Exercise session 2: multivariate extremes

Anna Kiriliouk and Johan Segers

UCLouvain, October 2024

To be installed: `copula`, `evd`, `ismev`, `maps`. You can install a package by `install.packages("copula")` etc.

```
library(ismev)
library(copula)
library(evd)
library(maps)
```

The data consist of annual maxima of daily maximum temperatures for the years 1950–2018 for $d = 54$ grid points in Belgium. Derived from the homogenized E-OBS dataset E-OBS v19.0eHOM data, made available in the supplementary material of Auld et al. (2023).

The aim is to estimate **joint probabilities** of high annual temperatures

- at multiple locations *simultaneously* (AND),
- at *at least* one location (OR).

We will compare various estimates:

1. the empirical probability
2. by applying univariate extreme-value theory (EVT) to the maximum of maxima
3. by combining a non-parametric estimate of the Pickands dependence function with parametric GEV margins
4. by fitting a max-stable distribution to the vector of component-wise maxima

1 Warm-up: the data

We load the data and ensure that the temperature data object is of class `data.frame`.

```
load("Temp.RData")
Temp <- as.data.frame(Temp)
```

This loads three objects: an $(n \times d)$ matrix of observations, a $(d \times 2)$ matrix of coordinates and a vector of years.

Let's plot the locations:

```
map("world", xlim = c(2,6.5), ylim = c(49,51.5))
points(TempCoord, col = "red")
```

For simplicity, we'll select three locations, in the neighbourhood of Middelkerke, Louvain-la-Neuve, and Virton:

```
cities <- as.data.frame(rbind(
  c("Middelkerke", 1),
  c("LLN", 30),
```

```
  c("Virton", 54)
))
names(cities) <- c("name", "ID")
cities$ID <- as.integer(cities$ID)
M <- Temp[,cities$ID]
names(M) <- cities$name
map("world", xlim = c(2,6.5), ylim = c(49,51.5))
points(TempCoord[cities$ID,], col = "red")
text(TempCoord[cities$ID,], col = "red", labels = cities$name, pos = 3)
```

Make a scatterplot displaying all three pairs of stations.

Compute some summary statistics of the data, such as the range of the annual maxima.

2 Empirical probability

We are interested in probabilities of the type

$$\begin{array}{ll} \mathsf{P}(M_i > u_i \text{ and } M_j > u_j) & \mathsf{P}(M_1 > u_1 \text{ and } M_2 > u_2 \text{ and } M_3 > u_3) \\ \mathsf{P}(M_i > u_i \text{ or } M_j > u_j) & \mathsf{P}(M_1 > u_1 \text{ or } M_2 > u_2 \text{ or } M_3 > u_3) \end{array}$$

Estimate the above probabilities for $u_1 = u_2 = u_3 = 35°$ by the empirical probabilities, that is, the frequency of the events in the sample. In R, the logical operators are `&` (AND) and `|` (OR).

3 Univariate EVT applied to the maximum of the maxima

For each year, extract the maximum of the maxima at two or three locations. Use the function `pmax()`. For our three cities, this gives four series of maxima (three pairs and one triple).

Fit a univariate GEV distribution to each of the series, by the function `gev.fit()` in `ismev`.

For the OR version, calculate the probabilities by the GEV fit. To do so, use

$$\mathsf{P}(M_i > u \text{ or } M_j > u) = \mathsf{P}[\max(M_i, M_j) > u].$$

The GEV cumulative distribution function is implemented in `pgev()` of the package `evd`.

For the AND version, use the additive law:

$$\mathsf{P}(M_i > u \text{ and } M_j > u) = \mathsf{P}(M_i > u) + \mathsf{P}(M_j > u) - \mathsf{P}(M_i > i \text{ or } M_j > j)$$

(Bonus theory question: how would you extend the formula for three locations? For m locations?)

The individual probabilities $\mathsf{P}(M_i > u)$ can also be estimated by fitting a univariate GEV.

4 Nonparametric estimation of a max-stable distribution

We will estimate the Pickands dependence function non-parametrically and then combine the estimated dependence structure with the parametric GEV margins to estimate the multivariate excess probabilities. For simplicity, we will limit ourselves to the bivariate case.

4.1 Estimation and visualization of the Pickands dependence function

For each of the three pairs of cities, make a graph of the Capéraà-Fougères-Genest estimator $\hat{A}$ of the Pickands dependence function A.

Use the function `An.biv()` of the package `copula` to compute the estimate and use `curve` to make the plot.

For each pair, also compute an estimate of the dependence coefficient χ. To do so, put

$$\hat{\chi} = 2\left\{1 - \hat{A}(0.5)\right\}.$$

Optionally: visualize $\hat{\chi}$ on the plot of $\hat{A}$.

Interpret the value of $\hat{\chi}$ in terms of the locations of the cities.

4.2 Estimation of the joint excess probability

For a bivariate max-stable distribution function G with margins G_1 and G_2 and Pickands dependence function A, we have

$$G(x_1, x_2) = (u_1 u_2)^{A(w)}, \qquad \begin{cases} u_1 = G_1(x_1) \\ u_2 = G_2(x_2) \\ w = \log(u_2)/\log(u_1 u_2) \end{cases}$$

For a pair of cities of your choice, combine the estimates of the marginal (non-)excess probabilities and the estimate of the Pickands dependence function to obtain the required excess probabilities.

Note that, for the pair (i, j), say, we have

$$\begin{aligned} \mathsf{P}(M_i > x \text{ or } M_j > x) &= 1 - G_{ij}(x, x) \\ \mathsf{P}(M_i > x \text{ and } M_j > x) &= \mathsf{P}(M_i > x) + \mathsf{P}(M_j > x) - \mathsf{P}(M_i > x \text{ or } M_j > x). \end{aligned}$$

Finally, compare the probabilities with the one of the previous approaches.

5 Parametric modelling of a max-stable distribution

Now let's assume a parametric family for the bivariate GEV distribution. We will first estimate the parameters by maximum likelihood and then compute the multivariate excess probabilities of interest in the fitted parametric model.

For the *marginal* GEV parameters, we have two choices:

- estimate the parameters jointly with the dependence parameter;
- use a two-stage procedure: in a first stage, we fit the marginal GEV parameters separately to each margin; in a second stage, we fix the marginal parameters and estimate the dependence parameter.

(A third option would be to use the pseudo-likelihood procedure, estimating the margins nonparametrically with the empirical cdf.)

For the parametric estimate, we will use the function `fbvevd()` from the package `evd`. To evaluate probabilities under the fitted model, use the function `pmvevd()` from the same package; alternatively, use the function `pCopula()` from the `copula` package, knowing that the logistic model in EVT is actually the same as the Gumbel copula. Recall that the link between a bivariate cdf G and its margins G_1, G_2 is given by

$$G(x_1, x_2) = C\left(G_1(x_1), G_2(x_2)\right),$$

where C is the copula of G.

5.1 Joint parameter estimation

Fit the bivariate GEV with the logistic dependence model, estimating all parameters jointly.

Estimate the probabilities $\mathsf{P}(M_i > u)$, $\mathsf{P}(M_j > u)$, $\mathsf{P}(M_i > u$ or $M_j > u)$ and $\mathsf{P}(M_i > u$ and $M_j > u)$. Use the functions `pgev()` for the marginal ones and `pmvevd()` for the bivariate joint distribution function $\mathsf{P}(M_i \le u \text{ and } M_j \le u)$.

5.2 Two-stage estimation procedure

Now do the same, but using the two-stage procedure.

Estimate the marginal parameters by the `gev.fit()` procedure from the `evd` package.

Estimate the dependence parameter by `fbvevd()`, fixing the marginal parameters.

Estimate the probabilities $\mathsf{P}(M_i > u)$, $\mathsf{P}(M_j > u)$, $\mathsf{P}(M_i > u$ or $M_j > u)$ and $\mathsf{P}(M_i > u$ and $M_j > u)$.

Optionally: **Fit the bivariate GEV with the Hüsler—Reiss dependence model.** To calculate the bivariate GEV, use the function `pcopula()` from the `copula` package. Among the two models, which one would you prefer? (Hint: inspect the value of the AIC.)