Exercise session 1: univariate extremes

Anna Kiriliouk and Johan Segers

UCLouvain, October 2024

You'll need to install and load the `ismev` package first using `install.packages("ismev")` and `library(ismev)`.

Exercise 1.

Load the `ismev` package and the data `Ex1uni.RData` as below. The `data` vector contains daily maximal speeds of wind gusts, measured in kilometers per hour in Eindhoven (the Netherlands) during extended winter (October–March), from October 2001 up to and including March 2022. The `times` vector contains the dates of the observations.

```r
library(ismev)
load("Ex1uni.RData")
```

a. Use the following command to select monthly maxima. Also select the yearly maxima (careful: one for each extended winter season, spanning from October to March).

```r
monthly <- as.vector(tapply(data, paste(times$year,times$mon), max))
```

b. Plot the two datasets. Do they look stationary?

c. Fit a GEV distribution to both the monthly and the yearly maxima and check the goodness-of-fit plots. You'll need to use the functions `gev.fit` and `gev.diag`. Would you choose to continue with the monthly or the yearly maxima?

d. For the chosen data (monthly or weekly), give the GEV parameter estimates with 95 % confidence intervals. Is the data bounded-tailed, light-tailed or heavy-tailed?

e. Using the function `gum.fit`, estimate the parameters of a GEV distribution with $\xi = 0$. Decide whether such a model suffices based on a likelihood ratio test. *Reminder: if* $L(\xi, \sigma, \mu)$ *denotes the log-likelihood function, then* $2\left\{L(\xi, \sigma, \mu) - L(0, \sigma, \mu)\right\} \to \chi^2_1$.

f. The return-level plot returned by `gum.diag` is not very clear. Using the formula seen this morning (slide 20), calculate the 10-year and the 1000-year return levels. Confidence intervals can be obtained using the delta method: if $\xi = 0$, $\nabla x_p^T = (1, -\log(-\log(1-p)))$, and the covariance matrix of $(\hat{\mu}, \hat{\sigma})$ can be obtained from the output of `gum.fit`. Is there a big difference between the 10-year and the 1000-year return levels? Is this surprising?

g. Fit a GP distribution to the daily maximal speeds of wind gusts that exceed 60 km/h and check the goodness-of-fit plots. You'll need to use the functions `gpd.fit` and `gpd.diag`.

h. Use the functions `gpd.fitrange` and `mrl.plot` to decide whether the 60 km/h threshold is appropriate: if not, what threshold would you suggest?

Next, estimate the probability of a wind gust exceeding 144 km/h (the harshest inland wind ever to be measured in the Netherlands, during storm Eunice in February 2022) using the semi-parametric model on slide 36.

Exercise 2.

Load the `ismev` package and the data `fremantle` from the same package. This gives a `data.frame` where the first column contains the years, the second column gives annual maximum sea levels recorded at Fremantle, Western Australia, and the third column gives annual mean values of the Southern Oscillation Index, which is a proxy for meteorological volatility.

```
library(ismev)
data(fremantle)
```

a. Plot the data. Would a stationary GEV model be appropriate?
b. Fit three non-stationary GEV models: for fixed scale and shape, try letting the location

- vary linearly with time
- vary linearly with the southern oscillation index
- vary linearly with both time and the southern oscillation index

Which model would you choose?

Exercise 3.

Simulate samples of size $n \in \{2000, 20000, 100000\}$ from the following two distributions:

- The inverse gamma distribution with shape $\alpha = 3$ and rate $\beta = 1$ (you can take the reciprocal of a gamma random variable with the same shape and rate).
- The log-Pareto distribution, whose cumulative distribution function is $F(x) = 1/\log(x)$ for $x \geq e$.

Using blocks of size 100, calculate the block maxima for each of the samples (leading to 20, 200, and 1000 block maxima respectively) and fit a GEV distribution (if possible!). Are these two distributions in the max-domain of attraction of a GEV? If yes, with what shape parameter?