Chapter 3: Application of extreme-value theory to extrapolation

Stéphane Girard

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

November-december 2018

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

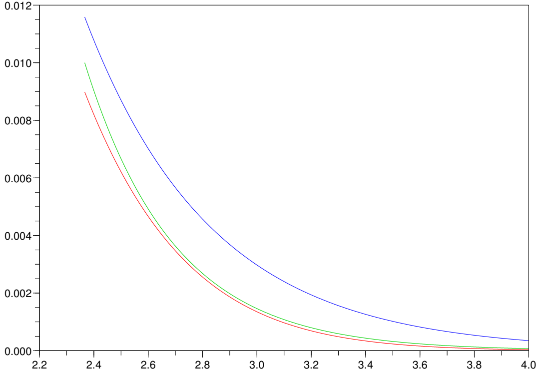
$$\begin{aligned}\bar{F}(x_n) &\simeq \frac{1}{n}\left[1+\gamma\left(\frac{x_n-b_n}{a_n}\right)\right]^{-1/\gamma} \quad \text{if} \quad \gamma \neq 0 \\ &\simeq \frac{1}{n}\exp\left(-\frac{x_n-b_n}{a_n}\right) \quad \text{if} \quad \gamma = 0\end{aligned}$$

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:

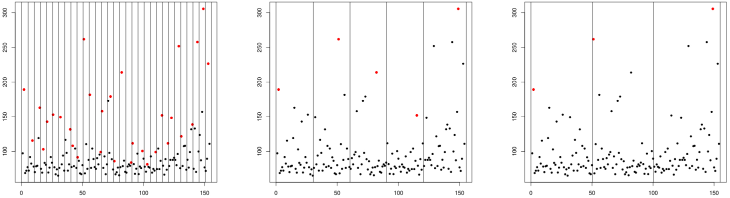
- Maximum Likelihood Estimators (MLE): not closed-form,
- Probability Weighted Moments (PWM).

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.

MLE

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 $i=1,\ldots,k$ 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

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

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 $H_{\gamma,a,b}$:

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

thanks to the change of variable

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

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 $F \in \mathrm{MDA}(H_\gamma)$, $\gamma < 1/2$.

1. Application of the Extreme-Value Theorem
2. Application of Pickands theorem
3. Semi-parametric approach in MDA(Fréchet)

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 $\gamma \neq 0$:

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

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

There are three parameters to be estimated:

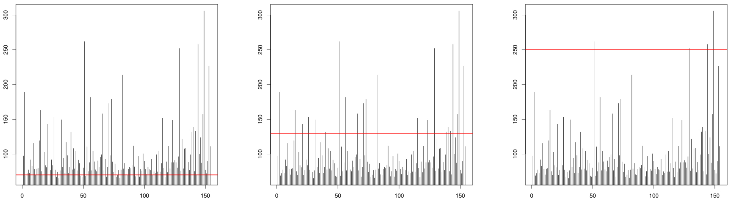
- the extreme value index γ (shape parameter),
- a scale parameter σ (GPD) or b_n (EVD),
- a position parameter $\bar{F}^{-1}(\alpha)$ (GPD) or a_n (EVD).

Inference (with the GPD approach)

Let $\{Y_1, \ldots, Y_k\}$ 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 $\{X_1, \ldots, X_n\}$ from an arbitrary cdf F. The common practice is to choose a number of k excesses and select $\{Y_1, \ldots, Y_k\} := \{X_{n-k+1,n} - X_{n-k,n}, \ldots, X_{n,n} - X_{n-k,n}\}$ (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 $i=1,\ldots,k$ 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

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

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 $F \in \mathrm{MDA}(H_\gamma)$ with $\gamma < 1/2$.