Exercise Set 04: Distributions and Extreme Values

BEE 4850/5850, Fall 2024

> **Due Date**
>
> Friday, 2/16/24, 9:00pm

> 💡 Tip
>
> You can find a Jupyter notebook, data, and a Julia 1.9.x environment in the exercise's Github repository. You should feel free to clone the repository and switch the notebook to another language, or to download the relevant data file(s) and solve the problems without using a notebook. In either of these cases, if you using a different environment, you will be responsible for setting up an appropriate package environment.
>
> Regardless of your solution method, make sure to include your name and NetID on your solution PDF for submission to Gradescope.

Overview

Instructions

The goal of this exercise is for you to explore how distributional assumptions can influence estimates of extreme values and return levels.

Load Environment

The following code loads the environment and makes sure all needed packages are installed. This should be at the start of most Julia scripts.

```julia
import Pkg
Pkg.activate(@__DIR__)
Pkg.instantiate()
```

The following packages are included in the environment (to help you find other similar packages in other languages). The code below loads these packages for use in the subsequent notebook (the desired functionality for each package is commented next to the package).

```julia
using DataFrames # tabular data structure
using Distributions # API to work with statistical distributions
using Plots # plotting library
using StatsBase # statistical quantities like mean, median, etc
using StatsPlots # some additional statistical plotting tools
using Optim # optimization tools
```

Problems

Problem 1

We represent an experimental data-generating process by finding the maximum value of 100 draws from $\mathcal{N}(2.5, 10)$. Repeat this experiment 1,000 times and plot the histogram of resulting maxima. What can you conclude about the distribution of the maxima?

Problem 2

Fit (through maximum likelihood) the following distributions to your data from Problem 1.

1. A Normal (`Normal`) distribution;
2. A Cauchy (`Cauchy`) distribution;
3. A Weibull (`Weibull`) distribution;
4. A Fréchet (`Frechet`) distribution.

> **ℹ Note**
>
> As we will talk about next week, the Weibull and Fréchet distributions are special cases of a Generalized Extreme Value (GEV) distribution. But more on this later!

How well do they fit the distribution of maxima (plot the fitted distributions over your histogram in one or more plots, use Q-Q plots, etc.)? What is the return level associated with an exceedance probability of 0.01, and how does this compare to the observed quantile? Which distributional fit do you prefer and why?

References