Exercice sheet

Maximum likelihood estimation

1. Using the **R** package `mev` or otherwise
   - Simulate $n = 15$ observations from a generalized Pareto model with unit scale and shape $\xi = -0.1$
   - Fit a generalized Pareto distribution.
2. Repeat these instructions $B = 1000$ times
   - What is the average shape parameter: does it match your intuition?
   - How often does the MLE lie on the boundary of the parameter space (i.e., $\hat{\xi} = -1$)?

Univariate extreme value modelling

Consider the data for station Montélimar from `frwind` (series `S4`).

1. Extract the three largest observations per year
2. Fit a generalized extreme value distribution to annual maxima.
3. Compare the parameter estimates with the fit of the r-largest order statistic, via `fit.rlarg`.
4. Compute a score test statistic for the hypothesis $\xi = 0$ with the GEV. To do so,
   - fit in addition the restricted model via `fit.gev` with fixed parameter `fpar = list(shape=0)`
   - compute the score and information matrices with the restricted parameters and form the statistic $\ell_\theta^\top(\theta_0)\jmath^{-1}(\theta_0)\ell_\theta(\theta_0)$
   - under the null hypothesis, the score statistic is distributed as χ^2_1.
   - compare with the likelihood ratio statistic, obtained by comparing models using the `anova` method.
5. Compute the 50-year return level using the generalized extreme value model and provide a 50% confidence interval for the latter.

Bayesian and nonstationary models

2. Using `evgam`, fit a generalized Pareto model to all four stations

- set thresholds at the site-wise 98% empirical percentile.
- consider site-specific scale
- and a common shape for all four stations. To this effect, create a new data frame by concatenating exceedances, with a factor (dummy) for the station identifier.

3. Use a likelihood ratio test to compare the models with different shape parameters for each station. Are there evidence of different shapes?
4. Based on the estimated coefficients, which model will yield the highest 50-year return level? Rank the stations from smallest to largest.
5. Using `revdbayes` and a binomial/generalized Pareto model with a Beta/maximal data information prior, obtain 50% credible interval for the model fitted to each station separately.
6. Compare these with approximate 90% credible interval for 50-year return levels obtained using the Gaussian approximation to the posterior from `evgam` for Montélimar.

Time series

1. Consider the temporal structure of the time series for Lyon Saint-Exupéry by plotting an extremogram: is there evidence of serial correlation in the extremes?
2. Using the `lite` package, estimate the extremal index and adjust return levels. Compare the estimates with the model that ignores clustering.

Conditional extremes

We could also consider a conditional model by taking as covariates humidity (`H2`) and mean temperature (`T2`). Fit the Heffernan–Tawn conditional extremes model to the trivariate data (`S2`, `H2`, `T2`), given that humidity is above 80%

1. preprocess the data to ensure approximate stationarity by keeping only data for the summer months (June, July and August)
2. are there evidence of asymptotic dependent variables?
3. produce threshold stability plots for the dependence parameters
4. estimate the probability that the temperature exceed it's 95% marginal summer quantile, given that humidity is above the 80% and wind speed is lower than the 50%.