INSTITUTE FOR MATHEMATICS
FREIE UNIVERSITÄT BERLIN

Bachelor's Thesis

Extreme Value Theory

Nathalie Lieckfeld

Primary Reviewer:	Prof. Dr. Peter Koltai
Secondary Reviewer:	Dr. Stefanie Winkelmann
Semester:	Winter 2019/20
Author:	Nathalie Lieckfeld
Field of Study:	Mathematics B.Sc.

Submission date: **11. Februrary 2020**

Statutory Declaration

I herewith declare that I have composed the Bachelor's thesis with the title:

Extreme Value Theory

myself and without use of any other than the cited sources and aids. Sentences or parts of sentences quoted literally are marked as such; other references with regard to the statement and scope are indicated by full details of the publications concerned.

The thesis in the same or similar form has not been submitted to any examination body and has not been published. This thesis was not yet, even in part, used in another examination or as a course performance.

Berlin (Germany), 11.02.2020

Nathalie Lieckfeld

Abstract

This Bachelor's thesis introduces classic extreme value theory. The main focus lies on the characterization and analysis of limiting distributions for normalized and centered maxima of independent and identically distributed random variables. Maxima are defined as block maxima or as exceedances over a high threshold. Depending on the definition, the limiting distribution coincides with the generalized extreme value (GEV) distribution or the generalized Pareto (GP) distribution. The Fisher-Tippett-Theorem and Pickands-Balkema-de Haan Theorem are stated and proved. These theorems provide information on the asymptotic behavior of distributions for normalized and centered maxima. To examine wether a given distribution function belongs to the Domain of Attraction of a GEV distribution or to the Domain of residual life time Attraction of the GP distribution, sufficient and necessary conditions on the distribution function are discussed. Finally, both the GEV distribution and the GP distribution are fitted to real world data of historic precipitation heights measured in Dahlem Berlin and verified as an appropriate model.

Contents

Nomenclature

$\doteq$	definition
$\mathcal{N}(\mu, \sigma^2)$	Gaussian or normal distribution with mean μ and variance σ^2
$\sim$	$f(x) \sim g(x)$ as $x \to x_0$ means that $\lim_{x \to x_0} \frac{f(x)}{g(x)} = 1$
$\xrightarrow{d}$	convergence in distribution
$\xrightarrow{w}$	convergence at continuity points
$DA(G_\alpha)$	domain of attraction of the α-stable distribution G_α
$DRA(H_{\xi;\mu,\sigma})$	domain of residual life time attraction of the GP distribution function $H_{\xi;\mu,\sigma}$
$G_{\xi;\mu,\sigma}$ and GEV distribution	generalized extreme value distribution
$H_{\xi;\mu,\sigma}$ and GP distribution	generalized Pareto distribution
M_n	maximum of $X_1, ..., X_n$
$MDA(G)$	maximum domain of attraction of the distribution function G
$o(1)$	small o, $f(x) = o(g(x))$ as $x \to \infty$ if for every $c > 0$ there exists a constant N such that $\mid f(x) \mid < cg(x)$ for all $x > N$
$X_1, ..., X_n$	sequence of random variables
x_F	upper endpoint of the distribution function F
i.i.d.	independent and identically distributed

1 Introduction and Motivation

Extreme value theory comprises the theory of analyzing observed extremes and forecasting further extremes [Gum12]. Extreme events are of interest in a variety of fields of science and technology reaching from hydrology, logistics and agriculture to financial markets and more [LFdF$^+$16]. The reason a whole theory around extreme events exists, is that they often relate to hazards or highly undesirable incidents impacting e.g. the environment, infrastructure or financial markets [LFdF$^+$16]. In particular climate and weather extremes can cost the life of thousands of people such as the European heat wave of 2003 which counted about 70,000 deaths [LFdF$^+$16]. It is clear that the prevention, assessment and handling of such extreme events, and therefore extreme value theory itself, is of great importance. The key idea behind extreme value theory is to focus on tail distributions of random variables rather than their average behavior. The goal is to construct a model that aims to describe only the extreme data [LFdF$^+$16].

In the 1920s, civil engineers were the first to apply extreme value theory, realizing that cotton fibre breaks when the weakest fiber breaks. But early history of extreme value theory reaches further back to 1709, when Nicolas Bernoulli considered the problem of the mean largest distance from the origin of n points lying on a straight line of length t [Gum12]. Two hundred years later in the fist half of the 20th century, Fréchet was the first to obtain a limiting distribution of the largest value. Shortly after, L.H.C. Tippett and R.A. Fisher published their works on extreme values, presenting in addition to Fréchet's result two other possible limiting distributions for extreme values [Gum12]. Their combining results are nowadays known as the so called *Fisher-Tippett Theorem*. In 1937, R. von Mises published his well known work on sufficient conditions under which the three distributions are a valid approximation for the distribution of extreme events [KN00]. Yet, the most significant historical work is *Statistics of Extremes* by Emil Gumbel which provides a large variety of graphs and tables visualizing the theoretical results. His contributions made extreme value theory accessible to a broader audience [Gum12, CBTD01, KN00].

In this thesis we will provide an introduction into extreme value theory. In Chapter 2, we start off by introducing some basic probabilistic and statistical definitions and concepts, which are needed throughout this work. Chapter 3 is then devoted to the asymptotic behavior of distributions of sums of random variables. This chapter builds a perfect preparation for the main part of this work which we cover in Chapter 4. There we will discuss two different characterizations of extremes and encounter the corresponding limiting distributions of normalized and centered maxima: the generalized extreme value (GEV) distribution and the generalized Pareto (GP) distribution. The *Fisher-Tippett Theorem* and *Pickand-Balkeman-de Haan Theorem* proving the limiting behavior will be discussed in detail. Furthermore, the sufficient and necessary conditions on a distribution function to show such behavior and belong to the maximum domain of attraction of the GEV distribution or to the domain of residual life time attraction of the GP distributions are provided. The final Chapter 5 includes the principles of point estimation and model diagnostics. In a concluding example the GEV and GP distribution will be fitted to a precipitation data set. All computations were done using Python and R and the code is accessible on Github[1].

[1] https://github.com/nathalieveronika/extreme_value_theory

2 Basic Probability and Statistical Concepts

This chapter can be seen as a “reference book”. We will refrain from proofs and examples since the focus lies on subsequent chapters. For more details refer to [CB02] or [Beh12].
Statistical procedures aim to draw inference about a population by conducting random experiments. The following definitions are essential for this purpose:

Definition 1. (sample space [CB02])
The set Ω of all possible outcomes of a particular experiment is called *sample space* for the experiment.

Definition 2. (event [CB02])
Any subset of possible outcomes of the sample space Ω, including Ω itself, is called an *event.*

Definition 3. (sigma algebra [CB02])
A collection of subsets of the non-empty sample space Ω is called a *sigma algebra* $\mathcal{B}$ if it satisfies the following properties:

- The empty set is an element of $\mathcal{B}$, i.e. $\emptyset \in \mathcal{B}$
- $\mathcal{B}$ is closed under complementation, i.e. if $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$
- $\mathcal{B}$ is closed under countable unions, i.e. if $A_1, A_2, ... \in \mathcal{B}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$

Definition 4. (measurable space [Bil95])
Let Ω be a sample space and $\mathcal{B}$ be the associated sigma algebra. Then the pair $(\Omega, \mathcal{B})$ is called *measurable space.*

Definition 5. (measurable function [Bil95])
Let $(\Omega, \mathcal{B})$ and $(\Omega', \mathcal{B}')$ be two measurable spaces. The function $f : \Omega \to \Omega'$ is called *measurable* if for each $E \subseteq \mathcal{B}'$

$$f^{-1}(E) = \{\omega \in \Omega : f(\omega) \in E\} \in \mathcal{B}.$$

When conducting a random experiment, the occurrences of events can be described by means of a *probability measure:*

Definition 6. (probability measure [CB02])
Let Ω be a sample space and $\mathcal{B}$ be the associated sigma algebra. A function $\mathbb{P} : \mathcal{B} \to [0, 1]$ is called a *probability measure* if it satisfies all of the following probability axioms:

- $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{B}$
- $\mathbb{P}(\Omega) = 1$

- $\mathbb{P}$ is a σ-additive function meaning that if $A_1, A_2, ... \in \mathcal{B}$ are pairwise disjoint i.e. if $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \mathbb{P}\Big(\sum_{i=1}^{\infty} A_i\Big)$

Furthermore, the probability measure satisfies the *normalizing property* $\sum_{A \in \Omega} \mathbb{P}(A) = 1$, since $\sum_{A \in \Omega} \mathbb{P}(A) = \mathbb{P}(\Omega)$.

Definition 7. (probability space [Beh12])
Let $\mathcal{B}$ be a sigma algebra on a non-empty sample space Ω and let $\mathbb{P}$ be the associated probability measure. Then the triple $(\Omega, \mathcal{B}, \mathbb{P})$ is called a *probability space.*

Theorem 1. *(Properties of Events [CB02])*
Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space and let $A, B \in \mathcal{B}$ be events. Then,

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- *If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$*
- $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \mathbb{P}\Big(\sum_{i=1}^{\infty} A_i\Big)$ *for any $A_1, A_2, ... \in \Omega$ (union bound/Boole's inequality)*

Next, the important notion of independence of events is introduced. To start with, we can think of two events being independent if the (non-)occurrence of one event does not change the probability of the other event. Naturally, more than two events can be independent as well:

Definition 8. (independence of events [CB02])

- Two events $A, B \in \mathcal{B}$ in a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ are called *independent events,* if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$
- Let $(E_i)_{i \in I} \subseteq \mathcal{B}$ be any collection of events in a probability space $(\mathcal{B}, \Omega, \mathbb{P})$. Then the events are called
  - *pairwise independent*, if
$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i)\mathbb{P}(E_j)$$
    for every indices $i, j \in I$, $i \neq j$
  - *k-wise independent,* if
$$\mathbb{P}\left(\bigcap_{j=1}^{\hat{k}} E_{i_j}\right) = \prod_{j=1}^{\hat{k}} \mathbb{P}\left(E_{i_j}\right)$$
    for every $\hat{k} \leq k$ and distinct indices $i_1, ..., i_{\hat{k}} \in I$

· *mutually independent*, if they are k-wise independent for every finite k

Definition 9. (conditional probability [CB02])
Let $A, B \in \mathcal{B}$ be two events in a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ where B is such that $\mathbb{P}(B) > 0$. Then the *conditional probability of A given B* is defined as

$$\mathbb{P}(A \mid B) \doteq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Note that if A and B are independent, then $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

Now, we consider (*discrete and continuous*) *random variables* which in fact are not variables as their name suggests, but functions mapping the sample space Ω to the real numbers $\mathbb{R}$.

Definition 10. (discrete random variable [CB02])
A *discrete real-valued random variable* (or simply *discrete random variable*) on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ is a function $X : \Omega \longrightarrow \mathbb{R}$ with discrete sample space Ω.

Example 1. Consider the experiment of rolling a fair, six-sided die. A corresponding random variable is discrete since the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ is discrete.

Formally, we defines a discrete random variable as a subset of the sample space Ω [Bil95]:

$$X^{-1}(x) \doteq \{\omega \in \Omega | X(\omega) = x\} \quad \text{with } x \in \mathbb{R}.$$

Thus, the discrete random variable $X^{-1}(x)$ is an event and therefore can be assigned a probability:

Definition 11. (cumulative distribution function [CB02])
The *cumulative distribution function* (or *cdf* or simply *distribution function*) $F_X(x)$ of a random variable X is defined by

$$F_X(x) \doteq \mathbb{P}(X \leq x) \quad \text{for all } x.$$

To satisfy the probability axioms, F must be a non-decreasing function of x such that $F(x_-) = 0$ and $F(x_+) = 1$, where x_- and x_+ are the lower and upper limits of Ω, respectively [CBTD01].

Notation. Technically, the distribution function is defined as

$$F_X(x) \doteq \mathbb{P}\left((X^{-1}(\hat{x}) : \hat{x} \in (-\infty, x]\right) = \mathbb{P}\left(\omega \in \Omega : X(\omega) \leq x\right),$$

but notation is commonly abused for the sake of simplicity [Geo15].

Similarly, for multiple random variables the *joint distribution function* is defined:

Definition 12. (joint distribution function [CB02])
Let $X_1, ..., X_n$ be random variables on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$. Then the *joint distribution function* $F_{X_1,...,X_n}$ of $X_1, ..., X_n$ is defined as

$$F_{X_1,...,X_n}(x_1, ..., x_n) \doteq \mathbb{P}(X_1 \leq x_1, ..., X_n \leq x_n).$$

Definition 13. (quantile function [RTR07])
Let F_X be the distribution of a random variable X. Then the inverse of F_X is the so called *quantile function* F_X^{-1} defined as

$$F_X^{-1}(q) \doteq \inf \{x : F(x) \geq q\}$$

for $0 < q < 1$.

Remark. The continuity of the distribution function defines the corresponding random variable in the sense that a random variable X is discrete if its corresponding distribution function $F_X(x)$ is a step function of x [CB02]. In fact, *continuous random variables* are defined by means of their distribution functions:

Definition 14. (continuous random variable [CB02])
A real-valued random variable X on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ is a *continuous random variable* if the corresponding distribution function is continuous.

Notation. Both discrete and continuous random variables are always denoted with upper case letters, whereas the realized value of the random variable is denoted with the corresponding lower case letter. This means that a random variable X can take a value x.

Both discrete and continuous random variables are not only associated with the distribution function but also with the *probability mass function* and the *probability density function*, respectively. [CB02].

Definition 15. (probability mass function [CB02])
The *probability mass function* (or *pmf*) f_X of a discrete random variable X on a probability space $(\mathcal{B}, \Omega, \mathbb{P})$ is defined as

$$f_X(x) \doteq \mathbb{P}(X = x) \quad \text{for all } x \text{ in } \Omega.$$

Thus, given the probability mass function f_X of a discrete random variable X, the corresponding distribution function F_X is determined by the following sum:

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{y \in \Omega : y \leq x} f_X(y).$$

Notation. Similar to the note on notation above, the probability mass function is really defined as $f_X(x) \doteq \mathbb{P}\left(X^{-1}(x)\right)$.

Now for more than one random variable we define similar functions for a random vector, a vector consisting of random variables.

Definition 16. (joint probability mass function [CB02])
Let (X, Y) be a discrete bivariate random vector. Then the function $f_{X,Y}(x, y) : \mathbb{R}^2 \to \mathbb{R}$ defined by $f_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$ is called *joint probability mass function* of (X, Y).

Definition 17. (marginal probability mass function [CB02])
Let (X, Y) be a discrete bivariate random vector with joint probability mass function $f_{X,Y}(x, y)$. Then the *marginal probability mass functions* of X and Y, $f_X(x) = \mathbb{P}(X = x)$ and $f_Y(y) = \mathbb{P}(Y = y)$ are given by

$$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x, y) \quad \text{and} \quad f_Y(y) = \sum_{x \in \mathbb{R}} f_{X,Y}(x, y).$$

Definition 18. (probability density function [CB02])
The *probability density function* or *pdf* $f_X(x)$ of a continuous random variable X on a probability space $(\mathcal{B}, \Omega, \mathbb{P})$ with distribution function F_X is defined as the function f_X satisfying

$$F_X(x) = \int\limits_{-\infty}^{x} f_X(t)dt \ \text{ for all } x.$$

Definition 19. (joint probability density function [CB02])
The function $f_{X,Y}(x, y) : \mathbb{R}^2 \to \mathbb{R}$ is called *joint probability density function* of the continuous bivariate random vector (X, Y) if, for every $A \subset \mathbb{R}^2$,

$$\mathbb{P}\left((X, Y) \in A\right) = \int_A \int f_{X,Y}(x, y) dx\, dy.$$

Definition 20. (marginal probability density function [CB02])
Let (X, Y) be a continuous bivariate random vector with joint probability density function $f_{X,Y}(x, y)$. Then the *marginal probability density function* of X and Y are given by

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy, & -\infty \leq x \leq \infty \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx, & -\infty \leq y \leq \infty. \end{aligned}$$

The distribution function and the probability mass or probability density function of a discrete or continuous random variable X can be used to specify the *distribution* of X:

Definition 21. (distribution [Geo15])
Let X be a random variable on a probability space $(\mathcal{B}, \Omega, \mathbb{P})$. The *distribution* or *law* of X is the function $(\mathbb{P} \circ X^{-1})$ defined by

$$(\mathbb{P} \circ X^{-1})(A) \doteq \mathbb{P}(X \in A) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in A\}\right) \ \text{ for } A \subseteq \mathbb{R}.$$

If ρ is the distribution of X, we write $X \sim \rho$.

Definition 22. (support [CB02])
Let X be a random variable on a probability space $(\mathcal{B}, \Omega, \mathbb{P})$ with probability mass function (or probability density function) f_X. The *support* of X is the set of all values that X takes with positive probability, i.e.

$$supp(X) \doteq \{x \in \mathbb{R} \mid f_X(x) > 0\}.$$

To visualize the above results, we will next consider two distributions, for both discrete and continuous random variables, and plot their corresponding distribution function, probability mass function and probability density function:

Example 2. (Poisson distribution [Beh12])
A discrete random variable X with sample space $\Omega = \mathbb{N}_0 = \{0, 1, 2, ...\} = supp(X)$ is said to be Poisson distributed with parameter $\lambda > 0$, if, for each $x \in \mathbb{N}_0$ the probability mass function of X is

$$\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

where $0! \doteq 1$. This is denoted as $X \sim \text{Pois}(\lambda)$. The Poisson distribution models the number of occurrences of events in a fixed time interval of length λ. Both the probability mass function and distribution function of the Poisson distribution are plotted in Figure 1 for $\lambda = 1, 5, 10$.

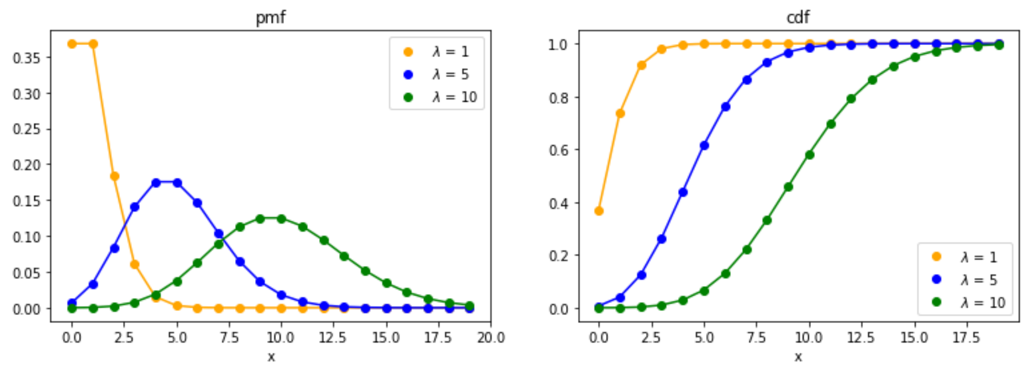

Figure 1: probability mass function (left) and distribution function (right) of Poisson distribution for $\lambda = 1, 5, 10$

Example 3. (exponential distribution [Beh12])
A continuous random variable X with $\Omega = [0, \infty) = supp(X)$ is said to be exponential distributed with parameter $\lambda > 0$, denoted as $X \sim Exp(\lambda)$, if the probability density function of X is

$$f_X(x) = \lambda e^{-\lambda x}.$$

The exponential distribution is often used to model waiting times such that for appropriate λ the waiting time between c and d is $\int_c^d \lambda e^{-\lambda x} dx$. Both the probability density function and the distribution function of the exponential distribution are plotted in Figure 2 for $\lambda = 0.5, 1, 1.5$.

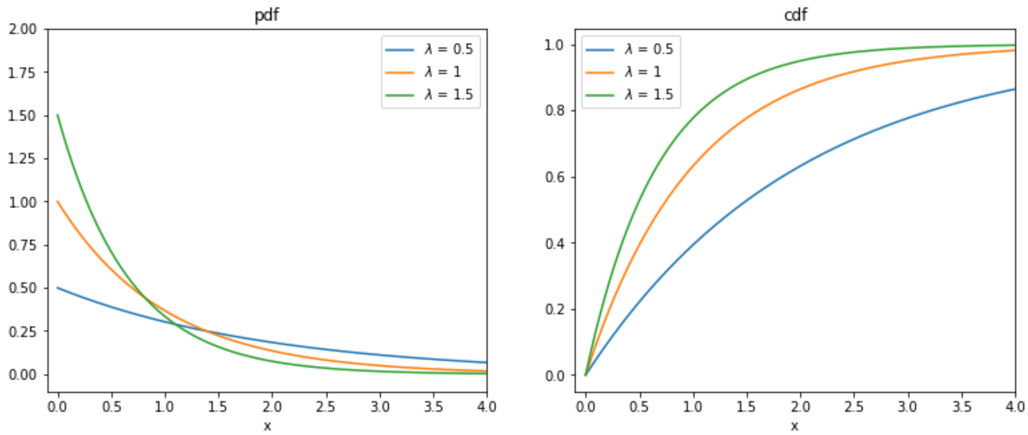

Figure 2: probability density function (left) and distribution function (right) of the exponential distribution for $\lambda = 0.5, 1, 1.5$

We continue with a theorem which shows that different combinations of random variables are again random variables.

Theorem 2. *[Beh12]*
Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space and $X, Y, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be random variables. Then the following functions are also random variables:

- cX *for every* $c \in \mathbb{R}$
- $X + Y$
- $X \cdot Y$
- $\frac{X}{Y}$ *if* $Y(\omega) \neq 0$ *for all* ω
- $\sup_{n\in\mathbb{N}} X_n : \omega \mapsto \sup_{n\in\mathbb{N}} X_n(\omega)$ *and* $\inf_{n\in\mathbb{N}} X_n : \omega \mapsto \inf_{n\in\mathbb{N}} X_n(\omega)$ *if the supremum and infimum are finite for all* ω
- $\lim_{n\to\infty} X_n : \omega \mapsto \lim_{n\to\infty} X_n(\omega)$ *if the limes in* $\mathbb{R}$ *exists for all* ω

Next, we transfer the previous results on independence of events to independence of random variables:

Definition 23. (independence of random variable [CB02])
Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space. Let $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ be random variables with joint probability density function or joint probability mass function $f(x, y)$. Let $f_X(x)$ and $f_Y(y)$ denote their marginal probability density function or marginal probability mass function. Then X and Y are called *independent,* if for every x and y

$$f(x, y) = f_X(x) f_Y(y).$$

Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space. Let $X_1, ..., X_n : \Omega \to \mathbb{R}$ be random variables with joint probability density function or joint probability mass function $f(x_1, ..., x_n)$. Let $f_{X_i}(x_i)$ denote the marginal probability density function or marginal probability mass function of X_i. Then $X_1, ..., X_n$ are *mutually independent,* if for every $(x_1, ..., x_n)$

$$f(x_1, ..., x_n) = f_{X_1}(x_1) \cdot ... \cdot f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

When conducting a random experiment, the random variable X is expected to take an average value amongst the realized values x. This value can be thought of as the weighted average according to the corresponding probability mass or probability density function. Differing between continuous and discrete random variables, the formal definitions of the *expected value* of a random variable X are as follows:

Definition 24. (expected value, expectation [CB02])
Let $(\Omega_1, \mathcal{B}_1, \mathbb{P}_1)$ be a probability space and $X : \Omega_1 \to \mathbb{R}$ be a discrete random variable with probability mass function $f_X(x)$. Then the *expected value* or *mean* of X is defined as

$$\mathbb{E}[X] \doteq \sum_{x \in \mathbb{R}} x f_X(x)$$

provided that the sum exists.
Let $(\Omega_2, \mathcal{B}_2, \mathbb{P}_2)$ be a probability space and $Y : \Omega_2 \to \mathbb{R}$ be a continuous random variable with probability density function $f_Y(y)$. Then the *expected value* or *mean* of Y is defined as

$$\mathbb{E}[Y] \doteq \int_{-\infty}^{\infty} Y(y) f_Y(y) dy$$

provided that the integral exists.
If $\mathbb{E}[X] = \infty$ or $\mathbb{E}[Y] = \infty$ we say that the expected value does not exist.

Notation. Once again, the more precise definition for the expected value of a discrete or continuous random variable would be $\mathbb{E}(X) \doteq \sum_{\omega \in \Omega_1} X(\omega) \mathbb{P}_1(\omega)$ and $\mathbb{E}(Y) \doteq \int_{\omega \in \Omega_2} Y(\omega) d\mathbb{P}_2(\omega)$, respectively.

Theorem 3. *(Properties of the Expectation [Beh12])*
Let X and Y be real-valued random variables on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ such that $\mathbb{E}(X) \neq \infty$ and $\mathbb{E}(Y) \neq \infty$. Then,

- $\mathbb{E}[cX] = c\mathbb{E}[X]$ *for* $c \in \mathbb{R}$

- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- *If* $X(\omega) \leq Y(\omega)$ *for all* ω, *then* $\mathbb{E}[X] \leq \mathbb{E}[Y]$
- *if* X *and* Y *are independent, then* $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

To describe how much the realized values x of a random variable X deviate from its mean $\mathbb{E}[X]$, *variance* is introduced next:

Definition 25. (variance and standard deviation [CB02])
Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space and let X be a (discrete or continuous) random variable. Then its *variance* $\mathbb{V}$ is defined as

$$\mathbb{V}[X] \doteq \mathbb{E}\left[(X - \mathbb{E}(X))^2\right].$$

and the *standard deviation* of X is defined as $\sigma \doteq \sqrt{\mathbb{V}(X)}$.

Theorem 4. *(Properties of the Variance [Beh12])*
Let X, Y, $X_1, ..., X_n$ *be random variables on the probability space* $(\Omega, \mathcal{B}, \mathbb{P})$. *Provided that both* $\mathbb{V}[X]$ *and* $\mathbb{V}[Y]$ *exist,*

- $\mathbb{V}[cX] = c^2\mathbb{V}[X]$ *for* $c \in \mathbb{R}$
- $\mathbb{V}[X + c] = \mathbb{V}[X]$ *for* $c \in \mathbb{R}$, *i.e. if* X *is a constant* c, *then* $\mathbb{V}[X] = 0$
- $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- *if* $X_1, ..., X_n$ *are independent then* $\mathbb{V}[X_1 + ... + X_n] = \mathbb{V}[X_1] + ... + \mathbb{V}[X_n]$

The variance belongs to a special class of expectations, the so called *moments.* In fact, the variance is equivalent to the *second central moment:*

Definition 26. (moment [CB02])
Let X be a random variable on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$. The n^{th} *moment of* X is defined as

$$\mathbb{E}[X^n]$$

for each $n \in \mathbb{N}$. The n^{th} *central moment of* X is defined as

$$\mathbb{E}\left[(X - \mathbb{E}(X))^n\right]$$

for each $n \in \mathbb{N}$.

Even if it is not the case in many real world scenarios, the observations made in an experiment are often assumed to be *independent and identically distributed* as it makes calculations easier.

Definition 27. (independent and identically distributed [CB02])
The random variables $X_1, ..., X_n$ are called *independent and identically distributed*, denoted by *i.i.d.,* if

- they are mutually independent
- AND they are identically distributed i.e. the distribution function $F(x)$ is the same for each X_i in the sense that $\mathbb{P}(X_i \in E) = \mathbb{P}(X_j \in E)$ for all $E \subseteq \mathbb{R}$ and distinct $i, j \in \{1, ..., n\}$

3 Fluctuations of Sums

Before the classic theory of extreme value theory is introduced, we first consider the limit distribution for sums $\bar{X}_n$ of i.i.d. random variables. Why? Because sums $\bar{X}_n$ and maxima M_n of i.i.d. random variables share one remarkable property: their limiting distributions coincide with the class of so called *(max-) stable distributions* [EKM13]. It is helpful to cover the theory behind limits of sums first because it happens in a more familiar environment and provides an excellent preparation for understanding the theory on limiting distributions of extreme values. Throughout this chapter we follow the structure and ideas of [EKM13, Chapter 2], complementing it with material from different sources. Again, proofs will be omitted as we reserve this for the most important part of this work in Chapter 4.
The main focus in Chapter 3 lies on the limit behavior of a sequence $(X_n)_{n\in\mathbb{N}}$ of i.i.d. random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with common *non-degenerate distribution function* F. Unfortunately, it is not possible to obtain information in regards to the convergence of this sequence as such since it usually does not converge [EKM13]. Instead, it is best to concentrate on the convergence of sums S_n and cumulative means $\bar{X}_n$, where

$$\begin{aligned} S_0 &\doteq 0,\ S_n \doteq X_1 + ... + X_n \quad n \geq 1 \\ \bar{X}_n &\doteq \frac{1}{n} S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad n \geq 1 \end{aligned}$$

for which desirable convergence results will be obtained. For this purpose we introduce different modes of convergence: *convergence in probability*, *almost sure convergence* and *convergence in distribution*. It will turn out that with the concept of *convergence in distribution* the most desirable results on limiting distributions for sums of i.i.d. random variables will be achieved.

Note. From this chapter on, random variables are introduced without specifying their corresponding probability space which of course does not imply they do not have one!

3.1 Convergence of Random Variables

Definition 28. (convergence in probability [CB02])
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. This sequence is said to *converge in probability* to a random variable X, if for every $\epsilon > 0$

$$\mathbb{P}(\mid X_n - X \mid \geq \epsilon) \xrightarrow{n\to\infty} 0$$

which can be expressed equivalently as

$$\mathbb{P}(\mid X_n - X \mid < \epsilon) \xrightarrow{n\to\infty} 1.$$

Convergence in probability allows to state a first (weak) result about limits of sums of i.i.d. random variables:

Theorem 5. *(Weak Law of Large Numbers [CB02])*
Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2 < \infty$ for all $i \in \{1, ..., n\}$. Then $\bar{X}_n$, defined as above, converges in probability to μ, i.e. for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\mid \bar{X}_n - \mu \mid < \epsilon) = 1.$$

It is clear that $\mu = \mathbb{E}[\bar{X}_n]$. This is easily checked since

$$\begin{aligned}
\mathbb{E}[\bar{X}_n] &= \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i] \\
&= \frac{\sum_{i=1}^{n} \mathbb{E}\left[X_i\right]}{n} \\
&= \frac{n\mu}{n} \\
&= \mu.
\end{aligned}$$

A stronger type of convergence leading to a stronger limit of sums of i.i.d. random variables is *almost sure convergence*:

Definition 29. (almost sure convergence [CB02])
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. This sequence is said to *converge almost surely* to a random variable X, denoted $X_n \xrightarrow{a.s.} X$, if, for every $\epsilon > 0$, the sequence converges in probability to X, i.e.

$$\lim_{n\to\infty} \mathbb{P}(\mid X_n - X \mid < \epsilon) = 1$$

and if, for every $\epsilon > 0$,

$$\mathbb{P}\left(\lim_{n\to\infty} \mid X_n - X \mid < \epsilon\right) = 1.$$

Theorem 6. *(Strong Law of Large Numbers (SLLN) [CB02])*
Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2 < \infty$. Let $\bar{X}_n$ be defined as before. Then $\bar{X}_n$ converges almost surely to μ, i.e. for every $\epsilon > 0$,

$$\mathbb{P}\left(\lim_{n\to\infty} \mid \bar{X}_n - \mu \mid < \epsilon\right) = 1.$$

Theorem 7. *(Refined Version of SLLN [CB02])*
Let $X_1, X_2, ..., X_n$ be i.i.d. random variables and S_n defined as before. Suppose that p∈(0,2). Then the refined version of the SLLN

$$n^{-1/p}(S_n - an) \xrightarrow{a.s.} 0 \tag{1}$$

holds for some real constant a if and only if $\mathbb{E}\left[|X_1|^p\right] < \infty$.

It can be shown that the SLLN and its refined version with normalizing constant yield the following almost sure approximation of first order, respectively [EKM13]:

$$\bar{X}_n = \mu + o(1) \text{ a.s.},$$

and

$$\bar{X}_n = \mu + o(n^{1/p-1}) \text{ a.s.}$$

with $\mathbb{E}[|X_1|^p] < \infty$ and for some $p \in [1, 2)$. However, the best possible almost sure approximation of first order of $\bar{X}_n$ by its mean μ is

$$\bar{X}_n = \mu + O\left((\ln \ln n/n)^{-1/2}\right) \text{ a.s.}.$$

For more details on these results, please refer to [EKM13, Chapter 2].
Note that (1) does not hold for $p \geq 2$, as $\limsup_{n\to\infty} n^{-1/2}|S_n - a_n| = \infty$. Fortunately, information on the growth of $n^{-1/2}S_n$ and second-order, non-trivial approximations on the limiting behavior of sums of i.i.d. random variables are still deducible when introducing another notion of convergence:

Definition 30. (convergence in distribution [CB02])
A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ with distribution function F_{X_n} *converges in distribution (converges weakly)* to a random variable X with distribution function F_X, denoted $X_n \xrightarrow{d} X$, if

$$F_{X_n}(x) \xrightarrow{n\to\infty} F_X(x),$$

for every x at which F_X is continuous.

Notation. For simplicity convergence at continuity points is denoted by $\xrightarrow{w}$ from now on, i.e. F_{X_n} converges in distribution to a random variable X, if $F_{X_n}(x) \xrightarrow[n\to\infty]{w} F_X(x)$.

With the definition of convergence in distribution in mind, we will now cover a class of distributions which is very helpful to fully characterize the limiting distribution of sums of i.i.d. random variables.

3.2 Stable Distributions

Definition 31. (stable distribution function and stable random variable [EKM13])
Both a random variable X and its corresponding distribution function F are called *stable* if

$$c_1X_1 + c_2X_2 \xrightarrow{d} a(c_1, c_2)X + b(c_1, c_2)$$

for i.i.d. random variables X, X_1, X_2, for all non-negative numbers c_1 and c_2 and appropriate real numbers $a(c_1, c_2) > 0$ and $b(c_1, c_2)$.

Now, assuming that $X_1, ...X_n$ are i.i.d. stable random variables. Then for S_n defined as before it holds that

$$S_n = X_1 + ... + X_n \xrightarrow{d} a_nX_1 + b_n$$

for some constants $a_n > 0$, b_n and for $n \geq 0$. This can be rewritten as

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} X_1.$$

Conclusion. Stable distributions are limit distribution for sums of i.i.d. random variables!

In fact, stable distributions are the only possible *non-degenerate* limiting distributions as the following *Limit Property of a Stable Law* states. But first let us understand what it means for a distribution function to be *non-degenerate.*

Definition 32. (non-degenerate and degenerate distribution function [Bil95, LFdF$^+$16])
A distribution function F is *non-degenerate* if it does not have a unit jump, i.e. if there exists no $y_0 \in \mathbb{R}$ such that $F(y_0) = 1$ and $F(y) = 0$, $\forall y < y_0$.
Conversely, a distribution function $\hat{F}$ is *degenerate* if it has the form $\Delta(x - x_0)$ for some x_0 where

$$\Delta(x) \doteq \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}.$$

Theorem 8. *(Limit Property of Stable Laws [EKM13])*
The class of stable non-degenerate distribution functions coincides with the class of all possible non-degenerate limit laws for properly normalized and centered sums of i.i.d. random variables.

Degenerate limit laws for sums of i.i.d. random variables are not much of interest as they would not provide much information about how the sums are distributed. The Limit Property of Stable Laws however, is quite remarkable as it provides a full characterization of the class of possible non-degenerate limiting distributions for sums of i.i.d. random variables. As we will soon see, under certain conditions and with appropriate choice of constants we can equate the limiting distribution with the *normal distribution.* First of all, it is helpful to characterize stable distributions by their *characteristic functions* since the density functions of stable distributions are in general very complicated [EKM13].

Definition 33. (characteristic function [Beh12])
Let X be a real-valued random variable. The *characteristic function* ϕ_X of X is defined as

$$\phi_X : \mathbb{R} \to \mathbb{C}, \quad \phi_X(t) \doteq \mathbb{E}(\exp iXt).$$

Theorem 9. *(Spectral representation of a stable law [EKM13])*
The characteristic function of a stable distribution is

$$\phi_X(t) = \exp\left(i\gamma t - c|t|^\alpha \left(1 - i\beta sign(t) z(t, \alpha)\right)\right) \tag{2}$$

where γ is a real constant, i is the imaginary unit, $c > 0$, $\alpha \in (0, 2]$, $\beta \in [-1, 1]$ and

$$z(t, \alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \text{if } \alpha \neq 1 \\ -\frac{2}{\pi} \ln|t| & \text{if } \alpha = 1 \end{cases}$$

and

$$sign(t) = \begin{cases} -1 & \textit{if } t < 0 \\ 0 & \textit{if } t = 0 \\ 1 & \textit{if } t > 0. \end{cases}$$

Note. For simplicity the location parameter γ will be set γ to zero from now on.

Definition 34. (characteristic exponent, α-stable distribution [EKM13])
Consider the characteristic function (2) of a stable distribution. Then the number α is called the *characteristic exponent* and the corresponding distribution is called an *α-stable distribution.*

For $\alpha = 2$ we obtain a special and important case of a stable distribution, the so called *normal* or *Gaussian distribution* [EKM13]. The *normal distribution* is a very popular distribution in statistics since its symmetric bell shape and tractability provide it with very convenient features [CB02].

Definition 35. (normal distribution/Gaussian distribution [CB02])
A random variable $X : \Omega \to \mathbb{R}$ is said to be *normally* or *Gaussian distributed* with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, denoted $X \sim \mathcal{N}(\mu, \sigma^2)$, if it has probability density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

i.e. if

$$F_X(y) = \mathbb{P}(X \leq y) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{y} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Remark. [EKM13] The characteristic function for a normally distributed random variable X with mean zero and variance $2c$ is

$$\begin{aligned} \phi_X(t) &= \exp\left(-c|t|^\alpha \left(1 - i\beta \mathrm{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right)\right)\right) \\ &= \exp\left(-c|t|^2 \left(1 - i\beta \mathrm{sign}(t) \tan\left(\frac{\pi 2}{2}\right)\right)\right), \quad \alpha = 2 \\ &= \exp\left(-c|t|^2 (1-0)\right) \\ &= \exp\left(-ct^2\right) \end{aligned}$$

provided γ is set to zero.
Finally, the most important result in this chapter can be presented:

3.3 Central Limit Theorem

The *Central Limit Theorem* proves that the normal distribution, i.e. the 2-stable distribution, is a limiting distribution for a large number of different distributions. Indeed, it enables to find the limit of $n^{-1/2} S_n$ which diverged under the notion of almost sure convergence.

Theorem 10. *(Central Limit Theorem/CLT [CB02])*
Let $X_1, ..., X_n$ be a sequence of i.i.d. random variables. Let $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2 > 0$, with both μ and σ^2 being finite. Then,

$$\frac{n^{-1/2}(S_n - \mu n)}{\sigma} = \sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1).$$

As announced before, the notion of *convergence in distribution* made it possible to state the powerful result, the CLT, on limiting distributions for normalized and centered sums X_n of i.i.d. random variables. The CLT states that $\frac{S_n - b_n}{a_n}$ converges in distribution to an α-stable distribution, with $\alpha = 2$ and appropriate constants $a_n = \sqrt{n}\sigma$ and $b_n = \mu n$, provided that $\mu, \sigma^2 < \infty$.
It is crucial to realize that the distribution of the $X_1, ..., X_n$ is of no concern in regards to their convergence to a normal distribution. For a better understanding of the importance of the theorem, the meaning of the CLT is visualized in the following example:

Example 4. Assume 1000 samples are drawn from the exponential distribution (Example 3) with $\lambda = 1$. With increasing sample size n the sample mean becomes approximately normal distributed with $\mu = 0$ and $\sigma^2 = 1$. See Figure 3.

3.4 Domains of Attraction

It is now of interest which conditions the normalized and centered sums of i.i.d. random variables must satisfy in order to converge to an α-stable distribution G_α with $\alpha \in (0, 2]$. Therefore, we introduce the *Domain of Attraction*:

Definition 36. (Domain of Attraction [EKM13])
Let F be the common distribution function of the i.i.d. random variables $X_1, ..., X_n$. Then F is in the *domain of attraction* of the α-stable distribution G_α if there exists $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} G_\alpha \tag{3}$$

holds. This is denoted as $F \in DA(G_\alpha)$.

Instead of considering the conditions the normalized and centered sums of i.i.d. random variables must satisfy to belong to an to an α-stable distribution, we are really interested in the conditions which a distribution function F needs to satisfy in order to be in the domain of attraction of an α-stable law. However, for a full characterization of the conditions we first need to introduce *regular varying functions*:

Definition 37. (regularly varying function [Res13])
A measurable function $U : \mathbb{R}_+ \to \mathbb{R}_+$ is *regularly varying* at ∞ with index ρ if

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\rho$$

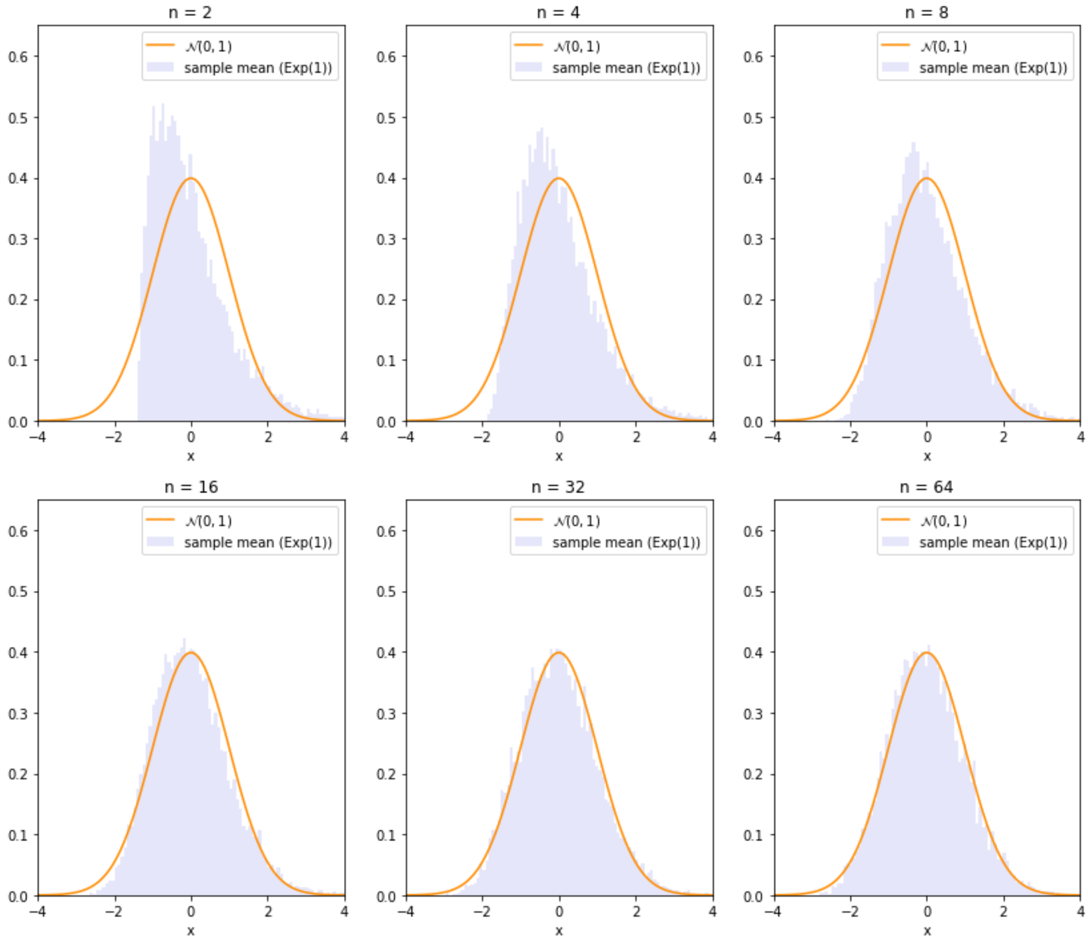

Figure 3: Visualization of the CLT. With increasing sample size n the sample mean of exponentially distributed random variables becomes approximately normal distributed.

for $x > 0$. If $\rho = 0$, then U is called a *slowly varying function* and is usually denoted by $L(x)$.

Theorem 11. *(Characterization of Domain of Attraction [EKM13])*

- *The distribution function F of a random variable X belongs to the domain of attraction of a normal distribution if and only if one of the following holds:*
  - *The integral*
$$\int_{|y| \leq x} y^2 dF(y)$$
    is slowly varying
  - $\mathbb{E}(X^2) < \infty$

· $\mathbb{E}(X^2) = \infty$ *and* $\mathbb{P}(|X| > x) = o\left(x^{-2} \int_{|y| \leq x} y^2 dF(y)\right), \ x \to \infty$

- *The distribution function F of a random variable belongs to the domain of attraction of an α-stable law for some $\alpha < 2$ if and only if*

$$F(-x) = \frac{c_1 + o(1)}{x^\alpha} L(x), \quad 1 - F(x) = \frac{c_2 + o(1)}{x^\alpha} L(x), \ x \to \infty,$$

where L is slowly varying and c_1, c_2 are non-negative constants such that $c_1 + c_2 > 0$.

This theorem suggests that the domain of attraction of a normal distribution comprises many more distributions than the domain of attraction of an α-stable law with $\alpha < 2$. It is still open to answer how to choose appropriate centering and normalizing constants:

Theorem 12. *(Centering and Normalizing Constants in the CLT [EKM13])*
Let $F \in DA(G_\alpha)$. Define the quantities

$$K(x) \doteq x^{-2} \int_{|y| \leq x} y^2 dF(y), \quad x > 0$$

and

$$Q(x) \doteq \mathbb{P}(|X| > x) + K(x), \quad x > 0.$$

Then the normalizing constants a_n in equation (3) can be chosen as the unique solution of the equation

$$Q(a_n) = \frac{1}{n}, \quad n \geq 1,$$

and the centering constants b_n in equation (3) can then be chosen as

$$b_n = n \int_{|y| \leq a_n} y^2 dF(y).$$

If $\alpha < 2$, then a_n can be alternatively chosen as

$$a_n = \inf\{y : \mathbb{P}(\mid X \mid > y) < \frac{1}{n}\}, \ n \geq 1.$$

Summary. After having introduced different notions of convergence for sequences of random variables, we considered both the Weak and Strong Law of Large Numbers and a refined version. Convergence in distribution made it possible to characterize the class of non-degenerate stable distributions which turned out to be the class of limit distributions for properly normalized and centered sums of i.i.d. random variables. This class also includes the limit distribution for $n^{-1/2} S_n$ which could not be determined under the notion of convergence in probability and almost sure convergence. The most essential result in this chapter was the Central Limit Law, stating that with $a_n = \sqrt{n}\sigma$ and $b_n = \mu n$ the normalized and centered sums of i.i.d. random variables converge to the special 2-stable distribution, the Gaussian distribution. Moreover, conditions characterizing the domains of attraction of an α-stable law were provided.

4 Classic Extreme Value Theory

We are now fully equipped to introduce classic extreme value theory. In the subsequent chapter we will consider two different characterizations of extreme event and encounter the corresponding limiting distributions of normalized and centered maxima. We will learn how the *Fisher-Tippett Theorem* and *Pickands-Balkeman-de Haan Theorem* prove the limiting behavior of extremes. Moreover, we will discuss conditions which distribution function have to satisfy such that their limit behaves as desired.

4.1 The i.i.d. Setting

In extreme value theory there are two common approaches to characterize an *extreme event*. One common approach is to divide the data into blocks of same size n. Then, the maximum of each block is considered as an extreme event. This is the so called *block maxima approach*. Instead, we can also set an appropriate threshold and define all observations exceeding it as extreme events. This approach is called the *peaks-over-threshold approach* [CBTD01]. In the course of this chapter we have a close look at both.

The main concern is to develop the respective distribution functions of extreme events. It turns out that distribution of extreme events can be approximated by the *Generalized Extreme Value Distribution* or the *Generalized Pareto Distribution,* depending on the chosen approach to define extreme events. In both cases, $(X_n)_{n\in\mathbb{N}}$ describes a sequence of i.i.d. random variables with common distribution function F i.e. $F_{X_1}(y) = F_{X_2}(y) = ... \doteq F(y)$.

4.2 Block Maxima and the Generalized Extreme Value Distribution

The i.i.d. random variables $(X_n)_{n\in\mathbb{N}}$ with distribution function F usually represent values which are measured on some time interval, e.g. hourly, weekly, monthly or yearly [CBTD01]. The main concern lies on the statistical behavior of M_n

$$M_n \doteq \max\{X_1, ..., X_n\}, \quad n \geq 2$$

which is defined as the maximum of the first n random variables [CBTD01].

Remark. Due to the following equation, it is obvious that all subsequent results for maxima lead to analogous results for minima [LLR12]:

$$m_n \doteq \min(X_1, ..., X_n) = -\max(-X_1, ..., -X_n)$$

For more details on the limiting behavior of minima, have a look at [LLR12, Section 1.8].

In theory, the distribution function of M_n is known since it can be derived from the distribution function of the underlying X_is:

$$\begin{aligned}
\mathbb{P}(M_n \leq y) &= \mathbb{P}(X_1 \leq y, ..., X_n \leq y) \\
&= \mathbb{P}(X_1 \leq y) \cdot .... \cdot \mathbb{P}(X_n \leq y) \qquad \text{independence of } X_1, ..., X_n \\
&= F_{X_1}(y) \cdot ... \cdot F_{X_n}(y) \\
&= (F(y))^n
\end{aligned} \tag{4}$$

Unfortunately, in practice the distribution function F is unknown, making it impossible to deduce the distribution function of M_n in the above manner. There are two possibilities to handle this problem: We could approximate F with standard statistical procedures. However, this arises another problem as even slight deviations for F can lead to significant deviations for F^n [CBTD01].
Another approach to tackle this issue, is to look for approximate families of models for F^n. To estimate these families, we only use the provided extreme data [CBTD01]. This work follows the latter approach.

4.2.1 Fisher-Tippett Theorem

Consider the limits of F^n at the *upper endpoint* of F defined as

$$x_F \doteq \sup\{x \in \mathbb{R} : F(x) < 1\}$$

since extremes will occur somewhere in the upper end of the support of the distribution i.e. in the upper end of $\{x \in \mathbb{R} : F(x) \neq 0\}$. This means that there is reason to believe a connection between the limiting distribution of M_n and the distribution function F^n in the right tail around x_F exists [EKM13].

Differing between two cases we obtain that

1. for any $x < x_F$
$$\mathbb{P}(M_n \leq x) = (F(x))^n \xrightarrow{n \to \infty} 0$$

2. for any $x \geq x_F$ and $x_F < \infty$
$$\mathbb{P}(M_n \leq x) = (F(x))^n \xrightarrow{n \to \infty} 1$$
implying that $M_n \xrightarrow{a.s.} x_F$ and that the limit distribution function of M_n is degenerated [EKM13].

To avoid a degenerate limiting distribution, it is necessary to apply a positive affine transformation. In fact, we consider convergence in distribution for normalized and centered maxima M_n^* [EKM13, CB02]:

$$M_n^* \doteq \frac{M_n - b_n}{a_n}$$

for sequences of constants $(a_n)_{n\in\mathbb{N}} > 0$ and $(b_n)_{n\in\mathbb{N}} \in \mathbb{R}$. Yet, it still remains unknown if the limit for $\mathbb{P}(\frac{M_n-b_n}{a_n} \leq x)$ exists with appropriate choices for the constants.
It turns out that the answer to this question is the *Fisher-Tippett Theorem*, revealing the fact that the family of limiting distributions for normalized and centered maxima consists of three limiting distribution types. The Fisher-Tippett Theorem can be seen as the extreme value analogue to the Central Limit Theorem.

Theorem 13. *(Fisher-Tippett Theorem [CBTD01])*
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables. If there exist normalizing sequences of constants $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, with $a_n > 0$ for all n, such that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \qquad \text{as } n \to \infty$$

where G is a non-degenerate distribution function, then G coincides with one of the following three families - defined as the three extreme value distributions*:*

Type I: Gumbel distribution

$$\Lambda(x) = \exp\Big(-\exp\Big(-\tfrac{x-\mu}{\sigma}\Big)\Big), \ -\infty < x < \infty$$

Type II: Fréchet distribution

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{if } x \leq \mu, \\ \exp\Big(-\Big(\tfrac{x-\mu}{\sigma}\Big)^{-\alpha}\Big) & \text{if } x > \mu \end{cases} \quad \text{with } \alpha > 0$$

Type III: Weibull distribution

$$\Psi_\alpha(x) = \begin{cases} \exp\Big(-\Big(-\Big(\tfrac{x-\mu}{\sigma}\Big)^{\alpha}\Big)\Big) & x < \mu \\ 1 & x \geq \mu \end{cases} \quad \text{with } \alpha > 0$$

All three families have a location parameter $\mu \in \mathbb{R}$ *and a* scale parameter $\sigma > 0$. *The Fréchet and Weibull families have an additional* shape parameter $\alpha > 0$.

Remark. As we will see later, the three extreme value distribution types can be combined in a one-parameter representation. For now, we will primarily use the standardized form of the extreme value distributions where μ is set to zero and $\sigma = 1$. Figure 4 depicts the density and distribution functions of the three (standardized) extreme value distribution types.

Note that even if the limiting families are quite different for modeling purposes they are very similar from a mathematical point of view [EKM13]:

$$\begin{aligned} & X \text{ has distribution function } \Phi_\alpha \\ \iff \quad & \log(X^\alpha) \text{ has distribution function } \Lambda \\ \iff \quad & -X^{-1} \text{ has distribution function } \Psi_\alpha. \end{aligned}$$

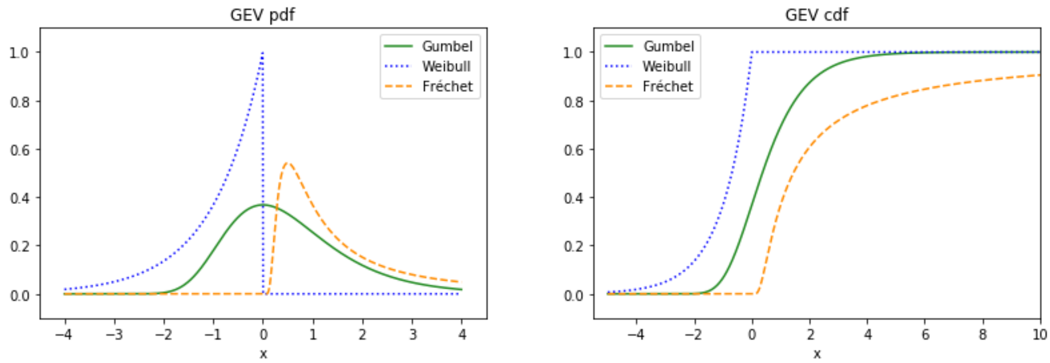

Figure 4: Density (left) and distribution (right) for the three standardized extreme value types, with $\sigma = 1$, $\mu = 0$ and with $\alpha = 1$ for Weibull and Fréchet.

To prove the Fisher-Tippett Theorem it is vital to know the main tool needed for the proof: the *Convergence to Types Theorem.* This is why we will first consider some definitions and lemmas which are necessary to prove this theorem. We will then point out similarities to sums of i.i.d. random variables in preparation for the proof of the Fisher-Tippett Theorem. The subsequent part mainly follows the structure of [LLR12, Chapter 1].

Preparing for the Proof of the Fisher-Tippett Theorem

4.2.1.1 Inverse functions and Convergence to Types Theorem

In the Fisher-Tippett Theorem we have already classified the extreme value distributions into different *types.* This term is often used in a very informal sense but it is indeed possible to state a formal definition [LLR12]:

Definition 38. (Convergence to Types [LLR12, EKM13])
Two distribution functions G_1 and G_2 *are of the same type* if there exist constants $a > 0$ and $b \in \mathbb{R}$ such that

$$G_1(x) = G_2(ax + b) \text{ for all } x. \tag{5}$$

Similarly, two random variables X and Y *belong to the same type* if there exist constants $a > 0$ and $b \in \mathbb{R}$ such that

$$X \stackrel{d}{=} aY + b,$$

where $\stackrel{d}{=}$ denotes the fact that X and $aY + b$ share the same distribution function.

By saying that G_1 and G_2 are equivalent if they satisfy (5) for constants $a > 0$, $b \in \mathbb{R}$, distribution functions can be divided into equivalence classes, which are called *types.*

A quick check shows that in fact, that (5) defines an equivalence relation $\sim$:

- (reflexive property)
  $G_1 \sim G_1$ since $G_1(x) = G_1(ax + b)$ with $a = 1, b = 0$

- (symmetric property)

$$\begin{aligned}
G_1 \sim G_2 &\Longleftrightarrow G_1(x) = G_2(ax+b) \\
&\Longleftrightarrow \mathbb{P}(X_1 \leq x) = \mathbb{P}(X_2 \leq ax+b) \\
&\Longleftrightarrow \mathbb{P}(X_1 \leq a'x'+b') = \mathbb{P}(X_2 \leq x') \text{ with } x' = ax+b,\ a' = \frac{1}{a},\ b' = -\frac{b}{a} \\
&\Longleftrightarrow G_1(a'x'+b') = G_2(x') \\
&\Longleftrightarrow G_2 \sim G_1
\end{aligned}$$

- (transitive property)
$G_1 \sim G_2 \Longleftrightarrow G_1(x) = G_2(ax+b)$ and $G_2 \sim G_3 \Longleftrightarrow G_2(x) = G_3(\tilde{a}x + \tilde{b})$
This implies that

$$G_1(x) = G_2(ax+b) = G_3(\tilde{a}(ax+b) + \tilde{b}) = G_3(\dot{a}x + \dot{b})$$

with $\dot{a} = \tilde{a}a$ und $\dot{b} = \tilde{a}b + \tilde{b}$, i.e. $G_1 \sim G_3$.

Theorem 14. *(Convergence to Types Theorem/Khintchine's Convergence Theorem [LLR12, EKM13])*
Let $(F_n)_{n\in\mathbb{N}}$ be a sequence of distribution functions and G be a non-degenerate distribution function. Let $a_n > 0$ and $b_n \in \mathbb{R}$ be constants such that for $n \geq 1$

$$F_n(a_n x + b_n) \xrightarrow{w} G(x). \tag{6}$$

Then, for some non-degenerate distribution function G_*, constants $\alpha_n > 0, \beta_n$ and $n \geq 1$

$$F_n(\alpha_n x + \beta_n) \xrightarrow{w} G_*(x), \tag{7}$$

if and only if

$$\frac{\alpha_n}{a_n} \xrightarrow{n\to\infty} a \quad \textit{and} \quad \frac{\beta_n - b_n}{a_n} \xrightarrow{n\to\infty} b \tag{8}$$

for some $a > 0$ and $b \in \mathbb{R}$, and then

$$G_*(x) = G(ax+b). \tag{9}$$

Before we can prove the Convergence to Types Theorem, it is necessary to introduce another definition and three lemmas:

Definition 39. (generalized inverse function, quantile function [LLR12, EH13])
Let $\Psi : \mathbb{R} \to \mathbb{R}$ be a non-decreasing right continuous function, i.e. no jump occurs when the limit point is approached from the right. Then we define the *generalized inverse function* $\Psi^- : \mathbb{R} \to \left[-\infty, +\infty\right]$ by

$$\Psi^-(y) \doteq \inf\Big\{x \in \mathbb{R} \mid \Psi(x) \geq y\Big\},\ y \in \mathbb{R}$$

with the convention that the generalized inverse function is $+\infty$ on an empty set. If $\Psi : \mathbb{R} \to [0,1]$ is a distribution function, then $\Psi^- : [0,1] \to \big[-\infty, +\infty\big]$ is called the *quantile function* of Ψ.

Figure 5 illustrates an example of an increasing function T and its corresponding generalized inverse function T^-. The following lemma shows that generalized inverse functions have some very practical properties:

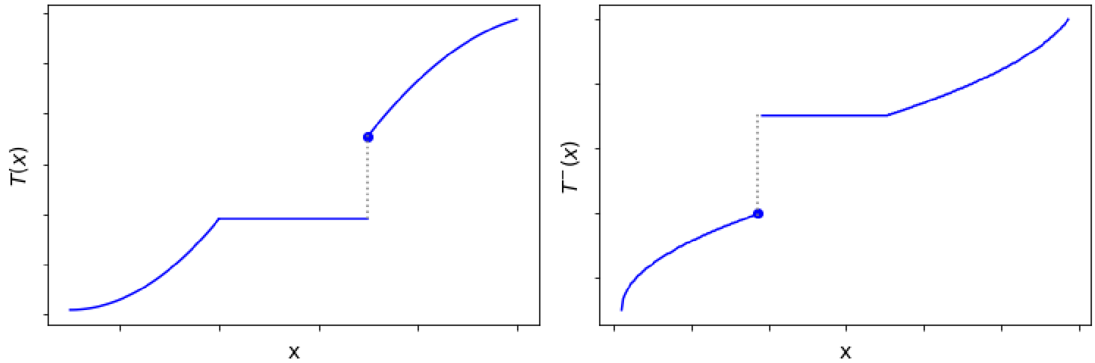

Figure 5: An increasing function T (left) and its corresponding generalized inverse function T^- (right)

Lemma 1. *(Properties of Generalized Inverse Functions [LLR12])*
Let ψ be a non-decreasing right continuous function and let ψ^- be its generalized inverse function. Then,

(a) $G^-(y) = a^{-1}(\Psi^-(y+c) - b)$, *where* $G(x) \doteq \Psi(ax+b) - c$ *and* $a > 0, b, c$ *are constants*

(b) $\Psi^-(\Psi(x)) = x$, *if* Ψ^- *is continuous*

(c) *if H is a non-degenerate distribution function, there exist $y_1 < y_2$ such that $H^-(y_1) < H^-(y_2)$ are well defined and finite*

Proof. [LLR12]
(a) Follows almost directly by definition:

$$\begin{aligned}
G^-(y) &= \inf\Big\{x \in \mathbb{R} \mid \Psi(ax+b) - c \geq y\Big\} \\
&= \inf\Big\{x \in \mathbb{R} \mid \Psi(ax+b) \geq y + c\Big\} \\
&= \inf\Big\{a^{-1}(ax+b-b) \mid \Psi(ax+b) \geq y + c\Big\} \\
&= a^{-1}\Big(\inf\Big\{ax+b \mid \Psi(ax+b) \geq y+c\Big\} - b\Big) \\
&= a^{-1}(\Psi^-(y+c) - b)
\end{aligned}$$

(b) Let Ψ^- be continuous. By definition of Ψ^- and since Ψ is non-decreasing it follows that

$$\Psi^-(\Psi(x)) = \inf\Big\{y \mid \Psi(y) \geq \Psi(x)\Big\} \leq x.$$

Assume $\Psi^-(\Psi(x)) < x$ for some x. Then again by definition of Ψ^- there must exist some $z < x$ such that $\Psi(z) \geq \Psi(x)$. Since by assumption Ψ is non-decreasing, $\Psi(z) = \Psi(x)$. The proof is now completed by contradiction since,

- $\Psi^-(y) \leq z$ for $y = \Psi(z) = \Psi(x)$
- and $\Psi^-(y) \geq x$ for $y > \Psi(z) = \Psi(x)$

Because $z < x$, the inequalities above imply that Ψ^- is not continuous since the limit from the right and from the left at y do not coincide, yet we assumed Ψ^- to be continuous. Hence, $\Psi^-(\Psi(x)) = x$.
(c) Let H be a non-degenerate distribution function. Then, there exist x_1 and x_2 such that $0 < H(x_1) < H(x_2) < 1$. It is clear that for $y_1 \doteq H(x_1)$ and $y_2 \doteq H(x_2)$ both $H^-(y_1)$ and $H^-(y_2)$ are well-defined. To prove $H^-(y_1) < H^-(y_2)$, note that $H^-(y_2) = H^-(H(x_2)) = \inf\Big\{ \text{y} \mid H(y) \geq H(x_2)\Big\} \geq x_1$. For $H^-(y_2)$ to be equal to x_1, it must be satisfied that $H(z) \geq H(x_2)$ for all $z > x_1$, implying that $H(x_1) = \lim_{\epsilon \searrow 0} H(x_1 + \epsilon) \geq H(x_2)$. Yet this contradicts the previous result $0 < H(x_1) < H(x_2) < 1$. Thus it follows that $H^-(y_2) = H^-(H(x_2)) > x_1 \geq H^-(H(x_1)) = H^-(y_1)$.

□

Remark. A non-degenerate distribution function G is non-decreasing and right continuous, and therefore has a generalized inverse function. This means that all properties listed in Lemma 1 also apply for G.

Lemma 2. *[LLR12] Let G be a non-degenerate distribution function and $a > 0$, $\alpha > 0$, $b, \beta \in \mathbb{R}$ be constants such that $G(ax + b) = G(\alpha x + \beta)$ for all x. Then $a = \alpha$ and $b = \beta$.*

Proof. [LLR12] By Lemma 1 (a), the generalized inverse of $G(ax + b)$ and $G(\alpha x + \beta)$ is given by

$$a^{-1}(G^-(y) - b) = \alpha^{-1}(G^-(y) - \beta) \text{ for all } y. \tag{10}$$

Lemma 1 (c) also shows that there exist $y_1 < y_2$ such that $-\infty < G^-(y_1) < G^-(y_2) < +\infty$. Substituting y_1 and y_2 into (10), we obtain

$$a^{-1}(G^-(y_1) - b) = \alpha^{-1}(G^-(y_1) - \beta) \text{ and } a^{-1}(G^-(y_2) - b) = \alpha^{-1}(G^-(y_2) - \beta),$$

which implies that $a = \alpha$ and $b = \beta$ (since $G^-(y_1) \neq G^-(y_2)$). □

Lemma 3. *[Bil95] Let $(F_n)_{n\in\mathbb{N}}$ be a sequence of distribution functions, F be a distribution function, (a_n) and (b_n) be a sequence of constants and a and b be two constants.
If $F_n \xrightarrow{n\to\infty} F$, $a_n \xrightarrow{n\to\infty} a$, and $b_n \xrightarrow{n\to\infty} b$, then $F_n(a_n x + b_n) \xrightarrow{n\to\infty} F(ax + b)$.*

Proof. [Bil95] Let x be a continuity point of $F(ax+b)$ and $\epsilon > 0$. Since distribution functions have only countably many discontinuities (see e.g. [Bil95, Section 14]) it is possible to choose two continuity points u and v of F such that $u < ax+b < v$ and $F(u) - F(v) < \epsilon$. Note that it is not necessary to write $\mid F(u) - F(v) \mid < \epsilon$ because F is a distribution function so it follows that $F(u) \geq F(v)$ for $u > v$. Since $F_n \xrightarrow{n\to\infty} F$, it follows that for n large enough $\mid F_n(u) - F(u) \mid < \epsilon$ and $\mid F_n(v) - F(v) \mid < \epsilon$. And since $a_n \xrightarrow{n\to\infty} a$, and $b_n \xrightarrow{n\to\infty} b$, we have again for n large enough $u < a_n x + b_n < v$. Putting this together:

$$\begin{aligned}
F(ax+b) - 2\epsilon &< F(u) - \epsilon && \text{since } F(ax+b) - F(u) < \epsilon \\
&< F_n(u) && \text{since } \mid F_n(u) - F(u) \mid < \epsilon \\
&\leq F_n(a_n x + b_n) && \text{since } u < a_n x + b_n \\
&\leq F_n(v) && \\
&< F(v) + \epsilon && \text{since } \mid F_n(v) - F(v) \mid < \epsilon \\
&< F(ax+b) + 2\epsilon
\end{aligned}$$

Because $F(ax+b) - 2\epsilon < F_n(a_n x + b_n) < F(ax+b) + 2\epsilon$ implies that for large enough n, $\mid F_n(x) - F(x) \mid < \epsilon$ for all x, we have $F_n(a_n x + b_n) \xrightarrow{n\to\infty} F(ax+b)$, as desired. □

Now all required tools have been covered which are necessary to prove the Convergence to Types Theorem:

Proof. (Convergence to Types Theorem [LLR12]) For the purpose of our proof, redefine $\tilde{\alpha}_n \doteq \frac{\alpha_n}{a_n}$, $\tilde{\beta}_n \doteq \frac{\beta_n - b_n}{a_n}$ and $\tilde{F}_n(x) \doteq F_n(a_n x + b_n)$. The proof is divided into two parts:

- (6) and (8) $\overset{!}{\Longrightarrow}$ (7) and (9)

  The definitions of $\tilde{\alpha}_n$, $\tilde{\beta}_n$ and $\tilde{F}_n(x)$ enable us to rewrite the above equations. Instead of (6), (7) and (8), we have

$$\tilde{F}_n(x) \xrightarrow{w} G(x), \tag{11}$$

$$\tilde{F}_n(\tilde{\alpha}_n x + \tilde{\beta}_n) \xrightarrow{w} G_*(x), \tag{12}$$

  since

$$\begin{aligned}
F_n(\alpha_n x + \beta_n) &= F_n(\tilde{\alpha}_n a_n x + \tilde{\beta}_n a_n + b_n) \\
&= F_n(a_n(\tilde{\alpha} x + \tilde{\beta}_n) + b_n) \\
&= \tilde{F}_n(\tilde{\alpha}_n x + \tilde{\beta}_n),
\end{aligned}$$

  and

$$\tilde{\alpha}_n \xrightarrow{n\to\infty} a \quad \text{and} \quad \tilde{\beta}_n \xrightarrow{n\to\infty} b \quad \text{for some } a > 0, b. \tag{13}$$

  With Lemma 3 it follows that if (11) and (13) are true, so is (12). This means that (6) and (8) imply (7) with $G_*(x) = G(ax+b)$.

- (6) and (7) $\stackrel{!}{\Longrightarrow}$ (8) and (9)

  By assumption, G_* is a non-degenerate distribution function. This implies that G_* does not have a unit jump i.e. there exist two distinct points $\tilde{x}$ and $\dot{x}$ such that $0 < G_*(\tilde{x}) < 1$ and $0 < G_*(\dot{x}) < 1$. To complete the proof we need to show that $\tilde{\alpha}_n \to a$ and $\tilde{\beta}_n \to b$:

  1. Proving that $(\tilde{\alpha}_n)$ and $(\tilde{\beta}_n)$ must be bounded:
     To prove this it suffices to show that $(\tilde{\alpha}_n\tilde{x} + \tilde{\beta}_n)$ and $(\tilde{\alpha}_n\dot{x} + \tilde{\beta}_n)$ must be bounded. Assuming $(\tilde{\alpha}_n\tilde{x} + \tilde{\beta}_n)$ is not bounded, we can choose a sequence (n_k) such that $\tilde{\alpha}_{n_k}\tilde{x} + \tilde{\beta}_{n_k} \to \pm\infty$. With (11) and the fact that G is a distribution function it follows that

     $$\tilde{F}_{n_k}(\tilde{\alpha}_{n_k}\tilde{x} + \tilde{\beta}_{n_k}) \xrightarrow{n_k \to \infty} \begin{cases} 0 & or \\ 1. \end{cases}$$

     Yet this contradicts (12) for $x = \tilde{x}$ since G_* is a non-degenerate distribution function. Therefore, $(\tilde{\alpha}_n\tilde{x} + \tilde{\beta}_n)$ must be bounded. It is easy to deduce with analogous arguments that $(\tilde{\alpha}_n\dot{x} + \tilde{\beta}_n)$ must also be bounded. Combining the boundedness of the two sequences $(\tilde{\alpha}_n\tilde{x} + \tilde{\beta}_n)$ and $(\tilde{\alpha}_n\dot{x} + \tilde{\beta}_n)$ shows that both $(\tilde{\alpha}_n)$ and $(\tilde{\beta}_n)$ must be bounded.
  2. Proving $\tilde{\alpha}_n \to a$ and $\tilde{\beta}_n \to b$:
     Since $(\tilde{\alpha}_n)$ and $(\tilde{\beta}_n)$ are bounded there must exist constants a, b and a sequence (n_k) such that

     $$\tilde{\alpha}_{n_k} \xrightarrow{n_k \to \infty} a \text{ and } \tilde{\beta}_{n_k} \xrightarrow{n_k \to \infty} b.$$

     Applying Lemma 3, it follows that

     $$\tilde{F}_{n_k}(\tilde{\alpha}_{n_k}\tilde{x} + \tilde{\beta}_{n_k}) \xrightarrow{w} G(ax + b).$$

     Now (12) implies that $G(ax + b) = G_*(x)$. Since a distribution function must be a non-decreasing function of x we can deduce that $a > 0$ as required.
     Lastly, consider a different sequence (m_k) such that

     $$\tilde{\alpha}_{m_k} \xrightarrow{m_k \to \infty} \tilde{a} \text{ and } \tilde{\beta}_{m_k} \xrightarrow{m_k \to \infty} \tilde{b}.$$

     It would follow again that $a > 0$ but also that $\tilde{F}_{m_k}(\tilde{\alpha}_{m_k}\tilde{x} + \tilde{\beta}_{m_k}) \xrightarrow{w} G(\tilde{a}x + \tilde{b})$ which is equal to $G_*(x) = G(ax + b)$. To complete the proof, apply Lemma 2 which implies that $a = \tilde{a}$ and $b = \tilde{b}$, thus $\tilde{\alpha}_n \to a$ and $\tilde{\beta}_n \to b$.

□

4.2.1.2 Max-stable distributions

In Section 3.2 we have identified the family of possible non-degenerate limit laws for properly normalized and centered sums of i.i.d. random variables with the class of stable distributions. Similarly, we can now identify the family of possible limiting distributions for maxima of i.i.d. random variables with the class of *max-stable distributions.* In fact, we will learn that a distribution function is max-stable if and only if it belongs to one of the extreme value distributions [LLR12].

Definition 40. (max-stable distribution [CBTD01, EKM13])
Let X be a non-degenerate random variable and F the corresponding distribution function. The random variable X as well as the distribution function F are called *max-stable* if, for every $n \geq 2$, there exist constants $\alpha_n > 0$ and β_n such that

$$(F(\alpha_n x + \beta_n))^n = F(x) \qquad x \in \mathbb{R}$$

or analogously, if

$$max(X_1, ..., X_n) \xrightarrow{d} \alpha_n X + \beta_n \tag{14}$$

holds for i.i.d. random variables $X, X_1, ..., X_n$, appropriate constants $\alpha_n > 0$, $\beta_n \in \mathbb{R}$ and every $n \geq 2$.

Having the Limit Property of Stable Laws (Theorem 8) in mind, let us now introduce an analogous result for max-stable laws:

Theorem 15. *(Limit Property of Max-Stable Laws [EKM13])*
The class of max-stable distributions coincides with the class of all possible non-degenerate limit laws for properly normalized and centered maxima of i.i.d. random variables.

Proof. [EKM13] Let X be any max-stable random variable. By definition this implies that (14) holds for i.i.d. random variables $X, X_1, ..., X_n$, appropriate constants $\alpha_n > 0$, $\beta_n \in \mathbb{R}$ and every $n \geq 2$ which is analogous to saying

$$\frac{M_n - \beta_n}{\alpha_n} \xrightarrow{d} X$$

must hold. It follows that every max-stable distribution is a limiting distribution for the normalized and centered maxima of i.i.d. random variables.
Now, it still needs to be shown that the only limiting distribution for maxima M_n of i.i.d. random variables with centering and normalizing constants are max-stable. Thus, assume for M_n with corresponding distribution function F and appropriate constants $\alpha_n > 0$, $\beta_n \in \mathbb{R}$ that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - \beta_n}{\alpha_n} \leq x\right) = \lim_{n\to\infty} \mathbb{P}(M_n \leq \alpha_n x + \beta_n) = \lim_{n\to\infty} (F(\alpha_n x + \beta_n))^n = G(x), \quad x \in \mathbb{R}$$

where G is some non-degenerate distribution function.
Then for every $k \in \mathbb{N}$

$$\lim_{n\to\infty} (F(\alpha_n x + \beta_n))^{nk} = \left(\lim_{n\to\infty} F^n(\alpha_n x + \beta_n)\right)^k = (G(x))^k \quad x \in \mathbb{R}$$

and

$$\lim_{n\to\infty} (F(\alpha_{nk}x + \beta_{nk}))^{nk} = G(x) \quad x \in \mathbb{R}.$$

Applying the previous results from Theorem 14 (Convergence to Types), we can deduce that there exist constants $a_k > 0$ and $b_k \in \mathbb{R}$ such that

$$\frac{\alpha_{nk}}{\alpha_n} \xrightarrow{n\to\infty} a_k \text{ and } \frac{\beta_{nk} - \beta_n}{\alpha_n} \xrightarrow{n\to\infty} b_k$$

and

$$(G(\alpha_n x + \beta_n))^n = G(x) \quad x \in \mathbb{R}$$

i.e. the limiting distribution G is max-stable, as required. $\square$

Up to this point, we have covered some important prerequisites to prove the *Fisher-Tippet Theorem*. However, there is still one major theorem missing for which we need to prove two lemmas first:

Lemma 4. *[LLR12] Let G be a non-degenerate max-stable distribution function. Then there exist real functions $a(s) > 0$ and $b(s)$ defined for $s > 0$ such that*

$$(G(a(s)x + b(s)))^s = G(x), \text{ for all } x \in \mathbb{R}, s > 0.$$

Proof. [LLR12] By definition of a max-stable distribution function G, for every $n \geq 2$, there exist constants $\alpha_n > 0$ and β_n such that

$$(G(\alpha_n x + \beta_n))^n = G(x) \qquad x \in \mathbb{R}.$$

This implies that

$$\left(G(\alpha_{[ns]}x + \beta_{[ns]})\right)^{[ns]} = G(x), \tag{15}$$

where $[t]$ denotes the greatest integer less than or equal to t, for $t \in \mathbb{R}$. And,

$$\begin{aligned}
\left(G(\alpha_{[ns]}x + \beta_{[ns]})\right)^n &= \left(G(\alpha_{[ns]}x + \beta_{[ns]})\right)^{n - \frac{[ns]}{s} + \frac{[ns]}{s}} \\
&= \left(G(\alpha_{[ns]}x + \beta_{[ns]})\right)^{n - \frac{[ns]}{s}} \left(G(\alpha_{[ns]}x + \beta_{[ns]})\right)^{\frac{[ns]}{s}} \\
&= \left(G(\alpha_{[ns]}x + \beta_{[ns]})\right)^{n - \frac{[ns]}{s}} (G(x))^{\frac{1}{s}}, \text{ applying eq.(15).}
\end{aligned}$$

Since $n - \frac{[ns]}{s}$ remains bounded, $\left(G(\alpha_{[ns]}x + \beta_{[ns]})\right)^{n - \frac{[ns]}{s}} \xrightarrow{n\to\infty} 1$. It follows that

$$\left(G(\alpha_{[ns]}x + \beta_{[ns]})\right)^n \xrightarrow[n\to\infty]{w} (G(x))^{\frac{1}{s}}.$$

Since $(G(\alpha_n x + \beta_n))^n \xrightarrow{w} G(x)$ (by definition) and $G^{\frac{1}{s}}$ must be non-degenerate, Theorem 14 (Convergence to Types) is applicable with $\frac{\alpha_{[ns]}}{\alpha_n} \xrightarrow{n\to\infty} a$ and $\frac{\beta_{[ns]} - \beta_n}{\alpha_n} \xrightarrow{n\to\infty} b$, i.e. $G(a(s)x + b(s)) = (G(x))^{\frac{1}{s}} \iff (G(a(s)x + b(s)))^s = G(x)$ for some $a(s) > 0, b(s)$. This completes the proof.

$\square$

Lemma 5. *[DH76] Let u be a non-decreasing function. If for all $t, s \in \mathbb{R}$*

$$u(t+s) = u(t)u(s),$$

then either $u(t) = 0$ for all t, or $u(t) = e^{\rho t}$ for some constant $\rho \in \mathbb{R}$.

Proof. [DH76] Assume that $u(t_*) \neq 0$ for some $t_* \neq 0$ and assume that w.l.o.g. $t_* > 0$. It follows that for all integers $m \in \mathbb{N}$

$$\begin{aligned} u(mt_*) &= u(mt_* + t_* - t_*) \\ &= u(t_*)u((m-1)t_*) \\ &= u(t_*)u(t_*)u((m-2)t_*) \\ &= \ldots \\ &= (u(t_*))^m . \end{aligned}$$

At the same time

$$\begin{aligned} u(mt_*) &= u\left(\frac{n}{n} mt_*\right) \\ &= \left(u\left(\frac{mt_*}{n}\right)\right)^n \end{aligned}$$

for all integers $n \in \mathbb{N}$, implying that

$$u\left(\frac{mt_*}{n}\right) = u\left(t_*\right)^{\frac{m}{n}} .$$

Note that the set

$$\left\{\frac{m}{n} t_* | m, n \in \mathbb{N}\right\}$$

is dense in $\mathbb{R}^+$ and as u was assumed to be entirely non-decreasing, i.e. monotone, it follows that

$$u(at_*) = (u(t_*))^a$$

for $a > 0$. Furthermore, $u(0) = (u(0))^2$ and $u(x) = u(x)u(0)$ for all x imply that $u(0) = 1$. Hence,

$$\lim_{\epsilon \to 0} u(x+\epsilon) = \lim_{\epsilon \to 0} u(x)u(\epsilon) = u(x)$$

for all x from which it follows that u is continuous. Finally, the proof concludes since $u(-t) = u\left(-t\frac{t_*}{t_*}\right) = (u(t_*))^{-\frac{t}{t_*}}$, as desired. $\square$

Theorem 16. *[LLR12] Every non-degenerate max-stable distribution is an extreme value distribution. Conversely, every extreme value distribution is max-stable.*

Proof. [LLR12, DH76] Let G be a non-degenerate max-stable distribution function. Then the previous result from Lemma 4 holds for all $x \in \mathbb{R}, s > 0$ and real functions $a(s) > 0$ and $b(s)$. Assume that $0 < G(x) < 1$, then

$$\begin{aligned}
&(G(a(s)x + b(s)))^s = G(x) \\
\iff & -\log\left((G(a(s)x + b(s)))^s\right) = -\log G(x) \\
\iff & -s\log\left(G(a(s)x + b(s))\right) = -\log G(x) \\
\iff & -\log\left(-s\log G(a(s)x + b(s))\right) = -\log\left(-\log G(x)\right) \\
\iff & -\log\left(-\log G(a(s)x + b(s))\right) - \log s = -\log\left(-\log G(x)\right) \qquad (16)
\end{aligned}$$

Since G is a max-stable distribution function there exist constants $\alpha_2 > 0$ and β_2 such that $(G(\alpha_2 x + \beta_2))^2 = G(x)$ for all $x \in \mathbb{R}$ and thus, G cannot have a jump at any finite (upper or lower) endpoint. This follows by contradiction: Assume that G has a jump at its finite upper endpoint x_G i.e. $G(x_{G-}) < 1 = G(x_G)$ where $x_{G-} = \lim_{x \nearrow x_G} G(x)$. Since $x^2 < x$ for all $x \in (0, 1)$, it follows that $G(\alpha_2 x_{G-} + \beta_2) > G(x_{G-})$. This implies that $(G(\alpha_2 x_{G-} + \beta_2))^2 = 1 = G(x_{G-})$ and contradicts the assumption. Similar arguments are applicable for the finite lower endpoint of G. Moreover, the non-decreasing function $\Psi(x) \doteq -\log\left(-\log G(x)\right)$ is such that $\inf_x \{\Psi(x)\} = -\infty$ and $\sup_x \{\Psi(x)\} = +\infty$. Hence, the generalized inverse function $\Psi^-(y)$ exists for all real y. Rewriting (16) in terms of $\Psi(x)$ gives

$$\Psi(a(s)x + b(s)) - \log s = \Psi(x).$$

By Lemma 1 (a)

$$\frac{\Psi^-(y + \log s) - b(s)}{a(s)} = \Psi^-(y).$$

Subtracting $\Psi^-(0) = \frac{\Psi^-(\log s) - b(s)}{a(s)}$ from both sides yields

$$\begin{aligned}
& \frac{\Psi^-(y + \log s) - b(s)}{a(s)} - \Psi^-(0) = \Psi^-(y) - \Psi^-(0) \\
\iff \quad & \frac{\Psi^-(y + \log s) - b(s)}{a(s)} - \frac{\Psi^-(\log s) - b(s)}{a(s)} = \Psi^-(y) - \Psi^-(0) \\
\iff \quad & \frac{\Psi^-(y + \log s) - \Psi^-(\log s)}{a(s)} = \Psi^-(y) - \Psi^-(0).
\end{aligned}$$

Now rewrite this by defining $z \doteq \log s$, $U(y) \doteq \Psi^-(y) - \Psi^-(0)$, $\tilde{a}(z) \doteq a(e^z)$, i.e.

$$\begin{aligned}
& \frac{\Psi^-(y+z) - \Psi^-(z)}{a(s)} = U(y) \\
\iff\quad & \frac{\Psi^-(y+z) - \Psi^-(0) + \Psi^-(0) - \Psi^-(z)}{a(e^{\log s})} = U(y) \\
\iff\quad & \frac{\Psi^-(y+z) - \Psi^-(0) - (\Psi^-(z) - \Psi^-(0))}{a(e^z)} = U(y) \\
\iff\quad & \frac{U(y+z) - U(z)}{\tilde{a}(z)} = U(y) \\
\iff\quad & U(y+z) - U(z) = U(y)\tilde{a}(z). \qquad (17)
\end{aligned}$$

Next, two cases are considered:

- $\boxed{\text{Case 1: } \tilde{a}(z) = 1 \text{ for all } z}$

  Then,

  $$U(y+z) = U(y) + U(z).$$

  The only possible solutions are

  $$U(y) = \rho y \text{ for some } \rho > 0$$

  which follows from Lemma 5 with a logarithmic transformation, so that

  $$\Psi^-(y) = \rho y + v, \quad v = \Psi^-(0).$$

  Since $\rho y + v$ is continuous, Lemma 1 (b) (Properties of Generalized Inverse Functions) is applicable, i.e.

  $$\begin{aligned}
  & \Psi^-(\Psi(x)) = \rho\Psi(x) + v = x \\
  \iff\quad & \Psi(x) = \frac{x - v}{\rho}.
  \end{aligned}$$

  Remembering the earlier definition of $\Psi(x)$, it follows that

  $$\begin{aligned}
  & \Psi(x) = -\log\left(-\log G(x)\right) \\
  \iff\quad & \frac{x - v}{\rho} = -\log\left(-\log G(x)\right) \\
  \iff\quad & \exp\left(-\exp\left(-\frac{x - v}{\rho}\right)\right) = G(x)
  \end{aligned}$$

  when $0 < G(x) < 1$. It was already mentioned before that G has no jump at any finite endpoint thus, $G(x) = \exp\left(-\exp\left(-\frac{x-v}{\rho}\right)\right)$ holds for all x and G coincides with the Type I Gumbel distribution.

○ $\boxed{\text{Case 2: } \tilde{a}(z) \neq 1 \text{ for some } z}$

Firstly, swap y and z in equation (17) and subtract this from the same equation, i.e.

$$\begin{aligned}
& U(y+z) - U(z) - U(y)\tilde{a}(z) - (U(z+y) - U(y) - U(z)\tilde{a}(y)) = 0 \\
\iff\ & U(y) - U(y)\tilde{a}(z) = U(z) - U(z)\tilde{a}(y) \\
\iff\ & U(y)\,(1-\tilde{a}(z)) = U(z)\,(1-\tilde{a}(y)) \\
\iff\ & U(y) = U(z)\frac{1-\tilde{a}(y)}{1-\tilde{a}(z)} \qquad (18)
\end{aligned}$$

with z such that $\tilde{a}(z) \neq 1$. Note that $\frac{U(z)}{(1-\tilde{a}(z))} \neq 0$, since $U(z) = 0$ would imply that $U(y) = 0$ for all y, and therefore $\Psi^-(y) = \Psi^-(0)$, constant. From (17) together with (18) it follows that

$$\begin{aligned}
& U(y+z) - U(z) = U(y)\tilde{a}(z) \\
\iff\ & U(z)\frac{(1-\tilde{a}(y+z))}{(1-\tilde{a}(z))} - U(z)\frac{(1-\tilde{a}(z))}{(1-\tilde{a}(z))} = U(z)\frac{(1-\tilde{a}(y))}{(1-\tilde{a}(z))}\tilde{a}(z) \\
\iff\ & \frac{U(z) - U(z)\tilde{a}(y+z) - U(z) + U(z)\tilde{a}(z)}{1-\tilde{a}(z)} = \frac{U(z)\tilde{a}(z) - U(z)\tilde{a}(y)\tilde{a}(z)}{1-\tilde{a}(z)} \\
\iff\ & \frac{-U(z)\tilde{a}(y+z) + U(z)\tilde{a}(z) - U(z)\tilde{a}(z) + U(z)\tilde{a}(y)\tilde{a}(z)}{1-\tilde{a}(z)} = 0 \\
\iff\ & \frac{U(z)}{(1-\tilde{a}(z))}\left(\tilde{a}(y)\tilde{a}(z) - \tilde{a}(y+z)\right) = 0.
\end{aligned}$$

Since $\frac{U(z)}{(1-\tilde{a}(z))} \neq 0$, this gives

$$\tilde{a}(y)\tilde{a}(z) = \tilde{a}(y+z).$$

From (18) it follows that $\tilde{a}$ is monotone and hence, the only possible solution is

$$\tilde{a}(y) = e^{\rho y}$$

for $\rho \neq 0$. Inserting this in (18) gives

$$\begin{aligned}
& U(y) = U(z)\frac{1-e^{\rho y}}{1-\tilde{a}(z)} \\
\iff\ & \Psi^-(y) - \Psi^-(0) = \frac{U(z)}{(1-\tilde{a}(z))}\left(1-e^{\rho y}\right) \\
\iff\ & \Psi^-(y) = v + c\left(1-e^{\rho y}\right)
\end{aligned}$$

where $v \doteq \Psi^-(0)$ and $c \doteq \frac{U(z)}{(1-\tilde{a}(z))}$. Since $\Psi(x) = -\log\left(-\log G(x)\right)$ is an increasing function, so is Ψ^-. Therefore, $c < 0$ if $\rho > 0$ and $c > 0$ if $\rho < 0$. Finally, it follows from Lemma 1 that

$$
\begin{aligned}
x = \Psi^{-}\left(\Psi(x)\right) =& v + c\left(1 - e^{\rho\Psi(x)}\right) \\
=& v + c\left(1 - e^{\rho(-\log(-\log G(x)))}\right) \\
=& v + c\left(1 - (-\log G(x))^{-\rho}\right).
\end{aligned}
$$

Solving this for $G(x)$, where $0 < G(x) < 1$,

$$
\begin{aligned}
& 1 - (-\log G(x))^{-\rho} = \frac{x-v}{c} \\
\iff \quad & \left(\log \frac{1}{G(x)}\right)^{-\rho} = 1 - \frac{x-v}{c} \\
\iff \quad & \frac{1}{G(x)} = \exp\left(1 - \frac{x-v}{c}\right)^{-1/\rho} \\
\iff \quad & G(x) = \exp\left(-\left(1 - \frac{x-v}{c}\right)^{-1/\rho}\right).
\end{aligned}
$$

As G has no jump at any finite (upper or lower) endpoint, G coincides with the Type II Fréchet or Type III Weibull distribution with $\alpha = +\frac{1}{\rho}$ or $\alpha = -\frac{1}{\rho}$, if $\rho > 0$ or $\rho < 0$, respectively.

Now that the first implication has been proven, it is still necessary to show that the converse is true., i.e. that every extreme value distribution is max-stable. This is quite straightforward, e.g. for the Type I Gumbel distribution

$$
\begin{aligned}
(G(\alpha_n x + \beta_n))^n &= \left(\exp\left(-\exp\left(-\left(\frac{\alpha_n x + \beta_n - b}{a}\right)\right)\right)\right)^n \\
&= \exp\left(-\exp\left(-\left(\frac{\alpha_n x + \beta_n - b}{a} - \log n\right)\right)\right) \\
&= G(x)
\end{aligned}
$$

for some constant $\alpha_n > 0$ and $\beta_n = a\log n + (1-\alpha_n)x$. Repeating this similarly for Type II and Type III distributions completes the proof. $\square$

Proof of the Fisher-Tippett Theorem Finally, to prove the Fisher-Tippet Theorem we only need to combine the work done so far:

Proof. (Fisher-Tippett Theorem [LLR12]) Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables. If there exist centering and normalizing sequences of constants $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, with $a_n > 0\ \forall n$, such that

$$
\mathbb{P}\left(\frac{M_n - b_n}{a_n} \le y\right) \xrightarrow{w} G(y) \qquad \text{as } n \to \infty
$$

then by Theorem 15 (Limit Property of Max-Stable Laws) G must be max-stable. And from Theorem 16 it follows that G is an extreme value distribution. $\square$

4.2.2 Generalized Extreme Value Distribution

Now a simplified version of the extreme value distribution is presented, combing all possible limiting distributions for normalized and centered maxima of i.i.d. random variables:

Definition 41. (the generalized extreme value distribution [CBTD01])
The combination of all three limiting distribution types for normalized and centered maxima of i.i.d. random variables is called *the generalized extreme value (GEV) distribution*:

$$G_{\xi;\mu,\sigma}(z) \doteq \begin{cases} \exp\left(-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right) & \xi \neq 0 \\ \exp\left(-e^{-\frac{z-\mu}{\sigma}}\right) & \xi = 0 \end{cases}$$

defined on the set $\left\{z : 1 + \frac{\xi(z-\mu)}{\sigma} > 0\right\}$. According to the value of ξ, the three distribution types are given:

- for $\xi > 0$, $G_{\xi;\mu,\sigma}$ is a Fréchet distribution
- for $\xi = 0$, $G_{\xi;\mu,\sigma}$ is a Gumbel distribution
- for $\xi < 0$, $G_{\xi;\mu,\sigma}$ is a Weibull distribution

As before, it is common to make use of the standardized form G_ξ and neglecting the location and scale parameter, i.e. setting μ to zero and σ to one, for the sake of simplicity. The probability density function for the GEV distribution is given by

$$g_{\xi;\mu,\sigma}(z) = \begin{cases} \frac{1}{\sigma}\exp\left(-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right)\left(1+\xi\left(\frac{z-\mu}{\sigma}\right)\right)^{-(1+1/\xi)} & \xi \neq 0 \\ \frac{1}{\sigma}\exp\left(-e^{-\frac{z-\mu}{\sigma}} - \frac{z-\mu}{\sigma}\right) & \xi = 0. \end{cases}$$

Note. The corresponding limiting distribution for normalized and centered *minima* of i.i.d. random variables is defined by

$$\tilde{G}_{\xi;\mu,\sigma}(z) = 1 - G_{\xi;\mu,\sigma}(-z)$$

and we commonly refer to the three cases for ξ as the *converse Gumbel, Fréchet and Weibull distribution* [RTR07].

4.2.3 Maximum Domains of Attraction

Even if in reality the underlying distribution of a given data set is unknown, assume for a moment that you are provided with the true distribution function F. So far it has not been answered how to classify F: Which conditions must F satisfy such that $\frac{M_n - b_n}{a_n}$ converges in distribution? How must we choose the normalizing and centering constants a_n and b_n? And in case of convergence, which GEV distribution type will be achieved? We will answer these question in this section and we will also consider examples for distribution functions which do *not* provide a limiting distribution for maxima M_n under any linear transformation. Similar to sums of i.i.d. random variables it is helpful to define the *maximum domain of attraction (MDA)* and thereafter determine the conditions which a distribution function must satisfy in order to belong to the MDA of the generalized extreme value distribution. This way the answers to the above questions will be obtained.

Definition 42. (Maximum Domain of Attraction (MDA) [EKM13])
Let F be the distribution function of the i.i.d. random variables $X_1, ...X_n$. Let G be one of the three extreme value distributions. Then F belongs to the *maximum domain of attraction* G if there exists $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G$$

holds. This is denoted as $F \in MDA(G)$.

Instead of only considering the convergence of $\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right)$ it is also of benefit to examine the convergence of $\mathbb{P}\left(M_n \leq u_n\right)$ where the sequence $(u_n)_{n\in\mathbb{N}}$ can be defined as $u_n \doteq a_n x - b_n$ but can equally be a sequence which does not depend on x or is much more complicated [LLR12]. This consideration leads to a standard tool stated below, the so called *Poisson approximation,* which is crucial for proving the conditions for a distribution function to belong to the MDA of one of the extreme values distributions.

Theorem 17. *(Poisson approximation [EKM13, LLR12])*
Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of random variables with common distribution function F and let $0 \leq \tau \leq \infty$. Suppose that $(u_n)_{n\in\mathbb{N}}$ is a sequence of real numbers such that

$$n(1 - F(u_n)) \xrightarrow{n\to\infty} \tau. \tag{19}$$

Then,

$$\mathbb{P}(M_n \leq u_n) \xrightarrow{n\to\infty} e^{-\tau}. \tag{20}$$

Conversely, assume that (20) holds for some $\tau \in [0, \infty]$, then so does (19).

Proof. [LLR12, EKM13] Let $\tau \in [0, \infty)$. If (19) holds, then

$$\begin{aligned}
\mathbb{P}\left(M_n \leq u_n\right) &= F^n(u_n) \\
&= \left(1 - \frac{1}{n} n(1 - F(u_n))\right)^n \\
&= \left(1 - \frac{\tau}{n} + o\left(\frac{1}{n}\right)\right)^n \\
&\xrightarrow{n\to\infty} e^{-\tau}
\end{aligned}$$

since $\exp x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n$, i.e. (20) follows.
Conversely, assume that (20) holds for $\tau \in [0, \infty)$. Then, $1 - F(u_n) \xrightarrow{n\to\infty} 0$ since otherwise $1 - F(u_{n_k})$ would be bounded away from zero for some subsequence (n_k). Yet, this contradicts the assumption (20) as it implies that

$$\mathbb{P}\left(M_n \le u_n\right) = (1 - (1 - F(u_n)))^n \xrightarrow{n\to\infty} 0.$$

Taking logarithms in the following way

$$\begin{aligned} & (1 - (1 - F(u_n)))^n \xrightarrow{n\to\infty} e^{-\tau} \\ \iff \quad & \log\left(1 - (1 - F(u_n))\right)^n \xrightarrow{n\to\infty} \log e^{-\tau} \\ \iff \quad & -n \log\left(1 - (1 - F(u_n))\right) \xrightarrow{n\to\infty} \tau \end{aligned}$$

and noting that

$$-\log(1 - x) \sim x \text{ for } x \to 0$$

implies

$$n(1 - F(u_n)) \xrightarrow{n\to\infty} \tau$$

as desired.
Now, consider $\tau = \infty$. Assume that (19) holds but (20) does not, i.e.

$$\mathbb{P}\left(M_n \le u_n\right) \nrightarrow 0 (= \lim_{k\to\infty} e^{-k})$$

as $n \to \infty$. Then, there must exist a subsequence (n_k) such that $\mathbb{P}\left(M_{n_k} \le u_{n_k}\right) \xrightarrow{k\to\infty} e^{-\tau'}$ for some $\tau' < \infty$. As $\tau' < \infty$, (20) implies (19) i.e. $n(1 - F(u_n)) \xrightarrow{n\to\infty} \tau' < \infty$. Yet, this contradicts the assumption that (19) holds with $\tau = \infty$. We proceed similarly to prove the converse for $\tau = \infty$. $\square$

Furthermore, it is worth mentioning the following corollary:

Corollary 1. *[LLR12] Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of random variables with common distribution function F. Let x_F be the upper endpoint of F defined as before and let $(u_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. Then,*

- $M_n \xrightarrow{a.s.} x_F$ *where* $x_F \le \infty$, *and*
- $\mathbb{P}\left(M_n \le u_n\right) \xrightarrow{n\to\infty} \rho$ *with* $\rho \in \{0, 1\}$ *provided that* $x_F < \infty$ *and* $F(x_{F-}) < 1$.

Proof. [LLR12] To prove the first claim, let $\lambda < x_F \le \infty$. Then, $1 - F(\lambda) > 0$ which implies that $n(1 - F(\lambda)) \xrightarrow{n\to\infty} \infty$. Therefore, applying Theorem 17 gives $\mathbb{P}(M_n \le \lambda) \xrightarrow{n\to\infty} 0$. It is obvious that

$$\mathbb{P}(M_n > x_F) = 0$$

for all n. Combining these considerations suggests that

$$\mathbb{P}(|M_n - x_F| \geq \epsilon) \xrightarrow{n\to\infty} 0$$

for every $\epsilon > 0$, i.e. M_n converges in probability to x_F. As (M_n) increases monotonically,

$$M_n \xrightarrow{a.s.} x_F$$

(see for instance in [Res03]).
To prove the second claim, let $x_F < \infty$ and $F(x_{F-}) < 1$. For a real-valued sequence $(u_n)_{n\in\mathbb{N}}$, we obtain $\mathbb{P}(M_n \leq u_n) \xrightarrow{n\to\infty} \rho$ with $0 \leq \rho \leq 1$. Thus, it is reasonable to write $\rho = e^{-\tau}$ with $0 \leq \tau \leq \infty$. Applying Theorem 17 implies $n(1-F(u_n)) \xrightarrow{n\to\infty} \tau$. Consider two cases:

- $u_n < x_F$ for infinitely many n, then $1 - F(u_n) \geq 1 - F(x_{F-}) > 0$ since $F(u_n) \leq F(x_{F-}) < 1$ and it follows that $\tau = \infty$
- $u_n \geq x_F$ for all sufficiently large values of n, then $n(1 - F(u_n)) = 0$ i.e. $\tau = 0$

As $\tau = \infty$ or 0, $\rho = e^{-\tau} = 0$ or 1 which proves the second part. □

Note that Corollary 1 proves that there exists no non-degenerate limiting distribution for M_n if the underlying distribution function F has a jump at its finite upper endpoint. Bearing Theorem 17 in mind, we are now ready to introduce and prove sufficient as well as necessary conditions which a distribution function has to satisfy in order to belong to the MDA of the generalized extreme value distribution.

Theorem 18. *(Characterization of the Maximum Domain of Attraction, sufficient conditions [LLR12])*
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with absolute continuous distribution function F and density f. The following are sufficient conditions for F to belong to the maximum domain of attraction of a Type I, II or III distribution.

Type I $\quad$ *$f'(x) < 0$ for all x in some interval (x_0, x_F) where $x_F \leq \infty$, $f(x) = 0$ for $x \geq x_F$, and*

$$\lim_{t\nearrow x_F} \frac{f'(t)(1-F(t))}{(f(t))^2} = -1. \tag{21}$$

Type II $\quad$ *$f(x) > 0$ for all $x \geq x_0$ finite, and*

$$\lim_{t\to\infty} \frac{tf(t)}{1-F(t)} = \alpha > 0. \tag{22}$$

Type III $\quad$ *$f(x) > 0$ for all x in some finite interval (x_0, x_F), $f(x) = 0$ for $x > x_F$, and*

$$\lim_{t\nearrow x_F} \frac{(x_F - t)f(t)}{1-F(t)} = \alpha > 0. \tag{23}$$

Remark. The proof of Theorem 18 as well as the proof of Theorem 19 will rely on the existence of some γ_n such that $1 - F(\gamma_n) \sim \frac{1}{n}$. However, it is not always possible to find such a γ_n satisfying this property for any distribution function F [LLR12]. But it is for sure that such a γ_n can be found for any distribution function belonging to the MDA of one of the extreme value distributions [LLR12]. To verify this consider

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{n\to\infty} G_\xi(x)$$

where G_ξ is the GEV distribution. As G_ξ is continuous it is possible to choose a x_* such that $G_\xi(x_*) = e^{-1}$. This choice implies that

$$\mathbb{P}\left(M_n \leq \gamma_n\right) \xrightarrow{n\to\infty} e^{-1}$$

with $\gamma_n = a_n x_* + b_n$. Therefore by Theorem 17 it follows that

$$n(1 - F(\gamma_n)) \xrightarrow{n\to\infty} 1,$$

i.e. $1 - F(\gamma_n) \sim \frac{1}{n}$. This suggests that if no γ_n with the desirable property can be found, the given distribution function does not belong to the MDA of any extreme value distribution.

The first sufficient condition is the most difficult case to prove because it allows the upper endpoint x_F to be both finite as well as equal to ∞ whereas conditions for Type II and Type III only allow $x_F = \infty$ and $x_F < \infty$, respectively.
The following Lemma is essential for the proof of the first condition. We will only state it, the proof can be found in [Res13, Lemma 1.3].

Lemma 6. *[DH76, Res13] Let $f(x) > 0$ be a differentiable function where $x \in (x_0, x_F)$ and $x_F \leq \infty$. Assume that*

$$\lim_{t \nearrow x_F} f'(t) = 0 \text{ and } \lim_{t \nearrow x_F} f(t) = 0.$$

Then

$$\lim_{t \nearrow x_F} \frac{f(t + xf(t))}{f(t)} = 1$$

uniformly in any bounded interval i.e. for every $\epsilon > 0$, there exists a $N \in \mathbb{N}$ such that for all $n \geq N$ and x in any bounded interval the inequality

$$\left|\frac{f(t + xf(t))}{f(t)} - 1\right| < \epsilon$$

holds.

Proof. (Lemma 6) Omitted. □

Proof. (sufficient condition for Type I convergence [DH76, Res13])
Assume (21) holds. Define

$$\alpha(t) \doteq \frac{1-F(t)}{f(t)}.$$

Then α satisfies the assumptions of Lemma 6 since, when applying (21) it follows that

$$\begin{aligned}\lim_{t\nearrow x_F} \alpha'(t) &= \lim_{t\nearrow x_F} \frac{-(f(t))^2 - f'(t)(1-F(t))}{(f(t))^2} \\ &= \lim_{t\nearrow x_F} -1 - \frac{f'(t)(1-F(t))}{(f(t))^2} \\ &= 0.\end{aligned}$$

As $\left|\frac{1-F(t)}{f(t)}\right| = \left|\frac{\int_t^{x_F} f(t)dt}{f(t)}\right| \leq \left|\frac{f(t)(x_F-t)}{f(t)}\right| = |x_F - t|$ and $\lim_{t\nearrow x_F}(x_F - t) = 0$ it follows by the squeeze theorem (e.g. in [Soh03, Theorem 3.3.6]) that for $x_F < \infty$, $\lim_{t\nearrow x_F} \alpha(t) = 0$. If $x_F = \infty$, $\lim_{t\nearrow\infty} \alpha(t) = 0$ follows from $\lim_{t\nearrow\infty} \alpha'(t) = 0$. Therefore,

$$\lim_{t\nearrow x_F} \frac{\alpha(t + x\alpha(t))}{\alpha(t)} = 1 \tag{24}$$

locally on $t \in \mathbb{R}$. Moreover, for any $x_*, \tilde{x}$ such that $x_0 < x_* \leq \tilde{x} < x_F$

$$\begin{aligned}\int_{x_*}^{\tilde{x}} \frac{1}{\alpha(t)} dt &= \int_{x_*}^{\tilde{x}} \frac{f(t)}{1-F(t)} dt \\ &= [-\log(1-F(t))]_{x_*}^{\tilde{x}} \\ &= -\log(1-F(\tilde{x})) + \log(1-F(x_*))\end{aligned}$$

as $F'(t) = f(t)$. It follows that

$$\begin{aligned}1 - F(\tilde{x}) &= (1-F(x_*))\frac{1-F(\tilde{x})}{1-F(x_*)} \\ &= (1-F(x_*)) \exp\left(-\left(-\log(1-F(\tilde{x})) + \log(1-F(x_*))\right)\right) \\ &= (1-F(x_*)) \exp\left(-\int_{x_*}^{\tilde{x}} \frac{1}{\alpha(t)} dt\right).\end{aligned}$$

Now, choose γ_n such that $1 - F(\gamma_n) = \frac{1}{n}$. Then, for $x \geq 0$ write $x_* = \gamma_n$ and $\tilde{x} = \gamma_n + x\alpha(\gamma_n)$, and the other way round if $x < 0$. Inserting this to the previous equation implies that

$$\begin{aligned}n(1 - F(\gamma_n + x\alpha(\gamma_n))) &= \exp\left(-\int_{\gamma_n}^{\gamma_n + x\alpha(\gamma_n)} \frac{1}{\alpha(t)} dt\right) \\ &= \exp\left(-\int_0^x \frac{\alpha(\gamma_n)}{\alpha(\gamma_n + u\alpha(\gamma_n))} du\right)\end{aligned}$$

with $u = \frac{t-\gamma_n}{\alpha(\gamma_n)}$. Since (24) holds, it follows that

$$n(1 - F(\gamma_n + x\alpha(\gamma_n))) = \exp\left(-\int_0^x \frac{\alpha(\gamma_n)}{\alpha(\gamma_n + u\alpha(\gamma_n))}du\right) \xrightarrow{u\to\infty} \exp -x.$$

Thus, by Theorem 17

$$\mathbb{P}(M_n \leq \gamma_n + x\alpha(\gamma_n)) \xrightarrow{n\to\infty} \exp(\exp -x)$$

and it concludes that a distribution function satisfying condition (21) belongs to the $MDA(\Lambda)$ with $b_n = \gamma_n$ and $a_n = \alpha(\gamma_n)$. $\square$

Proof. (sufficient condition for Type II convergence [LLR12, DH76])
Assume that the assumption (22) holds, i.e.

$$\lim_{t\to\infty} \frac{tf(t)}{1 - F(t)} = \alpha > 0$$

where $f(x) > 0$ for all $x \geq x_0$ finite. Define

$$\alpha_1(t) \doteq \frac{tf(t)}{1 - F(t)}.$$

As before, for any x_1, x_2 such that $x_2 \geq x_1 \geq x_0$,

$$\begin{aligned}\int_{x_1}^{x_2} \frac{\alpha_1(t)}{t}dt &= \int_{x_1}^{x_2} \frac{f(t)}{1 - F(t)}dt \\ &= -\log(1 - F(x_2)) + \log(1 - F(x_1)),\end{aligned}$$

and

$$1 - F(x_2) = (1 - F(x_1)) \exp\left(-\int_{x_1}^{x_2} \frac{\alpha_1(t)}{t}dt\right).$$

Find γ_n such that $1 - F(\gamma_n) = \frac{1}{n}$ which implies that $\gamma_n \xrightarrow{n\to\infty} x_F$ i.e. in this case $\gamma_n \xrightarrow{n\to\infty} \infty$. Then for $x \geq 1$ write $x_1 = \gamma_n$, $x_2 = \gamma_n x$ and the other way round for $0 < x < 1$. Inserting this in the above equation results in

$$\begin{aligned}1 - F(\gamma_n x) &= (1 - F(\gamma_n)) \exp\left(-\int_{\gamma_n}^{\gamma_n x} \frac{\alpha_1(t)}{t}dt\right) \\ \iff n(1 - F(\gamma_n x)) &= \exp\left(-\int_{\gamma_n}^{\gamma_n x} \frac{\alpha_1(t)}{t}dt\right) \\ &= \exp\left(-\int_1^x \frac{\alpha_1(\gamma_n s)}{\gamma_n s}\gamma_n ds\right) \text{ by substitution with } s = \frac{t}{\gamma_n} \\ &= \exp\left(-\int_1^x \frac{\alpha_1(\gamma_n s)}{s}ds\right).\end{aligned}$$

Since $\gamma_n \xrightarrow{n\to\infty} \infty$ and, by assumption $\lim_{t\to\infty} \alpha_1(t) = \alpha$,

$$\exp\left(-\int_1^x \frac{\alpha_1(\gamma_n s)}{s} ds\right) \xrightarrow{n\to\infty} \exp\left(-\int_1^x \frac{\alpha}{s} ds\right) = e^{-\alpha \log x} = x^{-\alpha}.$$

Then, Theorem 17 implies for $x > 0$

$$\mathbb{P}(M_n \leq \gamma_n x) \xrightarrow{n\to\infty} \exp(-x^{-\alpha}).$$

To show convergence to zero for $x \leq 0$, note that

$$\mathbb{P}(M_n \leq \gamma_n x) \leq \mathbb{P}(M_n \leq \gamma_n y) \text{ for } y > 0$$

and,

$$\mathbb{P}(M_n \leq \gamma_n y) \xrightarrow{n\to\infty} \exp(-y^{-\alpha})$$

as before. Now let $y \to 0$, then $\mathbb{P}(M_n \leq \gamma_n x) \xrightarrow{n\to\infty} 0$. Therefore, a distribution function satisfying condition (22) belongs to the $MDA(\Phi_\alpha)$ with $\alpha > 0$ and normalizing and centering constants $a_n = \gamma_n$, $b_n = 0$. □

Proof. (sufficient condition for Type III convergence [DH76])
The proof is very similar to the Type II case and is therefore omitted. □

Theorem 19. *(Characterization of the Maximum Domain of Attraction, necessary and sufficient conditions [LLR12])*
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with common distribution function F. Necessary and sufficient conditions for F to belong to the maximum domain of attraction of one of the extreme value distribution Types I, II or III are given as follows:

Type I *There exists some strictly positive function $\alpha(t)$ such that*

$$\lim_{t\nearrow x_F} \frac{1-F(t+x\alpha(t))}{1-F(t)} = e^{-x} \tag{25}$$

for all $x \in \mathbb{R}$.

Type II $x_F = \infty$ *and*

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha} \tag{26}$$

for each $x \in \mathbb{R}$ with $\alpha > 0$.

Type III $x_F < \infty$ *and*

$$\lim_{h\searrow 0} \frac{1-F(x_F - xh)}{1-F(x_F - h)} = x^{\alpha} \tag{27}$$

for each $x \in \mathbb{R}$ with $\alpha > 0$.

Remark. Since it would exceed the purpose of this work, we will only prove sufficiency of the above conditions. For the proof of necessity refer to [deH70, Section 2.3 and 2.4].

Proof. (sufficient (& necessary) condition for Type I convergence [LLR12])
Assume F satisfies the first condition. Let γ_n be such that $1 - F(\gamma_n) = \frac{1}{n}$. Then for $t = \gamma_n$, which converges to $x_F \leq \infty$ as $n \to \infty$,

$$\begin{aligned}\lim_{t \nearrow x_F} \frac{1 - F(t + x\alpha(t))}{1 - F(t)} &= \lim_{n\to\infty} \frac{1 - F(\gamma_n + x\alpha(\gamma_n))}{1 - F(\gamma_n)} \\ &= \lim_{n\to\infty} n(1 - F(\gamma_n + x\alpha(\gamma_n))) \\ &= e^{-x}\end{aligned}$$

for all x (by condition (25)). Therefore, Theorem 17 suggests that

$$\mathbb{P}(M_n \leq \gamma_n + x\alpha(\gamma_n))) \xrightarrow{n\to\infty} \exp\left(-e^{-x}\right).$$

This means that under condition (25), Type I convergence is guaranteed with $a_n = \alpha(\gamma_n)$ and $b_n = \gamma_n$. □

Proof. (sufficient (& necessary) condition for Type II convergence [LLR12])
Assume F satisfies condition (26). Let γ_n be such that $1 - F(\gamma_n) = \frac{1}{n}$. Then for $t = \gamma_n$,

$$\lim_{n\to\infty} \frac{n(1 - F(\gamma_n x))}{n(1 - F(\gamma_n))x^{-\alpha}} = 1$$

for each $x > 0$. Since $n(1 - F(\gamma_n))x^{-\alpha} \xrightarrow{n\to\infty} x^{-\alpha}$ it follows that

$$n(1 - F(\gamma_n x)) \xrightarrow{n\to\infty} x^{-\alpha}.$$

Theorem 17 implies for $x > 0$

$$\mathbb{P}(M_n \leq \gamma_n x) \xrightarrow{n\to\infty} \exp\left(-x^{-\alpha}\right).$$

Note that γ_n is positive for large enough n and $\exp\left(-x^{-\alpha}\right) \xrightarrow{x\searrow 0} 0$, therefore $\mathbb{P}(M_n \leq 0) \xrightarrow{n\to\infty} 0$ and for $x < 0$,

$$\mathbb{P}(M_n \leq \gamma_n x) \leq \mathbb{P}(M_n \leq 0) \xrightarrow{n\to\infty} 0$$

which concludes the proof that the limiting distribution of a distribution function satisfying (26) is of Type II with $a_n = \gamma_n$ and $b_n = 0$. □

Proof. (sufficient (& necessary) condition for Type III convergence [LLR12])
Proving Type III convergence under condition (27) is very similar to the two cases before. Let γ_n be such that $1 - F(\gamma_n) = \frac{1}{n}$. Define $h_n \doteq x_F - \gamma_n$ and note that $h_n \xrightarrow{n\to\infty} 0$ i.e. it is reasonable to substitute h with h_n such that for $x > 0$

$$\begin{aligned}\lim_{h\searrow 0} \frac{1 - F(x_F - xh)}{1 - F(x_F - h)} &= \lim_{n\to\infty} \frac{1 - F(x_F - x(x_F - \gamma_n))}{1 - F(x_F - x_F + \gamma_n)} \\ &= \lim_{n\to\infty} n\,(1 - F(x_F - x(x_F - \gamma_n))) \\ &= x^\alpha.\end{aligned}$$

Thus, for $x < 0$, replacing x with $-x$ implies

$$\lim_{n\to\infty} n\,(1 - F(x_F + x(x_F - \gamma_n))) = (-x)^\alpha.$$

By Theorem 17, Type III convergence follows with $a_n = x_F - \gamma_n$ and $b_n = x_F$. □

Corollary 2. *[LLR12] Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with common distribution function F. For convergence*

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n}\right) \xrightarrow{n\to\infty} G_\xi(x)$$

choose the constants a_n and b_n for each type as follows:

Type I $\quad a_n = \alpha(\gamma_n),\ b_n = \gamma_n$

Type II $\quad a_n = \gamma_n,\ b_n = 0$

Type III $\quad a_n = (x_F - \gamma_n)^{-1},\ b_n = x_F$

with $\gamma_n = \inf\left\{x : F(x) \geq 1 - \frac{1}{n}\right\}$ and $\alpha(t) = \frac{1-F(t)}{f(t)}$.

Proof. These constants already appeared in the proofs of Theorem 18 and Theorem 19. □

4.2.4 Examples

In the following subsection we will explore examples of distribution functions belonging to the MDA of each extreme value distribution type. In particular, we will show that each extreme value distribution belongs to its own MDA.

Type I - Gumbel

Example 5. (Exponential distribution [LLR12])
Let $X_1, ...X_n \sim Exp(1)$ be i.i.d. random variables, i.e. the common distribution function F is given by $F(x) = 1 - e^{-x}$ for $x > 0$. Instead of applying Theorem 18, it is expedient

to work with Theorem 17 as it directly provides appropriate constants a_n and b_n. So let $\tau > 0$ and define the sequence $(u_n)_{n\in\mathbb{N}}$ as

$$u_n \doteq -\log\frac{\tau}{n} = -\log\tau + \log n$$

such that

$$1 - F(u_n) = 1 - \left(1 - e^{-(-\log\tau + \log n)}\right) = e^{\log\tau + \log\frac{1}{n}} = \frac{\tau}{n}.$$

Then with $\tau \doteq e^{-x}$ it follows from Theorem 17

$$\begin{aligned}\mathbb{P}(M_n \le u_n) &= \mathbb{P}(M_n \le -\log\tau + \log n)\\ &= \mathbb{P}(M_n - \log n \le x)\\ &\xrightarrow{n\to\infty} \exp(-\tau) = \exp(-e^{-x})\end{aligned}$$

i.e. $a_n = 1$, $b_n = \log n$ and $G_\xi(x) = \Lambda(y) = \exp\left(-e^{-x}\right)$. Figure 6 visualizes this result. When drawing 1000 samples of the exponential distribution with $\lambda = 1$, the density of the normalized and centered maxima approximates the Gumbel distribution as the sample size n increases.

Example 6. (Type I Gumbel distribution itself [LLR12])
From Theorem 15 it is known that every extreme value distribution is max-stable, i.e. for every $n \ge 2$ there exist constants $\alpha_n > 0$, β_n such that $F^n(\alpha_n x + \beta_n) = F(x)$. As the Type I distribution function is given by $F(x) = \Lambda(x) = \exp\left(-e^{-x}\right)$ it follows that

$$\begin{aligned}(F(x))^n &= \left(\exp\left(-e^{-x}\right)\right)^n\\ &= \exp\left(-ne^{-x}\right)\\ &= \exp\left(-e^{\log n}e^{-x}\right)\\ &= \exp\left(-e^{-x+\log n}\right).\end{aligned}$$

Therefore,

$$(F(x + \log n))^n = F(x).$$

Hence, for all n

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \le x\right) = F^n(x + \log n) = F(x)$$

where $a_n = 1$ and $b_n = \log n$. Thus, the Type I Gumbel distribution belongs to its own MDA.

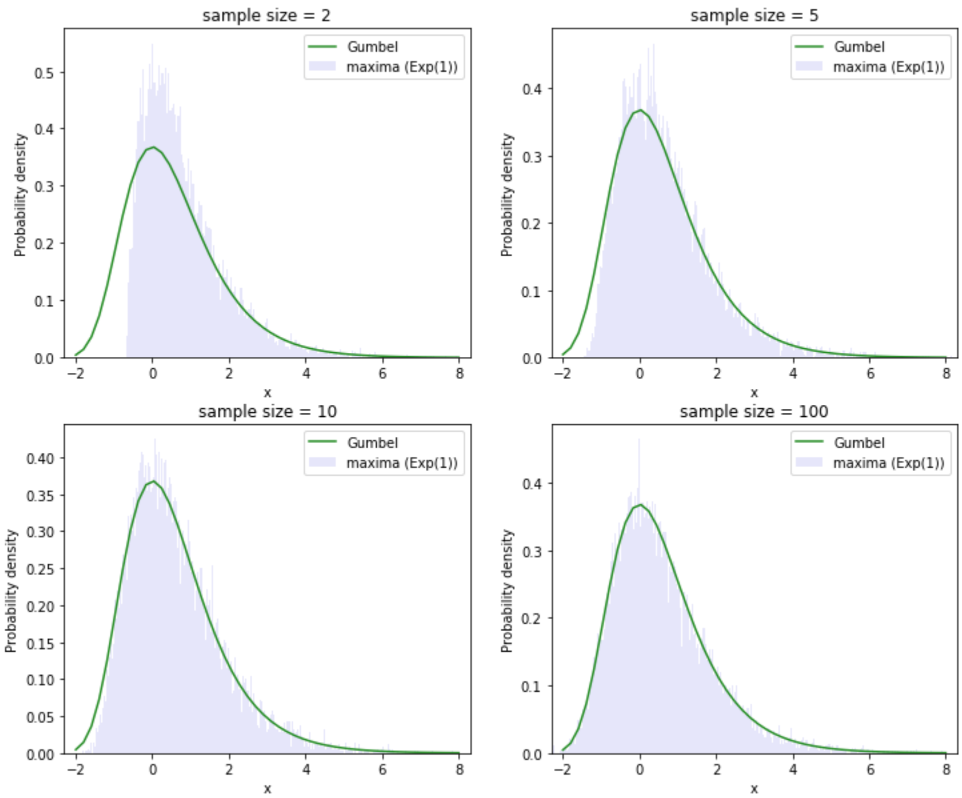

Figure 6: normalized and centered maxima of exponential distributed random variables with $\lambda = 1$ approximating Gumbel distribution as the sample size n increases

Type II - Fréchet

Example 7. (Cauchy distribution [EKM13, LLR12])
Let $X_1, ...X_n$ be i.i.d. random variables and assume them to be standard Cauchy distributed, i.e. the common distribution function F is

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} x.$$

As $\tan^{-1}$ can be rewritten as $\tan^{-1} t = \frac{\pi}{2} - \cot^{-1} t$ it follows that

$$\lim_{t \to \infty} t \left(\frac{\pi}{2} - \tan^{-1} t \right) = \lim_{t \to \infty} t \left(\frac{\pi}{2} - \frac{\pi}{2} + \cot^{-1} t \right) = \lim_{t \to \infty} t \cot^{-1} t = \lim_{\theta \to 0} \theta \cot \theta = 1$$

where we use $\theta = \cot^{-1} t$ as a substitution. When replacing t with tx the limit remains

1. Then condition 26 holds and

$$\begin{aligned}
\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} &= \lim_{t\to\infty} \frac{1-\left(\frac{1}{2}+\frac{1}{\pi}\tan^{-1} tx\right)}{1-\left(\frac{1}{2}+\frac{1}{\pi}\tan^{-1} t\right)} \\
&= \lim_{t\to\infty} \frac{\frac{1}{2}-\frac{1}{\pi}\tan^{-1} tx}{\frac{1}{2}-\frac{1}{\pi}\tan^{-1} t} \\
&= \lim_{t\to\infty} \frac{\frac{1}{2}-\frac{1}{\pi}\tan^{-1} tx}{\frac{1}{2}-\frac{1}{\pi}\tan^{-1} t} \frac{\pi tx}{\pi tx} \\
&= \lim_{t\to\infty} \frac{tx\left(\frac{\pi}{2}-\tan^{-1} tx\right)}{xt\left(\frac{\pi}{2}-\tan^{-1} t\right)} \\
&= \frac{1}{x} = x^{-1}
\end{aligned}$$

and therefore implies a Type II limiting distribution with $\alpha = 1$. It follows from Corollary 2 that $b_n = 0$ and $a_n = \gamma_n$ with γ_n such that $1 - F(\gamma_n) = \frac{1}{n}$. Solving for γ_n:

$$\frac{1}{2} - \frac{1}{\pi}\tan^{-1}\gamma_n = \frac{1}{n} \Longleftrightarrow \gamma_n = \tan\left(\frac{\pi}{2} - \frac{\pi}{n}\right) = \cot\left(\frac{\pi}{2} - \left(\frac{\pi}{2} - \frac{\pi}{n}\right)\right) = \cot\left(\frac{\pi}{n}\right).$$

Figure 7 helps to visualize the result. When drawing 1000 samples of the standard Cauchy distribution, the density of the normalized and centered maxima approximates a Fréchet distribution as the sample size n increases.

Example 8. (Type II Fréchet distribution itself [LLR12])
As before, the Type II distribution is max-stable. Remember that its standardized distribution function is

$$F(x) = \Phi_\alpha(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \exp\Big(-x^{-\alpha}\Big) & \text{if } x > 0 \end{cases} \text{ with } \alpha > 0.$$

For $x > 0$

$$(F(x))^n = \Big(\exp\Big(-x^{-\alpha}\Big)\Big)^n = \exp\Big(-nx^{-\alpha}\Big).$$

Therefore,

$$\Big(F(n^{\frac{1}{\alpha}}x)\Big)^n = F(x)$$

and for all n

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \Big(F(n^{\frac{1}{\alpha}}x)\Big)^n = \Phi_\alpha(x)$$

where $a_n = n^{\frac{1}{\alpha}}$ and $b_n = 0$ with $\alpha > 0$. This shows that the Type II Fréchet distribution itself belongs to its own MDA.

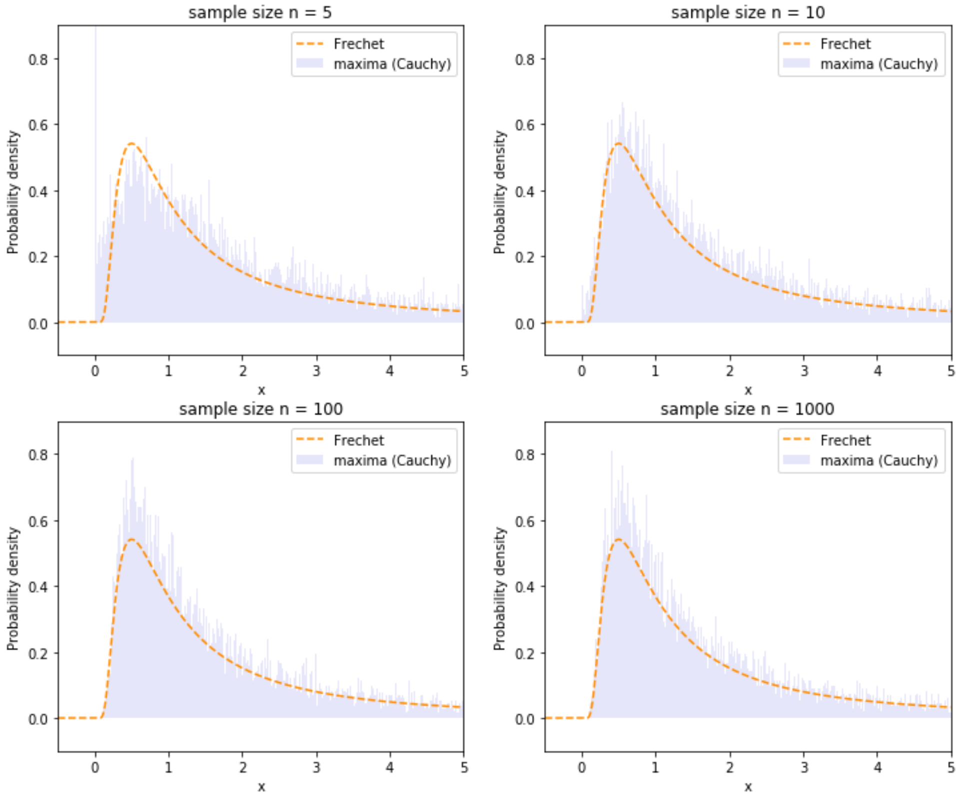

Figure 7: normalized and centered maxima of standard Cauchy distributed random variables approximating Fréchet distribution

Type III - Weibull

Example 9. (Uniform distribution [LLR12])
Let $X_1, ..., X_n$ be i.i.d. random variables and assume they are uniformly distributed on $[0, 1]$. By definition of a uniform distribution this means that the common distribution function F is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1. \end{cases}$$

Again it is helpful to use Theorem 17 like in Example 5 in order to prove that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{n\to\infty} \Psi_\alpha(x)$$

for some constants $a_n > 0$ and b_n. Let $\tau > 0$ and $u_n \doteq 1 - \frac{\tau}{n}$. Then

$$n(1 - F(u_n)) = n\left(1 - F\left(1 - \frac{\tau}{n}\right)\right) = n\left(1 - 1 + \frac{\tau}{n}\right) = \tau$$

and hence,

$$\mathbb{P}\left(M_n \leq u_n\right) = \mathbb{P}\left(M_n \leq 1 - \frac{\tau}{n}\right) \xrightarrow{n\to\infty} e^{-\tau}.$$

Then for $x < 0$ and $\tau \doteq -x$ (still satisfying the assumption $\tau > 0$) the following holds:

$$\mathbb{P}\left(M_n \leq 1 - \frac{\tau}{n}\right) = \mathbb{P}\left(M_n \leq 1 + \frac{x}{n}\right) = \mathbb{P}\left(n\left(M_n - 1\right) \leq x\right) \xrightarrow{n\to\infty} e^{x}$$

which implies a Type III limiting distribution with $\alpha = 1$ and constants $a_n = \frac{1}{n}$, $b_n = 1$. See Figure 8 for a visualization of the result.

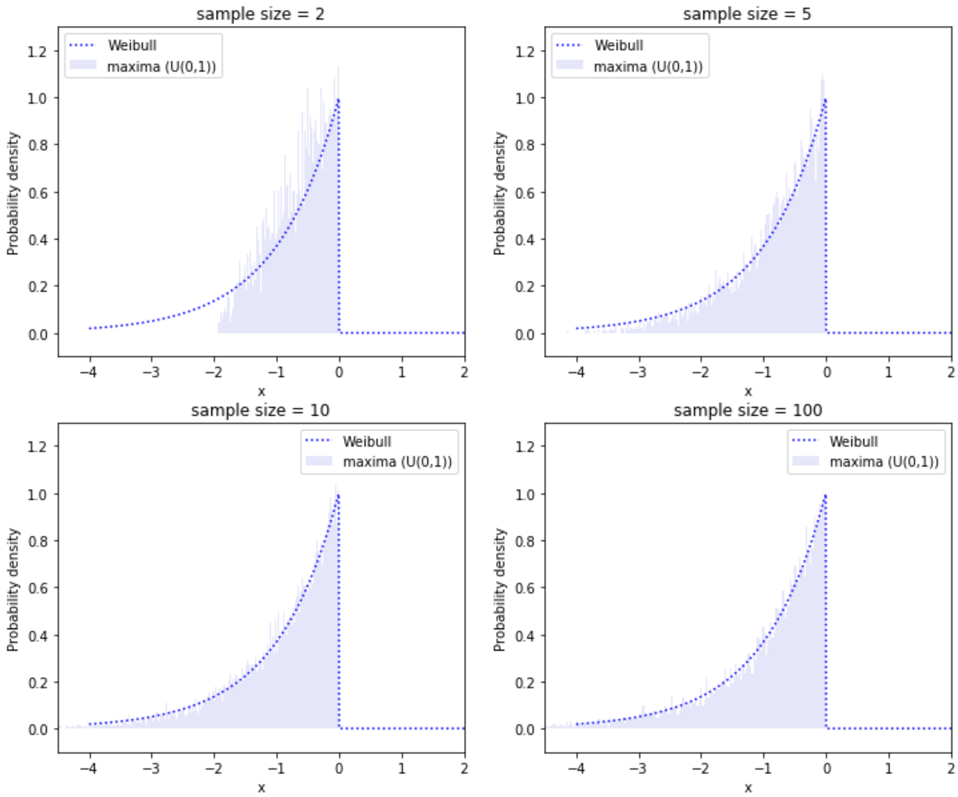

Figure 8: normalized and centered maxima of uniformly distributed random variables approximating Weibull distribution

Example 10. (Type III Weibull distribution itself [LLR12])
The standardized Weibull distribution

$$F(x) = \Psi_\alpha(x) = \begin{cases} \exp\Big(-\left(-x^{\alpha}\right)\Big) & x < 0 \\ 1 & x \geq 0 \end{cases} \text{with } \alpha > 0$$

is max-stable and since for $x < 0$

$$(F(x))^n = \Big(\exp\Big(-(-x^\alpha)\Big)\Big)^n = \exp\Big(-n(-x^\alpha)\Big)$$

it follows that

$$\Big(F(n^{-\frac{1}{\alpha}}x)\Big)^n = F(x).$$

Similar to the Type I and Type II case,

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \le x\right) = \Big(F(n^{-\frac{1}{\alpha}}x)\Big)^n = F(x)$$

for all n where $a_n = n^{-\frac{1}{\alpha}}$ and $b_n = 0$ with $\alpha > 0$, proving that the Type III Weibull distribution itself belongs to its own MDA.

No Type As promised, we will now conclude this section with examples of distribution functions which do not belong to any MDA of the three extreme value distributions. Indeed, for most discrete random variables there exists no non-degenerated limiting distribution for the maxima under any linear transformation [LLR12]. To understand why, consider the following Theorem:

Theorem 20. *[LLR12] Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with common distribution function F and let x_F be its upper endpoint. Assume that $0 < \tau < \infty$. Then, there exists a sequence $(u_n)_{n\in\mathbb{N}}$ such that (19) holds, i.e.*

$$n(1 - F(u_n)) \xrightarrow{n\to\infty} \tau$$

if and only if

$$\frac{1 - F(x)}{1 - F(x_-)} \xrightarrow{x\to x_F} 1 \tag{28}$$

or equivalently, if and only if

$$\frac{p(x)}{1 - F(x_-)} \xrightarrow{x\to x_F} 0 \tag{29}$$

with $p(x) \doteq F(x) - F(x_-)$. By Theorem 17 this is the same as saying, if $0 \le \rho \le 1$, there exists a sequence $(u_n)_{n\in\mathbb{N}}$ such that

$$\mathbb{P}(M_n \le u_n) \xrightarrow{n\to\infty} \rho$$

if and only if 28 or 29 holds. It is always possible to find such a sequence for $\rho = 0$ or 1.

Remark 1. It is crucial to note that a limit of $\mathbb{P}(M_n \le u_n)$, different to 0 or 1, exists only in the case that $n(1 - F(u_n)) \xrightarrow{n\to\infty} \tau$. This condition fails in particular for many discrete distribution functions which therefore cannot belong to the MDA of any extreme value distribution [LLR12]. For more details on how to tune the parameters of these distributions in order to achieve convergence to a non-degenerate limit distribution, refer to [NM02].

Proof. [LLR12] Assume that (19) is true for some $0 < \tau < \infty$ but (29) does not hold. Then by the negation of the (ϵ, δ)-definition of the limit there exists an $\epsilon > 0$ and a sequence $(x_n)_{n\in\mathbb{N}}$ such that $x_n \xrightarrow{n\to\infty} x_F$ and

$$p(x_n) \geq 2\epsilon(1 - F(x_{n-})). \tag{30}$$

Then choose a sequence $(n_j)_{j\in\mathbb{N}}$ of integers so that

$$1 - \frac{\tau}{n_j} \leq \frac{F(x_{j-}) + F(x_j)}{2} \leq 1 - \frac{\tau}{n_j + 1}, \tag{31}$$

i.e. such that the midpoint of the jump of F at x_j is approximately $1 - \frac{\tau}{n_j}$. Note that from this particular choice of $(n_j)_{j\in\mathbb{N}}$ it follows that $n_j \to \infty$.
It is true for every sequence but in particular for (u_n) that either $u_{n_j} < x_j$ for infinitely many values of j or $u_{n_j} \geq x_j$ for infinitely many values of j. Firstly, consider the first case i.e. $u_{n_j} < x_j$. Then clearly

$$n_j\left(1 - F(u_{n_j})\right) \geq n_j\left(1 - F(x_{j-})\right). \tag{32}$$

Expanding the latter term gives

$$\begin{aligned}
n_j\left(1 - F(x_{j-})\right) &= \tau + n_j\left(\left(1 - \frac{\tau}{n_j}\right) - \frac{F(x_j) + F(x_{j-})}{2} + \frac{F(x_j) - F(x_{j-})}{2}\right) \\
&= \tau + n_j\left(\left(1 - \frac{\tau}{n_j}\right) - \frac{F(x_j) + F(x_{j-})}{2} + \frac{p(x_j)}{2}\right) \\
&\geq \tau + n_j\left(\left(1 - \frac{\tau}{n_j}\right) - \left(1 - \frac{\tau}{n_j + 1}\right) + \frac{p(x_j)}{2}\right) && \text{applying (31)} \\
&= \tau - n_j\left(\frac{\tau}{n_j} - \frac{\tau}{n_j + 1}\right) + n_j\frac{p(x)}{2} \\
&\geq \tau - \frac{\tau}{n_j + 1} + \frac{n_j 2\epsilon(1 - F(x_{j-}))}{2} && \text{applying (30)} \\
&= \tau - \frac{\tau}{n_j + 1} + n_j\epsilon(1 - F(x_{j-}))
\end{aligned}$$

and therefore

$$(1 - \epsilon)n_j(1 - F(x_{j-})) \geq \tau - \frac{\tau}{n_j + 1}.$$

As $n_j \xrightarrow{j\to\infty} \infty$ and as $0 < \tau < \infty$ by assumption

$$\limsup_{j\to\infty} n_j(1 - F(x_{j-})) > \tau.$$

Moreover, from equation (32) it follows that

$$\limsup_{j\to\infty} n_j(1-F(u_n)) > \tau$$

yet this contradicts the assumption (19). The calculations for the second case, $u_{n_j} \geq x_j$ for infinitely many values of j, are very similar to the first case and also lead to a contradiction. Therefore, if (19) holds (29) must hold, too.
To prove the converse, assume that (28) holds. Let $(u_n)_{n\in\mathbb{N}}$ be any sequence such that

$$F(u_{n-}) \leq 1 - \frac{\tau}{n} \leq F(u_n),$$

from which follows that

$$1 - F(u_n) \leq \frac{\tau}{n} \leq 1 - F(u_{n-})$$

and by further rearrangements that

$$\frac{1-F(u_n)}{1-F(u_{n-})}\tau \leq n(1-F(u_n)) \leq \tau.$$

As $u_n \xrightarrow{n\to\infty} x_F$ the last inequality implies (19). This completes the proof. □

The following examples make use of Theorem 20. Note that both are discrete random variables and be reminded of the remark above.

Example 11. (Poisson distribution [LLR12])
Let $X_1, ..., X_n$ be i.i.d. Poisson distributed random variables, i.e. the common probability mass function is $f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $\lambda > 0$ and $x \in \mathbb{N}_0$ with $x_F = \infty$. To check if a non-degenerate limit distribution for the maxima exists, use (29):

$$\begin{aligned}
\frac{F(n)-F(n-1)}{1-F(n-1)} &= \frac{\sum_{x=0}^{n} \frac{e^{-\lambda}\lambda^x}{x!} - \sum_{x=0}^{n-1} \frac{e^{-\lambda}\lambda^x}{x!}}{1-\sum_{x=0}^{n-1} \frac{e^{-\lambda}\lambda^x}{x!}} \\
&= \frac{\frac{e^{-\lambda}\lambda^n}{n!}}{\sum_{x=\mathrm{n}}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}} \\
&= \frac{\frac{\lambda^n}{n!}}{\sum_{x=\mathrm{n}}^{\infty} \frac{\lambda^x}{x!}} \\
&= \frac{\lambda^n}{n!\left(\frac{\lambda^n}{n!} + \sum_{x=\mathrm{n}+1}^{\infty} \frac{\lambda^x}{x!}\right)} \\
&= \frac{\lambda^n}{\lambda^n\left(1 + \frac{n!}{\lambda^n} \sum_{x=\mathrm{n}+1}^{\infty} \frac{\lambda^x}{x!}\right)} \\
&= \frac{1}{1+\sum_{x=\mathrm{n}+1}^{\infty} \frac{n!}{x!}\lambda^{x-n}}
\end{aligned}$$

Rewrite the sum in denominator such that

$$\begin{aligned}
\sum_{x=\mathrm{n}+1}^{\infty} \frac{n!}{x!}\lambda^{x-n} &= \sum_{t=1}^{\infty} \frac{n!}{(n+t)!}\lambda^{t} \\
&= \sum_{t=1}^{\infty} \frac{\lambda^{t}}{(n+1)(n+2)\cdot\dots\cdot(n+t)} \\
&\leq \sum_{t=1}^{\infty} \left(\frac{\lambda}{n}\right)^{t} \\
&= \frac{\frac{\lambda}{n}}{1-\frac{\lambda}{n}} \qquad \text{geometric series from 1.}
\end{aligned}$$

Since

$$\frac{\frac{\lambda}{n}}{1-\frac{\lambda}{n}} \xrightarrow{n\to\infty} 0$$

it follows that

$$\frac{F(n)-F(n-1)}{1-F(n-1)} \geq \frac{1}{1+\frac{\frac{\lambda}{n}}{1-\frac{\lambda}{n}}} \xrightarrow{n\to\infty} 1$$

and therefore by Theorem 20 no limit (other than 0 or 1) exists for $\mathbb{P}(M_n \leq u_n)$.

Example 12. (Geometric distribution [LLR12])
Let $X_1, ..., X_n$ be i.i.d. Geometric distributed random variables, i.e. the common probability mass function is $f_X(x) = q^{x-1}(1-q)$ for $0 < q < 1$ and $x \in \mathbb{N}$ with $x_F = \infty$. Very similar to the previous example it follows that

$$\begin{aligned}
\frac{F(n)-F(n-1)}{1-F(n-1)} &= \frac{(1-p)^{n-1}p}{\sum_{x=n}^{\infty}(1-p)^{x-1}p} \\
&= \frac{(1-p)^{n-1}}{\sum_{x=n}^{\infty}(1-p)^{x-1}}.
\end{aligned}$$

Rewriting the sum in the denominator,

$$\begin{aligned}
\sum_{x=n}^{\infty}(1-p)^{x-1} &= \sum_{x=0}^{\infty}(1-p)^{x} - \sum_{x=0}^{n-2}(1-p)^{x} \\
&= \frac{1}{1-(1-p)} - \frac{1-(1-p)^{n-1}}{1-(1-p)} \\
&= \frac{(1-p)^{n-1}}{p}
\end{aligned}$$

from which concludes that

$$\frac{(1-p)^{n-1}}{\sum_{x=n}^{\infty}(1-p)^{x-1}} = \frac{(1-p)^{n-1}}{\frac{(1-p)^{n-1}}{p}} = p \neq 1.$$

and hence, no non-degenerate limit distribution exists for the maxima of Geometric distributed random variables exists.

With these two examples we conclude the section on MDA and begin with a different method to characterize extreme events.

4.3 Peaks-Over-Threshold and the Generalized Pareto Distribution

In this section we introduce the *peaks-over-threshold* approach which entails some benefits in contrast to the block maxima approach covered in section 4.2. For one thing, the block maxima approach does not necessarily take all values into account that actually correspond to an extreme event as only the maximum of a block is considered and other high values within the same block are discarded [LFdF+16]. So in any practical situation we should be able to cover every large value with a block, not only the maximum one. Yet, this makes the block maxima approach very data intensive [LFdF+16]. Therefore, it is particularly advisable to refer to the POT approach when an entire time series of observations (e.g. hourly, daily, monthly) is available as it characterizes extreme events differently and this characterization makes more efficient use of the provided data [CBTD01, LFdF+16]: in the POT approach extreme values correspond to those X_is among the i.i.d. sequence of random variables $(X_n)_{n\in\mathbb{N}}$ with common distribution function F that exceed some high threshold u. Thus, the statistical behavior of

$$\mathbb{P}(X_i - u \leq x | X_i > u), \quad x > 0 \tag{33}$$

where X_i denotes an arbitrary term in the sequence $(X_n)_{n\in\mathbb{N}}$ now becomes of interest. Similar to the beginning of section 4.2, equation (33) provides a straight forward solution to calculate the distribution of threshold exceedances. But as before, F is usually unknown and thus, the goal is to find limit distributions, which are generally applicable for high enough threshold values [CBTD01]. It will turn out that the solution to this goal is the *Pickands-Balkema-de Haan Theorem* (Theorem 21) which states that if the limiting distribution of block maxima coincides with the GEV distribution G_ξ, then the limiting distribution of observations exceeding u coincides with the so called *generalized Pareto family* [CBTD01].

4.3.1 Generalized Pareto Distribution

Definition 43. (the generalized Pareto distribution (GP distribution) [EKM13, CBTD01]) The *generalized Pareto distribution* $H_{\xi;\mu,\sigma}(x)$ is defined as

$$H_{\xi;\mu,\sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{if } \xi = 0 \end{cases}$$

for some location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\xi \in \mathbb{R}$ where

$$\begin{aligned} x \geq \mu \quad & \text{if } \xi \geq 0 \\ \mu \leq x \leq \mu - \frac{\sigma}{\xi} \quad & \text{if } \xi < 0. \end{aligned}$$

The corresponding probability density function of the GP distribution is

$$h_{\xi;\mu,\sigma}(x) = \begin{cases} \frac{1}{\sigma}\left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-(1+1/\xi)} & \text{if } \xi \neq 0 \\ \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right) & \text{if } \xi = 0 \end{cases}$$

again with the same parameters as above. See Figure 9 for examples of the probability density function and the distribution function of the GP distribution.

Remark. The GPD with μ set to zero is of particular interest. This case will be referred to as the generalized Pareto distribution $H_{\xi,\sigma} = H_{\xi;0,\sigma}$.

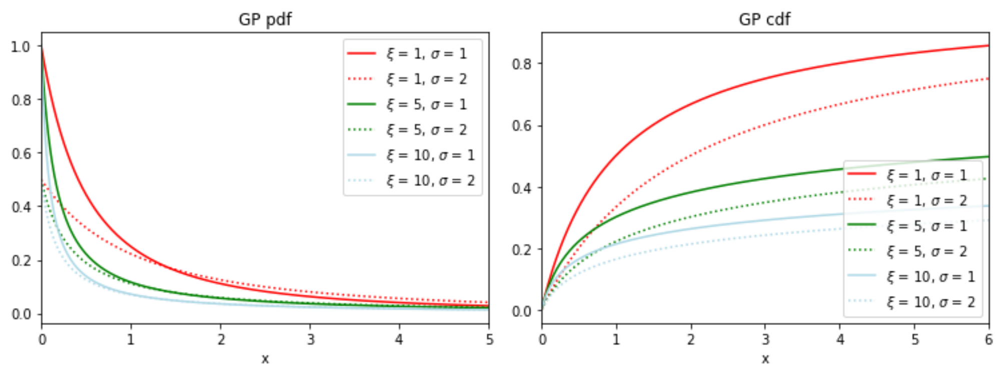

Figure 9: probability density function (left) and distribution (right) of the generalized Pareto with $\mu = 0$

Before continuing with Theorem 21 let us formally define equation (33):

Definition 44. (Excess distribution function, mean excess function, residual life time [RTR07, BDH74, EKM13])
Let X be a random variable with distribution function F_X and let x_F be the upper endpoint of X. Then the *excess distribution function* $F_u(x)$ of X *over the threshold* u is defined for a fixed $u < x_F$ as

$$\begin{aligned} F_u(x) &\doteq \mathbb{P}(X - u \leq x | X > u) \\ &= \frac{\mathbb{P}(X \leq x + u, X > u)}{\mathbb{P}(X > u)} \\ &= \frac{F_X(x + u) - F_X(u)}{1 - F_X(u)} \end{aligned}$$

for $x \geq 0$. In other words, the excess distribution function gives the probability that the exceedance of a random variable over the threshold u is less or equal than x, given that the random variable in fact exceeds the threshold u. The so called *residual life time* (r.l.t.) is defined by

$$\begin{aligned} R_u(x) &\doteq 1 - F_u(x) \\ &= \mathbb{P}(X - u > x | X > u) \\ &= \frac{1 - F_X(x + u)}{1 - F_X(u)}. \end{aligned}$$

Furthermore, the *mean excess function* $e(u)$ *of* X is defined as

$$e(u) \doteq \mathbb{E}(X - u | X > u).$$

4.3.2 Pickands-Balkema-de Haan Theorem

Now follows the previously announced *Pickands-Balkema-de Haan Theorem* which, along the Fisher-Tippet Theorem, is one of the most significant results of extreme value theory.

Theorem 21. *(Pickands-Balkema-de Haan Theorem [LFdF+16, EKM13, PI+75])*
Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of i.i.d. random variables with common continuous distribution function F, let

$$M_n = \max\{X_1, ..., X_n\}, \quad n \geq 2,$$

as before and let x_F be the upper endpoint. Then, $F \in MDA(G_\xi)$, i.e. F lies in the maximum domain of attraction of the generalized extreme value distribution, if and only if

$$\lim_{u \to x_F} \sup_{0 \leq z < x_F - u} | F_u(z) - H_{\xi, \sigma(u)}(z) | = 0$$

where, for $0 \leq z < x_F - u$, the excess distribution function $F_u(z)$ is defined as before and, for $\sigma(u) > 0$, $H_{\xi,\sigma(u)}(z) = H_\xi(z/\sigma(u))$, where $H_\xi(z)$ is the (standardized) generalized Pareto distribution.

In other words, $F \in MDA(G_\xi)$ if and only if, F_u is generalized Pareto distributed for $u \to x_F$, i.e.

$$\lim_{u \to x_F} F_u(x) = \lim_{u \to x_F} \frac{F(x+u) - F(u)}{1 - F(u)} = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma(u)}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma(u)}\right) & \text{if } \xi = 0. \end{cases}$$

Remark. The Pickands-Balkema-de Haan Theorem also suggests that there exists a close connection between the parameters of a GEV distribution and the the parameters of the corresponding GP distribution. In fact, they share the same shape parameter ξ [LFdF$^+$16]. And the other two parameters μ and σ are uniquely determined by the location and scale parameters of the GEV distribution [CBTD01]. Furthermore, note that the following relation

$$1 + \log(G_\xi(y)) = H_\xi(y)$$

holds, if $\log G_\xi(x) > -1$ [RTR07, LFdF$^+$16].

Proof. [EKM13] To prove the implication, consider the following three cases:

- Case 1: $F \in MDA(G_\xi)$ for $\xi = 0$

  Since by Theorem 19, $F \in MDA(G_0) = MDA(\Lambda)$ implies that

  $$\lim_{t \nearrow x_F} R_u(x\alpha(t)) = \lim_{t \nearrow x_F} \frac{1 - F(t + x\alpha(t))}{1 - F(t)} = e^{-x}$$

  for all $x \in \mathbb{R}$ where $\alpha(t) > 0$, it clearly follows that

  $$\lim_{u \nearrow x_F} 1 - R_u(x\alpha(u)) \lim_{u \nearrow x_F} 1 - \frac{1 - F(u + x\alpha(u))}{1 - F(u)} = \lim_{u \nearrow x_F} F_u(x\alpha(u)) = 1 - e^{-x}$$

  i.e.

  $$\lim_{u \nearrow x_F} F_u(x) = 1 - e^{-\frac{x}{\alpha(u)}}$$

  and then,

  $$\lim_{u \to x_F} \mid F_u(z) - H_{\xi,\sigma(u)}(z) \mid = 0$$

  where $\xi = 0$ and $\sigma(u) = \alpha(u)$.

- Case 2: $F \in MDA(G_\xi)$ for $\xi > 0$

  For $\xi > 0$, $MDA(G_\xi) = MDA(\Phi_\alpha)$ and therefore by Theorem 19

  $$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}.$$

  This implies that $1 - F$ is regular varying at ∞ with index $-\alpha$ and thus, the Representation Theorem for Regular Varying Functions is applicable (see for instance

[EKM13] or [BGST06]). It suggests that for some $z > 0$

$$1 - F(x) = c(x) \exp\left(- \int_z^x \frac{1}{a(t)} dt \right), \qquad z < x < \infty$$

where $c(x) \xrightarrow{x \to \infty} c \in (0, \infty)$ and $\frac{a(x)}{x} \xrightarrow{x \to \infty} \frac{1}{\alpha}$ locally uniformly. Hence,

$$\begin{aligned}
\lim_{u \to \infty} R_u(xa(u)) &= \lim_{u \to \infty} \frac{1 - F(u + xa(u))}{1 - F(u)} \\
&= \lim_{u \to \infty} \frac{c(u + xa(u)) \exp\left(- \int_z^{u+xa(u)} \frac{1}{a(t)} dt \right)}{c(x) \exp\left(- \int_z^x \frac{1}{a(t)} dt \right)} \\
&= \lim_{u \to \infty} \frac{c \exp\left(- \int_z^{u+xa(u)} \frac{1}{\frac{a(t)t}{t}} dt \right)}{c \exp\left(- \int_z^x \frac{1}{\frac{a(t)t}{t}} dt \right)} \\
&= \lim_{u \to \infty} \frac{\exp\left(- \int_z^{u+xa(u)} \frac{\alpha}{t} dt \right)}{\exp\left(- \int_z^x \frac{\alpha}{t} dt \right)} \\
&= \lim_{u \to \infty} \exp\left(-\alpha \int_u^{u+xa(u)} \frac{1}{t} dt \right) \\
&= \lim_{u \to \infty} \exp\left(\log((u + xa(u))^{-\alpha}) + \log u^{\alpha} \right) \\
&= \lim_{u \to \infty} \left(\frac{u + xa(u)}{u} \right)^{-\alpha} \\
&= \left(1 + \frac{x}{\alpha} \right)^{-\alpha}
\end{aligned}$$

i.e.

$$\lim_{u \to \infty} F_u(xa(u)) = \lim_{u \to \infty} 1 - R_u(xa(u)) = \lim_{u \to \infty} 1 - \frac{1 - F(u + xa(u))}{1 - F(u)} = 1 - \left(1 + \frac{x}{\alpha} \right)^{-\alpha}$$

and then,

$$\lim_{u \to x_F} \mid F_u(z) - H_{\xi, \sigma(u)}(z) \mid = 0$$

where $\xi = \frac{1}{\alpha}$ and $\sigma(u) = a(u)$.

- Case 3: $F \in MDA(G_\xi)$ for $\xi < 0$

  The proof for Case 3 is very similar to the proof of Case 2 and is therefore omitted.

Now we assume that F_u is generalized Pareto distributed for $u \to x_F$. Then choosing γ_n

such that $\gamma_n \doteq \inf\left\{x \in \mathbb{R} : \frac{1}{n} \geq 1 - F(x)\right\}$ implies that

$$\frac{1}{1 - F(\gamma_n)} \sim n.$$

Hence, when substituting $u = \gamma_n$ in $F_u(x)$ it follows that

$$\left(1 + \frac{x}{\alpha}\right)^{-\alpha} = \lim_{n\to\infty} \frac{1 - F(\gamma_n + xa(\gamma_n))}{1 - F(\gamma_n)} = \lim_{n\to\infty} n(1 - F(\gamma_n + xa(\gamma_n)))$$

and by Theorem 17, $F \in MDA(G_\xi)$ with $\xi = \frac{1}{\alpha}$.
So far, it has been proven that $F \in MDA(G_\xi)$ if and only if

$$\lim_{u\to x_F} \mid F_u(z) - H_{\xi,\sigma(u)}(z) \mid = 0$$

with $\sigma(u) = a(u)$. Since both the distribution function F as well as the generalized Pareto distribution G_ξ are continuous, pointwise convergence generalizes to uniform convergence and therefore,

$$F \in MDA(G_\xi)$$

if and only if

$$\lim_{u\to x_F} \sup_{0\leq z<x_F-u} \mid F_u(z) - H_{\xi,\sigma(u)}(z) \mid = 0.$$

This is due to Dini's theorem, see for instance [EKM13] or [Kön13], and completes the proof. □

4.3.3 Domain of Residual Life Time Attraction

For completeness, it is appropriate to state some results on the *domain of residual life time attraction*. It is plausible that the results on maximum domain of attraction of the GEV distribution are transferable to the GP distribution due to the Pickands-Balkema-de Haan Theorem. Only a small subsection is devoted to this matter, as most details have already been covered in 4.2.3. Detailed proofs are found in [BDH74].

Definition 45. (Domain of Residual Life Time Attraction of H_ξ [BDH74])
Let F be the distribution of i.i.d random variables $(X_n)_{n\in\mathbb{N}}$. Let $H_{\xi;\mu,\sigma}$ be the generalized Pareto distribution. Then F belongs to the *domain of residual life time attraction (DRA) of $H_{\xi;\mu,\sigma}$* if and only if there exist $a(u) > 0$ and $b(u)$ such that

$$F_u(a(u)x + b(u)) = \mathbb{P}(X_i - u \leq a(u)x + b(u)|X_i > u) \xrightarrow[u\to\infty]{d} H_{\xi;\mu,\sigma}$$

where X_i denotes an arbitrary term in the sequence $(X_n)_{n\in\mathbb{N}}$. This is denoted as $F_u \in DRA(H_{\xi;\mu,\sigma})$.

For the following theorems, the set of all distribution functions F such that $F(x) < 1$ for all x is denoted by D_0. Furthermore, consider the following definitions

$$\Gamma_\alpha(x) \doteq 1 - (1+x)^{-\alpha} \text{ for } x \geq 0$$
$$\Pi(x) \doteq 1 - \exp(-x) \text{ for } x \geq 0$$

with $\alpha > 0$. They are useful to consider the GP distribution H_ξ for $\xi \neq 0$ and $\xi = 0$ separately. Also remember that Λ and Φ_α denote the Gumbel and Fréchet distribution, respectively. Then:

Theorem 22. *[BDH74]*
$DRA(\Gamma_\alpha) = MDA(\Phi_\alpha)$ for all $\alpha > 0$.

Proof. Omitted. □

Corollary 3. *[BDH74]*
$F \in DRA(\Gamma_\alpha)$ if and only if

$$F_u(xt) \xrightarrow{u\to\infty} \Gamma_\alpha(x)$$

for all x.

Proof. Omitted. □

Theorem 23. *[BDH74]*
$DRA(\Pi) = MDA(\Lambda) \cap D_0$.

Proof. Omitted. □

Corollary 4. *[BDH74]*
$F \in DRA(\Pi)$ if and only if $F(x) < 1$ for all x and

$$\lim_{u\to\infty} \mathbb{P}\left(\frac{X-u}{a(u)} \leq x \mid X > u\right) = \Pi(x)$$

with $a(u) = \mathbb{E}[X - u \mid X > u]$.

Proof. Omitted. □

4.3.4 Examples

In this subsection we will revisit two examples of section 4.2.4 which belong to the MDA of the GEV distribution and whose excess distribution functions must therefore converge in distribution to the GP distribution.

Example 13. (exponential distribution [CBTD01])
For i.i.d. $X_1, ..., X_n \sim Exp(1)$ the common distribution function is given by $F(x) = 1 - e^{-x}$ for $x > 0$. Then,

$$F_u(x) = \frac{F(x+u) - F(u)}{1 - F(u)} = \frac{1 - e^{-(x+u)} - (1 - e^{-u})}{1 - (1 - e^{-u})} = 1 - e^{-x}$$

for all $x > 0$. Therefore, we obtain an exact result for all thresholds $u > 0$ and the limiting distribution of threshold exceedances coincides with the GP distribution with $\xi = 0$ and $\sigma = 1$ (and $\mu = 0$). This result is not surprising as from Example 3 it is known that $F \in MDA(\Lambda)$ and by Theorem 23 it follows that $F_u \in DRA(\Pi)$.

Example 14. (Uniform distribution [CBTD01])
Let $X_1, ..., X_n$ be i.i.d. uniformly distributed random variables on $[0, 1]$, i.e. the common distribution function is given by $F(x) = x$ for $0 \leq x \leq 1$. Then,

$$F_u(x) = \frac{F(x+u) - F(u)}{1 - F(u)} = \frac{x + u - u}{1 - u} = \frac{x}{1 - u}$$

for $0 \leq x \leq 1 - u$. Therefore, the limiting distribution of threshold exceedances coincides with the GP distribution with $\xi = -1$ and $\sigma = 1 - u$ (and $\mu = 0$).

Summary. Extreme events are characterized in two different ways. Either we consider extreme events as the maxima of blocks consisting of n i.i.d. random variable or as the exceedances of i.i.d. random variables above a high enough threshold u. The corresponding limiting distribution differ but due to Pickands-Balkema-De Haan Theorem there exists a close relation between them. In particular, with the knowledge about the MDA of the GEV distribution we can draw conclusions about the DRA of the GP distribution. In summary, the GEV distribution

$$G_\xi, \quad \text{with } \xi \in \mathbb{R}$$

comprises the limiting distributions for properly normalized and centered block-maxima of i.i.d. random variables whereas the GP distribution

$$H_{\xi,\mu} \text{ with } \xi \in \mathbb{R} \text{ and } \mu > 0$$

describes the limiting distributions for the scaled exceedances of i.i.d. random variables over a high enough threshold u.

5 Statistical Inference in Parametric Models and Application of Extreme Value Theory

This last chapter is devoted to the application of the theory covered so far. We will be considering daily precipitation heights measured at the Freie Universität in Berlin Dahlem between 1950 and 2020. The goal is to find an appropriate threshold and fit the exceedances to a generalized Pareto distribution. For this purpose, we will introduce *maximum likelihood estimation*, a method to estimate the unknown parameters of a distribution which can in particular be used to estimate the unknown shape and scale parameter ξ and σ of the generalized Pareto distribution. We will also learn how to check wether the distribution functions with estimated parameters provide a good fit to the data. The structures of this chapter mainly follows [CBTD01, Section 4].

5.1 Principles of Point Estimation

Assume you are given data $x_1, ..., x_n$ which consist of independent realizations of a random variable X with unknown probability density function f_X. But assume that it is in fact known to which family of probability functions, with density functions $\mathcal{F} = \{f(x, \theta : \theta \in \Theta)\}$, f_X belongs to. It is therefore of interest to estimate the point θ in the *parameter space* Θ, the so called *point estimator,* since knowing θ implies full knowledge of the population [CB02]. Note that the parameter θ can be both scalar or a vector but is assumed to be scalar in the following. We denote the true value of the parameter θ by θ_* and its estimate by $\hat{\theta}_*$. To estimate θ_* and therefore the distribution there exists both a parametric and a non-parametric approach but in this work only a parametric approach will be presented.

Definition 46. (point estimator and estimate [CB02])
A *point estimator* (or simply *estimator*) is a function $T(X_1, ..., X_n)$ of a sample that aims to estimate the true value of the parameter θ_*, whereas an *estimate* is the realization of an estimator for an actual data set.

Two common estimators are the *sample mean* and the *sample variance.* The sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

estimates the population mean $\mu = \mathbb{E}[X]$. It already appeared when considering sums of random variables in Chapter 3. The sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

is as an estimator for the population variance $\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right]$ and is not to be confused with S_n defined in Chapter 3 as the sum of random variables.

The most popular parametric approach for estimation of the true yet unknown parameter value θ_* in $\mathcal{F}$ is the *maximum likelihood* technique which develops the so called *maximum likelihood estimator* [CB02].

Definition 47. ((log-)likelihood function [CBTD01, CB02])
Assume that the observations $x_1, ..., x_n$ of a random variable X are independent as before with density function $f(x_i; \theta_*)$. Let f_X be the joint density function of the sample $X_1, ..., X_n$. Then the *likelihood function* $L(\theta)$ is given as follows:

$$L(\theta) = f_X(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

where factorization is possible because of the independence of observations. Often it is useful to take the logarithm of the likelihood function, the *log-likelihood function* $l(\theta)$ is

$$l(\theta) \doteq \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta).$$

The likelihood function states the probability of the observed data as a function of θ [CBTD01]. To make this more clear, if $L(\theta_1) = f_X(x_1, ..., x_n \mid \theta_1) > f_X(x_1, ..., x_n \mid \theta_2) = L(\theta_2)$ it suggests that the observations $x_1, ..., x_n$ are more likely to have occurred from a distribution using $\theta = \theta_1$ as its parameters than from a distribution using $\theta = \theta_2$ as its parameters.

Definition 48. (maximum likelihood estimator (MLE) [CBTD01])

The maximum likelihood estimator $\hat{\theta}_*$ of θ_* is the value θ that globally maximizes the likelihood function.

As the logarithm is a monotonically increasing function the maximum likelihood estimator $\hat{\theta}_*$ also maximizes the corresponding log-likelihood function. To find the maximum likelihood estimator it is expedient to determine the first derivative of the log-likelihood function, set it to zero and solve for the parameter θ. Unfortunately, this analytical approach is not always possible and numerical optimization algorithms have to be used instead.

Example 15. (MLE of the mean of a normal distribution with known variance [CB02])
Let $X_1, ...X_n \sim \mathcal{N}(\mu, \sigma^2)$ and assume that the variance σ^2 is known and μ is unknown. Then the likelihood function is given by

$$\begin{aligned} L(\mu) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right). \end{aligned}$$

Therefore the log-likelihood function is

$$l(\mu) = \log(L(\mu)) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

By differentiating with respect to μ it follows that

$$l'(\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu).$$

The maximum of the log likelihood function can be found by solving $l'(\mu) = 0$ which leads to $\hat{\mu} = \bar{X}_n$. It is only left to check that this is indeed a maximum. Hence, calculate the second derivative

$$l''(\mu) = -\frac{n}{\sigma^2}$$

and note that this must always be negative since $\sigma^2 > 0$.

Example 16. (MLE for GP and GEV distribution [CBTD01])
Let $X_1, ...X_n$ be GP distributed with shape parameter ξ and scale parameter $\sigma > 0$. Then for $\xi \neq 0$, the likelihood function is given by

$$L(\sigma,\xi) = \prod_{i=1}^{n}\frac{1}{\sigma}\left(1+\frac{\xi x_i}{\sigma}\right)^{-(1+1/\xi)}$$

and therefore

$$\begin{aligned}
l(\sigma,\xi) &= \log(L(\sigma,\xi)) \\
&= \sum_{i=1}^{n}\log\left(\frac{1}{\sigma}\left(1+\frac{\xi x_i}{\sigma}\right)^{-(1+1/\xi)}\right) \\
&= -n\log\sigma - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\log\left(1+\frac{\xi x_i}{\sigma}\right)
\end{aligned}$$

provided $1 + \frac{\xi x_i}{\sigma} > 0$ for all x_i. Otherwise $l(\sigma,\xi) = -\infty$.
For $\xi = 0$,

$$L(\sigma,0) = \prod_{i=1}^{n}\frac{1}{\sigma}\exp\left(-\frac{x_i}{\sigma}\right)$$

hence,

$$\begin{aligned} l(\sigma, 0) &= \log(L(\sigma, 0)) \\ &= \sum_{i=1}^{n} \log\left(\frac{1}{\sigma}\exp\left(-\frac{x_i}{\sigma}\right)\right) \\ &= -n\log\sigma - \frac{1}{\sigma}\sum_{i=1}^{n} x_i. \end{aligned}$$

Unfortunately, there exists no analytical solution for $l(\sigma, \xi) = 0$, however it is possible to find the maximum likelihood estimator $(\hat{\sigma}, \hat{\xi})$ by using numerical methods. Similarly, the log-likelihood function for the GEV distribution can be derived and for $\xi \neq 0$ is given by

$$l(\xi, \sigma, \mu) = -n\log\sigma - \left(1 + \frac{1}{\xi}\right)\sum_{i=1}^{n}\log\left(1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right) - \sum_{i=1}^{n}\left(1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}$$

provided $1 + \xi\left(\frac{x_i - \mu}{\sigma}\right) > 0$ for all x_i. Otherwise $l(\xi, \sigma, \mu) = -\infty$. And for $\xi = 0$ the log-likelihood function is

$$l(0, \sigma, \mu) = -n\log\sigma - \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^{n}\exp\left(-\left(\frac{x_i - \mu}{\sigma}\right)\right).$$

But again, no analytical solution for its maximum exists.

5.2 Model Diagnostics

Assume that we are given independent realizations $x_1, ..., x_n$ from a population with unknown distribution function F. In the previous section, we encountered ML estimation, a method to estimate the unknown distribution of the population. But how can we be sure that the estimated distribution $\tilde{F}$ is a good fit to the population from which the data was drawn? To answer this question, we commonly refer to graphical techniques which compare the estimated distribution function $\tilde{F}$ to the *empirical distribution function* $\hat{F}$.

Definition 49. (empirical distribution function [CBTD01])
Let $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ be an ordered sample of independent observations from a population with distribution function F. Then the *empirical distribution function* $\hat{F}$ is defined as

$$\hat{F}(x) = \frac{i}{n+1}, \quad \text{with } i \text{ such that } x_{(i)} \leq x < x_{(i+1)}.$$

Definition 50. (probability plot [CBTD01])
Let $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ be an ordered sample of independent observations from a population and let $\tilde{F}$ be the estimated distribution function. The corresponding *probability*

plot contains all points in

$$\left\{\left(\tilde{F}\left(x_{(i)}\right), \hat{F}\left(x_{(i)}\right)\right) : i = 1, ..., n\right\}.$$

Definition 51. (quantile plot [CBTD01])
Let $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ be an ordered sample of independent observations from a population and let $\tilde{F}$ be the estimated distribution function. Then the *quantile plot* consists of all point in

$$\left\{\left(\tilde{F}^{-1}\left(\frac{i}{n+1}\right), x_{(i)}\right) : i = 1, ..., n\right\}.$$

If the estimated distribution function $\tilde{F}$ is a reasonable estimate then both the probability plot and the quantile plot should consist of points close to the unit diagonal [CBTD01]. It is advisable to always consider both plots even tough they make use of the same information. However, the scale difference can be decgisive when making conclusions about the fit of the distribution: one plot may indicate a well fitted distribution but the other may not suggest the same [CBTD01]. It is also sensible to compare the estimated density to a histogram of the data.

5.3 Application

In this final section, we will consider a data set containing historic measurements of precipitation heights recorded between 1950 and 2020 in Berlin Dahlem. The data set comprises $n = 25583$ daily measurements and was made available by the German Meteorological Service (Deutscher Wetterdienst). Figure 10 visualizes this time series. Our objective is to fit a GEV distribution as well as a GP distribution to the data. Beginning with the block maxima approach, we first choose an appropriate block size n. A reasonable choice is to consider a whole year, i.e. $n = 365$, as it prevents irregularities in the analysis due to different seasons. To divide our data into 70 even sized blocks of length $n = 365$, we cut off the 33 most recents data points. In Figure 10, the 70 block maxima were made visible as red dots.

Next, we determine the estimated distribution function $\hat{G}$. Due to the Fisher-Tippett Theorem it will be of the following form:

$$\hat{G}_{\hat{\xi};\hat{\mu},\hat{\sigma}}(x) = \begin{cases} \exp\left(-\left[1+\hat{\xi}\left(\frac{z-\hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right) & \hat{\xi} \neq 0 \\ \exp\left(-e^{-\frac{z-\hat{\mu}}{\hat{\sigma}}}\right) & \hat{\xi} = 0 \end{cases}$$

where $\hat{\xi}$, $\hat{\mu}$ and $\hat{\sigma}$ are the estimated shape, location and scale parameters, respectively. Unfortunately, there exists no analytical way to determine them, but using the *eva* R-package, we can calculate a numerical solution for the ML estimators: $\hat{\xi} \approx 0.23$, $\hat{\mu} \approx 28.16$ and $\hat{\sigma} \approx 9.79$. To check wether the estimated distribution function provides a good fit, we plot both a probability and a quantile plot in Figure 11.

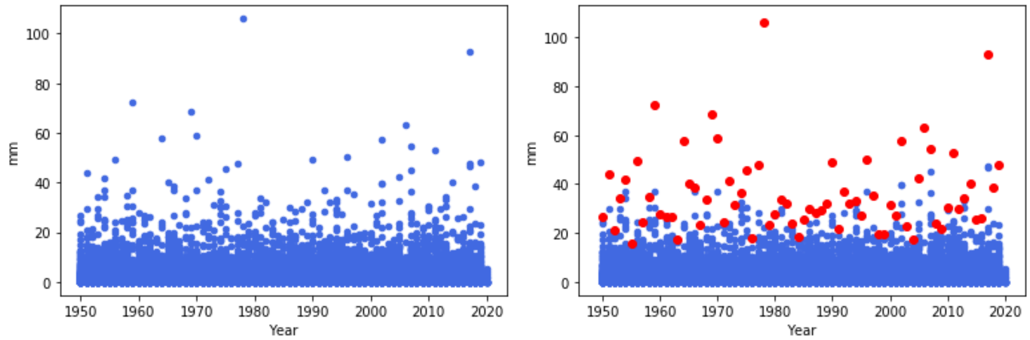

Figure 10: Daily rainfall heights (left) and daily rainfall heights with block maxima (right)

Both the probability plot and the quantile plot show deviations from the unit diago-

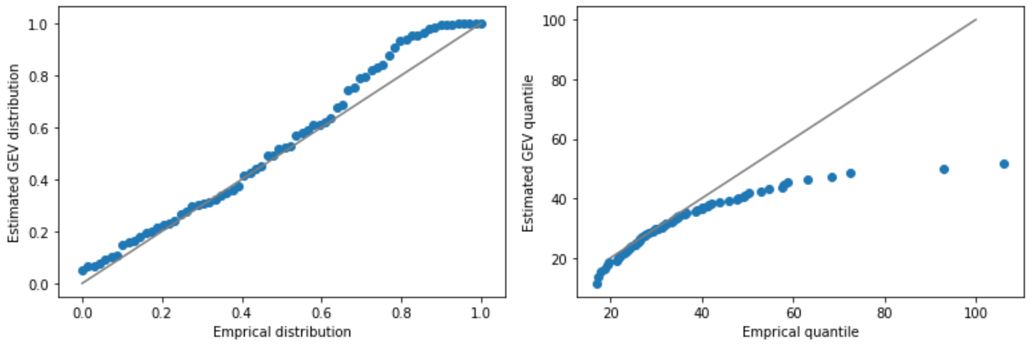

Figure 11: Probability plot (left) and quantile plot (right) for the block maxima approach

nal. Figure 12 provides a direct comparison between the empirical distribution function and the GEV distribution with estimated parameters as well as the histogram of block maxima plotted against the GEV density with estimated parameters.
In our next approach we will be using the peaks-over-threshold method and fit a GP distribution to the data set. Hence, we need to set an appropriate threshold u and determine a distribution function for the exceedances over this threshold. That is an approximation for $F_u = \mathbb{P}(X - u \leq x \mid X > u)$. The estimated distribution function $\hat{H}$ has the following form:

$$\hat{H}_{\hat{\xi},\hat{\sigma}}(x) = \begin{cases} 1 - \left(1 + \frac{\hat{\xi}x}{\hat{\sigma}}\right)^{-1/\hat{\xi}} & \text{if } \hat{\xi} \neq 0 \\ 1 - \exp\left(-\frac{x_i}{\hat{\sigma}}\right) & \text{if } \hat{\xi} = 0 \end{cases}$$

where $\hat{\xi}$ and $\hat{\sigma}$ are the estimated shape and scale parameters, respectively.
However, the choice of an appropriate threshold u for which the GP distribution is a valid approximation of F_u can be quite difficult. It is helpful to take advantage of a property

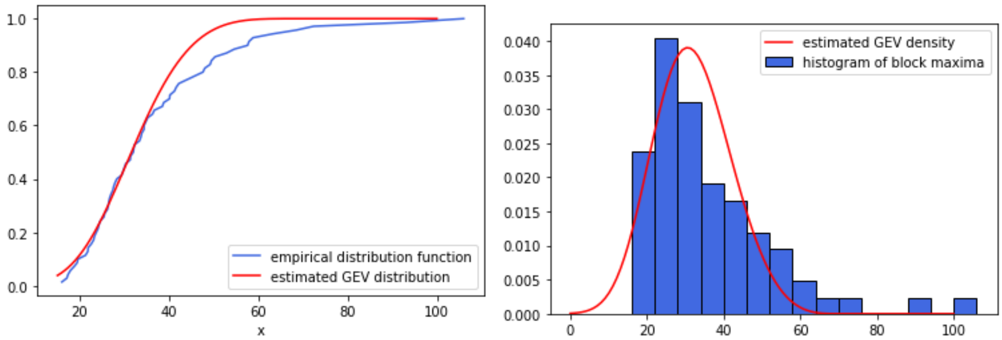

Figure 12: Direct comparison of the empirical distribution function with estimated GEV distribution (left) and comparison of histogram of data with estimated GEV density (right)

of the GP distribution:

Theorem 24. *(Property of the Generalized Pareto distribution [EKM13])*
Let the random variable X be GP distributed with shape parameter $\xi < 1$ and scale parameter $\sigma_u > 0$. Then for $u < x_F$,

$$e(u) = \mathbb{E}\left[X - u | X > u\right] = \frac{\sigma_u + \xi u}{1 - \xi}$$

where $\sigma_u - \xi u > 0$.

Proof. Omitted. □

This result implies that the mean excess function of a GP distribution is linear. Suppose that the GP distribution is a valid approximation of the exceedances of i.i.d. random variables $X_1, ..., X_n$ over a threshold u. Then by Theorem 24

$$\mathbb{E}\left[X_i - u | X_i > u\right] = \frac{\sigma_u + \xi u}{1 - \xi}$$

for any $i \in \{1, .., n\}$ with $\xi < 1$ and $\sigma_u > 0$. The GP distribution should still be a reasonable approximation of the exceedances when choosing a higher threshold $u_* > u$ and adjusting the scale parameter in accordance with u_*. Hence, for $u_* > u$

$$\mathbb{E}\left[X_i - u_* | X_i > u_*\right] \tag{34}$$

is a linear function of u. Using the sample mean as an estimator of (34), the u-region from which on the *mean residual life plot* becomes linear is a reasonable choice for the threshold u [CBTD01].

Definition 52. (mean residual life plot [CBTD01])
The *mean residual life plot* is given by the points

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : u < x_{max} \right\}$$

where $x_{(1)}, ..., x_{(n_u)}$ are the n_u observations that exceed the threshold u and x_{max} denotes the largest value among the $x_{(i)}$'s.

At first sight, the mean residual life plot for the daily rainfall data in Figure 13 looks linear between $u \approx 25$ and $u \approx 75$. From $u \approx 60$ on the mean residual life plot fluctuates. This is explainable due to fewer and fewer data points exceeding u. Hence, we should not choose a threshold above 60 as we can not draw significant conclusions of a small set of exceedances. Therefore, an appropriate threshold is any $u \geq 25$ such that enough exceedances exist.

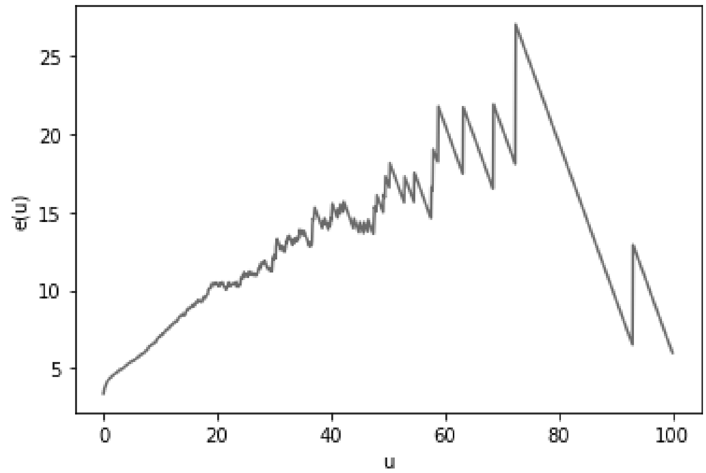

Figure 13: Mean residual life plot

As already mentioned in Example 16, there exists no analytical way to determine the ML estimators for the GP distribution. But using the *gpdFit* function in the *eva* R-package we can calculate a numerical approximation for the ML estimator. Choosing $u = 25$ as a threshold, the ML estimates are $\hat{\xi} \approx 0.18$ and $\hat{\sigma} \approx 9.05$. It is now of interest wether or not the GP distribution $\hat{H}_{\hat{\xi},\hat{\sigma}}$ provides an good fit to the precipitation data set. Both the probability plot and the quantile plot are shown in Figure 14.

It is noticeable that less deviations from the unit diagonal occur than in the block maxima approach. The quantile plot shows downwards outbreaks at the end of the line indicating a deviation from the original quantile. Figure 15 shows a direct comparison between the empirical distribution function and the GP distribution with estimated parameters as

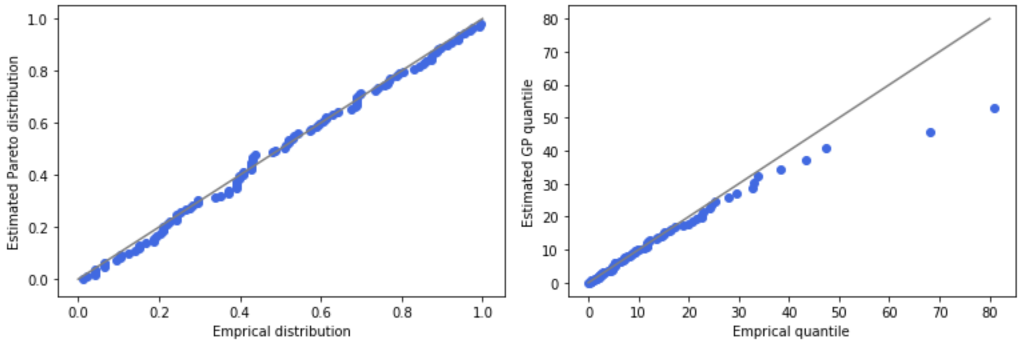

Figure 14: Probability plot (left) and quantile plot (right) for the peaks-over-threshold approach

well as a histogram of the exceedances plotted against the estimated GP density. Both

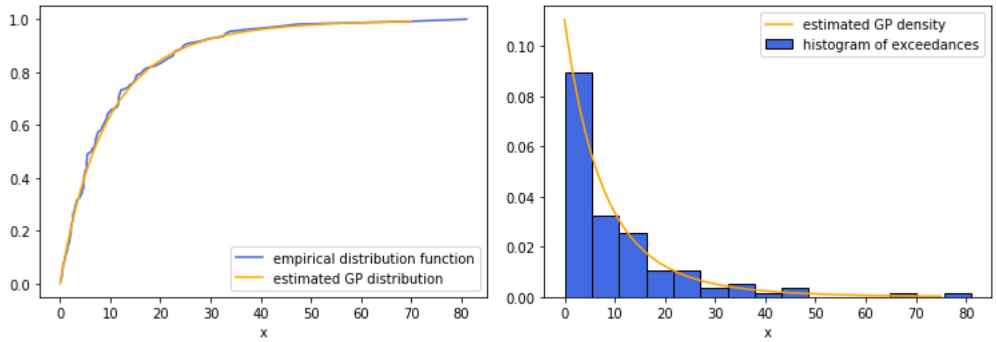

Figure 15: Direct comparison of the empirical distribution function with estimated GP distribution (left) and comparison of histogram of data with estimated GP density (right)

plots seem to indicate that the GP distribution with estimated parameters $\hat{\xi} \approx 0.18$ and $\hat{\sigma} \approx 9.05$ is a good fit to the data set with our chosen threshold $u = 25$.

Summary. In this chapter we showed how extreme value theory can be applied to a real world data set. The maximum likelihood method was used to fit the GEV and GP distribution to the data. Probability and quantile plots visualized deviations of the estimated distribution functions from the empirical distribution function.

6 Outlook

In this thesis, we have introduced the foundations of classic extreme value theory. As the main contribution we provide thorough proofs for both the Fisher-Tippett Theorem and the Pickands-Balkema-de Haan Theorem. Finally, we apply the theory to a real world data set and show how to fit the introduced GEV and GP distribution to data using maximum likelihood estimators.

Estimation could be improved by using profile likelihood estimators instead. These often provide more accurate parameter estimates. For further works on this see [CBTD01, Section 2.6.6].

Another interesting concept in extreme value theory - out of scope for this thesis - is that of return periods. These describe the levels (e.g. precipitation levels) which are expected to be exceeded every n years. Return levels find application in many real-world applications such as hydrology and finance. For more information refer to [CBTD01, Section 3.3.3 and 4.3.3] and [LFdF$^+$16, Chapter 5].

Moreover, we assumed i.i.d. data to ground our theory. In practice, as we often work with time-series data, this can be an unrealistic assumption. Consider our precipitation data set, where individual data points are not independent of their close neighbors. To account for temporal dependence, [CBTD01, Chapter 5] and [LLR12, Part II] provide an extension of extreme value theory to stationary sequences .

Also, point processes can be used to unite the different characterization of extreme event behavior and provide an elegant unification of the block maxima and peaks over threshold approach. Work on this can be found in [CBTD01, Chapter 7], [EKM13, Chapter 5] and [Res13, Chapter 3].

Finally, we only considered univariate underlying distributions in this thesis. However most of the theory can also be derived for the multivariate case. Work on multivariate extreme value theory can be found in [CBTD01, Chapter 8], [KN00, Chapter 3], [Res13, Chapter 5] and [RTR07, Part III].

References

[BDH74] A. A. Balkema and L. De Haan. Residual life time at great age. *The Annals of probability*, pages 792–804, 1974.

[Beh12] E. Behrends. *Elementare Stochastik: Ein Lernbuch-von Studierenden mitentwickelt*. Springer-Verlag, 2012.

[BGST06] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

[Bil95] P. Billingsley. *Probability and measure*. John Wiley & Sons, 1995.

[CB02] G. Casella and R. L. Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

[CBTD01] S. Coles, J. Bawa, L. Trenner, and P. Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

[deH70] L. F. M. deHaan. On regular variation and its application to the weak convergence of sample extremes. 1970.

[DH76] L. De Haan. Sample extremes: an elementary introduction. Technical report, 1976.

[EH13] P. Embrechts and M. Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432, 2013.

[EKM13] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 2013.

[Geo15] H.-O. Georgii. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Walter de Gruyter GmbH & Co KG, 2015.

[Gum12] E. J. Gumbel. *Statistics of extremes*. Courier Corporation, 2012.

[KN00] S. Kotz and S. Nadarajah. *Extreme value distributions: theory and applications*. World Scientific, 2000.

[Kön13] K. Königsberger. *Analysis 2*. Springer-Verlag, 2013.

[LFdF$^+$16] V. Lucarini, D. Faranda, J. M. M. de Freitas, M. Holland, T. Kuna, M. Nicol, M. Todd, S. Vaienti, et al. *Extremes and recurrence in dynamical systems*. John Wiley & Sons, 2016.

[LLR12] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Science & Business Media, 2012.

[NM02] S. Nadarajah and K. Mitov. Asymptotics of maxima of discrete random variables. *Extremes*, 5(3):287–294, 2002.

[PI+75] J. Pickands III et al. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1):119–131, 1975.

[Res03] S. I. Resnick. *A probability path.* Springer, 2003.

[Res13] S. I. Resnick. *Extreme values, regular variation and point processes.* Springer, 2013.

[RTR07] R.-D. Reiss, M. Thomas, and R. Reiss. *Statistical analysis of extreme values*, volume 2. Springer, 2007.

[Soh03] H. H. Sohrab. *Basic real analysis*, volume 231. Springer, 2003.