ENSAI - GS & GR 3A 2023-2024 ACADEMIC YEAR
Gilles Stupfler & José Gregorio Gomez Garcia Email : gilles.stupfler@univ-angers.fr

---

Extreme Value Theory - Practical session 2: The general case

Supporting data for this practical session are:

- Temperature data from Perth, Western Australia, see `Perth_min_temp.csv` and `Perth_max_temp.csv`,
- Wind data at 12 weather stations in the Republic of Ireland from the R package `gstat`.

Of course, feel free to revisit the datasets used for the first session too. This can help in further justifying that a heavy tail is present in these sets of data.

In each case, it will be important to code the estimators yourself, and to compare your procedures with those of R packages such as `evir`, `evd`, `evt0` and `extRemes`. As a preliminary step, you may therefore want to do the following things:

1. For a fixed sample of data of size n and $2 \leq k < n$, implement the Pickands and moment estimators. The result should be a pair of vectors of size $n-1$.
2. Implement likewise the maximum likelihood estimators. This is more difficult but closer to what practitioners typically use. The choice of the optimisation method (gradient descent, Newton-Raphson, Nelder-Mead, brute force...) is up to you, but you should check that you indeed find a maximum of the likelihood. As a starting point to the algorithm, you may want to use probability weighted moment estimators.
3. Implement diagnostic tools (QQ-plots...) to judge whether the fittings are reasonable.
4. Download a few packages. Good choices in this practical session are `evir`, `evd` and `extRemes`. These will help to see if your results make sense (and you might find that one of the packages does something wrong!)

1. Temperature data

The variables of interest are average monthly minimum temperature and maximum temperature in Perth, Western Australia, between 1944 and 2019. Accompanying these variables are the month and year of each recording.

1. Represent the data, first ignoring the time series structure, and then by plotting the data as a time series. Do you notice seasonality, autocorrelation, nonstationarity?
2. Can you find a reasonable statistical model for the whole of the data?
3. Calculate an estimate for the extreme value index of the data. Justify your choice of tuning parameters.
4. Calculate an estimate for the extreme quantiles at level 0.99, 0.995 and 0.999. Do these make sense?
5. Can you provide a confidence interval for the extreme value index and for these extreme quantiles? If so, why? If not, can you suggest a method that would allow you to do so?

2. Wind data

The variable of interest is average daily wind speed at weather stations in Ireland. It is sufficient to pick a single station for the analysis, although you may want to compare your results across stations (see final question). Accompanying this variable are the day, month and year of each recording. This data set can be loaded by typing `data(wind)` after having loaded the `gstat` package.

1. Represent the data, first ignoring the time series structure, and then by plotting the data as a time series. Do you notice seasonality, autocorrelation, nonstationarity?
2. Can you find a reasonable statistical model for the whole of the data?
3. Calculate an estimate for the extreme value index of the data. Justify your choice of tuning parameters.
4. Calculate an estimate for the extreme quantiles at level 0.995, 0.999 and $1-1/n$. Do these make sense?
5. Can you provide a confidence interval for the extreme value index and for these extreme quantiles? If so, why? If not, can you suggest a method that would allow you to do so?
6. Can you think of a way to assess the correlation between extreme values at different stations? If so, can you use this in order to analyze/infer joint extremes of wind speed?