there are only two parameters to be estimated.

- As already seen, letting $\alpha = k/n$, the position parameter $\bar{F}^{-1}(\alpha)$ can be estimated by the ordered statistics $X_{n-k,n}$.
- It only remains to estimate the extreme-value index γ.

Semi-parametric estimation of γ

The idea is to make us of the approximation

$$\bar{F}^{-1}(p) \simeq \bar{F}^{-1}(\alpha) \left(\frac{p}{\alpha}\right)^{-\gamma}.$$

Taking the logarithm yields

$$\log \bar{F}^{-1}(p) - \log \bar{F}^{-1}(\alpha) \simeq \gamma \log(\alpha/p).$$

Letting $\alpha = k/n$ and considering several values of $p = i/n$, $i = 0, \dots, k-1$, we have

$$\log \bar{F}^{-1}(i/n) - \log \bar{F}^{-1}(k/n) \simeq \gamma \log(k/i).$$

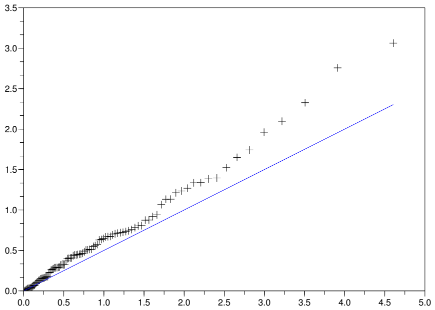
Now, the unknown cdf F is estimated by the empirical cdf which entails

$$\log X_{n-i,n} - \log X_{n-k,n} \simeq \gamma \log(k/i),$$

for $i = 0, \dots, k-1$. The quality of this approximation can be assessed graphically.

Quantile-quantile plot

Simulation of a sample of size $n = 500$ from a t_2 distribution ($\gamma = 1/2$). Here, $k = 100$.



Horizontally: $\log(k/i)$. Vertically: $y = x/2$ and $\log X_{n-i,n} - \log X_{n-k,n}$ for $i = 0, \ldots, k-1$.

Summing the previous approximations for $i = 0, \dots, k-1$, we get

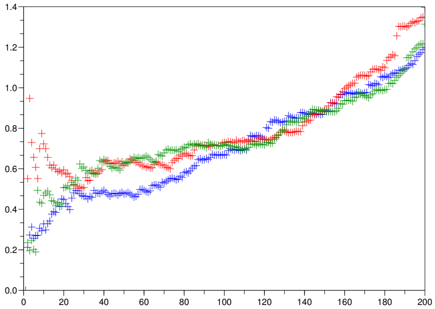
$$\gamma \simeq \frac{\displaystyle\sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}}{\displaystyle\sum_{i=0}^{k-1} \log(k/i)}.$$

The lower part can be rewritten as $\log[k^{k-1}/(k-1)!]$. Thanks to Stirling's formula, one may show that is is asymptotically equivalent to k when $k \to \infty$. We thus obtain:

Definition 1 (Hill estimator [7])

$$\hat{\gamma}(k) = \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n}).$$

Estimation of γ on three different samples of size $n = 500$ from a t_2 distribution ($\gamma = 1/2$).



Horizontally: k. Vertically: $\hat{\gamma}(k)$ for $k = 1, \ldots, 200$.

In practice ...

The choice of k is difficult :

- If k is small, $\hat{\gamma}(k)$ is based on few observations, it has therefore a **large variance**.
- If k is large, then $X_{n-k,n}$ is no longer in the distribution tail and the approximation of the survival function by a power function is no longer true, $\hat{\gamma}(k)$ has a **large bias**.

This balance between bias and variance also appears in the asymptotic normality result ...

Second order condition

- Recall that, $F \in \mathrm{DA}(\text{Fréchet})$ is equivalent to $\bar{F} \in \mathcal{RV}_{-1/\gamma}$. Introducing the tail quantile function $U(\cdot) = \bar{F}^{\leftarrow}(1/\cdot)$, the above properties are also equivalent to $U \in \mathcal{RV}_{\gamma}$, *i.e.* $U(tx)/U(t) \to x^{\gamma}$ for all $x > 0$ as $t \to \infty$. This property is sometimes referred to as a first order condition in the extreme-value literature. It is a sufficient condition to prove the consistency of extreme-value estimators.
- To establish the asymptotic distribution of an estimator, a second order condition is usually introduced: There exist $\gamma > 0$, $\rho \leq 0$ and a positive or negative function A with $A(t) \to 0$ as $t \to \infty$ such that

$$\frac{1}{A(t)}\left(\frac{U(tx)}{U(t)} - x^{\gamma}\right) \to x^{\gamma}\frac{x^{\rho}-1}{\rho} \tag{5}$$

for all $x > 0$ as $t \to \infty$. The limit should be understood as $x^{\gamma}\log x$ when $\rho = 0$. Since $A(t) \to 0$ as $t \to \infty$, the limit (5) can be equivalently rewritten as

$$\frac{1}{A(t)}\left(\log U(tx) - \log U(t) - \gamma\log x\right) \to \frac{x^{\rho}-1}{\rho} \tag{6}$$

for all $x > 0$ as $t \to \infty$. Moreover, it can be shown that $|A|$ is regularly varying with index ρ.

Theorem 1 (Consistency)

Let $X_1, \ldots, X_n$ be i.i.d. random variables with cdf $F \in MDA(H_\gamma)$, $\gamma > 0$. If $k \to \infty$ with $k/n \to 0$ as $n \to \infty$ then $\hat{\gamma}_n \xrightarrow{\mathrm{P}} \gamma$.

Theorem 2 (Asymptotic normality)

Suppose the assumptions of Theorem 1 hold with the second order condition (6). If, moreover, $\sqrt{k}A(n/k) \to \lambda < \infty$, then

$$\sqrt{k}(\hat{\gamma}_n - \gamma) \xrightarrow{d} \mathcal{N}(\lambda/(1-\rho), \gamma^2).$$

Theorem 2 shows that the asymptotic variance of Hill estimator is γ^2/k (a decreasing function of k) and that the asymptotic bias is $A(n/k)/(1-\rho)$ (an increasing function of k).

The asymptotic mean-squared error is thus given by

$$AMSE(k) = \frac{\gamma^2}{k} + \frac{A^2(n/k)}{(1-\rho)^2}.$$

This quantity is difficult to compute in practice since γ, ρ and $A(\cdot)$ are unknown. Nevertheless, its possible to derive the theoretical optimal value of k in the simplified case where $A(t) = \beta t^{\rho}$:

$$k(n) = \left[\left(\frac{\gamma^2(1-\rho)^2}{-2\rho\beta} \right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)} \right].$$

The optimal rate of convergence of Hill estimator is of order $n^{-\rho/(1-2\rho)}$. It approaches the parametric rate $n^{1/2}$ when $\rho \to -\infty$.

Both proofs of Theorem 1 and Theorem 2 rely on some common tools. First, let us note that, for all $i = 1, \ldots, n$, $X_i \stackrel{d}{=} U(Y_i)$ where $Y_1, \ldots, Y_n$ are i.i.d. from a standard Pareto distribution *i.e.* with cdf $1 - 1/y$, $y \leq 1$. It follows that

$$\hat{\gamma}_n \stackrel{d}{=} \frac{1}{k} \sum_{i=0}^{k-1} \log U(Y_{n-i,n}) - \log U(Y_{n-k,n}).$$

Moreover, remarking that, for all $i = 1, \ldots, n$, $\log Y_i$ follows a standard exponential distribution, the following lemma will reveal useful.

Lemma 3

Let $E_1, \ldots, E_n$ be i.i.d. random variables with standard exponential distribution. Let $E_{1,n} \leq \cdots \leq E_{n,n}$ be the associated order statistics and introduce $k \to \infty$ such that $k/n \to 0$ as $n \to \infty$.

- $E_{n-k,n} \stackrel{\mathrm{P}}{\to} \infty$,
- *If f is a function such that $var(f(E_1)) < \infty$, then*

$$D_n := \sqrt{k} \left(\frac{1}{k} \sum_{i=0}^{k-1} f(E_{n-i,n} - E_{n-k,n}) - \mathbb{E}(f(E_1)) \right)$$

is independent of $E_{n-k,n}$ and is asymptotically $\mathcal{N}(0, f(E_1))$.

Proof of Lemma 3. The independence and $E_{n-k,n} \stackrel{\mathrm{P}}{\to} \infty$ are direct consequences of Rényi representation which also implies

$$\{E_{n-i,n} - E_{n-k,n}\}_{i=0,\dots,k-1} \stackrel{d}{=} \{E_{k-i,k}\}_{i=0,\dots,k-1}.$$

It follows that

$$D_n \stackrel{d}{=} \sqrt{k}\left(\frac{1}{k}\sum_{i=0}^{k-1} f(E_{k-i,k}) - \mathbb{E}(f(E_1))\right) = \sqrt{k}\left(\frac{1}{k}\sum_{i=0}^{k-1} f(E_i) - \mathbb{E}(f(E_1))\right),$$

and the central limit theorem concludes the proof. ■

Proof of Theorem 1. From the first order condition and Potter bounds, we have for all $\varepsilon > 0$, $\varepsilon' > 0$, $x \geq 1$ and $t \geq t_0$,

$$(1-\varepsilon)x^{\gamma-\varepsilon'} < \frac{U(tx)}{U(t)} < (1+\varepsilon)x^{\gamma+\varepsilon'},$$

or equivalently

$$\log(1-\varepsilon)+(\gamma-\varepsilon')\log x < \log U(tx) - \log U(t) < \log(1+\varepsilon)+(\gamma+\varepsilon')\log x.$$

Applying this inequality for $t = Y_{n-k,n} \stackrel{\mathrm{P}}{\to} \infty$ (Lemma 3) and $x = Y_{n-i,n}/Y_{n-k,n} \geq 1$ yields

$$\begin{aligned} \log(1-\varepsilon)+(\gamma-\varepsilon')\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) &< \log\left(\frac{U(Y_{n-i,n})}{U(Y_{n-k,n})}\right) \\ &< \log(1+\varepsilon)+(\gamma+\varepsilon')\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right), \end{aligned}$$

for all $i = 0, \ldots, k-1$.

As a consequence, it follows that

$$\log(1-\varepsilon)+(\gamma-\varepsilon')S_n < \hat{\gamma}_n < \log(1+\varepsilon)+(\gamma+\varepsilon')S_n$$

where we have defined

$$S_n := \frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) \stackrel{d}{=} \frac{1}{k}\sum_{i=0}^{k-1}E_{n-i,n}-E_{n-k,n}.$$

Lemma 3 thus entails $S_n \xrightarrow{\mathrm{P}} 1$ and the result is proved. ■

Proof of Theorem 2. [5] has shown that the second order condition (6) implies

$$\left| \frac{\log U(tx) - \log U(t) - \gamma \log x}{A_0(t)} - \frac{x^\rho - 1}{\rho} \right| \leq \varepsilon x^{\rho+\varepsilon}, \tag{7}$$

where $A_0(t) \sim A(t)$ as $t \to \infty$, $\varepsilon > 0$, $t \geq t_0$, $x \geq 1$ see also (3.2.7) in [3]. Applying this inequality for $t = Y_{n-k,n} \xrightarrow{\mathrm{P}} \infty$ (Lemma 3) and $x = Y_{n-i,n}/Y_{n-k,n} \geq 1$ yields

$$\hat{\gamma}_n \stackrel{d}{=} S_n + A_0(Y_{n-k,n}) \frac{1}{k} \sum_{i=0}^{k-1} \frac{1}{\rho} \left(\left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^\rho - 1 \right) + R_n,$$

where S_n is defined in the proof of Theorem 1 and

$$|R_n| \leq \varepsilon |A_0(Y_{n-k,n})| \frac{1}{k} \sum_{i=0}^{k-1} \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^{\rho+\varepsilon}.$$

Introducing for all $a < 1$,

$$T_n(a) = \frac{1}{k} \sum_{i=0}^{k-1} \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^a,$$

it follows

$$\sqrt{k}(\hat{\gamma}_n - \gamma) \stackrel{d}{=} \gamma \sqrt{k}(S_n - 1) + \sqrt{k} A_0(Y_{n-k,n})(T_n(\rho) - 1)/\rho + \sqrt{k} R_n,$$

with $|R_n| \leq \varepsilon |A_0(Y_{n-k,n})| T_n(\rho + \varepsilon)$. Lemma 3 shows that $\sqrt{k}(S_n - 1)$ is asymptotically standard Gaussian. Besides,

$$\begin{aligned}
T_n(a) &= \frac{1}{k} \sum_{i=0}^{k-1} \exp\left(a(\log Y_{n-i,n} - \log Y_{n-k,n})\right) \\
&\stackrel{d}{=} \frac{1}{k} \sum_{i=0}^{k-1} \exp\left(a(E_{n-i,n} - E_{n-k,n})\right) \\
&\stackrel{d}{=} \frac{1}{k} \sum_{i=0}^{k-1} \exp\left(a E_{k-i,k}\right) = \frac{1}{k} \sum_{i=0}^{k-1} \exp\left(a E_i\right) \\
&\stackrel{\mathrm{P}}{\to} \mathbb{E}(\exp(a E_1)) = 1/(1-a),
\end{aligned}$$

from the law of large numbers.

Finally, properties of regularly-varying functions and $kY_{n-k,n}/n \xrightarrow{\mathrm{P}} 1$ show that

$$|A_0(Y_{n-k,n})| \sim |A(Y_{n-k,n})| \sim |A(n/k)|,$$

as $n \to \infty$. As a consequence,

$$\sqrt{k} A_0(Y_{n-k,n})(T_n(\rho) - 1)/\rho \to \lambda/(1-\rho)$$

and

$$\sqrt{k}|R_n| \leq 2\lambda\varepsilon/(1-\rho+\varepsilon)$$

eventually and the conclusion follows. ■

Estimation of extreme quantiles

Basing on the approximation

$$x_{p_n} := \bar{F}^{-1}(p_n) \simeq \bar{F}^{-1}(\alpha_n) \left(\frac{p_n}{\alpha_n} \right)^{-\gamma},$$

the following estimator has been introduced:

Definition 2 (Weissman estimator [12])

$$\hat{x}_{p_n} = X_{n-k_n,n} \left(\frac{k_n}{np_n} \right)^{\hat{\gamma}_n(k_n)},$$

where $\hat{\gamma}_n(k_n)$ is Hill estimator.

Theorem 3 (Asymptotic normality)

Suppose the assumptions of Theorem 2 hold. If moreover $np_n/k_n \to 0$ *as* $n \to \infty$ *then*

$$\frac{\sqrt{k_n}\left(\log \hat{x}_{p_n} - \log x_{p_n}\right)}{\log(k_n/(np_n))} \xrightarrow{d} \mathcal{N}(\lambda/(1-\rho), \gamma^2).$$

Proof. Recalling that $x_{p_n} = U(1/p_n)$, $X_{n-k_n,n} \stackrel{d}{=} U(Y_{n-k_n,n})$ and $U \in \mathcal{RV}_\gamma$ yields

$$\begin{aligned}
\log \hat{x}_{p_n} &= \gamma \log Y_{n-k_n,n} + \log \ell(Y_{n-k_n,n}) + \hat{\gamma}_n \log(k_n/(np_n)), \\
\log x_{p_n} &= \gamma \log(n/k_n) + \log \ell(1/p_n) + \gamma \log(k_n/(np_n)),
\end{aligned}$$

and therefore

$$\begin{aligned}
\frac{\sqrt{k_n}\,(\log \hat{x}_{p_n} - \log x_{p_n})}{\log(k_n/(np_n))} &\stackrel{d}{=} \gamma \frac{\sqrt{k_n}\,(\log Y_{n-k_n,n} - \log(n/k_n))}{\log(k_n/(np_n))} \\
&+ \frac{\sqrt{k_n}\,(\log \ell(Y_{n-k_n,n}) - \log \ell(1/p_n))}{\log(k_n/(np_n))} \\
&+ \sqrt{k_n}(\hat{\gamma}_n - \gamma) \\
&= T_{1,n} + T_{2,n} + T_{3,n}.
\end{aligned}$$

The three terms are studied separately.

- From [3], Corollary 2.2.2, $\sqrt{k_n}(k_n Y_{n-k_n,n}/n - 1) \xrightarrow{d} \mathcal{N}(0,1)$. The delta-method yields that $\sqrt{k_n}(\log Y_{n-k_n,n} - \log(n/k_n))$ has the same asymptotic distribution and thus $T_{1,n} \xrightarrow{\mathrm{P}} 0$.
- The consequence of the second-order condition (7) can be written as

$$\left| \frac{\log \ell(tx) - \log \ell(t)}{A_0(t)} - \frac{x^\rho - 1}{\rho} \right| \leq \varepsilon x^{\rho+\varepsilon}.$$

Letting $t = Y_{n-k_n,n}$ and $x = 1/(p_n Y_{n-k_n,n})$, it follows that $x = (k_n/(np_n))(1 + o_P(1)) \xrightarrow{\mathrm{P}} \infty$ and consequently,

$$\begin{aligned} \log \ell(Y_{n-k_n,n}) - \log \ell(1/p_n) &= -\frac{1}{\rho} A_0(Y_{n-k_n,n})(1 + o_P(1)) \\ &= -\frac{1}{\rho} A(n/k_n)(1 + o_P(1)). \end{aligned}$$

This entails that

$$T_{2,n} = -\frac{\lambda}{\rho} \frac{1}{\log(k_n/(np_n))}(1 + o_P(1)) \xrightarrow{\mathrm{P}} 0.$$

- Finally, $T_{3,n} \xrightarrow{d} \mathcal{N}(\lambda/(1-\rho), \gamma^2)$ from Theorem 2. ■

References

Cai, J.-J., de Haan, L. and Zhou, C. (2013). Bias correction in extreme value statistics with index around zero. *Extremes*, **16**, 173–201.

Davison, A. C., and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society*. Series B, **52**, 393–442.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer-Verlag, New York.

Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, **21**, 420–436.

Drees, H. (1998). On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, **25**, 187–210.

Ferreira, A., and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of Statistics*, **43**, 276–298.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, **3**, 1163–1174.

Hosking, J.R.M., Wallis, J.R., and Wood, E.F. (1985). Estimation of the Generalized Extreme-Value distribution by the method of probability-weighted moments. *Technometrics*, **27**, 251–261.

Hosking, J.R.M., and Wallis, J.R. (1987). Parameter and quantile estimation for the Generalized Pareto Distribution. *Technometrics*, **29**, 1339–1349.

Prescott, P., and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, **67**, 723–724.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67–90.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, **73**, 812–815.

Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, **101**, 971–983.