Statistical Inference 2 Summer Term 2025

Lecture: Prof. Dr. Göran Kauermann
Exercises: Sergio Buttazzo, Jan Anders
Tutorial: Eugen Gorich

Exercise Sheet 6

Exercise 6.1 - Generalized Extreme Value Distribution

The following table, taken from the *Danish fire insurance* dataset, contains data regarding the yearly maximal inter-arrival times (measured in days) of insurance claims, between the years 1980 and 1990. The inter-arrival time is defined as the number of days passing between two subsequent claims.

Year	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Max inter-arrival time	11	12	10	22	16	19	14	8	12	16	9

Assume that in each year the inter-arrival times are exponentially distributed with intensity $\lambda > 0$, and that the yearly maximal inter-arrival times are *i.i.d.*

(a) Show that the maximum of an *i.i.d.* sequence of exponentially distributed random variables $\{X_1, \dots, X_n\}$, with intensity $\lambda > 0$, can be approximated with a Gumbel distribution. How do the location and scale parameters of the resulting Gumbel distribution relate to λ and n?

Hint: use the convergence properties for $\max\{X_1, \dots, X_n\} - \frac{\log n}{\lambda}$.

(b) Use MLE (derive formulas as far as possible) to estimate the extreme value distribution associated with the yearly maximum inter-arrival times.

(c) Based on your estimation of the parameters, evaluate the probability for the yearly maximum inter-arrival time to exceed 25 days.

Exercise 6.2 - Extreme Value Analysis

Climate scientists are interested in modeling extreme precipitation (rainfall) events. In this exercise, we want to assess how the intensity and frequency of extreme rainfall events in Germany have evolved over time, based on historical daily precipitation data.

(a) Import the dataset for the daily precipitation between 1950 and 2022 from Weather Station No. 21 in Germany (Original source: DWD). Plot the distribution of daily rainfall, and provide basic descriptive statistics. Comment on its shape and the presence of extreme values.

(b) Estimate the distribution parameters `location`, `scale` and `shape` for the monthly precipitation maxima using `fevd` from the `extRemes` package, and plot the estimated curves over the data.

(c) Plot a time series of the yearly daily rainfall maxima (i.e. a series of 73 points, one per year, depicting the maximum daily rainfall for each year). Is there evidence of an increasing trend in the intensity of extreme rainfall?

(d) Let us now shift our focus from yearly maxima to overall tail or extreme events. Define an "extreme rainfall day" as a day with precipitation in the upper 5% quantile of all non-zero daily rainfall values. Using this definition, count and plot the number of days with extreme precipitation per year. Compare this to the yearly maxima from (c), e.g. by visual analysis. What do we observe?

(e) Discuss how the definition of "extreme" (e.g. based on fixed thresholds vs. relative quantiles) can affect real world conclusions. What are the implications for statistical inference and climate science?

Exercise 6.3 - Extreme Value Inference

In subjects like engineering and insurance, predicting the extrema is often of great importance. We will now consider the yearly maximum wind speed in miles per hour taken from "Extreme Value and Related Models in Engineering and Science Applications" by Castillo et al.

(a) Plot the different extreme value distributions (Gumbel, (Inverted) Weibull, Frechet-Pareto) for multiple parameter values. How do they relate to the generalized extreme value distribution?

(b) Read the data (`evo.txt`, found on the moodle page) and visualize it.

(c) Fit a generalized extreme value distribution to the data using Maximum Likelihood. Interpret the results. What distribution do you get? Does this make sense in this context?

You can use the `fevd` function in the `extRemes` package.

(d) Use the model found in point c) to estimate the probability of a yearly maximum wind speed above 50/75/100/150 miles per hour. Interpret the results.