# Statistical Models for Exceedances with Applications to Finance and Environmental Statistics

Dissertation
zur Erlangung des Grades eines Doktors
der Naturwissenschaften

vorgelegt von
M.Sc. Ulf Cormann
geb. am 25.01.1982 in Menden

eingereicht beim Fachbereich Mathematik
der Universität Siegen

Siegen 2009

TAG DER MÜNDLICHEN PRÜFUNG:

28.04.2010

BETREUER UND GUTACHTER:

Prof. Dr. Reiss, Universität Siegen
Prof. Dr. Scheffler, Universität Siegen

## Abstract

This thesis concerns statistical analysis for upper tails of distribution functions. Firstly, we derive asymptotic distributions of exceedances under general monotone transformations and analyze the pertaining domains of attraction. It turns out that all possible limiting distributions satisfy a certain form of a generalized pot–stability. We give a complete characterization of all strictly increasing, continuous limiting distributions. Further, we deduce the class of all limiting distributions under power–normalization and characterize the pertaining domains of attraction. The limiting distributions are identified as generalized log–Pareto and negative generalized log–Pareto distributions as well as a certain class of discrete distributions. Moreover, we introduce and study an extended class of generalized log–Pareto distributions and provide a hybrid Maximum–Likelihood estimator. These distributions can serve as a parametric asymptotic model for super–heavy tailed distributions.

In the second part of this thesis statistical inference for the upper tail of the conditional distribution of a response variable $Y$ given a covariate $X = x$ within the framework of asymptotic distributions is considered as well. We propose to base the inference on the conditional distribution of the point process of exceedances given the point process of covariates. The results are valid within a model where the response variables are conditionally independent given the covariates.

Both parts of the thesis are linked to each other by the fact that that a Pareto modeling of the conditional distribution leads to super–heavy upper tailed unconditional distributions.

## Kurzzusammenfassung

Diese Arbeit thematisiert die statistische Analyse der oberen Flanken einer Verteilungsfunktion. Zuerst werden asymptotische Verteilungen von Exzedenten unter monotonen Normalisierungen hergeleitet sowie die zugehörigen Anziehungsbereiche analysiert. Es stellt sich heraus, dass alle stetigen Grenzverteilungen verallgemeinert pot–stabil sind. Für den Fall, dass die Normalisierung einer zusätzlichen Bedingung genügt, gilt dies für alle nicht–degenerierten Grenzverteilungen, einschließlich der diskreten. Die Familie der streng monoton steigenden, stetigen Grenzverteilungen sowie die Familie der Grenzverteilungen unter Power–Normalisierung werden vollständig charakterisiert. In letzterem Fall treten verallgemeinerte log–Pareto Verteilungen, negative verallgemeinerte log–Pareto Verteilungen sowie bestimmte p–pot stabile, diskrete Familien auf. Weiterhin wird eine erweiterte Familie von verallgemeinerten log–Pareto Verteilungen eingeführt und untersucht. Letztere können als asymptotisches Modell für Verteilungen mit super–schweren Flanken dienen.

Im zweiten Teil der Arbeit wird die obere Flanke der bedingten Verteilung einer Response–Variable $Y$ gegeben einer Kovariablen $X = x$ untersucht. Hierzu werden die asymptotischen Ergebnisse für Exzedentenverteilungen aus dem ersten Teil der Arbeit herangezogen. In der vorliegenden Arbeit wird ein bedingtes Punktprozessmodell eingeführt und für die statistische Analyse verwendet. Die zugehörigen Ergebnisse sind gültig, falls die Response-Variablen bedingt unabhängig unter den Kovariablen sind.

Eine Verbindung der beiden Teile der Arbeit wird durch die Tatsache hergestellt, dass eine Modellbildung mit Pareto–Verteilungen im bedingten Fall zu unbedingten Verteilungen mit super–schweren Flanken führt.

# Danksagung

Die vorliegende Arbeit ist während meiner Zeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Statistik, Risikoanalyse und Computing von Herrn Prof. Dr. R.–D. Reiß entstanden. Ihm gilt mein besonderer Dank für die Unterstützung und Betreuung während der Entstehung dieser Dissertation.
Weiterhin möchte ich meinen Dank dem Zweitgutachter Herrn Prof. Dr. Scheffler für vielfältige Anregungen und Kritik aussprechen, sowie Petra, Herrn Prof. Dr. Müller und Herrn Prof. Dr. Kaufmann für zahlreiche wertvolle Hinweise. Ebenfalls gilt mein Dank dem Team der Fachgruppe Statistik, Risikoanalyse und Computing, deren Diskussionsbereitschaft wesentlich zur Entstehung dieser Arbeit beigetragen haben. Insgesamt bot die offene und freundliche Atmosphäre des Fachbereichs Mathematik an der Universität Siegen ein sehr angenehmes Arbeitsumfeld, in dem fachlicher Austausch jederzeit möglich war.
Schließlich gilt mein herzlicher Dank Sabrina, meiner Familie sowie Uwe, die mich immer unterstützt haben.

# Contents

## List of Special Symbols

| | |
|---|---|
| $\alpha(F)$ | left endpoint of the support of a distribution function (df) $F$, p. 8 |
| $\varepsilon_x$ | Dirac measure at $x$ with mass 1 |
| $\Phi$, $\Phi^{-1}$, $\varphi$ | distribution–, quantile– and density function of the standard normal distribution |
| $\omega(F)$ | right endpoint of the support of a df $F$, p. 8 |
| $\mathbb{B}(\mathbb{R})$ | Borel $\sigma$-field on the real numbers |
| $C(f)$ | set of points of continuity of a function $f$ |
| $\tilde{C}(F)$ | set of points of continuity of a df $F$ with $0 < F(x) < 1$ |
| $\mathcal{D}_{T-pot}(L)$ | pot–domain of attraction of a df $L$ under a monotone transformation $T$, p. 22 |
| $\mathcal{D}_{\mathcal{T}-pot}(L)$ | pot–domain of attraction of a df $L$ under a family of monotone transformations $\mathcal{T}$, p. 22 |
| $\mathcal{D}_{p-pot}(L)$ | pot–domain of attraction of a df $L$ under power–normalization, p. 42 |
| $\mathcal{D}_{pot}(W)$ | pot–domain of attraction of a df $W$ under linear normalization, p. 42 |
| $F^{[u]}$ | exceedance df above the threshold $u$ pertaining to a df $F$, p. 8 |
| $\bar{F}$ | survival function pertaining to a df $F$ ($\bar{F} = 1 - F$) |
| $F(\cdot\|x)$ | conditional distribution function of $Y$ given $X = x$ |
| $F^{[u]}(\cdot\|x)$ | exceedance df above the threshold $u$ pertaining to $F(\cdot\|x)$, p. 69 |
| iid | independent, identically distributed (random variables) |
| $K\nu$ | measure induced by the Markov kernel $K$ and the measure $\nu$ with $K\nu(B) = \int K(B\|x)d\nu(x)$ |
| $L_{\gamma,\beta,\sigma}$ | df of the generalized log–Pareto distribution (GLPD) with shape parameters $\gamma$ and $\beta$ and scale parameter $\sigma$, p. 55 |
| $\mathcal{L}(X)$ | distribution of a random variable $X$ |
| $\mathcal{L}(N)$ | distribution of a point process $N$ |
| lim inf | limes inferior |
| lim sup | limes superior |
| $m_F$ | median excess function pertaining to a df $F$, p. 57 |
| $\mathbb{M}(S)$ | space of point measures on $S$, p. 71 |
| $\mathcal{M}(S)$ | $\sigma$–field on $\mathbb{M}(S)$, p. 71 |
| $\mathbb{N}$ | set of positive integers |

| | |
|---|---|
| innid | independent, not necessarily identically distributed (random variables), p. 90 |
| $\mathcal{P}_n$ | set of permutations of $\{1,\dots,n\}$ |
| $Q_{q,F}$ | $q$–excess function pertaining to a df $F$, p. 58 |
| $\mathbb{R}$ | set of all real numbers |
| $\tilde{S}(F)$ | $\{x : 0 < F(x) < 1\}$ for a df $F$, p. 8 |
| $W_{\gamma,\mu,\sigma}$ | df of the generalized Pareto distribution (GPD) with shape parameter $\gamma$, location parameter $\mu$ and scale parameter $\sigma$, p. 9 |
| $W$ | generalized Pareto function, p. 29 |
| $\boldsymbol{X}$ | random vector $(X_1,\dots,X_n)$ |
| $\boldsymbol{x}$ | vector $(x_1,\dots,x_n)$ |

# 1 Introduction

## 1.1 Subject and Background

Starting with articles by Fisher and Tippett [34] in 1928 and Gnedenko [36] in 1943 extreme value theory (EVT) has become an important and active area of research in the field of mathematical statistics as well as in many applications such as finance, insurance, hydrology and other fields of environmental research. In many of these applications extreme events are of particular interest, especially if one is involved with risk management.

If one is planning coast protection it is important to quantify the behavior of the most severe storm floods rather than the mean sea level. For that purpose hydrologists have introduced the $t$–year return level as a measure for the threat due to flooding in a certain area. The $t$–year return level is the flood level which is expected to be exceeded once in a $t$–year period. From a statistical point of view this is the $1 - 1/t$ quantile of the distribution of the flood level. Since $t$ is usually chosen as fifty or one hundred it is obvious that the pertaining quantile, which has to be estimated from historical flood data, is far in the upper tail and determined by the extremal behavior.

The recent financial crisis exhibited that also in the field of finance and insurance one has to pay special attention to the quantification of the extremal behavior of financial markets. In recent years the Value at Risk (VaR) has become the most important risk measure to quantify the risk of a financial product as for example a portfolio of stocks (see, e.g., [40]). Statistically speaking, the VaR is the analogue to the return level in hydrology. A (small) probability $\alpha$, usually 0.1%, 1% or 5%, is fixed and the Value at Risk at the level $\alpha$ is the loss which is only exceeded with probability $\alpha$ in a given period. A bank which invests in a speculative asset is committed to hold enough equity to cover losses from this asset up to the amount of the pertaining VaR without going bankrupt. As in the case of flood levels, the VaR is determined by the distribution of extreme rather than average losses.

Many other applications call for a special statistical theory for extremes too, e.g. insurance and in particular re–insurance or engineering.

The basic result of Fisher and Tippett [34] and Gnedenko [36] concerns linearly normalized maxima of $n$ independent and identically distributed (iid) random variables (rvs). If the pertaining distribution function (df) converges weakly to a non–degenerate distribution for $n$ tending to infinity, then this limiting distribution is an extreme value distribution (EVD). The family of EVDs is, except for location and scale shift, given by

$$G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right), \quad \gamma \neq 0, \tag{1.1}$$

and

$$G_0(x) = \exp\left(-e^{-x}\right).$$

If the maxima pertaining to a df $F$ converge to an EVD $G$ then $F$ is said to belong to the max–domain of attraction of $G$.

The normalizing constants can be included in the EVD model as location and scale parameters. Therefore, one can assume a parametric statistical model with only three parameters in case of maxima of iid random variables.

A crucial property of EVDs is the max–stability. The maximum of $n$ iid random variables, for which the common distribution is an EVD $G$ has again an EVD with identical shape parameter. Thus the maximum has again the df $G$, if appropriate linearly normalized. Max–stability is a characteristic property of EVDs, they are the only non–degenerate max–stable distributions. Particularly, this entails that there exists no non–degenerate discrete max–stable distributions.

The major shortcoming of using maxima to draw conclusions on the extremal behavior of a distribution is, that only a smaller part of the data can be utilized. A new direction of EVT research started in the 1970s with articles by Balkema and de Haan [3] and Pickands [53]. The major attention turned from maxima to exceedances over high thresholds which also describe the extremal behavior of a distribution. These authors derived in analogy to Fisher and Tippett [34] and Genedenko [36] limiting distributions of linearly normalized exceedances over thresholds tending to the right endpoint of the underlying distributions. As in the case of EVDs it turns out that the non–degenerate limiting distributions constitute a parametric model, yet also discrete limiting distributions occur. The continuous limiting distributions are closely related to EVDs and form the family of generalized Pareto distributions (GPDs), namely,

$$W_\gamma(x) = 1 + \log\left(G_\gamma(x)\right), \quad \log\left(G_\gamma(x)\right) > -1.$$

Again, the unknown normalizing constants can be included in the statistical model as location and scale parameters.

In analogy to max–domains of attraction one can define the pot–domain of attraction of a GPD $W$ by the family of distributions whose pertaining distribution of exceedances converges linearly normalized to $W$. In fact, the pot–domains of attraction coincide with the max–domains of attraction. More precisely, the pot–domain of attraction of a GPD $W$ is equal to the max–domain of attraction of the EVD $W = 1 + \log(G)$.

Therefore, for every df $F$ in the domain of attraction of an EVD one may replace the pertaining df of exceedances over a threshold $u$, $F^{[u]}$, by a GPD. This yields a parametric/nonparametric model for $F$ in the upper tail because

$$F(x) = F(u) + (1 - F(u))F^{[u]}(x), \quad x > u.$$

This is of particular importance if we focus on the extreme upper tail of a distribution, since we usually have only very few observations in this area making purely non–parametric approaches less reliable or even impossible. EVT suggests a parametric/nonparametric approach by

- deriving a parametric estimation for the model of the exceedances using a GPD modeling;
- replacing $F(u)$ by an empirical estimate.

Piecing both components together we can derive an estimate for the upper tail of $F$ and, therefore, for high quantiles which are of particular importance for many applications as already mentioned at the beginning.

A large number of articles are dealing with various aspects of extreme value theory in statistics, for example the estimation of the underlying shape parameter (see e.g. [20], [18], [39], [59]) or the choice of the threshold (see e.g. [5], [15], [16], [19], [24], [47], [56]) as well as in probability theory, for example extensions of the limiting results to the multivariate case (e.g. [27], [26], [28], [29], [31], [30], [32], [57], [64]).

The present thesis picks up two recent developments, namely,

- the characterization of limiting distributions of exceedances under non–linear transformations;
- the investigation of extremes under covariate information.

Motivation for that endeavor may be gained by the following example. Consider the conditional distribution of a random variable $Y$, subsequently addressed as response variable, given another random variable $X = x$ and assume that it is in the pot–domain of attraction of a GPD with shape parameter $1/x$. If the distribution of $X$, in the following addressed as covariate, is a gamma distribution then the unconditional distribution of $Y$ is not in the pot–domain of attraction of any GPD, see Theorem 3.8. Roughly speaking, if the conditional response variable and the covariate are heavy tailed, then the unconditional distribution of the response variable is "super–heavy" tailed. The latter term will be introduced in Chapter 4. This simple example has two consequences.

- If the covariates cannot be observed, one has to take into account distributions different from GPDs as a model for the pertaining exceedances. The use of other distributions should be justified by asymptotic considerations as well.
- If the covariates are observable one can make inference for the upper tail of the conditional distribution.

We will investigate the first problem in greater generality. We study asymptotic distributions of exceedances under general monotone transformations. We identify these limiting distributions as the class of general pot–stable distributions. This enables us to derive a representation of such distributions and, furthermore, some important properties of their domains of attraction. Yet it remains a difficult task to provide the class of limiting distributions for a given transformation. We propose a certain method using the theory of functional equations.

Apart from the traditional linear normalization we study two further classes of normalizations in detail, namely power–normalizations and exponential normalizations. The resulting limiting model is constituted by the family of generalized log–Pareto distributions (GLPDs) which was introduced in [14].

The second problem is dealt with using the concept of extremes under covariate information. A wide range of statistical methods dealing with the tail of a conditional distribution has been developed in recent years. For a comprehensive treatment we refer to [11], [56], [60] and [61]. The main tool at this approach is the concept of point processes. For more details we refer to Chapter 5. We introduce a unified point process model and identify the recent estimation procedures proposed in [61] and [25] as a MLE and, respectively, conditional MLE in our model. We demonstrate that the conditional MLE exhibits a performance superior to the unconditional one.

## 1.2 Organization

In Chapter 2 we deduce asymptotic distributions of exceedances under certain monotone transformations. We introduce the concept of general pot–domains of attraction and derive some important properties. As in the linear case it turns out that general max–domains of attraction are a subset of their peaks–over–threshold counterparts. Our main focus aims at continuous limiting distributions, yet some results are also derived in greater generality. We show how our general result can be used if a special family of transformations is considered using the example of linear normalization. Finally, we shortly indicate how these results can be extended to the multivariate case.

Chapter 3 concerns limiting distributions within the context of two special normalizations, namely, power– and exponential normalizations. We derive the class of limiting distributions of exceedances under power–normalization including the discrete ones. Hereby we make use of a certain relationship of power– and linear normalizations which was observed and used in a similar manner in the context of maxima in [10]. We also include a section showing how these results can be obtained using the theory from Chapter 2. In the case of exponential normalizations we confine ourselves to continuous limiting distributions. We also present the relation of linear and power–normalization in a more general framework and remark how one can use this relation to derive limiting distributions for a large class of monotone transformations. This leads to the notion of iterated heavy tails.

The limiting model which is related to power–normalization has certain drawbacks when exceedances are modeled. A possible improvement is achieved by the model of generalized log–Pareto distributions (GLPDs) which is introduced in Chapter 4. The GLPD model contains the GPD model and can, therefore, be regarded as an extension of the latter. For practical applications the GLPD family is a natural model for super–heavy tailed distributions which have found an increasing interest in the statistical literature in recent years, see e.g [14]. We propose a hybrid estimator within this model and use it as initial estimate for the Maximum–Likelihood (ML) estimation. As in the case of GPDs we have no closed form of the Maximum–Likelihood estimator (MLE).

The numerical problems are more difficult in the GLPD case because of an additional parameter. We also include the analysis of real data to study the relevance of the GLPD model in applications.

In Chapter 5 we turn the focus of our analysis on upper tails of distributions in more complex systems. We assume that we observe a covariate in addition to a response variable and we are primarily interested in estimation of the upper tail of the conditional distribution of the response variable given the covariate. We propose a conditional approach within a new model. We do not require independent and identically distributed observations to apply our estimation procedure. A considerably weaker condition is formulated which is basically an extended form of conditional independence of the response variables given the covariates. We also include some remarks concerning the applicability of our condition and provide some real data examples. Technical results concerning conditional distributions which are extensively used throughout this chapter are postponed until the appendix.

The thesis is concluded by some remarks about open questions and future research work in this area.

# 2 Limiting Distributions of Exceedances Under Monotone Transformations

In the stochastic theory as well as in many applied fields, limit theorems play a crucial role. The general setting of this limit theorems is usually the following. One considers a sequence of dfs $F_n$ which satisfy a certain condition. This condition might be, for example, that $F_n$ is a linearly normalized df of the maximum or the sum of $n$ independent and identically distributed (iid) rvs. Usually these conditions are quite weak and satisfied for a large, non–parametric class of distributions. In a next step it is assumed that the sequence $F_n$ converges weakly to a non–degenerate df $L$. Hereby, a df is called degenerate if the pertaining probability measure is a Dirac measure. It is often possible to prove that $L$ has to satisfy a certain condition, such as a stability property.

Two special cases are extensively studied in the statistical literature.

- linearly normalized maxima: we have $F_n(x) = F^n(a_n + b_n x)$ for a df $F$ and sequences $(a_n)$ and $(b_n) > 0$. The class of limiting dfs $L$ turns out to be max–stable, there exists sequences $(c_n)$ and $(d_n) > 0$ such that $L^n(c_n + d_n x) = L(x)$. As already indicated above, the class of max–stable dfs constitutes the parametric family of EVDs.
- linearly normalized sums: $F_n(x) = F^{*n}(a_n + b_n x)$ for a df $F$ and sequences $(a_n)$ and $(b_n) > 0$. Here, $F^{*n}$ denotes the $n$–fold convolution of the df $F$. Again, a limiting df $L$ satisfies the pertaining sum–stability property $L^{*n}(c_n + d_n x) = L(x)$. We do not have a closed parametric form of the family of sum–stable dfs, but one can derive a parametric representation of the pertaining Fourier transformations.

Such results have very important implications. Assume that the random experiment or the pertaining df, which is under investigation, can be interpreted as part of a sequence $(F_n)$ converging weakly to a df $L$. Within certain error bounds one may assume that under weak assumptions on $(F_n)$, for large enough $n$, $F_n$ is close to $L$ of which a parametric form is known. Thus, a non–parametric statistical problem can be approximated by a parametric one. The most prominent case of an approximate distribution is the normal distribution. The error of a measurement may be assumed to be normal as the sum of several errors.

In this thesis we will consider a related problem, namely limiting distributions of exceedances over thresholds $u$ tending to the right endpoint of the pertaining distribution under general monotone transformations. Our study will also work in two steps. First we will derive a certain stability property which is then used to derive the form of the limiting distribution.

We denote for a rv $X$ with df $F$ by

$$F^{[u]}(x) = P(X \le x | X > u) \tag{2.1}$$

the common df of the exceedances pertaining to $F$. We have

$$F^{[u]}(x) = \begin{cases} \frac{F(x)-F(u)}{1-F(u)}, & x \ge u, F(u) < 1; \\ & \text{if} \\ 0, & x < u \text{ or } F(u) = 1. \end{cases}$$

Throughout this thesis we use the notation

$$\omega(F) := \sup\{x : F(x) < 1\} \quad \text{and} \quad \alpha(F) := \inf\{x : F(x) > 0\} \tag{2.2}$$

for the left and right endpoint of the support of a df $F$ and

$$\tilde{S}(F) := (\alpha(F), \omega(F)) = \{x : 0 < F(x) < 1\}.$$

The aim of this chapter is to deduce limiting distributions of exceedances over high thresholds under general monotone transformations. Subsequently, a function $T : \mathbb{R}^2 \to \mathbb{R}$ will be addressed as monotone transformation if it is strictly increasing in the second argument. We will also use the term normalization as a synonym for transformation.

We consider non–degenerate dfs $F$ and $L$ and a monotone transformation $T$ such that

$$\lim_{u \to \omega(F)} F^{[u]}(T(u, x)) = L(x). \tag{2.3}$$

for all points of continuity of $L$. Well known are linear transformations $T(u, x) = a(u) + b(u)x$, where $a$ and $b$ are real valued functions with $b(u) > 0$ for all $u \in \mathbb{R}$. We will also consider power transformations $T(u, x) = \operatorname{sign}(x)a(u) + |x|^{b(u)}$ and exponential transformations $T(u, x) = a(u) \exp(b(u)x)$.

Let us assume that, for a fixed family of transformations $T$, the limiting df $L$ belongs to a parametric family. In that case the meaning of a limit relation as given in (2.3) is, that statistical inference for exceedances from a possibly complicated and only partly known underlying model $F$ can be based on a sufficiently simple and parametric asymptotic model $L$.

The main results stated in this chapter are

- all continuous limiting dfs in (2.3) satisfy a certain form of pot–stability, subsequently labeled $g$–pot stability (cf. Theorem 2.3);
- we deduce an explicit representation of all strictly increasing, continuous $g$–pot–stable dfs (cf. Theorem 2.7);
- if the transformation $T$ satisfies an additional Condition (Condition 2.8) all limiting dfs in (2.3), including the discrete ones, are $g$–pot–stbale (cf. Lemma 2.9).

The remainder of this chapter can be outlined as follows. In Section 2.1 we deal with the well known case that $T$ belongs to the family of linear transformations. Section 2.2 contains a basic result concerning continuous limiting dfs in (2.3). It is proven that these dfs satisfy a certain pot–stability property with respect to a certain function $g$ which is related to $T$. This property is subsequently labeled $g$–pot–stability. More details about continuous $g$–pot stable dfs are given in Section 2.3. We include some remarks concerning discrete $g$–pot stable dfs in Section 2.4. Relations to the asymptotic theory of maxima under general monotone transformations are addressed in Section 2.5. We conclude this chapter by pointing out consequences of the general theory for the case of linear transformations, see Section 2.6, and extensions to the multivariate case, see Section 2.7.

## 2.1 The Traditional POT–Approach

The traditional peaks–over–threshold (pot) approach within extreme value theory (EVT) started with the meanwhile classical results in Balkema and de Haan [3] and Pickands [53]. These authors proved that, if for functions $a(\cdot)$ and $b(\cdot) > 0$ and dfs $F$ and $W$

$$F^{[u]}(a(u) + b(u)x) \xrightarrow[u \to \omega(F)]{} W(x),$$

holds, then $W$ is pot–stable.

We recall some basic facts about pot–stability. A df $F$ is pot–stable if for each $y \in \tilde{C}(F)$ there are constants $a(y) \in \mathbb{R}$ and $b(y) > 0$ such that

$$F^{[y]}(a(y) + b(y)x) = F(x) \tag{2.4}$$

for all $x \in \tilde{C}(F)$. A df $F_1$ is said to be of the same linear type as the df $F_2$ if there are $a \in \mathbb{R}$ and $b > 0$ such that $F_1(x) = F_2(a + bx)$. Within the class of dfs of the same type one may select a standard version.

Due to a classical result in [3] one knows that a df $W$ is pot–stable if, and only if, it is of the linear type of one of the following dfs (with $[x]$ denoting the integer part of the real number $x$), namely, a continuous df

$$W_\gamma(x) \quad = \quad 1 - (1 + \gamma x)^{-1/\gamma}, \quad \gamma \in \mathbb{R}, \tag{2.5}$$

and dicrete dfs

$$\Pi_{\gamma,\alpha}(x) \quad = \quad 1 - \exp\left(-\alpha\left[\gamma^{-1}\log(1 + \gamma x)\right]\right), \quad \gamma \in \mathbb{R},\, \alpha > 0, \tag{2.6}$$

for $0 < x$ if $\gamma \geq 0$ and $0 < x < 1/|\gamma|$ if $\gamma > 0$. For $\gamma = 0$ choose the limits for $\gamma \to 0$, getting

$$W_0(x) = 1 - e^{-x}, \tag{2.7}$$

and

$$\Pi_{0,\alpha}(x) = 1 - \exp\left(-\alpha\,[x]\right). \tag{2.8}$$

Thus, the family of standard pot–stable dfs consists of the generalized Pareto dfs (GPD) $W_\gamma$ and of certain discrete dfs $\Pi_{\gamma,\alpha}$.

In the linear setup, limiting distributions of exceedances are closely related to limiting distributions of maxima. As already mentioned in the introduction, statistical extreme value theory started with articles by Fisher and Tippett [34] and Gnedenko [36]. These authors proved that if for a df $F$ and a non–degenerate df $G$ and suitably chosen norming sequences $a_n$ and $b_n > 0$

$$F^n(a_n + b_n x) \xrightarrow[n\to\infty]{} G(x) \tag{2.9}$$

holds, then $G$ is necessarily an extreme value distribution (EVD), cf. (1.1).

Recall that

$$W_\gamma(x) = 1 - \log\left(G_\gamma(x)\right), \quad \log(G_\gamma(x)) > -1. \tag{2.10}$$

It turns out that (2.10) is a kind of universal relation of limiting distribution of maxima and exceedances, also if non–linear transformations are considered. Notice that in contrast to the existence of discrete limiting distributions of exceedances only continuous distributions can appear in (2.9). This result carries over to more general transformations, too.

## 2.2 A Basic Result Concerning Limiting Distributions

In this section we extend the results concerning the linear pot–approach in the pot method to a more general framework. Let $g$ be a monotone transformation. A df $F$ is called pot–stable with respect to $g$ if the pertaining exceedances df $F^{[y]}$ satisfies

$$F^{[y]}(g(y,x)) = F(x).$$

In terms of survivor functions $\bar{F} = 1 - F$ this is

$$\frac{\bar{F}(g(y,x))}{\bar{F}(y)} = \bar{F}(x).$$

It will be shown that every continuous limiting df $L$ in (2.3) is pot–stable with respect to a transformation $g$ according to the following definition.

**Definition 2.1** *Let $F$ be a df and $g : \mathbb{R}^2 \to \mathbb{R}$. Then, $F$ is said to be $g$–pot–stable if*

$$\bar{F}(g(y,x)) = \bar{F}(x)\bar{F}(y)$$

*for all* $y, x \in \tilde{S}(F)$, *whereby*

$$\tilde{S}(F) := \{x \in \mathbb{R} : 0 < F(x) < 1\}.$$

For that purpose we introduce some basic notation. The set of points of continuity of a function $f$ is denoted by $C(f)$. For a monotone transformation $T$ put $T_u(x) := T(u,x)$ and $T^{(-1)}(u,x) := T_u^{-1}(x)$. Moreover, let

$$S_{u,y}(x) := T^{(-1)}\left(u, T\left(T(u,y), x\right)\right). \tag{2.11}$$

Notice that $S_{u,y}$ is also strictly increasing. It will turn out that $S_{u,y}$ is crucial for the construction of the function $g$ in Definition 2.1.

We start with an auxiliary result for the transformations $T$. It follows from (2.3) that $T(u,x) > u$ if $x \in \tilde{S}(L)$ for $u$ sufficiently close to $\omega(F)$. Otherwise, we have

$$\lim_{u \to \omega(F)} F^{[u]}(T(u,x)) = 0 < L(x).$$

If $T(u,x) > u$, equation (2.3) is equivalent to

$$\lim_{u \to \omega(F)} \frac{\bar{F}(T(u,x))}{\bar{F}(u)} = \bar{L}(x).$$

Since $T(u,x) > u$ for $u$ sufficiently close to $\omega(F)$ we necessarily have

$$T(u,x) \xrightarrow[u \to \omega(F)]{} \omega(F),$$

and, therefore, $T(u,x) > x$ for large $u$.

We require the following technical result concerning weak convergence.

**Lemma 2.2** *Let* $H_n$, $n \in \mathbb{N}$ *and* $H$ *be non–increasing (non–decreasing) functions such that*

$$H_n(x) \xrightarrow[n \to \infty]{} H(x), \quad \textit{for all } x \in C(H)$$

*and let* $(x_n)$ *a sequence converging to some* $x \in \mathbb{R}$.

*Then,*

$$H_n(x_n) \xrightarrow[n \to \infty]{} H(x)$$

*if* $x \in C(H)$.

Proof. The proof is merely carried out for non–increasing functions. The proof for non–decreasing functions runs in a similar manner. First notice that $C(H)$ is dense in $\mathbb{R}$ since $H$ is a monotone function. Therefore, we find for each $\varepsilon > 0$ some $0 < \varepsilon' < \varepsilon$ and $0 < \varepsilon'' < \varepsilon$ such that $x + \varepsilon'$ and $x - \varepsilon''$ both belong to $C(H)$. Since $x_n \to x$ we find some $n_0 \in \mathbb{N}$ such that for all $n > n_0$

$$x + \varepsilon' > x_n > x - \varepsilon''.$$

Making use of the monotonicity of $H_n$ one gets

$$H_n(x - \varepsilon'') \geq H_n(x_n) \geq H_n(x + \varepsilon').$$

Letting $n \to \infty$ yields

$$H(x - \varepsilon'') \geq \limsup_{n\to\infty} H_n(x_n) \geq \liminf_{n\to\infty} H_n(x_n) \geq H(x + \varepsilon').$$

Now with $\varepsilon \to 0$ (and, therefore, $\varepsilon' \to 0$ and $\varepsilon'' \to 0$) one gets the assertion since $x \in C(H)$. □

The following theorem essentially states that every continuous limiting distribution of exceedances under a monotone transformation $T$ is pot–stable with respect to a transformation which can be deduced from $T$. To prove this theorem we need an auxiliary result stated in Lemma 2.4 below.

**Theorem 2.3** *Consider a monotone transformation $T : I \times \mathbb{R} \to \mathbb{R}$, where $I \subset \mathbb{R}$ is an interval. Furthermore, let $F$ be a df and $L$ a continuous df such that*

$$\frac{\bar{F}(T(u,x))}{\bar{F}(u)} \overset{u\to\omega(F)}{\longrightarrow} \bar{L}(x) \tag{2.12}$$

*for all $x \in \mathbb{R}$. Then $L$ satisfies*

$$\bar{L}(g(y,x)) = \bar{L}(x)\bar{L}(x), \quad x, y \in \tilde{S}(L)$$

*where for each $y \in \tilde{S}(L)$ there exists a sequence $(u_n)$ such that*

$$g(y,x) = \lim_{u_n\to\omega(F)} S_{u_n,y}(x).$$

Proof. Put $g(y,x) := g_y(x)$ where $g_y$ is given in Lemma 2.4 below. Then the asserted $g$–pot–stability of $L$ follows immediately from Lemma 2.4. □

The interval $I$ in the preceding theorem may be chosen as $I = \mathbb{R}$, yet is is not necessary that $T$ is defined for all $u \in \mathbb{R}$ if $\omega(F)$ is finite.

**Lemma 2.4** *Suppose the assumptions in Theorem 2.3 hold. Then, for each $y \in \tilde{S}(F)$ and each sequence $(u_n)$ with $u_n \xrightarrow[n\to\infty]{} \omega(F)$ there exists a subsequence $(u_n^*)$ such that*

$$g_y(x) = \lim_{n\to\infty} S_{u_n^*,y}(x)$$

*exists for all* $x \in \tilde{S}(L)$. *Moreover,* $\bar{L} = 1 - L$ *satisfies*

$$\bar{L}\left(g_y(x)\right) = \bar{L}(x)\bar{L}(y), \quad x \in \tilde{S}(L). \tag{2.13}$$

Proof. First observe that (2.12) yields

$$\begin{aligned}\lim_{u \to \omega(F)} \frac{\bar{F}\left(T(u,x)\right)}{\bar{F}\left(T(u,y)\right)} &= \lim_{u \to \omega(F)} \frac{\bar{F}\left(T(u,x)\right)}{\bar{F}(u)} \frac{\bar{F}(u)}{\bar{F}\left(T(u,y)\right)} \\ &= \frac{\bar{L}(x)}{\bar{L}(y)}. \end{aligned} \tag{2.14}$$

Let

$$S_{u,y}(x) := T^{(-1)}\left(u, T\left(T(u,y), x\right)\right)$$

as in (2.11). Moreover, put

$$\tilde{F}_y^{(u)}(x) := \frac{\bar{F}\left(T(u,x)\right)}{\bar{F}\left(T(u,y)\right)}.$$

We get from (2.14)

$$\tilde{F}_y^{(u)}(x) \xrightarrow[u \to \omega(F)]{} \frac{\bar{L}(x)}{\bar{L}(y)}.$$

We also have

$$\begin{aligned}\tilde{F}_y^{(u)}\left(S_{u,y}(x)\right) &= \frac{\bar{F}\left(T\left(T(u,y),x\right)\right)}{\bar{F}\left(T(u,y)\right)} \\ &\xrightarrow{u \to \omega(F)} \bar{L}(x)\end{aligned}$$

since $T(u,y) \to \omega(F)$ if $u \to \omega(F)$. Thus, (2.13) can be considered a convergence to type theorem.

Let $x, y \in \tilde{S}(L)$. Because $\bar{L}(y) < 1$ and $\bar{L}(x) > 0$ we have

$$\lim_{u \to \omega(F)} \tilde{F}_y^{(u)}(S_{u,y}(x)) = \bar{L}(x) < \frac{\bar{L}(x)}{\bar{L}(y)} = \lim_{u \to \omega(F)} \tilde{F}_y^{(u)}(x).$$

Because $\tilde{F}_y^{(u)}$ is non–increasing for all $u$ we get $S_{u,y}(x) > x$ for $u$ sufficiently close to $\omega(F)$. Thus $S_{u,y}(x)$ is bounded from below by $x$.

Now choose an arbitrary sequence $(u_n)$ with $u_n \to \omega(F)$ as $n \to \infty$. Then, $S_{u_n,y}(x)$ is also bounded from above for $x, y \in \tilde{S}(L)$, because, otherwise, we find a sub–sequence $(a_n)$ of $(u_n)$ such that $S_{a_n,y}(x) \to \infty$. Thus, for any $m \in \mathbb{R}$ we have for sufficiently large $n$

$$\tilde{F}_y^{(a_n)}\left(S_{a_n,y}(x)\right) \le \tilde{F}_y^{(a_n)}\left(m\right).$$

Moreover,

$$\tilde{F}_y^{(a_n)}(m) \xrightarrow[n\to\infty]{} \bar{L}(m)/\bar{L}(y).$$

Letting $m \to \infty$ one gets

$$\tilde{F}_y^{(a_n)}(S_{a_n,y}(x)) \xrightarrow[n\to\infty]{} 0$$

which contradicts the assumption that $x \in \tilde{S}(L)$.

Now fix $y \in \tilde{S}(L)$. The sequence $S_{u_n,y}(x)$ is bounded for each $x \in \tilde{S}(L)$. Fixing some $x \in \tilde{S}(L)$, we find a subsequence $(u_n^{(1)})$ of $(u_n)$ such that $S_{u_n^{(1)},y}(x)$ converges to some real number. Next choose an arbitrary countable dense subset $A$ of $\tilde{S}(L)$. Using a diagonal sequence we find a subsequence $(u_n^*)$ of $(u_n)$ such that

$$S_{u_n^*,y}(x) \xrightarrow[n\to\infty]{} g_y'(x) \quad \text{for all } x \in A. \tag{2.15}$$

Now let $x \in \tilde{S}(L)\backslash A$. Since $A$ is dense we find a sequence $(x_n) \in A$ with $x_n \downarrow x$. Define

$$g_y'(x) := \lim_{n\to\infty} g_y'(x_n). \tag{2.16}$$

This limit exists since $S_{u_n^*,y}$ is strictly increasing and, therefore, $g_y'$ is non–decreasing on $A$ and $(g_y'(x_n))$ is a monotone sequence. Moreover, it is bounded. We find some $n_0(x) \in \mathbb{N}$ and $x'$, $x'' \in A$ such that $x' < x_n < x''$ for $n > n_0(x)$ and, thus,

$$g_y'(x') \le g_y'(x_n) \le g_y'(x''), \quad n > n_0(x).$$

Notice that Definition (2.16) ensures that $g_y'$ is also non–decreasing on $\tilde{S}(L)$. Now the convergence in (2.15) also holds for all $x \in \tilde{S}(L) \cap C(g_y')$. To see this let $x \in (\tilde{S}(L) \cap C(g_y'))\backslash A$. Since $A$ is dense we find for each positive $\varepsilon$ sufficiently small $\varepsilon', \varepsilon'' < \varepsilon$ such that $x + \varepsilon'$ and $x - \varepsilon''$ are both in $A$. From (2.15) and the monotonicity of $S_{u_n^*,y}$ and $g_y'$ one gets for $n \ge n_0(x,\varepsilon)$,

$$\begin{aligned} g_y'(x-\varepsilon) - \varepsilon \le g_y'(x-\varepsilon'') - \varepsilon \le S_{u_n^*}(x-\varepsilon'') \le S_{u_n^*,y}(x) \\ \le S_{u_n^*,y}(x+\varepsilon') \le g_y'(x+\varepsilon') + \varepsilon \le g_y'(x+\varepsilon) + \varepsilon. \end{aligned}$$

Now letting $\varepsilon \to 0$ the convergence in (2.15) is immediate since $x$ is assumed to be a point of continuity of $g_y'$.

We are now going to construct a function $g_y$ such that convergence in (2.15) holds for all $x \in \tilde{S}(L)$. Notice that $g_y'$ is a monotone function and, therefore, has at most countable many points of discontinuity. Now let $B$ be the set of all such points. Since $B$ is countable we find again by the boundedness of $S_{u_n^*,y}(x)$ and a diagonal sequence argument a subsequence $(u_n^{**})$ of $(u_n^*)$ such that

$$S_{u_n^{**},y}(x) \xrightarrow[n\to\infty]{} g_y''(x) \quad \text{for all } x \in B.$$

Put

$$g_y(x) := \begin{cases} g'_y(x), & x \in \tilde{S}(L)\backslash B, \\ & \text{if} \\ g''_y(x), & x \in B. \end{cases}$$

We get

$$S_{u_n^{**},y}(x) \xrightarrow[n\to\infty]{} g_y(x) \quad \text{for all } x \in \tilde{S}(L). \tag{2.17}$$

Since $L$ is continuous we immediately get from Lemma 2.2 that

$$\bar{L}(g_y(x)) = \bar{L}(x)\bar{L}(y) \quad \text{for all } x \in \tilde{S}(L),$$

which is the desired equation. □

Let $L$ be a continuous $g$–pot stable df with $g$ as given in Theorem 2.3. Then, apparently

$$g : \tilde{S}(L) \times \tilde{S}(L) \to \tilde{S}(L).$$

According to Theorem 2.3 the class of $g$–pot–stable dfs plays a central role in the theory of exceedances under general monotone transformations. In the following sections these dfs will investigated more closely.

## 2.3 Continuous g–POT–Stable Distributions

We now turn our attention to the limiting distribution $L$. The results of the foregoing section yield that the class of limiting distributions under a transformation $T$ consists of certain $g$–pot–stable distributions. We completely characterize the class of strictly increasing, continuous limiting distributions. These distributions are the most relevant for applications. Let $g(y, x)$ be as in the foregoing section, and let $L$ be a df which is $g$–pot–stable. If the limiting df $L$ is strictly increasing we get immediately the following sharper version of Lemma 2.4.

**Corollary 2.5** *Let $T$, $F$ and $L$ be as in Theorem 2.3. In addition assume that $L$ is strictly increasing on $\tilde{S}(L)$. Then,*

$$\bar{L}(g(y,x)) = \bar{L}(x)\bar{L}(y)$$

*for all $x, y \in \tilde{S}(L)$ where*

$$g(y,x) = \lim_{u\to\omega(F)} S_{u,y}(x).$$

Proof. With $L$, $g$ and $S_{u,y}$ as in Lemma 2.4 and Theorem 2.3 we get

$$g(y,x) = \bar{L}^{-1}(\bar{L}(x)\bar{L}(y))$$

since $L$ is strictly increasing. Therefore, $g$ is uniquely determined for all $x, y \in \tilde{S}(L)$. It remains to prove that $S_{u_n,y}$ converges pointwise on $\tilde{S}(L)$ for all sequences $u_n \to \omega(F)$ and $y \in \tilde{S}(L)$. Fix some $y \in \tilde{S}(L)$. Let $(u_n)$ be an arbitrary sequence converging to $\omega(F)$. From Lemma 2.4 we know that there exists a subsequence $(u_n^{**})$ such that

$$S_{u_n^{**},y}(x) \xrightarrow[u\to\omega(F)]{} g_y(x) \text{ for all } x \in \tilde{S}(L)$$

and

$$\bar{L}(g_y(x)) = \bar{L}(x)\bar{L}(y).$$

Now let $x \in \tilde{S}(L)$ and $(a_n)$ be a subsequence of $(u_n)$ such that

$$S_{a_n,y}(x) \xrightarrow[n\to\infty]{} a.$$

Using Lemma 2.2 and Lemma 2.4 we get

$$\frac{\bar{L}(a)}{\bar{L}(y)} = \lim_{n\to\infty} \tilde{F}_y^{(a_n)}(S_{a_n,y}(x)) = \bar{L}(x).$$

Thus, we have

$$\bar{L}(a) = \bar{L}(x)\bar{L}(y) = \bar{L}(g_y(x)).$$

Since $L$ is strictly increasing we get $a = g_y(x)$.

Therefore, $g_y(x)$ is the limit of each converging subsequence of $S_{u_n,y}(x)$. Recall that $S_{u_n,y}(x)$ is bounded for all $x$, $y \in \tilde{S}(L)$. Hence, all accumulation points of $S_{u_n,y}$ coincide. Consequently, $S_{u_n,y}(x)$ converges to $g_y(x)$ for all sequences $u_n \to \omega(F)$ .

□

So far we established a relationship of the limiting df $L$ and the pertaining transformation $T$ via the $g$–pot–stability of $L$. This result raises further questions:

- is it possible to get further insight in the structure of $g$, or in other words, can any monotone transformation $g$ appear in Corollary 2.5?
- suppose $g$ is known, can $L$ be derived from $g$?

We provide an answer to both questions, under the condition that the limiting df $L$ is strictly increasing. The first question is solved in the subsequent lemma, which also gives a special representation of $g$.

**Lemma 2.6** *Let $L$, $F$ and $T$ be as in Corollary 2.5. Then the function $g$ satisfies*

$$g\left(g(x,y),z\right) = g\left(x,g(y,z)\right), \quad x,y \in \tilde{S}(L). \tag{2.18}$$

*Moreover, $g$ is of the form*

$$g(y,x) = h^{-1}\left(h(x)+h(y)\right)$$

*for some continuous and strictly increasing function $h$.*

Proof. First observe that $\tilde{S}(L)$ is an interval. Moreover, for $x,y,z \in \tilde{S}(L)$ one gets

$$\begin{aligned} \bar{L}\left(g\left(x,g(y,z)\right)\right) &= \bar{L}(x)\bar{L}\left(g(y,x)\right) \\ &= \bar{L}(x)\bar{L}(y)\bar{L}(z) \\ &= \bar{L}\left(g(x,y)\right)\bar{L}(z) \\ &= \bar{L}\left(g\left(g(x,y),z\right)\right). \end{aligned}$$

Because $\bar{L}$ is strictly increasing this yields the associativity equation

$$g\left(g(x,y),z\right) = g\left(x,g(y,z)\right).$$

Furthermore, we get

$$g(x,y) = \bar{L}^{-1}\left(\bar{L}(x)\bar{L}(y)\right)$$

and, thus, $g$ is continuous and strictly increasing in both components. Consequently, $g(y_1,x) = g(y_2,x)$ or $g(x,y_1) = g(x,y_1)$ for any $x \in \tilde{S}(L)$ implies $y_1 = y_2$. Hence $g$ is cancellative in the sense of [1], Section 7, Theorem 1. Therefore,

$$g(x,y) = h^{-1}\left(h(x)+h(y)\right)$$

for some continuous and strictly increasing function $h : I \to J$, where $I, J \subset \mathbb{R}$ are appropriate intervals. □

Equation (2.18) is also known as associativity equation. It is closely investigated in [1], Section 7. Apparently, Lemma 2.6 implies that $g$ is also commutative, that is,

$$g(y,x) = g(x,y), \quad x,y \in \tilde{S}(L). \tag{2.19}$$

This property may as well be deduced from the equality

$$\bar{L}(g(x,y)) = \bar{L}(x)\bar{L}(y) = \bar{L}(y)\bar{L}(x) = \bar{L}(g(y,x))$$

if $L$ is strictly increasing.

The next theorem gives the answer to the second of the questions above. We provide a representation of $g$–pot–stable dfs, where $g$ is as given in Lemma 2.6.

**Theorem 2.7** *Let $L$ be a df, pot–stable with respect to $g$ with*

$$g(y,x) := h^{-1}\left(h(x) + h(y)\right)$$

*for $h : \tilde{S}(L) \to \tilde{S}(L)$ continuous and strictly increasing. Then*

$$L(x) = 1 - \exp(-\beta h(x))$$

*for $\beta > 0$.*

Proof. Because $L$ is $g$–pot–stable one gets

$$\bar{L}\left(g(y,x)\right) = \bar{L}(x)\bar{L}(y) \Leftrightarrow \bar{L}\left(h^{-1}(h(x) + h(y))\right) = \bar{L}(x)\bar{L}(y)$$

Define $V := \bar{L} \circ h^{-1}$, then the above equation becomes

$$V\left(h(x) + h(y)\right) = V(h(x))V(h(y)).$$

Because $V$ is decreasing and positive it can only be of he form

$$V(x) = \exp(-\beta x), \quad \beta > 0.$$

Therefore,

$$L(x) = 1 - V \circ h(x) = 1 - \exp(-\beta h(x))$$

which is the desired representation. □

Notice that the function $h$ in the proceeding lemma might be replaced with $\tilde{h}(x) := ah(x)$ for some $a > 0$ without changing $g$. In addition there exists $a \in [-\infty, \infty)$ and $b \in (-\infty, \infty]$ such that $h(x) \to 0$ if $x \to a$ and $h(x) \to \infty$ if $x \to b$.

If a special transformation $T$ (for example linear, power transformations or exponential transformations, cf. Chapter 3) is given and the pertaining strictly increasing, continuous limiting dfs of exceedances are to be determined the previous results can be utilized. If $L$ is not strictly increasing we can merely deduce $g$–pot–stability.

## 2.4 Some Remarks on Discrete POT-Stable Dfs

We include some remarks concerning discrete pot–stable distribution, which may also occur as limiting distributions of exceedances under monotone transformations. We first prove an extension of Lemma 2.4 to discontinuous limiting dfs. To obtain the pertaining result we have to impose an additional condition on the transformation $T$.

**Condition 2.8** *We assume that if $S_{u_n,y}$ (cf. (2.11)) converges pointwise to some function $g_y$ for some sequence $(u_n) \to U \in (-\infty, \infty]$ then $g_y$ is strictly increasing.*

Condition 2.8 seems a little bit technical and restrictive at first sight, yet it will turn out that it is quite easy to verify for a lot of transformations $T$.

The role of the interior of the support of the limiting df $L$, $\tilde{S}(L)$ will be played by the set of points of continuity of $L$ in the interior of its support

$$\widetilde{C}(L) := \{x \in C(L) : 0 < L(x) < 1\}$$

if $L$ is discontinuous.

**Lemma 2.9** *Consider a monotone transformation $T : I \times \mathbb{R} \to \mathbb{R}$, where $I \subset \mathbb{R}$ is an interval. Assume that Condition 2.8 is satisfied. Furthermore, let $F$ and $L$ be non–degenerate dfs such that*

$$\frac{\bar{F}(T(u,x))}{\bar{F}(u)} \overset{u\to\omega(F)}{\longrightarrow} \bar{L}(x) \tag{2.20}$$

*for all $x \in C(L)$.*
*Then, for each $y \in \widetilde{C}(L)$ and each sequence $(u_n)$ with $u_n \xrightarrow[n\to\infty]{} \omega(F)$ there exists a subsequence $(u_n^{**})$ such that*

$$g_y(x) = \lim_{n\to\infty} S_{u_n^{**},y}(x)$$

*exists for all $x \in \widetilde{C}(L)$. Moreover, $\bar{L} = 1 - L$ satisfies*

$$\bar{L}\left(g_y(x)\right) = \bar{L}(x)\bar{L}(y), \quad x \in \widetilde{C}(L). \tag{2.21}$$

Proof. First notice that the proof of Lemma 2.4 can be repeated up to equation (2.17) with $\tilde{S}(L)$ replaced by $\widetilde{C}(L)$ and $\mathbb{R}$ replaced with $C(L)$. It remains to prove that

$$\bar{L}(g_y(x)) = \bar{L}(x)\bar{L}(y) \quad \text{for all } x \in \widetilde{C}(L).$$

If Condition 2.8 holds $g_y$ is strictly increasing. Let $x \in \widetilde{C}(L)$. Hence, for each $\varepsilon > 0$ we find $\varepsilon', \varepsilon'' < \varepsilon$ and some $\delta_0 > 0$ such that $x - \varepsilon''$ and $x + \varepsilon'$ are in $\widetilde{C}(L)$ and for all $0 < \delta \leq \delta_0$ it holds that

$$g_y(x - \varepsilon'') < g_y(x) - \delta < g_y(x) < g_y(x) + \delta < g_y(x + \varepsilon').$$

Therefore, we have for $n > n(x, \varepsilon)$ and $(u_n^{**})$ as in the proof of Lemma 2.4

$$S_{u_n^{**},y}(x - \varepsilon'') < S_{u_n^{**},y}(x) - \delta < S_{u_n^{**},y}(x) < S_{u_n^{**},y}(x) + \delta < S_{u_n^{**},y}(x + \varepsilon').$$

Applying the monotonicity of $\tilde{F}_y^{(u_n^{**})}$ to the preceding inequalities one receives

$$\begin{aligned}\tilde{F}_y^{(u_n^{**})}\left(S_{u_n^{**},y}(x - \varepsilon'')\right) \geq \tilde{F}_y^{(u_n^{**})}\left(S_{u_n^{**},y}(x) - \delta\right) \geq \tilde{F}_y^{(u_n^{**})}\left(S_{u_n^{**},y}(x)\right) \\ \geq \tilde{F}_y^{(u_n^{**})}\left(S_{u_n^{**},y}(x) + \delta\right) \geq \tilde{F}_y^{(u_n^{**})}\left(S_{u_n^{**},y}(x + \varepsilon')\right).\end{aligned}$$

Now choose $\delta'$ and $\delta''$ both smaller than $\delta_0$ such that $g_y(x)+\delta'$ and $g_y(x)-\delta''$ are both in $\widetilde{C}(L)$. Letting $n \to \infty$ and applying Lemma 2.2 one gets

$$\bar{L}(x-\varepsilon'') \geq \frac{\bar{L}(g_y(x)-\delta'')}{\bar{L}(y)} \geq \bar{L}(x) \geq \frac{\bar{L}(g_y(x)+\delta')}{\bar{L}(y)} \geq \bar{L}(x+\varepsilon').$$

Fix some arbitrary $\varepsilon > 0$. Since $\widetilde{C}(L)$ is a dense subset of an open interval in $\mathbb{R}$ and the above inequality holds for all $\delta', \delta'' < \delta_0$ satisfying the above condition we might let $\delta', \delta'' \to 0$ in the above inequality. If $\varepsilon > 0$ is sufficiently small we, thus, find $\varepsilon', \varepsilon'' < \varepsilon$ such that

$$\bar{L}(x-\varepsilon'') \geq \lim_{\delta''\to 0} \frac{\bar{L}(g_y(x)-\delta'')}{\bar{L}(y)} \geq \bar{L}(x) \geq \lim_{\delta'\to 0} \frac{\bar{L}(g_y(x)+\delta')}{\bar{L}(y)} \geq \bar{L}(x+\varepsilon').$$

Notice that both limits exist since they concern bounded and monotone sequences. Because $\bar{L}$ is continuous from the right we have in addition

$$\bar{L}(x-\varepsilon'') \geq \lim_{\delta''\to 0} \frac{\bar{L}(g_y(x)-\delta'')}{\bar{L}(y)} \geq \bar{L}(x) \geq \frac{\bar{L}(g_y(x))}{\bar{L}(y)} \geq \bar{L}(x+\varepsilon').$$

Since this holds for all $\varepsilon > 0$ and $x$ is a point of continuity of $L$ we get

$$\bar{L}(x) = \frac{\bar{L}(g_y(x))}{\bar{L}(y)}$$

which proves the assertion. □

The last lines of the proof of the foregoing lemma also entail that for all $x, y \in \widetilde{C}(L)$ $g_y(x)$ is a point of continuity of $L$.

An important class of such discrete distributions is closely related to the class of pot–stable distributions introduced in the foregoing section. Consider the transformation

$$g(y,x) = h^{-1}\left(h(x) + [h(y)]\right), \tag{2.22}$$

where $h$ denotes, as in the foregoing sections, a positive and strictly increasing function with $h(x) \to 0$ for $x \to a \in [-\infty, \infty)$ and $h(x) \to \infty$ for $x \to b \in (-\infty, \infty]$. Then we may define the discrete df

$$L(x) = 1 - \exp\left(-c[h(x)]\right), \quad c > 0.$$

Check that

$$\bar{L}(g(y,x)) = \exp(-(c[h(x)] + c[h(y)])) = \bar{L}(x)\bar{L}(y).$$

It is an open question whether transformations as in (2.22) are the only possible ones in the relation

$$\bar{L}(g(y,x)) = \bar{L}(x)\bar{L}(y), \quad x, y \in \widetilde{C}(L),$$

where $L$ denotes a discontinuous df. Notice that $g$ is associative (cf. 2.18) but not commutative (cf. 2.19).

## 2.5 Relations to the Limit Theory of Maxima

In this section we study the relation of limiting dfs of maxima of iid rvs under monotone transformations as introduced in [49] and limiting distributions of exceedances. In the linear setup it is well known that a continuous limiting dfs of exceedances $W$ can be derived from an EVD $G$ by the relation

$$W(x) = 1 + \log\left(G(x)\right), \quad \log\left(G(x)\right) > -1.$$

We will derive the analog relation for limiting dfs under general monotone transformations.

Moreover, we recall the concept of general max–domains of attraction (see, e.g., [49]) and introduce general pot–domains of attraction. Again, it turns out that both concepts are closely related to each other.

### 2.5.1 Some Remarks on Limiting Dfs of Maxima

Assume for dfs $F$ and $H$ the relation

$$F^n\left(\Delta_n(x)\right) \xrightarrow[n\to\infty]{} H(x), \quad x \in C(H) \tag{2.23}$$

holds, where $(\Delta_n)$ is a sequence of strictly increasing functions. Due to a result in [49] we know that for sequences $(m_n)$ with $m_n/n \xrightarrow[n\to\infty]{} \lambda > 0$ one gets

$$\Delta_n\left(\Delta_{m_n}^{-1}(x)\right) \xrightarrow[n\to\infty]{} k(\lambda, x). \tag{2.24}$$

If, in addition, $k$ is invertible as a function of $\lambda$, that is, e.g., strictly increasing in the first component, then

$$k(\lambda, x) = h^{-1}\left(h(x) + c\log(\lambda)\right)$$

for some positive, strictly increasing function $h$ and $c > 0$. Moreover, $H$ is generalized max–stable with respect to $k$, that is, we have

$$H^{\lambda}\left(k(\lambda, x)\right) = H(x), \quad \lambda > 0. \tag{2.25}$$

From this one may deduce the representation

$$H(x) = \exp\left(-e^{-\tilde{c}h(x)}\right) \tag{2.26}$$

for $\tilde{c} > 0$ (see, e.g., [49]). In [62] the relations (2.25) and (2.26) are utilized to deduce the function $h$ directly from $k$ under some additional conditions. This is important in practical applications since it is often straightforward to derive $k$ from the sequence of functions $(\Delta_n)$, but it is usually a difficult task to derive $h$.

### 2.5.2 Domains of Attraction

If (2.3) holds for $F$, $L$ and some monotone transformation $T$, then $F$ is said to be in the $T$–pot–domain of attraction of the df $L$. In the following definition we distinguish between three different forms of domains of attraction.

**Definition 2.10** *Let $F$ and $L$ be dfs and $T : \mathbb{R}^2 \to \mathbb{R}$ a monotone transformation.*

*(i) $F$ is in the $T$–pot–domain of attraction of $L$ ($F \in \mathcal{D}_{T\text{–}pot}(L)$) if (2.3) holds for all $x \in C(L)$.*

*(ii) Let $\mathcal{T}$ be a family of transformations $T : \mathbb{R}^2 \to \mathbb{R}$. We say that $F$ is in the $\mathcal{T}$–pot–domain of attraction of $L$ ($F \in \mathcal{D}_{\mathcal{T}\text{–}pot}(L)$) if there exists some $T \in \mathcal{T}$ such that $F \in \mathcal{D}_{T\text{–}pot}(L)$.*

*(iii) $F$ is in the general pot–domain of attraction of $L$ if $F \in \mathcal{D}_{T\text{–}pot}(L)$ for some monotone transformation $T$.*

In the classical EVT literature (see e.g. [3] and [36]) the term domain of attraction is assigned to the second case of the preceding definition, and a special family of transformation, namely linear transformations, is considered. In [49] the third meaning of a domain of attraction is used in conjunction with limiting dfs of maxima of iid rvs.

For statistical applications domains of attraction play a crucial role. If $F \in \mathcal{D}_{T\text{–pot}}(L)$ we have

$$F^{[u]}\left(T(u, x)\right) \approx L(x)$$

or, equivalently,

$$F^{[u]}(x) \approx L(T^{(-1)}(u, x)) \tag{2.27}$$

for $u$ sufficiently close to $\omega(F)$. If nothing is known about the transformation $T$ the estimation of $F^{[u]}$ via $L \circ T^{(-1)}(u, \cdot)$ remains a non–parametric problem and not much is gained. But if $T$ belongs to a known parametric family $\mathcal{T}$ we may utilize (2.27) to derive a parametric estimation of $F^{[u]}$ and, therefore, the upper tail of $F$. This approach reflects the spirit of traditional EVT, namely replacing a non–parametric model by a parametric one which is justified by asymptotic relations.

We get from the results in Sections 2.2 and 2.3 that the limiting df $L$ is $g$–pot–stable. It will become apparent from the subsequent section that it is often possible to derive the transformation $g$ from $T$, yet it remains a difficult task to compute all possible functions $h$ in Lemma 2.6 for a given transformation $g$.

It is not within the scope of this thesis to derive characterizations of all $T$–pot–domains of attraction, but we will establish relationships to max–domains of attraction. We start with some remarks concerning general max–domains of attraction. There has been some confusion about the concrete form of general $T$–max–domains of attraction introduced in [49]. In this article the author claimed that there exists a sequence of monotone functions $(\Delta_n)$ such that

$$F^n\left(\Delta_n(x)\right) \xrightarrow[n\to\infty]{} H(x) = \exp\left(-e^{-h(x)}\right) \tag{2.28}$$

if, and only if, the df $F$ satisfies

$$1 - F(x) \sim U\left(h(x)\right)\exp\left(-h(x)\right) \tag{2.29}$$

as $x \to \omega(F)$, where $U$ is a regularly varying function. The sequence of functions $(\Delta_n)$ can be chosen as

$$\Delta_n = h^{-1}\Bigg(h(x) + \log\left(n\log\left(U(\log(n))\right)\right)\Bigg),$$

see [49] and [62]. Notice that these considerations concern the third meaning of the term domain of attraction as given in Definition 2.10. It is shown in Examples 1 and 2 of [62] that this result is not correct. In fact, the stated condition (2.29) is sufficient but not necessary for (2.28). A correct necessary and sufficient condition for (2.29) is stated in Theorem 4.1 of [62]. We give the result without proof.

**Theorem 2.11 (Sreehari 2009)** *Let $F$ be a non–degenerate df such that (2.28) holds for some general max–stable df $H = \exp(-e^{-h(x)})$ and a sequence of monotone functions $(\Delta_n)_{n\in\mathbb{N}}$, then there exists a sequence of positive functions $(\Delta_n^*)_{n\in\mathbb{N}}$ such that*

$$\frac{K\left(h(x)\right) + \log\left(n\Delta_n^*(x)\right)}{\Delta_n^*(x)} \xrightarrow[n\to\infty]{} 1 \tag{2.30}$$

*for all $x \in (\alpha(F), \omega(F))$, where*

$$K(x) = \left(1 - F\left(h^{-1}(x)\right)\right)\exp(x).$$

*Conversely, if (2.30) holds for some strictly increasing continuous function $h$ and sequences of positive function $(\Delta_n^*)$, then there exists a sequence of monotone increasing functions $(\Delta_n)$ such that (2.28 ) holds with $H(x) = \exp\left(-e^{-h(x)}\right)$. The sequence of functions can be chosen as*

$$\Delta_n(x) = h^{-1}\left(h(x) + \log\left(\Delta_n^*(x)\right)\right).$$

The incorrect characterization of general max–domains of attraction given in [49] still yields a sufficient condition for a df $F$ such that (2.28) holds for some $(\Delta_n)$, namely if $F$ has the representation

$$F(x) = 1 - (1 + O(1))\, R\,(h(x)) \exp\left(-h(x)\right), \quad x \to \omega(F)$$

where $R$ is a regularly varying function. We now turn again to general pot–domains of attraction and their relations to general max–domains of attraction.

**Lemma 2.12** *Let $F, H, \Delta$ and $h$ be as in (2.28). Then, there exists a mapping $n^* : \mathbb{R} \to \mathbb{N}$ such that $T(u,x) := \Delta_{n^*(u)}(x)$ is a monotone transformation, and*

*(i)*

$$\frac{F\left(T(u,x)\right)}{\bar{F}(u)} \xrightarrow[u\to\omega(F)]{} \bar{L}(x), \quad x \in C(L),$$

*(ii)*

$$L(x) = 1 + \log(H(x)) = 1 - e^{-h(x)}, \quad h(x) > 0,$$

*(iii)*

$$S_{u,y}(x) \xrightarrow[u\to\omega(F)]{} h^{-1}\left(h(x) + h(y)\right)$$

*where $S_{u,y}$ is defined as in Lemma 2.9.*

Proof. We get from [49] that (2.23) is equivalent to

$$n\bar{F}\left(\Delta_n(x)\right) \xrightarrow[n\to\infty]{} -\log\left(H(x)\right), \quad x \in C(H), \tag{2.31}$$

and this convergence necessarily entails $\Delta_n(x) \xrightarrow[n\to\infty]{} \omega(F)$. Define $n^*(u)$ such that

$$n^*(u) = \min_{n\in\mathbb{N}} \left\{ \frac{1}{n+1} \le \bar{F}(u) < \frac{1}{n} \right\}.$$

Obviously, $n^*(u) \xrightarrow[u\to\omega(F)]{} \infty$ and

$$n^*(u) < \frac{1}{\bar{F}(u)} \le n^*(u) + 1. \tag{2.32}$$

Using the last relation we get

$$n^*(u)\bar{F}\left(\Delta_{n^*(u)}(x)\right) < \frac{\bar{F}\left(\Delta_{n^*(u)(x)}\right)}{\bar{F}(u)} \le (n^*(u)+1)\bar{F}\left(\Delta_{n^*(u)}(x)\right). \tag{2.33}$$

Furthermore, since $n^*(u) \to \infty$ as $u \to \omega(F)$ (2.33) together with (2.31) yields

$$\frac{\bar{F}\left(\Delta_{n^*(u)}(x)\right)}{\bar{F}(u)} \xrightarrow[u \to \omega(F)]{} -\log(H(x)), \quad x \in C(H), \tag{2.34}$$

notice that $\bar{F}\left(\Delta_{n^*(u)}(x)\right) \xrightarrow[u \to \omega(F)]{} 0$. Put $T(u,x) := \Delta_{n^*(u)}(x)$ and

$$L(x) := \begin{cases} 1 + \log(H(x)) = 1 - e^{-h(x)} & \text{if } -\log(H(x)) > -1; \\ 0, & \phantom{\text{if }} -\log(H(x)) \leq -1. \end{cases}$$

Then we may rewrite (2.34) by

$$\frac{\bar{F}\left(T(u,x)\right)}{\bar{F}(u)} \xrightarrow[u \to \omega(F)]{} \bar{L}(x), \quad x \in \widetilde{C}(L).$$

Consequently (i) and (ii) hold. It remains to prove part (iii). First observe that

$$S_{u,y}(x) = T^{(-1)}\left(u, T\left(T(u,y), x\right)\right) = \Delta^{-1}_{n^*(u)}\left(\Delta_{n(T(u,y))}\right).$$

Using a result in [49], see also (2.24), one gets

$$S_{u,y}(x) \xrightarrow[u \to \omega(F)]{} h^{-1}\left(h(x) + c\log(\lambda)\right)$$

for some $c > 0$, if

$$\lim_{u \to \omega(F)} \frac{n^*(u)}{n^*\left(T(u,y)\right)} = \lambda > 0. \tag{2.35}$$

It follows from (2.32) that

$$\frac{1}{\bar{F}(u)} - 1 \leq n^*(u) \leq \frac{1}{\bar{F}(u)}$$

and

$$\bar{F}\left(T(u,y)\right) \leq \frac{1}{n^*\left(T(u,y)\right)} \leq \frac{1}{1/\bar{F}\left(T(u,y)\right) - 1}.$$

Thus, we get

$$\bar{F}\left(T(u,y)\right)\left(\frac{1}{\bar{F}(u)} - 1\right) \leq \frac{n^*(u)}{n^*\left(T(u,y)\right)} \leq \left(\frac{\bar{F}(u)}{\bar{F}\left(T(u,y)\right)} - \bar{F}(u)\right)^{-1}. \tag{2.36}$$

Check that the right hand as well as the left hand side of (2.36) tend to $e^{-h(y)}$ for $u \to \omega(F)$. Therefore, (2.35) holds with $\lambda = e^{-h(y)}$ and $c = 1$ for all $y \in \widetilde{C}(L)$. □

Basically, Lemma 2.12 yields that the relation $F \in \mathcal{D}_{max}(H) \Rightarrow F \in \mathcal{D}_{pot}(1 + \log(H))$, which is well known from the linear case, is still valid in a more general context. The transformation $T(u, \cdot)$, which is used to derive the limiting distribution of exceedances, is in some sense of the same type as the transformations for the maxima. The converse implication is not proven in the framework of this thesis. Nevertheless, one may conjecture that it is also true. The max–domain of attraction of $H$ under the function $\Delta_n$ is a subset of the pot–domain of attraction of $L = 1 + \log(H)$ under the transformation $T$.

## 2.6 Deriving the Result of Balkema and de Haan

In this section we use the results of the foregoing sections to derive some of the famous result of Balkema and de Haan [3] and Pickands [53]. We deduce that GPDs are the only strictly increasing, continuous limiting distributions of exceedances under linear transformations. Assume that

$$\frac{\bar{F}(T(u,x))}{\bar{F}(u)} \xrightarrow[u \to \omega(F)]{} \bar{L}(x)$$

for

$$T(u,x) = a(u) + b(u)x, \quad a(u) \in \mathbb{R},\ b(u) > 0.$$

One immediately gets

$$T^{(-1)}(u,x) = \frac{x - a(u)}{b(u)}$$

and

$$S_{u,y}(x) = \frac{a(T(u,y)) - a(u)}{b(u)} + \frac{b(T(u,y))}{b(u)}x.$$

Let $y \in \tilde{S}(L)$ be arbitrary but fixed. We get from Lemma 2.4 that there exists a sequence $(u_n)$, $u_n \to \omega(F)$ such that $S_{u_n,y}$ converges for all $x \in \tilde{S}(L)$, thus

$$\frac{a(T(u_n,y)) - a(u_n)}{b(u_n)} \xrightarrow[n \to \infty]{} A(y) \in \mathbb{R}$$

and

$$\frac{b(T(u_n,y))}{b(u_n)} \xrightarrow[n \to \infty]{} B(y) \geq 0.$$

The first convergence follows from the fact that one can assume without loss of generality that $0 \in \tilde{S}(L)$, since, otherwise, we just change $T(u,x)$ by an additive constant. The second convergence is then a direct consequence of the first one. Since we find such a sequence $(u_n)$ for each $y \in \tilde{S}(L)$ one may conclude that $L$ is $g$–pot–stable with

$$g(y,x) = A(y) + B(y)x.$$

We now determine the functions $A(y)$ and $B(y)$ if $L$ is strictly increasing and continuous. We may choose $T$ such that $0,\ 1 \in \tilde{S}(L)$. Changing $T$ by an additive and a multiplicative constant does only change the limiting df by a location and scale shift. Furthermore, we know that

$$g(y,x) = h^{-1}\left(h(x) + h(y)\right) \tag{2.37}$$

for some continuous and strictly increasing function $h$. Thus, $g$ is commutative according to (2.19) and, therefore,

$$A(y) = g(y,0) = g(0,y) = A(0) + B(0)y. \tag{2.38}$$

We get from (2.38) and again by the commutativity of $g$ that for $x$, $y_1$ and $y_2$ all in $\tilde{S}(L)$

$$\begin{aligned} g(x, y_1 + y_2) &= A(x) + B(x)(y_1 + y_2) \\ &= g(x, y_1) + g(x, y_2) - A(x) \\ &= g(y_1, x) + g(y_2, x) - A(x) \\ &= A(y_1) + B(y_1)x + A(y_2) + B(y_2)x - A(0) - B(0)x \\ &= A(0) + B(0)(y_1 + y_2) + B(y_1) + B(y_2) - B(0)x \end{aligned} \tag{2.39}$$

and also

$$\begin{aligned} g(x, y_1 + y_2) &= g(y_1 + y_2, x) \\ &= A(y_1 + y_2) + B(y_1 + y_2)x \\ &= A(0) + B(0)(y_1 + y_2) + B(y_1 + y_2)x. \end{aligned} \tag{2.40}$$

Now combining (2.39) and (2.40) and plugging in $x = 1$ we receive

$$B(y_1 + y_2) = B(y_1) + B(y_2) - B(0). \tag{2.41}$$

Since (2.41) holds for all $y_1, y_2 \in \tilde{S}(L)$ and $\tilde{S}(L)$ contains an interval we get

$$B(y) = \alpha + \beta y, \quad \alpha \in \mathbb{R}, \beta \in \mathbb{R}.$$

We now use the representation (2.37) to determine the function $h$ and the limiting df $L(x) = 1 - \exp(-h(x))$. We have

$$\begin{aligned} g(y,x) &= h^{(-1)}(h(x) + h(y)) \\ &= A(0) + \alpha x + \alpha y + \beta yx, \end{aligned}$$

and, therefore,

$$h(x) + h(y) = h\left(A(0) + \alpha x + \alpha y + \beta yx\right).$$

If $\beta = 0$, this immediately yields

$$h(x) = \frac{x-\mu}{\sigma},$$

where $\mu = -A(0)$ and $\sigma > 0$ and, thus, $L$ is an exponential distribution

$$W_{0,\mu,\sigma}$$

with location ans scale parameters $\mu$ and $\sigma$.

If $\beta > 0$ it is more difficult to determine $h$. Put $y = x$, we get

$$2h(x) = h\left(A(0) + 2\alpha x + \beta x^2\right).$$

Notice that a solution of this functional equation is given by

$$h(x) = a \log\left(b + cx\right)$$

for adequate $a$, $b$, $c \in \mathbb{R}$. Due to [3] we know that this is the only solution which yields a df. Nevertheless, it should be possible to prove this without relying on results in [3], but to use the theory of functional equations. Notice that similar problems occur in the context of Archimedian copulas. We may also parametrize $h$ in a different form, namely by

$$h(x) = \frac{1}{\gamma} \log\left(1 + \gamma\frac{x-\mu}{\sigma}\right), \quad \gamma, \mu \in \mathbb{R}\,, \sigma > 0$$

and, thus, the limiting df is a GPD $W_{\gamma,\mu,\sigma}$ as in (2.5).

## 2.7 The Multivariate Case

We shortly mention extensions of the foregoing results to the multivariate case.

Traditionally multivariate EVT has concentrated on limiting distributions of componentwise taken maxima. We refer to [27], Chapter 4, for a broad discussion of this topic. A systematic treatment of multivariate exceedances started not until the last decade with a PhD–thesis by Tajvidi [64] and a pertaining article [57] as well as a series of articles by Reiss and Falk ([26], [28], [29], [31], [30], [32]) and Kaufmann and Reiss [43]. A summarized overview may be found in [27], Chapters 5 and 6.

All these investigations rely on the above mentioned well developed theory for multivariate maxima. The basic result of the latter theory is, that if for non–degenerate multivariate dfs $F$ and $G$ and adequate vectors of constants $\boldsymbol{a}_n$, $\boldsymbol{b}_n \in \mathbb{R}^d$

$$F^n(\boldsymbol{a}_n + \boldsymbol{b}_n \cdot \boldsymbol{x}) \xrightarrow[n\to\infty]{} G(\boldsymbol{x}) \tag{2.42}$$

holds, then $G$ is a multivariate extreme value distribution. Here the operations "+" and "·" are meant componentwise. If (2.42) holds we say, as in the univariate case, that $F$ is in the max–domain of attraction of the EVD $G$ under linear transformation.

Furthermore, (2.42) entails in particular, that $G$ is max–stable. For each $n \in \mathbb{N}$ there exists some $\boldsymbol{a}_n$, $\boldsymbol{b}_n \in \mathbb{R}^d$ such that

$$G^n(\boldsymbol{a}_n + \boldsymbol{b}_n \cdot \boldsymbol{x}) = G(\boldsymbol{x}).$$

All univariate margins of $G$ are max–stable and, therefore, univariate EVDs. If the marginal distributions $G_i$, $i = 1, \ldots, d$, of $G$ are transformed to standard reverse exponential margins

$$G_i(x) = e^x, \quad x < 0$$

then $G$ has the Pickands representation

$$G(\boldsymbol{x}) = \exp\left(D(\boldsymbol{z}) \sum_{i=1}^{d} x_i\right)$$

where $D$ is a Pickands dependence function and

$$\boldsymbol{z} = \frac{1}{\sum_{i=1}^{d} x_i}(x_1, \ldots, x_{d-1}).$$

For more details we refer to [27], Section 4.3.

In [27] the authors introduce a multivariate generalized Pareto (GP) function defined by

$$W(\boldsymbol{x}) = 1 + \log(G(\boldsymbol{x})), \quad \log(G(\boldsymbol{x})) > -1 \tag{2.43}$$

extending this well–known relationship from the univariate case to higher dimensions. In the bivariate case a GP function is a df, this is not necessarily the case in higher dimensions. Yet there exists dfs which coincide with a GP function in the upper tail. If the EVD $G$ has reverse exponential margins the pertaining GP function is the upper tail of the uniform distribution on $[-1, 0]$.

It is shown that a GP function is an adequate model for multivariate distribution tails as it is in the univariate case. A first important property of a GP function in connection with multivariate exceedances is stated in the following lemma (part (iii) of Lemma 5.1.3 in [27]).

**Lemma 2.13 (Falk and Reiss (2004))** *Let $\boldsymbol{X}$ be a random vector distributed ac-*

*cording to a multivariate EVD with reverse exponential margins. Then for* $\boldsymbol{x}$, $\boldsymbol{t} \le 0$

$$\lim_{r\downarrow 0} P\left(\boldsymbol{X} \le r\boldsymbol{x} \,|\boldsymbol{X} \in (-\infty, r\boldsymbol{t}]^c\right) = F_{\boldsymbol{t}}(\boldsymbol{x})$$

*where* $F_{\boldsymbol{t}}$ *is a df satisfying*

$$F_{\boldsymbol{t}}(\boldsymbol{x}) = W(\boldsymbol{x}), \quad \boldsymbol{a} < \boldsymbol{x} < 0$$

*for some* $\boldsymbol{a} \in \mathbb{R}^d$ *and some GP function* $W$*.*

In the foregoing lemma a vector $\boldsymbol{x}$ is considered as an exceedance over the threshold $\boldsymbol{u} = r\boldsymbol{t}$ if at least one component of $\boldsymbol{x}$ exceeds the pertaining component of $\boldsymbol{u}$. The appropriate limit distribution coincides with a GP function for all such $\boldsymbol{x}$ where all components of $\boldsymbol{x}$ exceed the pertaining component of $\boldsymbol{u}$. In addition a GP function satisfies various versions of multivariate pot–stability (see [27], Section 5.2).

In [57] the authors generalize the above result. They study multivariate generalized Pareto distributions of the form

$$\widetilde{W}(\boldsymbol{x}) = \frac{1}{-\log\left(G\left(\boldsymbol{0}\right)\right)} \log\left(\frac{G\left(\boldsymbol{x}\right)}{G\left(\boldsymbol{x} \wedge \boldsymbol{0}\right)}\right), \quad 0 < G\left(\boldsymbol{0}\right) < 1$$

where $\wedge$ denotes the componentwise taken minimum. Notice that $G$ cannot be taken as an EVD with reverse exponential margins since otherwise we have $G\left(\boldsymbol{0}\right) = 1$. Yet we get for

$$\widetilde{W}(\boldsymbol{x}) = 1 - \frac{\log\left(G(\boldsymbol{x})\right)}{\log\left(G(\boldsymbol{0})\right)}, \quad \boldsymbol{x} > \boldsymbol{0}, \tag{2.44}$$

that

$$\widetilde{W}(\boldsymbol{x}) = W(\boldsymbol{x}), \quad \boldsymbol{x} > 0,$$

if $G$ is normalized such that $G\left(\boldsymbol{0}\right) = e^{-1}$. In particular, $G$ does not possess reverse exponential margins.

When dealing with limiting distributions of exceedances one has to determine how the multivariate threshold $\boldsymbol{u}$ approaches the right endpoint of a given distribution. In Lemma 2.13 the authors choose some direction $\boldsymbol{t}$ and then let $\boldsymbol{u}$ approach zero along this direction. In general one may define (cf. [57]) a $\mathbb{R}^d$–valued threshold curve $\{\boldsymbol{u}(t) | t \in [1, \infty)\}$ such that $F\left(\boldsymbol{u}(t)\right) \to 1$ if $t \to \infty$. It is proven in [57] that, if $\boldsymbol{X}$ is a $\mathbb{R}^d$–valued random vector with df $F$ and

- there exists a threshold curve $\boldsymbol{u}(t)$
- a function $\boldsymbol{b} : \mathbb{R}^d \to \mathbb{R}^d$
- and a df $L$ with non–degenerate margins

such that

$$P\left(\frac{\boldsymbol{X}-\boldsymbol{u}(t)}{\boldsymbol{b}\left(\boldsymbol{u}(t)\right)} \leq \boldsymbol{x} \middle| \boldsymbol{X} \in (-\boldsymbol{\infty}, \boldsymbol{u}(t)]^c\right) \xrightarrow[t\to\infty]{} L(\boldsymbol{x}),$$

then $L$ is a multivariate GPD $\widetilde{W}$ as defined in (2.44). Because only a scale transformation is used and the shift is confined to be equal to the threshold, only continuous pot–stable marginal distributions, thus GPDs, occur in the above limit relation. The question whether different dependence structures appear in the limit if an arbitrary shift function $\boldsymbol{a}$ is considered is still unsolved.

A natural question arises in this context. Consider a general monotone transformation and the pertaining limiting distribution in the multivariate case where the pertaining margins are uniform distributions on $[-1, 0]$. Do other dependence structures different from that induced by the Pickands dependence function arise?

We will confine us to transformations $\boldsymbol{T} : \mathbb{R}^d \to \mathbb{R}^d$ which apply to each component separately, that is

$$\boldsymbol{T}(\boldsymbol{x}) = (T_1(x_1), \ldots, T_d(x_d)), \quad \boldsymbol{x} = (x_1, \ldots, x_d), \tag{2.45}$$

where the marginal transformations $T_i$ are as in Condition 2.8. We only consider the construction (2.43). Thus, the question is reduced to multivariate max–stable distributions under general, yet componentwise monotone transformations. Max–stable distributions in greater generality are studied in [62], [50] and [51]. The following lemma shows that there are no other dependence structures than in the case of linear transformations.

**Lemma 2.14** *Let $\boldsymbol{K} : (0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d$, $\boldsymbol{K}(\boldsymbol{x}) = (K_1(r, x_1), \ldots, K_d(r, x_d))$ be a transformations where $K_i : (0,\infty) \times \mathbb{R} \to \mathbb{R}$ is strictly increasing in both components for, $i = 1, \ldots, d$. Furthermore, let $H$ be a df satisfying*

$$H^r\left(\boldsymbol{K}(r, \boldsymbol{x})\right) = H\left(\boldsymbol{x}\right), \quad \boldsymbol{x} \in \mathbb{R}^d, \, r > 0. \tag{2.46}$$

*Then there exists a transformation $\boldsymbol{K}^* : \mathbb{R}^d \to \mathbb{R}^d$, $\boldsymbol{K}^*(\boldsymbol{x}) = (K_1^*(x_1), \ldots, K^*(x_d))$ where $K_i^* : \mathbb{R} \to \mathbb{R}$ is strictly increasing for $i = 1, \ldots, d$ such that*

$$H\left(\boldsymbol{K}^*(\boldsymbol{x})\right) = G(\boldsymbol{x})$$

*where $G$ is an EVD with reverse exponential margins. In particular $H$ and $G$ have the same copula.*

Proof. First observe that (2.46) yields that the margins $H_i$ of $H$ are generalized max–stable with respect to $K_i$. This entails in particular that there exists strictly monotone functions $h_i$ and some $c_i > 0$ (see again [51] and [62]) such that

$$K_i(r, x) = h_i^{-1}\left(h_i(x) + c_i \log(r)\right).$$

We get

$$H^n\left(h_1^{-1}\left(h_1(x_1)+c_1\log(n)\right),\dots,h_d^{-1}\left(h_d(x_d)+c_d\log(n)\right)\right)=H\left(x_1,\dots,x_d\right)$$

for all $n\in\mathbb{N}$. Now choose $\boldsymbol{y}$ such that $x_i=h_i^{-1}(y_i)$, $i=1,\dots,d$. It follows that

$$\begin{aligned}H^n\left(h_1^{-1}\left(y_1+c_1\log(n)\right),\dots,h_d^{-1}\left(y_d+c_d\log(n)\right)\right)&\\ =H\left(h_1^{-1}(y_1),\dots,h_d^{-1}(y_d)\right).&\end{aligned}$$

Putting

$$\widetilde{H}(x_1,\dots,x_d):=H\left(h_1^{-1}(x_1),\dots,h_d^{-1}(x_d)\right),$$

one gets

$$\widetilde{H}^n\left(x_1+c_1\log(n),\dots,x_d+c_d\log(n)\right)=\widetilde{H}(x_1,\dots,x_d)$$

and, thus, $\widetilde{H}=G$ for some EVD $G$. Apparently, this yields that the copulas of $H$ and $G$ coincide. □

It is not in the scope of this thesis to investigate whether limiting distributions of exceedances appear in the multivariate framework which are not derived from a max–stable distribution $G$. If one also considers transformations which mix the marginal distributions, thus not being of the form given in (2.45), one has to develop a new theory. First attempts into this direction are made in [4] where matrix–transformations are applied to the exceedances.

# 3 Limiting Distributions Under Special Monotone Transformations

In this chapter we will study some special non–linear normalizations in detail, namely power–normalizations and exponential normalizations.

Our main focus lies on power–normalizations. The concept of power–normalization was first introduced in [49] in conjunction with limiting distributions of maxima of independent rvs. In contrast to linear normalizations $T(u,x) = a(u) + b(u)x$ the term power–normalization denotes transformations of the type

$$T(u,x) = \operatorname{sign}(x)\alpha(u)\,|x|^{\beta(u)}\,. \tag{3.1}$$

In [49] all possible limiting distributions of maxima of iid rvs under power normalization are derived. These limiting dfs constitute the class of p–max stable laws. More details about p–max stable laws are indicated in Section 3.3. P–max stable laws and their pertaining p–max–domains of attraction are thoroughly studied in [46] and [63] using the general result of maxima under monotone transformations in [49] which are also summarized in Section 2.5. Finally, these results are derived in [10] without relying on the results in [49] but using relations to the limit theory for maxima under linear normalization. It turns out that all results concerning limiting dfs and domains of attraction of maxima under power normalization can be derived from traditional EVT. The basic result is, that every p–max stable df $H$ can be represented in the form

$$H(x) = G(\log(x)), \quad x > 0 \tag{3.2}$$

or

$$H(x) = G\left(-\log(-x)\right), \quad x < 0, \tag{3.3}$$

where $G$ denotes an EVD. An analog result holds for the p–max–domains of attraction, that is, roughly speaking, a df $F$ belongs to the p–max–domain of attraction of $H(x) = G\left(\log(x)\right)$ if $F \circ \exp$ belongs to the max–domain of attraction of $G$. A similar result holds for $H$ as in (3.3). We will derive the analog result for limiting distributions of exceedances under linear and power normalization. An interesting but until recently rarely studied aspect of limit theory for maxima under power normalization is that it yields approximate distributions for maxima of certain distributions which have slowly varying distribution tails and are, therefore, not in the max–domain of attraction of any EVD. This result carries over to exceedances. Models for distributions with slowly varying tails are also studied in Chapter 4.

As already indicated in Theorem 2.3 limiting dfs of exceedances under certain transformations are closely related to the pertaining pot–stable dfs.

One may speak of p–pot stability if for a df $F$ and all $u \in \tilde{C}(F)$ there exist positive functions $\alpha$ and $\beta$ such that

$$F^{[u]}(\operatorname{sign}(x)\alpha(u)|x|^{\beta(u)}) = F(x)$$

for all $x \in \tilde{S}(F)$. In the following section we characterize all p–pot–stable dfs. Special attention is paid to the family of continuous p–pot–stable dfs which coincides with the parametric family of generalized log–Pareto dfs (GLPDs). In Chapter 4 we introduce a different form of a generalized Pareto distribution which can be deduced from the one used in this chapter by an additional scale shift.

An outline of this chapter is as follows. We deduce the class of p–pot–stable dfs in Section 3.1. In Section 3.2 we derive all limiting distributions of exceedances in the sense of the foregoing chapter, if the transformations $T$ are chosen as power–transformations as given in (3.1). Section 3.4 is concerned with the pertaining domains of attractions, in Section 3.5 we deal with a particular subset of these domains of attraction which is established by mixtures of regularly varying dfs. We discuss an extension of power normalization in Section 3.6. This chapter is concluded by some remarks concerning a different normalization, namely exponential transformations, in Section 3.7.

## 3.1 P–POT Stable Distributions

As in [46] we call a df $F_1$ a p–type of $F_2$, if $F_1(x) = F_2\left(\operatorname{sign}(x)\alpha|x|^{\beta}\right)$ for positive constants $\alpha$ and $\beta$. Check that $F_1$ is p–pot–stable if, and only if, $F_2$ is p–pot–stable.

In the subsequent lines we note standard versions of p–pot–stable dfs. The pertaining proof is given in Theorem 3.1 (Theorem 1 of [14]). The first family of p–pot–stable laws consists of generalized log–Pareto dfs

$$\tilde{L}_\gamma(x) = W_\gamma(\log(x)) = 1 - (1 + \gamma \log(x))^{-1/\gamma}, \quad \gamma \in \mathbb{R}, \tag{3.4}$$

for $1 < x$ if $\gamma \geq 0$, and $1 < x < \exp(1/|\gamma|)$ if $\gamma < 0$. Every df which is a p–type of $\tilde{L}_\gamma$ has its total mass on the positive half–line. Specifically, for $\gamma \to 0$ one gets

$$\tilde{L}_0(x) = 1 - 1/x, \quad x \geq 1,$$

which is the unique standard Pareto df (under power normalization).

All continuous p–pot–stable dfs with mass on the negative half–line are p–types of the negative log–Pareto df

$$V_\gamma(x) = W_\gamma(-\log(-x)) = 1 - (1 - \gamma \log(-x))^{-1/\gamma} \tag{3.5}$$

for $-1 < x < 0$ if $\gamma \geq 0$ and $-1 < x < -\exp(1/\gamma)$ if $\gamma < 0$. Notice that $V_0$ is the uniform distribution on the interval $[-1, 0]$.

The discrete p–pot–stable laws are given by

$$\Psi_{\gamma,\alpha}(x) = \Pi_{\gamma,\alpha}(\log(x))$$

and

$$\Upsilon_{\gamma,\alpha}(x) = \Pi_{\gamma,\alpha}(-\log(-x)),$$

cf. also (2.6).

**Theorem 3.1** *Let $F$ be a df such that for each $y \in \widetilde{C}(F)$ there exist functions $\alpha(\cdot) > 0$ and $\beta(\cdot) > 0$ such that*

$$F^{[y]}\left(sign(x)\alpha(y)|x|^{\beta(y)}\right) = F(x) \tag{3.6}$$

*for all $x \in \tilde{C}(F)$. Then,*

$$F(x) = W\left(\log(x)\right),$$

*or*

$$F(x) = W\left(-\log(-x)\right),$$

*where $W$ denotes a pot–stable df.*

Proof. Let $y \in \tilde{C}(F)$. Apparently, (3.6) is equivalent to

$$\frac{\bar{F}(\text{sign}(x)\alpha(y)|x|^{\beta(y)})}{\bar{F}(y)} = \bar{F}(x).$$

Let $F(0) > 0$. Then,

$$\frac{\bar{F}(0)}{\bar{F}(y)} = \bar{F}(0)$$

and $F(0) = 1$ because $0 < F(y) < 1$. Thus, we have $F(0) = 0$ or $F(0) = 1$ and, consequently, $F$ has all mass either on the positive or negative half–line.

(a) Let $F(0) = 0$ and, therefore, $F(x) = 0$ for all $x < 0$. It suffices to consider $x, y > 0$. Let $x > 0$, $x, y \in \tilde{C}(F)$. Then, (3.6) yields

$$\frac{\bar{F}\left(\alpha(y)x^{\beta(y)}\right)}{\bar{F}(y)} = \bar{F}(x).$$

It follows that

$$\frac{\bar{F}\left(\alpha(\exp(y))\exp(x)^{\beta(\exp(y))}\right)}{\bar{F}(\exp(y))} = \bar{F}(\exp(x))$$

for all $x, y \in \tilde{C}(F \circ \exp)$. Furthermore,

$$\frac{\bar{F}\left(\alpha(\exp(y))\exp(x)^{\beta(\exp(y))}\right)}{\bar{F}(\exp(y))} = \bar{F}(\exp(x))$$
$$\Leftrightarrow \quad \frac{\bar{F}\left(\exp\left(\log(\alpha(\exp(y))) + \beta(\exp(y))x\right)\right)}{\bar{F}\left(\exp(y)\right)} = \bar{F}\left(\exp(x)\right).$$

Put $F^* := F(\exp(\cdot))$. The above computations yield

$$\frac{1 - F^*\left(\tilde{\alpha}(y) + \tilde{\beta}(y)x\right)}{1 - F^*(y)} = 1 - F^*(x)$$

with $\tilde{\alpha}(y) = \log\left(\alpha(\exp(y))\right)$ and $\tilde{\beta}(y) = \beta(\exp(y))$. Consequently, $F^* = W$ for some pot–stable df $W$ and $F(\cdot) = W(\log(\cdot))$.

(b) Next assume that $F(0) = 1$. Let $x \leq 0$ and $x, y \in \tilde{C}(F)$. Then similar arguments as in part (a) yield that (3.6) is equivalent to

$$\frac{\bar{F}_*(\tilde{\alpha}(y) + \tilde{\beta}(y)x)}{\bar{F}_*(y)} = \bar{F}_*(x)$$

with $F_*(x) := F(-\exp(-x))$ for $x < 0$ and $F_*(x) = 1$ for $x \geq 0$, where $\tilde{\alpha}(y)$ can be chosen as $\alpha\left(-\exp(-y)\right)$ and $\tilde{\beta}(y) = \beta\left(-\exp(-y)\right)$. Thus, $F_*$ is a pot–stable df $W$ and $F(x) = W\left(-\log(-x)\right)$. □

The class of pot–stable dfs $W$ is well known due to [3] and [53]. If $W$ is continuous, then $W$ is a GPD

$$W_{\gamma,\beta,\mu} := 1 - \left(1 + \gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}$$

thus a linear type of the GPD given in (2.5) with additional scale and location parameter $\sigma > 0$ and $\mu$. In addition there are two classes of discrete pot–stable dfs and thus, the mentioned p–pot–stable dfs.

In Section 3.2 we identify the p–pot–stable dfs as the only possible limiting dfs of exceedances under power–normalization.

## 3.2 Limiting Distributions

Next we study limiting dfs $L$ of exceedances above high thresholds under power normalization, that is, for some df $F$ we have

$$F^{[u]}\left(\operatorname{sign}(x)\alpha(u)\left|x\right|^{\beta(u)}\right) \xrightarrow[u\to\omega(F)]{} L(x). \tag{3.7}$$

Notice that (3.7) is equivalent to

$$\bar{F}\left(\operatorname{sign}(x)\alpha(u)\,|x|^{\beta(u)}\right)\Big/\bar{F}(u) \xrightarrow[u\to\omega(F)]{} \bar{L}(x), \tag{3.8}$$

where, again, $\bar{F} := 1 - F$ is the survivor function of $F$.

We use two approaches for the derivation of our results. The first one uses results from traditional extreme value theory, the second one utilizes the results from Chapter 2.

### 3.2.1 Derivation from Traditional EVT

Given a df $F$ we define auxiliary dfs $F^*$ and $F_*$ by

$$F^*(x) = \frac{F(\exp(x)) - F(0)}{1 - F(0)}, \quad x \in \mathbb{R},$$

if $\omega(F) > 0$, and

$$F_*(x) = F(-\exp(-x)), \quad x \in \mathbb{R},$$

if $\omega(F) \le 0$. We start with a technical lemma concerning $F^*$ and $F_*$.

**Lemma 3.2** *Let $L$ be a non–degenerate limiting df in (3.7) for some df $F$. Then,*

*(i) there are functions $a(\cdot)$ and $b(\cdot) > 0$ such that*

$$\bar{F}^*(a(u) + b(u)x)\big/\bar{F}^*(u) \xrightarrow[u\to\omega(F^*)]{} \bar{L}^*(x)$$

*if $\omega(F) > 0$, and*

$$\bar{F}_*(a(u) + b(u)x)\big/\bar{F}_*(u) \xrightarrow[u\to\omega(F_*)]{} \bar{L}_*(x)$$

*if $\omega(F) \le 0$.*

*(ii) $L^*$ and, respectively, $L_*$ are pot–stable dfs.*

Proof. First we prove that the total mass of $L$ is either concentrated on the positive or negative half–line and, therefore,

$$L(\exp(x)) = L^*(x), \quad x \in \mathbb{R},$$

if $\omega(L) > 0$, and

$$L(-\exp(-x)) = L_*(x), \quad x \in \mathbb{R},$$

if $\omega(L) < 0$.

The remainder of the proof is merely outlined for $\omega(F) > 0$. The case of $\omega(F) \leq 0$ follows by similar arguments. If $x < 0$, we have

$$F^{[u]}\left(\mathrm{sign}(x)\alpha(u)\,|x|^{\beta(u)}\right) \leq F^{[u]}(0) \xrightarrow[u\to\omega(F)]{} 0$$

because $\omega(F) > 0$. This implies $L(x) = 0$ for all $x \leq 0$.

Next consider $x > 0$. From (3.8) one gets

$$\bar{F}\left(\alpha(u)x^{\beta(u)}\right)/\bar{F}(u) \xrightarrow[u\to\omega(F)]{} \bar{L}(x).$$

Moreover, by straightforward computations,

$$\frac{\bar{F}\left(\exp\left(a(u)+b(u)x\right)\right)}{\bar{F}(\exp(u))} \xrightarrow[\exp(u)\to\omega(F)]{} \bar{L}(\exp(x)) \tag{3.9}$$

for all $x \in \mathbb{R}$ with $a(u) = \log(\alpha\left(\exp(u)\right)) \in \mathbb{R}$ and $b(u) = \beta\left(\exp(u)\right) > 0$. Therefore,

$$\frac{\bar{F}^*(a(u)+b(u)x)}{\bar{F}^*(u)} \xrightarrow[u\to\omega(F^*)]{} \bar{L}(\exp(x)) = \bar{L}^*(x), \tag{3.10}$$

and assertion (i) is verified. This also implies (ii) because limiting dfs under linear normalization are pot–stable. □

Lemma 3.2 now offers the prerequisites to prove the main result of this section.

**Theorem 3.3** *Every non–degenerate limiting df $L$ in (3.7) is p–pot–stable.*

Proof. Again, we merely prove the case $\omega(F) > 0$. From Lemma 3.2(ii) we know that $L^*$ is pot–stable. Thus, there are $a(y) \in \mathbb{R}$ and $b(y) > 0$ such that

$$\bar{L}^*(a(y)+b(y)x)/\bar{L}^*(u) = \bar{L}^*(x)$$

for all $x, y \in \tilde{C}(L^*)$. This yields for $x, y > 0$,

$$\bar{L}^*(a(y)+b(y)\log(x))/\bar{L}^*(\log(y)) = \bar{L}^*(\log(x)).$$

Choosing $\alpha(y)$ and $\beta(y)$ as in the proof of Lemma 3.2 one gets from the equation $\bar{L}^*(a(y)+b(y)\log(x)) = \bar{L}^*\left(\log\left(\alpha(y)x^{\beta(y)}\right)\right)$ that

$$\bar{L}\left(\alpha(y)x^{\beta(y)}\right)/\bar{L}(y) = \bar{L}(x)$$

for all $x, y \in \tilde{C}\left(L^* \circ \log\right)$. Notice that $L(x) = L^*(\log(x))$ if $x > 0$, and $L(x) = 0$ if $x \leq 0$. This yields the p–pot stability of $L$ according to the preceding equation. □

It is evident that the converse implication is also true, that is, every p–pot–stable df $L$ is a limiting df in (3.7) by choosing $F = L$. Summarizing the previous results we get that $L$ is a limiting df of exceedances pertaining to a df $F$ under power-normalization if, and only if, $L^*$ (if $\omega(F) > 0$) or $L_*$ (if $\omega(F) \leq 0$) are pot–stable.

### 3.2.2 Derivation Using the General Result

In the case of continuous, strictly increasing limit dfs, the results of the foregoing lines can also be obtained by using the general result of Chapter 2. We merely consider continuous limiting dfs. We have

$$\frac{\bar{F}\left(T(u,x)\right)}{\bar{F}(u)} \xrightarrow[u\to\omega(F)]{} \bar{L}(x), \quad x \in S(L),$$

where

$$T(u,x) = \text{sign}(x)\alpha(u)\,|x|^{\beta(u)}$$

and $\alpha$ and $\beta$ are functions with values in the positive real numbers. We have

$$T^{(-1)}(u,x) = \text{sign}(x)\left(\frac{|x|}{\alpha(u)}\right)^{1/\beta(u)}$$

and, therefore,

$$S_{u,y}(x) = \text{sign}(x)\left(\frac{\alpha\left(T(u,y)\right)}{\alpha(u)}\right)^{1/\beta(u)} |x|^{\beta(T(u,y))/\beta(u)}.$$

In the following lines we will use arguments similar to those in Section 2.6. As in Section 2.6 we know from Lemma 2.4 that for each $y \in \tilde{S}(L)$ there exists a sequence $(u_n)$, $u_n \xrightarrow[n\to\infty]{} \omega(F)$ such that $S_{u_n,y}(\cdot)$ converges for all $x \in \tilde{S}(L)$. Obviously convergence of $S_{u_n,y}(x)$ is equivalent to

$$\left(\frac{\alpha\left(T(u_n,y)\right)}{\alpha(u_n)}\right)^{1/\beta(u_n)} \xrightarrow[n\to\infty]{} A(y)$$

and

$$\beta\left(T(u_n,y)\right)/\beta(u_n) \xrightarrow[n\to\infty]{} B(y)$$

since we may assume, without loss of generality, that $1 \in \tilde{S}(L)$. Thus

$$S_{u_n,y}(x) \xrightarrow[n\to\infty]{} g_y(x) = \text{sign}(x)A(y)|x|^{B(y)}, \quad \text{for all } x \in \tilde{S}(L)$$

and also by Lemma 2.4

$$\bar{L}\left(\text{sign}(x)A(y)|x|^{B(y)}\right) = \bar{L}(x)\bar{L}(y), \quad \text{for all } x \in \tilde{S}(L).$$

In the next step we prove that $g_y$ is strictly increasing for all $y \in \tilde{S}(L)$. First notice that $A(y), B(y) \geq 0$ for all $y \in \tilde{S}(L)$ because they are limits of non–negative sequences. If $A(y) = 0$ then $g_y(x) = 0$ for all $x \in \tilde{S}(L)$ and, therefore,

$$\bar{L}(0) = \bar{L}(y)\bar{L}(x)$$

for all $x \in \tilde{S}(L)$ and, therefore, $\bar{L}$ is constant. Correspondingly, $B(y) = 0$ yields

$$\bar{L}(\mathrm{sign}(x)A(y)) = \bar{L}(y)\bar{L}(x)$$

for all $x \in \tilde{S}(L)\setminus\{0\}$. Therefore we get $A(y), B(y) > 0$ for all $y \in \tilde{S}(L)$ and, thus, $g_y$ is strictly increasing.

Now we merely consider those limiting dfs $L$ which are strictly increasing on $\tilde{S}(L)$. It remains to determine the concrete form of the functions $A$ and $B$ to derive $g$ and the strictly increasing function $h$ which then gives $L$. First note that $\tilde{S}(L)$ is necessarily either a subset of the positive or negative real line. This can be easily seen using arguments as in the proof of Lemma 3.2. Therefore, we have to determine $g(y, x)$ for the cases $x, y > 0$ and $x, y < 0$. We only consider the case where $\tilde{S}(L)$ is contained in the positive real numbers. Therefore, let $x, y > 0$. First note that, since $g(y, x) = h^{-1}(h(x) + h(y))$, we have that $g(y, x) = g(x, y)$. Secondly we may change $A$ and $B$ by multiplicative constants because this only leads to a power–transformation of the limiting distribution. Since we have two free parameters we may assume $1 \in \tilde{S}(L)$ and $\exp(1) \in \tilde{S}(L)$. This yields

$$A(y) = g(y, 1) = g(1, y) = A(1)y^{B(1)} \tag{3.11}$$

and furthermore for $x, y_1, y_2 > 0$

$$g(x, y_1y_2) = A(1)x^{B(1)}(y_1y_2)^{B(x)} = \left(x^{B(1)}A(1)\right)^{-1} g(y_1, x)g(y_2, x). \tag{3.12}$$

From the commutativity of $g$ (cf. (2.19)) we also get

$$g(x, y_1y_2) = g(y_1y_2, x) = A(y_1y_2)x^{B(y_1y_2)}. \tag{3.13}$$

Combining equations (3.11) to (3.13), plugging in $\exp(1)$ for $x$, and again using the commutativity of $g$ one gets

$$\begin{aligned} &\left(x^{B(1)}A(1)\right)^{-1} g(y_1, x)g(y_2, x) = A(y_1y_2)x^{B(y_1y_2)} \\ \Leftrightarrow\quad &B(y_1) + B(y_2) - B(1) = B(y_1y_2). \end{aligned} \tag{3.14}$$

Since we know that, in addition, $B$ is positive on $\tilde{S}(L)$ the only solution of equation (3.14) is

$$B(y) = a + b\log(y), \quad a, b > 0, \tag{3.15}$$

which gives

$$g(y,x) = \exp\left(\log\left(A(1)\right) + a\log(y) + a\log(x) + b\log(x)\log(y)\right).$$

From this we get the following functional equation for $h$ (and $x, y > 0$)

$$h(x) + h(y) = h\left(\exp\left(\log\left(A(1)\right) + a\log(y) + a\log(x) + b\log(x)\log(y)\right)\right),$$

with $\tilde{h}(x) := h\left(\exp(x)\right)$. Plugging in $y = \exp(y')$ and $x = \exp(x')$ this yields

$$\tilde{h}(x') + \tilde{h}(y') = \tilde{h}\left(c + ax' + ay' + bx'y'\right)$$

for some constant $c \in \mathbb{R}$. Thus, $\tilde{h}$ is some function as derived in Section 2.6.

If $b = 0$ we have $h(x) = \alpha\log(\beta x)$ for $\alpha \in \mathbb{R}, \beta > 0$. If $b > 0$ the solution of this functional equation is given by

$$h(x) = a\log\left(b\log(cx)\right), \quad a \in \mathbb{R}, b, c > 0.$$

Now it is an immediate consequence of this equation that $L$ is a log–Pareto df as given in (3.4). Similar arguments yield the corresponding result for negative log–Pareto dfs if the support of $L$ is contained in the negative half–line.

## 3.3 Relations to P–Max Stable Laws

We start with a representation of log–Pareto dfs by means of p–max–stable dfs. Recall that a df $F$ is p–max–stable if there exist sequences $\alpha_n$, $\beta_n > 0$ such that

$$F^n(\mathrm{sign}(x)\alpha_n|x|^{\beta_n}) = F(x), \quad x \in \mathbb{R},$$

and all positive integers $n$, cf. [49] or [27], Section 2.6.

For the special p–max stable df

$$H_{1,\gamma}(x) = \exp(-(\log x)^{-\gamma}), \quad x \geq 1,$$

with $\gamma > 0$, define

$$\begin{aligned} F_\gamma(x) &= 1 + \log H_{1,\gamma}(x) \\ &= 1 - (\log x)^{-\gamma}, \quad x \geq \exp(1), \end{aligned} \tag{3.16}$$

which is a log–Pareto df with shape parameter $1/\gamma$.

In analogy to (3.16), the whole family of distributions in (3.4) can be deduced from p–max stable laws $H_{i,\gamma,\beta,\sigma}(x) = H_{i,\gamma}((x/\sigma)^\beta)$, for $i = 1, 2, 3$ (for the definition of general p–max stable laws see [27], 2nd edition, Section 2.6). This relationship makes the theory of p–max dfs applicable to log–Pareto dfs to some extent (cf. also Theorem 2.12).

## 3.4 Domains of Attraction of P–POT Stable Distributions

Within the linear framework, a df $F$ belongs to the pot–domain of attraction of a df $W$, denoted by $F \in \mathcal{D}_{pot}(W)$, if there are functions $a(\cdot)$ and $b(\cdot) > 0$ such that

$$F^{[u]}(a(u) + b(u)x) \xrightarrow[u \to \omega(F)]{} W(x).$$

Correspondingly, if relation (3.7) holds for dfs $F$ and $L$, then $F$ belongs to the p–pot–domain of attraction of $L$ denoted by $F \in \mathcal{D}_{p-pot}(L)$. Note that this refers to the second notion of a domain of attraction in Definition 2.10.

### 3.4.1 A General Result

We characterize p–pot–domains of attraction of a p–pot–stable df $L$ by means of pot–domains of attraction of $L^*$ or $L_*$ which are pot–stable according to Theorem 3.3. As a direct consequence of Lemma 3.2(i) one gets the following theorem.

**Theorem 3.4** *For the p–pot–domain of attraction $\mathcal{D}_{p-pot}(L)$ of a p–pot stable law $L$ we have*

$$\mathcal{D}_{p-pot}(L) = \{F : F^* \in \mathcal{D}_{pot}(L^*)\},$$

*if $\omega(L) > 0$, and*

$$\mathcal{D}_{p-pot}(L) = \{F : F_* \in \mathcal{D}_{pot}(L_*)\}$$

*if $\omega(L) \leq 0$.*

P–pot–domains of attraction of continuous p–pot–stable laws can be deduced from p–max–domains of attractions, given for example in [27], due to the identity of pot– and max–domains of attraction in the linear setup. The domains of attraction of the discrete p–pot–stable laws have no counterpart in the framework of max-stable dfs. Their domains of attraction can be derived from the preceding theorem and Section 3 of [3].

### 3.4.2 Special Conditions in the Log-Pareto Case

We make use of a parametrization of log–Pareto dfs which is different from that in (3.4). Let

$$\hat{L}_\gamma(x) = 1 - (\log(x))^{-1/\gamma}, \quad \gamma > 0, \, x \geq \exp(1). \tag{3.17}$$

It is apparent that $\hat{L}_\gamma$ is a p–type of $\tilde{L}_\gamma$ in (3.4). Such dfs can be regarded as prototypes of p–pot–stable dfs with slowly varying tails.

**Corollary 3.5** *We have $F \in \mathcal{D}_{p-pot}(\hat{L}_\gamma)$ if, and only if, there is a slowly varying function $U$ and some $a > 1$ such that*

$$F(x) = 1 - (\log(x))^{-1/\gamma}\, U(\log(x)), \quad x > a. \tag{3.18}$$

Proof. This is a direct consequence of Theorem 3.4. We have for $x > 0$ that $\bar{F}(x) = \bar{F}(0)\bar{F}^*(\log(x))$ for the df $F^*$ which is in the pot-domain of attraction of a Pareto df and, therefore, $\bar{F}^*$ is regularly varying at infinity. □

The p–pot–domain of attraction of a log–Pareto df $\hat{L}_\gamma$ can as well be characterized by a property with is deduced from regular variation which characterizes the pot–domain of attraction of Pareto dfs under linear transformation. Observe that

$$\bar{\hat{L}}_\gamma\left(x^{\log(y)}\right)/\bar{\hat{L}}_\gamma(y) = (\log(x))^{-1/\gamma}$$

which is the p–pot stability of $\hat{L}_\gamma$. For the domain of attraction this relation holds in the limit and, furthermore, this yields a characterization of the domain attraction.

**Corollary 3.6** *We have $F \in \mathcal{D}_{p-pot}(\hat{L}_\gamma)$ if, and only if,*

$$\bar{F}\left(x^{\log(u)}\right)/\bar{F}(u) \xrightarrow[u\to\infty]{} (\log(x))^{-1/\gamma}, \quad x > 1. \tag{3.19}$$

Proof. If $F \in \mathcal{D}_{p-pot}(\hat{L}_\gamma)$ we have

$$\bar{F}(x) = (\log(x))^{-1/\gamma}\, U\left(\log(x)\right), \quad x > a$$

for some slowly varying function $U$ and some $a > 1$. Therefore,

$$\begin{aligned}
\frac{\bar{F}\left(x^{\log(u)}\right)}{\bar{F}(u)} &= \frac{\left(\log\left(x^{\log(u)}\right)\right)^{-1/\gamma} U\left(\log\left(x^{\log(u)}\right)\right)}{(\log(u))^{-1/\gamma}\, U\left(\log(u)\right)} \\
&= (\log(x))^{-1/\gamma}\frac{U\left(\log(u)\log(x)\right)}{U\left(\log(u)\right)} \\
&\to (\log(x))^{-1/\gamma} \quad \text{for } u \to \infty.
\end{aligned}$$

Conversely, let

$$\lim_{u\to\infty} \bar{F}(x^{\log(u)})/\bar{F}(u) = (\log(x))^{-1/\gamma}$$

for $x > 1$. It follows that

$$\lim_{u\to\infty} \bar{F}\left(\exp(uy)\right)/\bar{F}\left(\exp(u)\right) = y^{-1/\gamma}$$

for all $y > 0$. Thus, $F^* \in \mathcal{D}_{pot}\left(W_\gamma\right)$ and, consequently, $F \in \mathcal{D}_{p-pot}\left(\hat{L}_\gamma\right)$. □

We include a result about the invariance of $\mathcal{D}_{p-pot}\left(\hat{L}_\gamma\right)$ under shift and power transformations.

**Corollary 3.7** *The following equivalences hold true for $\mu \in \mathbb{R}$ and $\beta$, $\sigma > 0$:*

$$F(\cdot) \in \mathcal{D}_{p-pot}\left(\hat{L}_\gamma\right) \Leftrightarrow F\left((\cdot - \mu)\right) \in \mathcal{D}_{p-pot}\left(\hat{L}_\gamma\right),$$

*and*

$$F(\cdot) \in \mathcal{D}_{p-pot}\left(\hat{L}_\gamma\right) \Leftrightarrow F\left(\sigma(\cdot)^\beta\right) \in \mathcal{D}_{p-pot}\left(\hat{L}_\gamma\right).$$

PROOF. We only prove part (i) because (ii) concerns a power–normalization and, therefore, it is straightforward. Putting $F_\mu(x) = F(x-\mu)$ for $F \in \mathcal{D}_{p-pot}\left(\hat{L}_\gamma\right)$ we get

$$\begin{aligned}
\frac{1-(F_\mu)^*(tx)}{1-(F_\mu)^*(t)} &= \frac{\bar{F}(\exp(tx)-\mu)}{\bar{F}(\exp(t)-\mu)} \\
&= \frac{\bar{F}\left(\frac{\exp(tx)-\mu}{\exp(tx)}\exp(tx)\right)}{\bar{F}\left(\frac{\exp(t)-\mu}{\exp(t)}\exp(t)\right)} \\
&= \frac{\bar{F}\left(\exp\left(tx+\log\left(\frac{\exp(tx)-\mu}{\exp(tx)}\right)\right)\right)}{\bar{F}\left(\exp\left(t+\log\left(\frac{\exp(t)-\mu}{\exp(t)}\right)\right)\right)} \\
&= \frac{F^*(tx+a_t)}{F^*(t+b_t)}
\end{aligned}$$

with

$$a_t = \log\left(\frac{\exp(tx)-\mu}{\exp(tx)}\right) \quad \text{and} \quad b_t = \log\left(\frac{\exp(t)-\mu}{\exp(t)}\right).$$

Obviously $a_t \to 0$ and $b_t \to 0$. Since $\bar{F}^*$ is regularly varying at infinity the convergence

$$\frac{\bar{F}^*(tx)}{\bar{F}^*(t)} \xrightarrow[t\to\infty]{} x^{-1/\gamma}$$

holds uniformly, hence

$$\frac{\bar{F}^*(tx+a_t)}{\bar{F}^*(t+b_t)} \xrightarrow[t\to\infty]{} x^{-1/\gamma}$$

and, thus, $F(\cdot-\mu) \in \mathcal{D}_{p-pot}$. □

The previous result yields that

$$\mathcal{D}_{p-pot}(L) = \mathcal{D}_{p-pot}(\hat{L}_\gamma)$$

for all p-types $L$ of $\hat{L}_\gamma$. It is easily seen that this result is valid for a p–pot–domain of attraction of an arbitrary p–pot–stable law. The result concerning location shifts cannot be extended to p–pot–stable laws with finite right endpoints.

## 3.5 Mixtures of Regularly Varying Distribution Functions

We start with a result in [56] concerning a relation of log–Pareto dfs

$$\hat{L}_\gamma(x) = 1 - (\log(x))^{-1/\gamma}$$

and Pareto dfs

$$\widetilde{W}_{\gamma,\sigma}(x) = 1 - (x/\sigma)^{-1/\gamma}$$

which corresponds to the example for super–heavy tailed dfs already addressed in the introduction. Log–Pareto dfs can be represented as mixtures of certain Pareto dfs with respect to gamma densities. We have

$$\hat{L}_\gamma(x) = \int_0^\infty \widetilde{W}_{1/z,e}(x) h_{1/\gamma}(z)dz \tag{3.20}$$

where $h_\alpha$ is the gamma density

$$h_\alpha(x) = \frac{1}{\Gamma(\alpha)} \exp(-x)x^{\alpha-1}, \tag{3.21}$$

with $e = \exp(1)$.

We prove that this result can be extended to dfs in the domains of attraction of log–Pareto and Pareto dfs under power and, respectively, linear normalization.

**Theorem 3.8** *The following properties hold for the p–pot domain of attraction of a log–Pareto df* $\hat{L}_\gamma$*:*

*(i) Let* $F \in \mathcal{D}_{p-pot}(\hat{L}_\gamma)$ *for some* $\gamma > 0$*. Then there is a family of dfs* $G_z$*, with* $G_z \in \mathcal{D}_{pot}(\widetilde{W}_{1/z})$*, such that*

$$F(x) = \int_0^\infty G_z(x)p(z)dz,$$

*where p is a density which is ultimately monotone (monotone on* $[x_0, \infty)$ *for some* $x_0 > 0$*) and regularly varying at zero with index* $1/\gamma - 1$*.*

*(ii) Let* $G_z$ *be a family of dfs with* $G_z \in \mathcal{D}(W_{1/z})$ *with representation*

$$G_z(x) = 1 - x^{-z} U(\log(x)), \quad x > a_1,$$

*for some slowly varying function U and some* $a_1 > 0$*. Then the mixture*

$$F(x) := \int_0^\infty G_z(x)p(z)dz,$$

*where $p$ is a density as in (i), has the representation*

$$F(x) = 1 - (\log(x))^{-1/\gamma} V(\log(x)), \quad x > a_2$$

*for some slowly varying function $V$ and some $a_2 > 1$, thus, $F \in \mathcal{D}_{p-pot}(\hat{L}_\gamma)$.*

PROOF. To prove (i) observe that the gamma density $h_{1/\gamma}$ in (3.21) satisfies the conditions imposed on $p$. Therefore, (i) is a direct consequence of (3.20) and Corollary 3.5. Therefore the statement is still true with $p$ replaced by $h_{1/\gamma}$.

Assertion (ii) is a modification and extension of Lemma 1 in [45] by Meerschaert and Scheffler. Notice that

$$1 - F(x) = \int_0^\infty e^{-z\log(x)} p(z) dz U(\log(x)).$$

The integral is now a function $\hat{p}(\log(\cdot))$ where $\hat{p}$ denotes the Laplace transform of $p$. Since $p$ is assumed to be ultimately monotone and regularly varying at zero with index $1/\gamma - 1$ one can apply Theorem 4 on page 446 of [33] getting

$$\int_0^\infty e^{-z\log(x)} p(z) dz = \log(x)^{-1/\gamma} \widetilde{V}(\log(x)), \quad x > a_3$$

for some slowly varying function $\widetilde{V}$ and $a_3 > 1$. Now $V(x) := U(x)\widetilde{V}(x)$ is again slowly varying which completes the proof. □

## 3.6 The Iterated Case

The concepts of linear and power normalization can be generalized as follows. For $k = 0, 1, 2, \ldots$ let $\mathcal{T}^{(k)}$ be a family of monotone transformations $T^{(k)}$ such that $T^{(k)}(u, 0) = 0$ for all $u$ and $k \in \mathbb{N}$ and there is some $T^{(k-1)} \in \mathcal{T}^{(k-1)}$ such that

$$T^{(k)}(u, x) = \operatorname{sign}(x) \exp\left(T^{(k-1)}(\operatorname{sign}(u)\log(|u|), \log(|x|))\right). \tag{3.22}$$

A df $L$ is called $k$–th order $\mathcal{T}$–pot–stable if there is a $T^{(k)} \in \mathcal{T}^{(k)}$ such that

$$\bar{L}\left(T^{(k)}(y, x)\right) = \bar{L}(x)\bar{L}(y) \tag{3.23}$$

for all $x, y \in \tilde{C}(L)$. We call $L$ $\mathcal{T}$–pot–stable if the latter relation holds for $k = 0$.

We impose the following condition for a family of monotone transformations $\mathcal{T} = \mathcal{T}^{(0)}$.

**Condition 3.9** *For a df $F$ there exists $T^{(0)} \in \mathcal{T}^{(0)}$ and a non–degenerate df $L$ such* *that*

$$F^{[u]}(T^{(0)}(u,x)) \xrightarrow[u \to \omega(F)]{} L(x) \tag{3.24}$$

*for all $x$ if, and only if, $L$ is $\mathcal{T}$–pot–stable.*

The following theorem extends the results of previous sections to a more general setting.

**Theorem 3.10** *Let $\mathcal{T}^{(0)}$ be a family of monotone transformations. Define $\mathcal{T}^{(k)}$, $k \in \mathbb{N}$, as above.*

*(i) A df $L$ is $k$-th order $\mathcal{T}$–pot–stable if, and only if,*

$$L(x) \quad = \quad \begin{cases} K(\log(x)) & x > 0, \\ & \text{if} \\ 0 & x \leq 0, \end{cases}$$

*or*

$$L(x) \quad = \quad \begin{cases} 1 & x > 0, \\ & \text{if} \\ K(-\log(-x)) & x \leq 0, \end{cases}$$

*for a $(k-1)$–th order $\mathcal{T}$–Pot–stable df $K$.*

*(ii) In addition, assume that Condition 3.9 holds. Let $F$ and $L$ be non–degenerate dfs such that*

$$F^{[u]}(T^{(k)}(u,x)) \xrightarrow[u \to \omega(F)]{} L(x) \tag{3.25}$$

*for all $x$ and a monotone transformations $T^{(k)} \in \mathcal{T}^{(k)}$. Then, $L$ is $k$–th order $\mathcal{T}$–pot–stable. Furthermore, (3.25) holds if, and only if, this relation is also true for some $T^{(k-1)} \in \mathcal{T}^{(k-1)}$ and $F^*$ and $L^*$ if $\omega(F) > 0$, and $F_*$ and $L_*$ if $\omega(F) \leq 0$.*

Proof. (i) Observe that (3.23) is equivalent to

$$\bar{L}\left(T^{(k)}(y,x)\right)/\bar{L}(y) = \bar{L}(x) \tag{3.26}$$

if $x, y \in \tilde{C}(L)$. Let $L(0) > 0$. This yields

$$\bar{L}(0)/\bar{L}(y) = \bar{L}(0),$$

and, therefore, $\bar{L}(0) = 0$ and consequently $L(0) = 1$. Thus, $L$ has all mass either on the positive or negative half–line.

We only consider $\omega(L) > 0$, the proof is similar for $\omega(L) \leq 0$. We have $L(x) = 0$ for $x < 0$. Now let $x > 0$, one gets from (3.26)

$$\frac{\bar{L}\left(\exp\left(T^{(k-1)}\left(\log(y), \log(x)\right)\right)\right)}{\bar{L}(y)} = \bar{L}(x), \quad x, y \in \tilde{C}\left(L\right).$$

Now straightforward computations yield

$$\frac{\bar{L}^*\left(T^{(k-1)}(y, x)\right)}{\bar{L}^*(y)} = \bar{L}^*(x), \quad x, y \in \tilde{C}\left(L^*\right).$$

Thus, $L^* = L \circ \exp$ is $(k-1)$–th order $\mathcal{T}$–pot–stable and we get the stated representation of $L$. The converse implication, that the stated representation yields a $k$–th order $\mathcal{T}$–pot–stable df, can be verified by straightforward computations.
(ii) The proof can be carried out in analogy to the proofs of Lemma 3.2 and Theorem 3.3. □

The results in the previous section fit into this context. Let $\mathcal{T}^{(0)}$ be the family of linear transformations $T^{(0)}_{\mu,\sigma}(u, x) = \mu(u) + \sigma(u)x$ with $\mu \in \mathbb{R}$ and $\sigma > 0$. Apparently Condition 3.9 holds for this choice of $\mathcal{T}^{(0)}$. It follows that $\mathcal{T}^{(1)}$ is the family of power–normalizations. From relation (3.22) we may define a second order power–normalization by

$$T^{(2)}(u, x) = \operatorname{sign}(x) \exp\Big(\operatorname{sign}\left(\log(|x|)\right) \alpha(u) \left|\log\left(|x|\right)\right|^{\beta(u)}\Big) \quad \alpha, \beta > 0,$$

and by iterating this procedure we derive a $k$-th order power–normalization by $T^{(k)}$. The results about the relation of domains of attraction of linear and power transformations analogously apply to the relation of domains of attraction under $k$–th and $(k+1)$-th order iterated power–normalizations.

A special case of the previous considerations concerns dfs with iterated heavy tails. Let again $T^{(0)}(x) = T^{(0)}_{\mu,\sigma}(u, x) = \mu(u) + \sigma(u)x$. We only consider continuous dfs with $\omega(F) = \infty$ which are in some sense the most heavy tailed dfs in each class of $k$-th order $\mathcal{T}$–pot–stable dfs. It is well known that in the case of heavy tailed dfs relation (3.24) holds with $T^{(0)}(u, x) = ux$ and $L = W_\gamma$ for some $\gamma > 0$, that is that $\bar{F}$ is regularly varying at infinity. Using (3.22) we define for $x > 0$

$$T^{(1)}(u, x) = \exp\Big(T^{(0)}\left(\log(u), \log(x)\right)\Big) = x^{\log(u)}.$$

Then (3.24) becomes (3.19) which is a necessary and sufficient condition for a df $F$ to belong to the p–pot–domain of attraction of a log–Pareto df $L_\gamma$ and is, therefore, super–heavy tailed. A df $F$ is said to be third order iterated heavy tailed if, and only if,

$$\frac{\bar{F}\left(T^{(2)}(u, x)\right)}{\bar{F}(u)} \xrightarrow[u \to \infty]{} \left(\log\log(x)\right)^{-1/\gamma}, \quad x > \exp(1)$$

with

$$T^{(2)}(u,x) = \exp\left(\log(x)^{\log\log(u)}\right).$$

Iterating this procedure one may use

$$T^{(k)}(u,x) = \exp\left(T^{(k-1)}\left(\log(u), \log(x)\right)\right)$$

to define $k$–th order iterated heavy tailed dfs. Apparently the results about mixtures of certain heavy tailed dfs can be extended to iterated heavy tailed dfs. This is also valid for representations in terms of regularly varying functions as given in Corollary 3.6.

## 3.7 Exponential Normalization

In this section we will study a transformation which is different from linear and power–normalizations. It turns out that the classes of GPDs and GLPDs which are the (continuous) limiting distributions of exceedances under linear and power–normalization, respectively, can also appear as limiting distributions under exponential normalization. Until now we have only studied classes of transformations which form a group and as a consequence $S_{u,y}$ and, therefore, $g$ have been of the same type as $T$, that is, they are also members of the group. In this section we will study a normalization for which such a relation is not valid.

We will consider continuous limiting distributions of exceedances under exponential normalizations

$$\frac{\bar{F}(T(u,x))}{\bar{F}(u)} \xrightarrow[u \to \omega(F)]{} \bar{L}(x), \quad x \in \tilde{C}(L) \tag{3.27}$$

with

$$T(u,x) = a(u)\exp\left(b(u)x\right),$$

$a(u)$, $b(u) > 0$. We use the same strategy as applied in Section 3.2.2. We have

$$T^{(-1)}(u,x) = \frac{1}{b(u)}\log\left(x/a(u)\right), \quad x > 0$$

and, therefore,

$$S_{u,y}(x) = \frac{1}{b(u)}\log\left(a\left(T(u,y)\right)/a(u)\right) + \frac{b\left(T(u,y)\right)}{b(u)}x, \quad x \in \mathbb{R}.$$

We know from Lemma 2.4 that $S_{u,y}(x)$ converges to some $g(y,x)$ for $u \to \omega(F)$ and, therefore, $g$ has the representation

$$g(y,x) = A(y) + B(y)x, \quad A(y) \in \mathbb{R},\ B(y) > 0.$$

To see that $B(y) > 0$ holds, repeat the arguments in Section 2.6. Hence $L$ has to be pot–stable under linear transformations and we know from [3] that $L$ is a GPD $W_\gamma$.

For statistical purposes the upper tail of a df $F$ which is in the pot–domain of attraction of some GPD $W_\gamma$ under exponential transformations can be approximated by a p–type of a GLPD $\hat{L}_\gamma$ as introduced in (3.4) since we have for $u$ sufficiently close to $\omega(F)$ (see also (2.27))

$$F^{[u]}(x) \approx W_\gamma\left(T^{(-1)}(u,x)\right) = \hat{L}_\gamma\left(\beta x^\alpha\right), \quad x > 0.$$

The domains of attraction under exponential normalizations are closely related to a certain subclass of the domains of attraction under power–normalization. First note that we necessarily have $\omega(F) > 0$ in (3.27) since, otherwise, we have $\bar{F}(0) = 0$ and, thus,

$$\frac{\bar{F}\left(a(u)\exp(b(u)x)\right)}{\bar{F}(u)} \leq \frac{\bar{F}(0)}{\bar{F}(u)} = 0$$

for all $x \in \mathbb{R}$. The domain of attraction of a GPD $W_\gamma$ under exponential normalization coincides with that of a GLPD $L_\gamma = W_\gamma \circ \log$ since we have, by setting $x = \log(y)$,

$$\begin{aligned}
&\frac{\bar{F}\left(a(u)\exp\left(b(u)x\right)\right)}{\bar{F}(u)} \xrightarrow[u\to\omega(F)]{} W_\gamma(x), \quad x \in (\alpha(W_\gamma), \omega(W_\gamma)) \\
\Leftrightarrow \quad &\frac{\bar{F}\left(a(u)y^{b(u)}\right)}{\bar{F}(u)} \xrightarrow[u\to\omega(F)]{} W_\gamma\left(\log(y)\right), \quad y \in \left(\alpha(W_\gamma \circ \log), \omega(W_\gamma \circ \log)\right).
\end{aligned}$$

Hence, considering exponential transformations does not offer additional statistical models. Moreover, the domains of attractions are contained in the domains of attractions under power–normalization.

# 4 The Log–Pareto Distribution

Recall the results stated in Theorem 3.8. These entail that certain mixtures of heavy–tailed distributions are not in the pot–domain of attraction of GPDs under linear normalization, and, therefore, not in the scope of the linear pot–approach. Yet these distributions might also occur in other contexts. Their main property is that they have more mass in the extreme upper tail than heavy–tailed distributions, thus distributions with regularly varying tails.

In this chapter we introduce statistical models for such distributions with representation

$$F(x) = 1 - (\log(x))^{-1/\gamma} U(\log(x)), \quad \gamma > 0,\, x > a, \tag{4.1}$$

for sufficiently large $a > 0$, where $U$ is a slowly varying function. Note the following properties of $F$. First, the tail of $F$ is slowly varying, that is,

$$\frac{1 - F(tx)}{1 - F(t)} \xrightarrow[t\to\infty]{} 1.$$

All moments of $F$ are infinite, the log moments exist only up to the order of $1/\gamma$ so one may label $F$ as super–heavy tailed.

It is an open question if such distribution occur in contexts of real life data, but treating data from a super–heavy tailed distribution as if they were only heavy tailed may lead to serious misjudgements, for example the underestimation of extreme quantiles which play an important role in risk management in financial as well as in environmental applications. Until recently super–heavy tailed distribution have been rarely studied in the statistical literature, but one can recognize an increasing interest. Notable references are [21], [22], [23] and [67].

In [35] the authors use the term super–heavy tailed distribution for distributions with slowly varying tails. They study a certain sub–class of such distributions in detail, namely distributions with survivor functions in the class $\Pi$. By definition a survivor function $\bar{F}$ belongs to the class $\Pi$ if there exists a function $a : \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\lim_{t\to\infty} \frac{\bar{F}(tx) - \bar{F}(t)}{a(t)} = \log(x), \quad x > 0.$$

We refer to [6] for a broad theoretical treatment of such functions and to [17] for a description in the framework of probability theory.

A test procedure is derived in [35] to distinguish between distributions which tail functions belong to the class $\Pi$ (thus a subclass of super–heavy tailed distributions) and distributions with regularly varying tails (thus heavy tailed distribution). These authors also apply this test to a real data set of certain internet traffic data. In [48] additional real data sets are tested for super–heavy tails. More details will be presented at the end of this chapter.

While the above mentioned literature deals with testing whether super–heavy tails are present in a certain data set, we are primarily concerned with statistical models for such data. The result in Corollary 3.5 suggests log–Pareto distributions as given in (3.17) with an additional scale parameter and shape parameter $\beta$ as an adequate asymptotic model for the tail of the df $F$ in (4.1). For convenience we include a scale parameter of the form $\tilde{\sigma} = \sigma \exp(\beta)$ which leads to a log–Pareto df

$$\tilde{L}_{\gamma,\beta,\sigma}(x) = 1 - \left(1 + \frac{1}{\beta}\log(x/\sigma)\right)^{-1/\gamma}, \quad x > \sigma. \tag{4.2}$$

The introduction of a scale parameter is necessary for finite sample considerations since the unknown constants of the power–normalization (see (3.1)) have to be accounted for. For a further explanation in a more general framework see also the lines around equation (2.27).

Yet the use of this model for the modeling of exceedances has certain drawbacks. The scale parameter $\sigma$ is also the left endpoint of $\tilde{L}_{\gamma,\beta,\sigma}(x)$ and, therefore, necessarily equal to $u$ if exceedances over a threshold $u$ are to be modeled. Thus the model $\hat{L}_{\gamma,\beta,\sigma}$ lacks a free scale parameter but includes only two shape parameters. For a df $F$ in the p–pot–domain of attraction of a log–Pareto df $\hat{L}_\gamma$ we may assume that

$$F^{[u]}(x) \approx \hat{L}_\gamma\left((x/u)^\beta\right), \quad x > u,$$

for a sufficiently high threshold $u$. This yields for the df of the pertaining excesses

$$F^{[u]}(x+u) \approx 1 - (1 + \beta\log(1 + x/u))^{-1/\gamma}, \quad x > 0,$$

a df with two shape parameters $\gamma$ and $\beta$ and a scale parameter $u$ which is determined by the threshold. In this chapter we introduce and study a slight generalization of this model by adding a free scale parameter $\sigma$. On the one hand this model is more flexible when applied to real data especially for moderate sample sizes. On the other hand it is also possible to treat a data set of excesses in its own right without knowing the pertaining threshold in the broader model.

The generalized log–Pareto model introduced in the subsequent sections is an extension of the model in (4.2) which corrects its described disadvantages. Moreover, this model exhibits several interesting relations to the GPD model. For the statistical considerations we concentrate on its properties as a model for super–heavy tailed distributions, yet it includes also heavy and even short tailed distributions.

In contrast to the previous chapters, this chapter is of a statistical nature. In Section 4.1 we recall the main steps which generally lead to a log–family of dfs, introduce log–Pareto dfs, and show in which manner Pareto dfs can be regained from log–Pareto dfs.

Certain generalized log–Pareto dfs are discussed in Section 4.2. Section 4.3 concerns the statistical inference within the 3–parameter model of log–Pareto dfs. We propose a hybrid estimator which is a combination of a quick estimator and a maximum likelihood estimator (MLE) in some 1–dimensional submodel. This estimator is adopted as an initial estimator for the MLE in the full log–Pareto model and its performance is illustrated in simulation studies. Visual tools and test procedures are shortly addressed. Applications to real data are added in Section 4.4. Sections 4.1 and 4.2 as well as Section 4.3.1 are mainly taken from [14].

## 4.1 The Log–Pareto Model as an Extension of the Pareto Model

We start with some general remarks about the construction of log–families of dfs. Let $X$ be a rv with df $H_\vartheta$ where $\vartheta$ is a shape parameter. Assume that the left endpoint of the support of $H_\vartheta$ is equal to zero. Notice that $\beta X$ — with $\beta > 0$ — has the df $H_{\vartheta,\beta}(x) = H_\vartheta(x/\beta)$ with shape and scale parameters $\vartheta$ and $\beta$. Then, the transformed rv $\exp(\beta X) - 1$ has the df $H_\vartheta\left(\frac{1}{\beta}\log(1+x)\right)$, $x > 0$, with shape parameters $\vartheta$ and $\beta$ and left endpoint of the support equal to zero. Adding a scale parameter $\sigma > 0$ one gets the df

$$F_{\vartheta,\beta,\sigma}(x) = H_\vartheta\left(\frac{1}{\beta}\log(1+x/\sigma)\right), \quad x > 0. \tag{4.3}$$

Putting $\sigma = \eta/\beta$ and letting $\beta \to 0$, one gets

$$F_{\vartheta,\beta,\eta/\beta}(x) \to H_{\vartheta,\eta}(x). \tag{4.4}$$

Therefore, the original dfs can be regained from the family of log–dfs.

A first example for this approach is provided by the exponential df $H_\gamma(x) = 1 - \exp(-x/\gamma)$, $x > 0$, with scale parameter $\gamma > 0$. In this case, we have an initial df without a shape parameter. The log–exponential df corresponding to (4.3) is given by

$$\widetilde{W}_{\gamma,\beta}(x) = 1 - (1 + x/\beta)^{-1/\gamma}, \quad x > 0, \tag{4.5}$$

which is a Pareto df with shape and scale parameters $\gamma$ and $\beta$.

The second example concerns log–Pareto dfs. The initial dfs are those in (4.5). One obtains log–Pareto dfs

$$\tilde{L}_{\gamma,\beta,\sigma}(x) = 1 - \left(1 + \frac{1}{\beta}\log\left(1 + \frac{x}{\sigma}\right)\right)^{-1/\gamma}, \quad x > 0, \tag{4.6}$$

with shape parameters $\gamma$, $\beta > 0$ and scale parameter $\sigma > 0$. Notice that all moments are infinite. The log–moments $\int(\log(1+x))^z dL_{\gamma,\beta,\sigma}(x)$ are infinite if $z \geq 1/\gamma$; that is, log–Pareto dfs possess super–heavy upper tails. Note that only the shape parameter $\gamma$

is crucial for the existence of finite log–moments.

According to (4.4),

$$\tilde{L}_{\gamma,\beta,\eta/\beta}(x) \to \widetilde{W}_{\gamma,\eta}(x), \quad \beta \to 0, \tag{4.7}$$

and, thus, one gets Pareto dfs in the limit.

If the transformation $(\exp(\xi X) - 1)\,\vartheta$ is applied to a log–Pareto rv $X$ with df $\tilde{L}_{\gamma,\beta,1}$ in (4.6) one receives a second order log–Pareto df with three shape and one scale parameter, namely

$$\tilde{L}^{(2)}_{\gamma,\beta,\xi,\vartheta}(x) = 1 - \left(1 + \frac{1}{\beta}\log\left(1 + \frac{1}{\xi}\log\left(1 + x/\vartheta\right)\right)\right)^{-1/\gamma}, \quad x > 0. \tag{4.8}$$

The log–transformation of such dfs leads to dfs with super–heavy tails, hence one may speak of second–order super–heavy tails. Apparently, this procedure can be iterated further on leading to dfs with more and more shape–parameters and higher order iterated super–heavy tails.

## 4.2 Generalized Log–Pareto Families

The preceding considerations can be extended to generalized Pareto dfs (GPDs) and generalized log–Pareto dfs (GLPDs). Recall that Pareto, exponential and certain beta dfs constitute the family of GPDs which is the basic family of dfs in the pot–approach within extreme value theory. For that purpose, a slightly different parametrization was introduced in (2.5).

Changing the parametrization of the Pareto df in (4.5) into the form

$$W_{\gamma,\beta}(x) = \widetilde{W}_{\gamma,\beta/\gamma}(x) = 1 - \left(1 + x\frac{\gamma}{\beta}\right)^{-1/\gamma}, \quad x > 0, \tag{4.9}$$

one gets the standard form of the GPD family (in the von Mises representation) which is valid for all real $\gamma$, if for negative $\gamma$ the additional restriction $0 < x < \beta/|\gamma|$ is included (see also (2.5)). The case $\gamma = 0$ is taken as the limit for $\gamma \to 0$ which yields exponential dfs

$$W_{0,\beta}(x) = 1 - \exp(-x/\beta), \quad x > 0.$$

The parametrization in (4.5) is preferable to that in (4.9) in certain statistical applications, see e.g. Section 4.3 of this thesis, [7] or the Bayesian inference in Section 5.1 of [56], 3rd edition.

Applying the slightly modified exponential transformation

$$x \to \exp(\sigma/\beta x) - 1$$

to the GPD $W_{\gamma,\beta}$ one receives a different form of the log–Pareto df, namely,

$$L_{\gamma,\beta,\sigma}(x) = 1 - \left(1 + \frac{\gamma}{\beta}\log\left(1 + \frac{\beta}{\sigma}x\right)\right)^{-1/\gamma}, \quad x > 0. \tag{4.10}$$

It is easy to verify that

$$L_{\gamma,\beta,\sigma}(x) \xrightarrow[\gamma \to 0]{} W_{\beta,\sigma}(x) = L_{0,\beta,\sigma}(x) \tag{4.11}$$

and as in (4.7)

$$L_{\gamma,\beta,\sigma}(x) \xrightarrow[\beta \to 0]{} W_{\gamma,\sigma}(x) = L_{\gamma,0,\sigma}(x). \tag{4.12}$$

Thus we get that one GPD $W_{\gamma,\sigma}$ is limit of two different sequences of GLPDs with considerably different properties. Note that for none of the GLPDs in (4.12) the log–moments of larger order than $1/\gamma$ exists while in the case of (4.11) log–moments of the order $1/\gamma$ exist when $\beta$ gets smaller then $\gamma$. This causes some problems if one wants to estimate the parameters of a GLPD which is close to a GPD as will be seen in Section 4.3.

Notice that the case $\gamma < 0$ yields finite right endpoints. More precisely, we have a right endpoint equal to $(\exp(\beta/|\gamma|) - 1)\,\sigma/\beta$ for $\gamma < 0$. GLPDs form a unified model for dfs of excesses over high thresholds if the underlying df $F$ has a right endpoint larger than zero. GPDs with negative shape parameter can also be obtained in the limit for $\beta \to 0$ in (4.12) if the parameter $\gamma$ is negative. Thus, GPDs are included in the model of GLPDs in the limit.

## 4.3 Statistical Inference Within the Log–Pareto Model

If we restrict ourselves to the case $\gamma > 0$, thus super–heavy tailed distributions, the parametrization in form of the family $\tilde{L}_{\gamma,\beta,\sigma}$ is easier to handle for statistical applications. It is obviously an easy exercise to transfer this parametrization into the GLPD family. For the log–Pareto family consisting of dfs $\tilde{L}_{\gamma,\beta,\sigma}$ in (4.6) (and, therefore, for the GLPD family) we do not find closed form estimators of the parameters $\gamma$, $\beta$ and $\sigma$. Especially, we do not achieve a closed form of the MLE . Therefore, MLEs have to be numerically computed and an initial estimator is required as a starting value for an iteration procedure. Our primary aim is to find an initial estimator where the computational work can be reduced to some extent.

The performance of an initial (hybrid) estimator and the MLE will be illustrated by means of simulations. We also include applications of the MLE to two real data sets.

### 4.3.1 Quick Estimators and MLEs

We proceed in analogy to [55], where a modified Pickands estimator for the Pareto model is introduced. Replacing theoretical quantiles by appropriate order statistics, one gets estimators of $\sigma$ and $\gamma$ by solving certain equations. Estimators of this type may be classified as quick or systematic estimators.

In the case of the log–Pareto family one gets from $\tilde{L}_{\gamma,\beta,\sigma}(\tilde{L}^{-1}_{\gamma,\beta,\sigma}(q)) = q$, $0 < q < 1$, the equations

$$\log\left(1 + \frac{1}{\beta}\log\left(1 + \tilde{L}^{-1}_{\gamma,\beta,\sigma}(q)/\sigma\right)\right) = -\gamma\log(1-q). \tag{4.13}$$

Plugging in the special values $q_1 = 1 - a$ and $q_2 = 1 - a^2$, $0 < a < 1$, it is possible to solve this system of equations in $\gamma$ and $\beta$.

Replacing the quantiles $\tilde{L}^{-1}_{\gamma,\beta,\sigma}(q_1)$ and $\tilde{L}^{-1}_{\gamma,\beta,\sigma}(q_2)$ by appropriate order statistics $x_{j(a,k):k}$ and $x_{i(a,k):k}$ one receives estimators

$$\hat{\beta}_a(\sigma) = \frac{\left(\log(1 + x_{i(a,k):k}/\sigma)\right)^2}{\log\left(1 + x_{j(a,k):k}/\sigma\right) - 2\log\left(1 + x_{i(a,k):k}/\sigma\right)} \tag{4.14}$$

and

$$\hat{\gamma}_a(\sigma) = \log\left(\frac{\log\left(1 + x_{j(a,k):k}/\sigma\right) - \log\left(1 + x_{i(a,k):k}/\sigma\right)}{\log\left(1 + x_{i(a,k):k}/\sigma\right)}\right) / \log(1/a), \tag{4.15}$$

of $\gamma$ and $\beta$ for each $0 < a < 1$ where $i(a,k) = [k(1-a)]$ and $j(a,k) = [k(1-a^2)]$, if $\sigma$ is known.

If $\sigma$ is unknown one can make use of an additional equation

$$\tilde{L}_{\hat{\gamma}_a(\sigma),\hat{\beta}_a(\sigma),\sigma}(x_{l:k}) = q_3, \tag{4.16}$$

with $l = [q_3 k]$. This equation must be solved numerically in $\sigma$. One gets an estimator $\hat{\sigma}_{a,l}$ for the scale parameter. Plugging in $\hat{\sigma}_{a,l}$ in (4.14) and (4.15) one gets estimators $\hat{\gamma}_{a,l}$ and $\hat{\beta}_{a,l}$ for the shape parameters.

This procedure yields quick estimators $\hat{\gamma}_{a,l}, \hat{\beta}_{a,l}, \hat{\sigma}_{a,l}$ of $\gamma$, $\beta$, $\sigma$ for each pair $(a,l)$ with $l \neq i(a,k)$ and $l \neq j(a,k)$. The quality of the estimators can be improved by using medians of several quick estimators (as it was done in [7]). Regrettably, these modified, quick estimators still have a less favorable performance in the present case.

We suggest to replace the quick estimator $\hat{\sigma}_{a,l}$ in (4.16) by a MLE $\hat{\sigma}_a$ in the 1–parameter model

$$\{\tilde{L}_{\hat{\gamma}_a(\sigma),\hat{\beta}_a(\sigma),\sigma} : \sigma > 0\} \tag{4.17}$$

of log–Pareto dfs. Consequently, combining quick estimators and the MLE in the 1–parameter model one gets the hybrid estimators $\hat{\gamma}_{\mathrm{hyp}}$, $\hat{\beta}_{\mathrm{hyp}}$, $\hat{\sigma}_{\mathrm{hyp}}$ of $\gamma$, $\beta$, $\sigma$.

Finally, the hybrid estimators may be used as initial estimators in the numerical calculations of MLEs $\hat{\gamma}_{\mathrm{ML}}$, $\hat{\beta}_{\mathrm{ML}}$, $\hat{\sigma}_{\mathrm{ML}}$ in the full 3–parameter log–Pareto model.

### 4.3.2 Visual Tools for Data Analysis

We will shortly address some visual tools to examine whether the log–Pareto model is appropriate for given data. Our main is emphasis laid on the question whether the given data can be modeled rather by a GPD or by a GLPD. A simple visual tool to get a first insight if some data stems rather from a Pareto or log–Pareto model is the median excess function,

$$m_F(u) := \left(F^{[u]}\right)^{-1}(1/2) - u.$$

The Pareto model yields a linear function while one gets a convex function in the log–Pareto model if the shape parameter $\gamma$ is larger than zero and a concave shape if $\gamma$ is smaller than zero. To be more precise, the median–excess function in the log–Pareto model is given by

$$m_{\gamma,\beta,\sigma}(u) = (u + \sigma/\beta)^{2^\gamma} \exp\left((2^\gamma - 1)(\beta/\gamma - \log(\sigma/\beta))\right) - (u + \sigma/\beta)$$

if $\gamma \neq 0$ and

$$m_{0,\beta,\sigma}(u) = (u + \sigma/\beta)\left(2^\beta - 1\right).$$

If the GLPD model is close to a GPD, that is that the GLPD shape–parameters $\gamma$ and/or $\beta$ are close to zero it is difficult to distinguish between the models as can be seen in Figure 4.1.

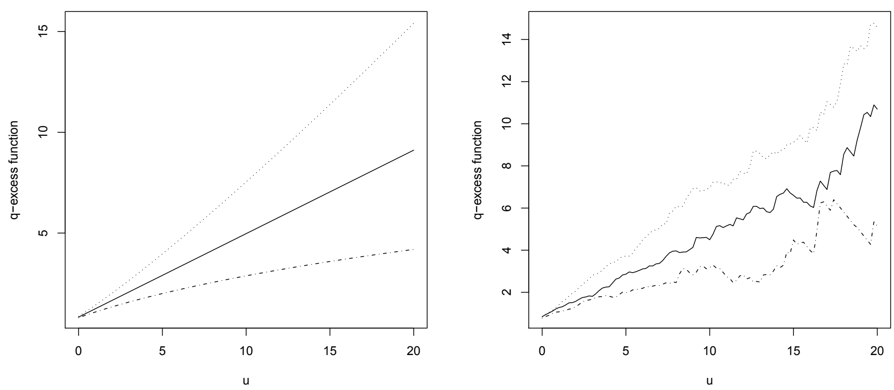



Figure 4.1: Median excess functions: Theoretical (left) and sample version based on 10000 simulations (right) for $\gamma = 0.1$ (dotted), $\gamma = 0$ (solid) and $\gamma = -0.1$ (dashed–dotted) for $\beta = 0.5$ and $\sigma = 1$.

More general one may use the $q$–excess function at level $q \in (0,1)$ defined by

$$Q_{q,F}(u) = \left(F^{[u]}\right)^{-1}(q) - u$$

for a df $F$. In the case of the GLPD family we get

$$Q_{q,\gamma,\beta,\sigma}(u) = (u + \sigma/\beta)^{(1-q)^{-\gamma}} \left(\exp\left(\frac{\beta}{\gamma} + \left((1-q)^{-\gamma} - 1\right)\frac{\sigma}{\beta}\right) - 1\right)$$

if $\gamma \neq 0$ and

$$Q_{q,0,\beta,\sigma}(u) = (u + \sigma/\beta)\left((1-q)^{-\beta} - 1\right)$$

in the GPD case. For $q$ tending to one we have an increasing difference of the $q$–excess function if $\gamma = 0$, where we still have a linear shape, and the cases where $\gamma < 0$ and $\gamma > 0$. In the latter case we have an increasing steepness for $q$ tending to one.

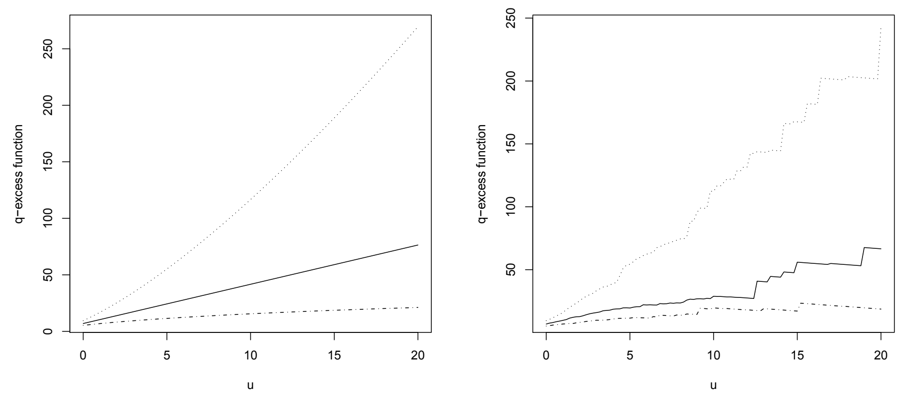



Figure 4.2: Q–excess functions at level 95%: theoretical (left) and sample version based on 10000 simulations (right) for $\gamma = 0.1$ (dotted), $\gamma = 0$ (solid) and $\gamma = -0.1$ (dashed–dotted) for $\beta = 0.5$ and $\sigma = 1$.

Figure 4.2 indicates that it is still not easy to decide visually based on a sample $q$–excess function whether the pertaining data comes rather from a GPD or a GLPD with shape parameter $\gamma \neq 0$. If we choose a high level for $q$ we have a larger difference of the shapes of the theoretical $q$–excess functions but on the other hand the sample versions are less reliable.

We propose to use a fraction of two $q$–excess functions of different levels as the most adequate tool to discriminate the cases $\gamma = 0$ and $\gamma \neq 0$. For $q_1$, $q_2 \in (0,1)$ we have

$$\frac{Q_{q_1,0,\beta,\sigma}(u)}{Q_{q_2,0,\beta,\sigma}(u)} = \frac{(1-q_1)^{-\beta} - 1}{(1-q_2)^{-\beta} - 1}$$

and thus the fraction does not depend on the threshold $u$ anymore. On the other hand, if $\gamma \neq 0$ we still have an increasing or decreasing shape of the fraction as a function of the threshold $u$

$$\frac{Q_{q_1,\gamma,\beta,\sigma}(u)}{Q_{q_2,\gamma,\beta,\sigma}(u)} = (u+\sigma/\beta)^{(1-q_1)^{-\gamma}-(1-q_2)^{-\gamma}} \frac{\exp\left(\frac{\beta}{\gamma} + ((1-q_1)^{-\gamma}-1)\frac{\sigma}{\beta}\right)-1}{\exp\left(\frac{\beta}{\gamma} + ((1-q_2)^{-\gamma}-1)\frac{\sigma}{\beta}\right)-1}.$$

We propose to use a relatively high quantile for $q_1$ (about 90% if the available sample size allows) and a low value for $q_2$ (such as 20%) resulting in the functions given in Figure 4.3.

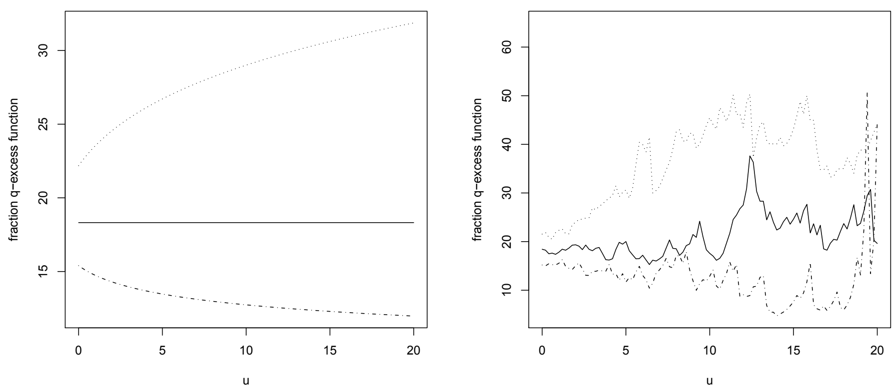



Figure 4.3: Fraction of $q$–excess functions at levels $q_1 = 90\%$ and $q_2 = 20\%$: Theoretical (left) and sample version based on 10000 simulations (right) for $\gamma = 0.1$ (dotted), $\gamma = 0$ (solid) and $\gamma = -0.1$ (dashed–dotted) for $\beta = 0.5$ and $\sigma = 1$.

### 4.3.3 Testing GLPDs versus GPDs

We include some remarks concerning testing the GPD against the GLPD model. Based on a dataset $x_1, \ldots, x_n$ the Neyman–Pearson test statistic for testing a GPD $L_{0,\beta,\gamma}$ against a GLPD $L_{\gamma,\beta,\sigma}$ with positive shape parameter $\gamma$, is given by

$$T_{\gamma,\beta,\sigma}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{l_{\gamma,\beta,\sigma}(x_i)}{l_{0,\beta,\sigma}(x_i)}.$$

Fix $\beta$ and $\sigma$ and put

$$g_{x,\beta,\sigma}(\gamma) = l_{\gamma,\beta,\sigma}(x), \quad x > 0.$$

Provided that $g$ is twice differentiable in a neighborhood of 0 and applying Taylors expansions at $\gamma = 0$ we receive

$$\log\left(T_{\gamma,\beta,\sigma}(x_1,\ldots,x_n)\right) \approx \sum_{i=1}^{n} \gamma \frac{g'_{x_i,\beta,\sigma}(\gamma)}{g_{x_i,\beta,\sigma}(\gamma)} =: \bar{T}_{\gamma,\beta,\sigma}(x_1,\ldots,x_n). \qquad (4.18)$$

Unfortunately $g_{x,\beta,\sigma}$ is not twice differentiable in 0, but nevertheless (4.18) still yields a reasonable approximation. We will construct a simple test statistic for testing $\gamma = 0$ against $\gamma > 0$, which is close to the Neyman–Pearson statistic based on the above extension. Let

$$c_{\beta,\sigma}(x) := -\log\left(1 - L_{0,\beta,\sigma}(x)\right), \quad x > 0.$$

The first derivative of $g$ satisfies

$$g'_{x,\beta,\sigma}(\gamma) = g_{x,\beta,\sigma}(\gamma)\left(\frac{\log\left(1 + c_{\beta,\sigma}(x)\gamma\right)}{\gamma^2} - \frac{c_{\beta,\sigma}(x)}{\gamma + c_{\beta,\sigma}(x)\gamma^2} - \frac{c_{\beta,\sigma}(x)}{1 + c_{\beta,\sigma}(x)\gamma}\right)$$

if $\gamma \neq 0$, and

$$g'_{x,\beta,\sigma}(0) = g_{x,\beta,\sigma}(0)\left(c_{\beta,\sigma}(x)^2/2 - c_{\beta,\sigma}(x)\right).$$

Note that $g'$ is continuous in 0. In what follows we consequently use the test statistic

$$\widetilde{T}_{\beta,\sigma}(x_1,\ldots,x_n) := \sum_{i=1}^{n}\left(1 + \log\left(\bar{L}_{0,\beta,\sigma}(x_i)\right)\right)^2$$

Notice that one gets for $\gamma > 0$

$$\widetilde{T}_{\beta,\sigma}(x_1,\ldots,x_n) = \frac{1}{2\gamma}\bar{T}_{\gamma,\beta,\sigma}(x_1,\ldots,x_n) - n.$$

Asymptotic normality of $\widetilde{T}$ under the null–hypothesis $\gamma = 0$ is easily established.

**Lemma 4.1** *If $X_1,\ldots,X_n$ are iid with common df $L_{0,\beta,\sigma}$, then*

$$\frac{1}{\sqrt{8n}}\left(\widetilde{T}_{\beta,\sigma}(X_1,\ldots,X_n) - n\right) \xrightarrow{d} X$$

*where $X \sim \mathcal{N}_{0,1}$.*

Proof. Notice that $\bar{L}_{0,\beta,\sigma}(X_i)$ is uniformly distributed on $[0,1]$ and, therefore, we get from Lemma A.10

$$\mathrm{E}\left(\left(1 + \log\left(\bar{L}_{0,\beta,\sigma}(X_i)\right)\right)^2\right) = 1 \quad \text{and} \quad \mathrm{Var}\left(\left(1 + \log\left(\bar{L}_{0,\beta,\sigma}(X_i)\right)\right)^2\right) = 8,$$

$i = 1,\ldots,n$. Consequently, the central limit theorem applies to $\widetilde{T}_{\beta,\sigma}(X_1,\ldots,X_n)$. □

The following test now suggests itself, we reject the null–hypothesis $\gamma = 0$ if

$$\frac{1}{\sqrt{8n}}(\widetilde{T}_{\beta,\sigma}(X_1,\ldots,X_n) - n) > \Phi^{-1}(1-\alpha), \quad 0 < \alpha < 1 \tag{4.19}$$

where $\Phi^{-1}$ is the qf pertaining the the standard normal distribution and $\alpha$ is the test level. Notice that if $\beta$ and $\sigma$ are known the performance of the test does not depend on these parameters in the null–hypothesis. Figure 4.4 indicates that the power of the test is quite good, but it heavily depends on the sample size (see also Figure 4.5).

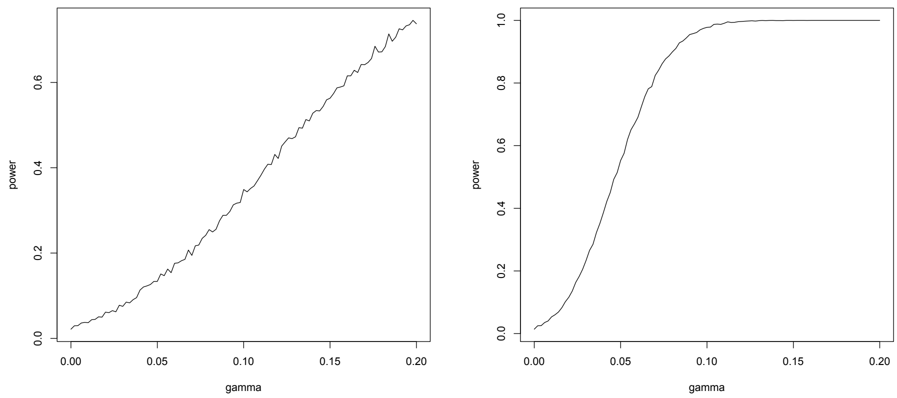



Figure 4.4: Power of the test (4.19) (with test level $\alpha = 0.01$) for varying $\gamma$ for a sample of size 100 (left) and 1000 (right), based on 4000 simulation runs.

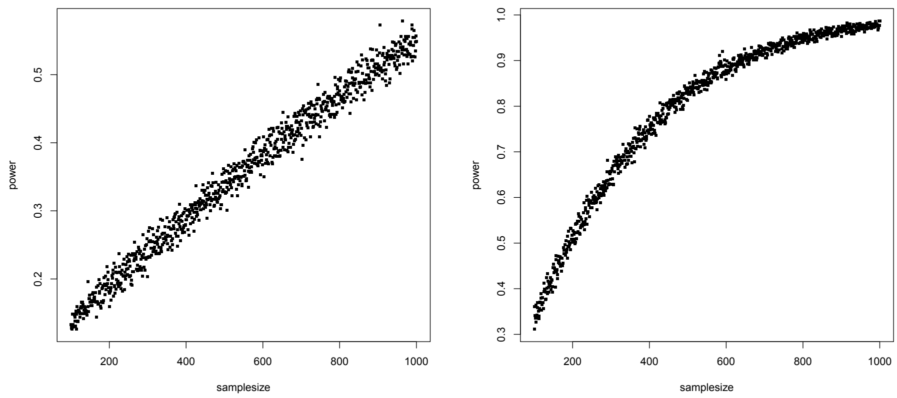



Figure 4.5: Power of the test (4.19) (with test level $\alpha = 0.01$) for varying sample size for shape parameters $\gamma = 0.05$ (left) and $\gamma = 0.1$ (right), based on 1000 simulation runs.

In this thesis we merely use the above test to exemplify the difficulty to recognize a super–tailed behavior in small to moderate samples. Roughly speaking the results illustrated in Figur 4.5 show that a small log–Pareto sample still looks like a Pareto sample.

Obviously the test (4.19) is not applicable in practical applications because one has to the fix the parameters $\beta$ and $\sigma$ of the null–hypothesis.

If one intends to test for an iid sample whether it stems rather from a GPD or GLPD we suggest to divide the sample into two disjoint sub–samples. The first one is then used to estimate the underlying GPD parameters for example by the ML–method. Then the test is applied to the second sub–sample where $\beta$ and $\sigma$ are replaced with the pertaining MLEs. Nevertheless, this will end in an increasing variance of the test statistic and, therefore, adulterate the test level.

### 4.3.4 Simulations

We shortly summarize the results of some simulation studies which were carried out using the statistical software **R**. The range of parameters $\gamma$ employed in the simulations was restricted to the interval $(0, 0.3)$ due the occurrence of large data which could not be handled by **R**.

First, we illustrate the performance of MLEs of $\gamma$ in the log–Pareto model and the Pareto model. Consider an iid sample $x_1, \ldots, x_k$ from a log–Pareto df $\tilde{L}_{\gamma,\beta,\sigma}$ and the transformed sample $y_1, \ldots, y_k$ with $y_i = \log(1 + x_i)$. The chosen parameters are $\gamma = 0.2$, $\beta = 0.5$ and $\sigma = 1, 5$. If $\sigma = 1$, then $y_1, \ldots, y_k$ is an iid sample from a Pareto df $\widetilde{W}_{\gamma,\beta}$. The MLEs in the log–Pareto and Pareto models are applied to $x_1, \ldots, x_k$ and, respectively, to $y_1, \ldots, y_k$ The illustrations in Figure 1 show the pertaining kernel densities of the MLEs of $\gamma$ based on $x_1, \ldots, x_k$ and respectively, on $y_1, \ldots, y_k$ based on 100 estimates, each taken for samples size $k = 1000$.

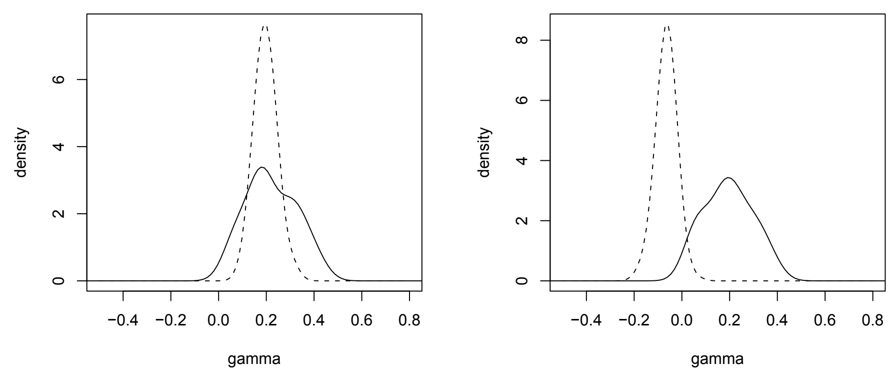



Figure 4.6: Kernel densities of the MLEs of $\gamma$ based on $x_i$ (solid) and $y_i$ (dashed) for $\sigma = 1$ (left) and $\sigma = 5$ (right).

The MLE in the Pareto model exhibits a better performance than the MLE in the log–Pareto model if the former one is applied within the correct model, that is, if the scale parameter $\sigma$ is equal to one. If this parameter differs from one, then the Pareto MLE is strongly biased.

Secondly, we compare the performance of the hybrid estimator and the MLE of $\gamma$ in the log–Pareto model. The corresponding kernel densities, based on simulated estimates, are displayed in Figure 4.7.

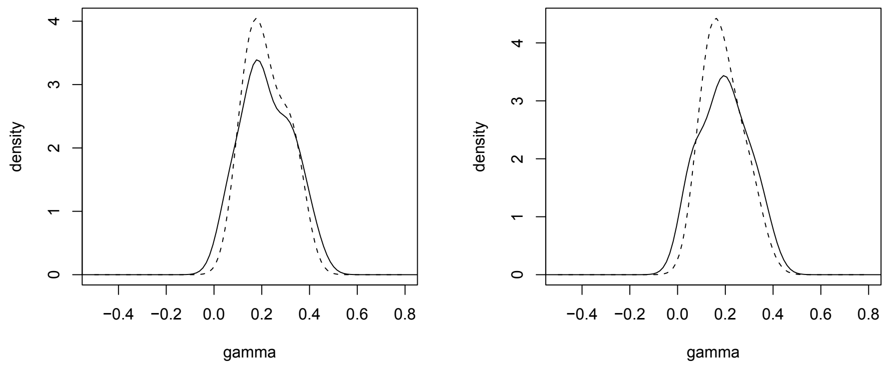



Figure 4.7: kernel densities of the hybrid estimator (dashed) and the MLE (solid) for $\sigma = 1$ (left) and $\sigma = 5$ (right)

The hybrid estimator exhibits a certain bias which is corrected by the MLE. This might be due to the fact, that the hybrid estimator is the median of several MLEs in different submodels. On the other hand, one observes a slightly higher variation of the MLE. The latter property can be influenced by numerical problems in the 3–dimensional maximization procedure.

## 4.4 Applications to Real Data

We present two data sets and apply the log–Pareto model $L_{\gamma,\beta,\sigma}$ as well as the Pareto model $W_{\gamma,\sigma}$ to the pertaining excesses. Initial estimates are obtained in the log–Pareto model $\tilde{L}_{\gamma,\beta,\sigma}$. Recall from (4.11) and (4.12) that $L_{\gamma,\beta,\sigma}$ is close to $W_{\beta,\sigma}$ or $W_{\gamma,\sigma}$ if $\gamma$ or $\beta$ is small.

Case Study 1. **[Plankton species data.]** The first example concerns long term copepod data first studied in [58] and also analyzed in [55].

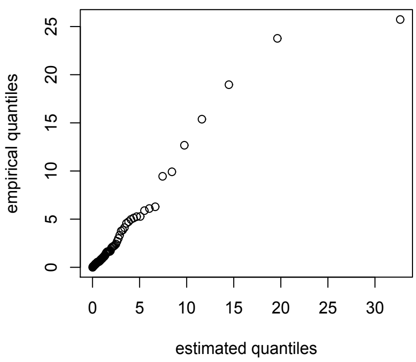
The data are deduced from weekly measurements of the abundance of a plankton species (a copepod called Centropages typicus) from January 1967 to December 1997 in the Villefranche Bay near Nice, France. Countings were separately made for females and males. The data set mentioned above concerns the product of the countings for females and males. This product is an index for the mating encounter rate which is regarded as a critical issue in plankton ecology.

The range of the data reaches from 0 to 908.782, the sample size is $n = 1353$. The models are adopted to the normalized data exceeding the threshold $u = 34000$. This choice of $u$ yields a number of k = 100 exceedances $x_1, ..., x_{100}$. The normalization is done by means of $y_i = x_i/34.000 - 1$.

We obtained 0.78 as estimate for $\gamma$ and 1.43 for $\sigma$ in the Pareto model as well as $\hat{\gamma} = 0.24$, $\hat{\beta} = 0.69$ and $\hat{\sigma} = 5.07$ in the log–Pareto model (4.6) as initial estimates the pertaining ML–estimates are $\hat{\gamma}_{ML} = 0.69$, $\hat{\beta}_{ML} = 0.43$ and $\hat{\sigma}_{ML} = 6.36$. This correponds to $\hat{\gamma}_{ML} = 0.69$, $\hat{\beta}_{ML} = 0.0014$ and $\hat{\sigma}_{ML} = 1.43$ in the GLPD model. Since the estimate of $\beta$ in GLPD model is small, the pertaining GLPD model is close to a GPD model.

As expected the difference between the q–q–plots is minute. Both plots show approximately a straight line, see Figure 4.8, so that both models can be adopted to the data.

The fraction of the $q$–excess function in Figure 4.9 shows an increasing form for smaller thresholds, so one may argue that the sample median excess function supports rather the log–Pareto than the Pareto model.



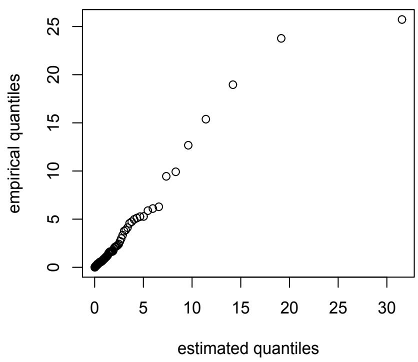




Figure 4.8: Plankton species data: q–q–plot, estimated log–Pareto df (left) and estimated Pareto df (right)

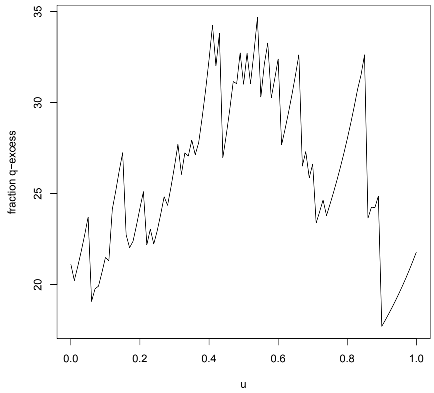



Figure 4.9: Plankton species data: fraction of $q$–excess functions.

We also include a list of some higher quantiles of the estimated Pareto and log–Pareto df.

| | Pareto model | log–Pareto model |
|---|---|---|
| 90% | 5.57 | 5.59 |
| 95% | 9.87 | 9.93 |
| 99% | 32.87 | 33.46 |
| 99.9% | 166.58 | 181.24 |
| 99.99% | 821.45 | 1263.56 |
| 99.999% | 4028.86 | 51815.42 |

Table 4.1: Plankton species data: Estimated extreme quantiles in the Pareto and log–Pareto model.

For moderately high quantiles the deduced values are nearly identical. Yet extreme quantiles are significantly larger in the log–Pareto approach.

As indicated in Table 4.1, the consequences of adopting the Pareto model or, alternatively, the extended model can be drastic, also, e.g., with respect to the existence of moments or log–moments.

**Case Study 2. [Internet traffic data.]** As a second example we consider internet traffic data (file lengths in bytes) included in the Internet Traffic Archive (http://ita.ee.lbl.gov/index.html) already studied in [48] in view of the presence of super–heavy tails. In this article the authors apply a test procedure introduced in [35] to this data set and reject the presence of super–heavy tails in their sense, that is, that the underlying df belongs to the class Π, cf. for example [35]. All log–Pareto dfs introduced in this article are members of the class Π, thus, among others, the family of log–Pareto dfs is rejected for this data set.

However, applying the log–Pareto model to normalized exceedances $y_1$, ..., $y_{51}$ over the threshold 994 143, where the normalization is this time done by $y_i = x_i/994\,143 - 1$, where

$x_1, ..., x_{51}$ denote the original exceedances, yields a reasonable fit as Figure 4.10 indicates. The plateau which is noticeable in both plots is due to the multiple occurrence of some values in the data set. The pertaining ML–estimates in the log–Pareto model are $\hat{\gamma}_{\mathrm{ML}} = 0.04$, $\hat{\beta}_{\mathrm{ML}} = 6.29$ and $\hat{\sigma}_{\mathrm{ML}} = 1.75$. Using the parametrization of the GLPD family these estimates correspond to a GLPD $L_{0.04,0.25,0.44}$. Applying the generalized Pareto model yields 0.30 as estimate for $\gamma$ and 0.43 for $\sigma$. According to the limit relation (4.11) and the small value of the MLE $\hat{\gamma}_{\mathrm{ML}}$ both estimated models are close to each other. This becomes also obvious by the hardly distinguishable q–q–plots, see Figure 4.10.

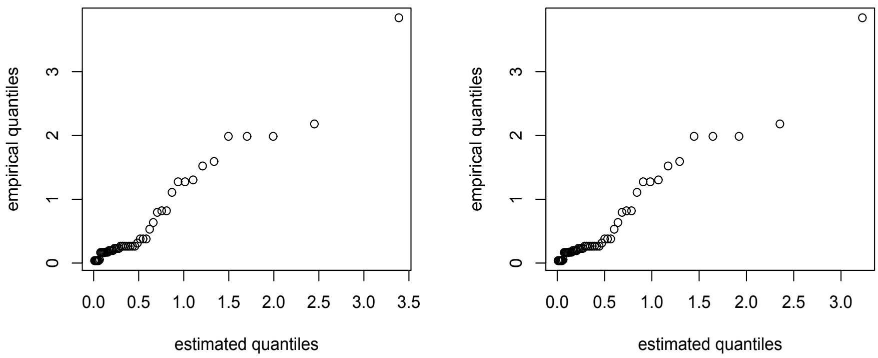



Figure 4.10: Internet traffic data: q–q–plot, estimated log–Pareto df (left) and estimated Pareto df (right)

**Case Study 3. [Negative log–returns of the Altana stock.]** The data set considered in this case study consists of negative log–returns of the closing prices of the Altana stock recorded from January 2nd 1990 to June 26th 2009 which yields a total sample size of 5062 observations. It is well known that log–returns are likely to be dependent which is also supported by Figure 4.11 but often one can assume an ergodic structure and thus treat the data as independent and distributed according to the pertaining stationary distribution. We chose the threshold $u = 0.0274$ which yields a number of 254 exceedances (about 5% of the data). The threshold was chosen by trial and error.

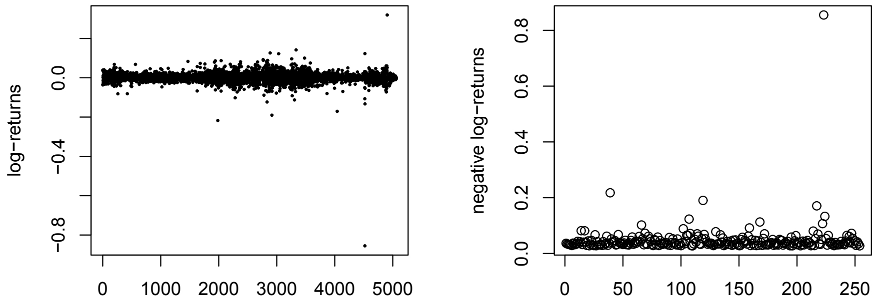



Figure 4.11: Log returns of the Altana stock (left) and pertaining exceedances of the negative log–returns over the threshold $u = 0.0274$ (right).

A notable feature of this data set is the very large loss which resulted in a log–return of about $-8.5$ which is far outside the range of the remaining observations. This indicates that the GLPD model might be a reasonable model for this data set. The received ML–estimates are $\hat{\gamma}_{\mathrm{ML}} = 0.17$, $\hat{\beta}_{\mathrm{ML}} = 0.12$ and $\hat{\sigma}_{\mathrm{ML}} = 0.013$ which indicates a significant tail weight because the estimated GLPD model is not close to the GPD model. The estimates of the shape parameters differ significantly from zero.

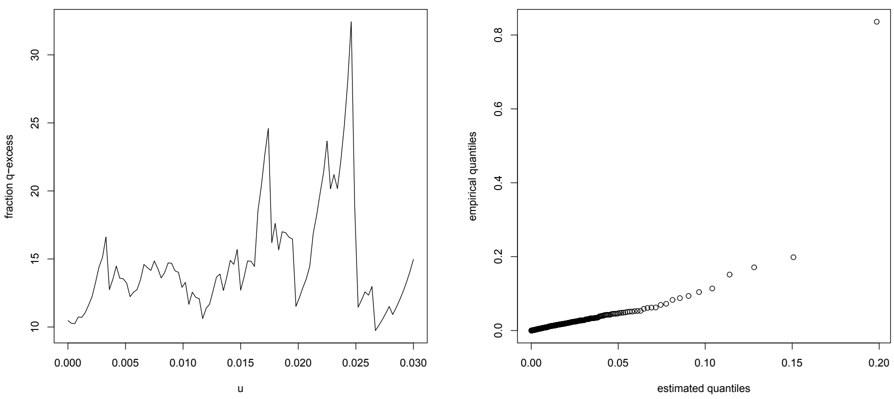



Figure 4.12: fraction of $q$–excess functions of the Altana negative log–returns data (left) and pertaining GLPD q–q-plot (right).

The q–q–plot in Figure 4.12 indicates that we have a good fit of the model if the largest observation is omitted. This is an unsatisfactory result since the GLPD model was constructed to account for exactly such extra ordinary large observations in a sample. But nevertheless the GPD model yields a worse fit, so we have at least an improvement compared to the standard model.

In [48] another data set of seismic data consisting of seismic moments for California seismicity recorded from 1800 to 1999 is studied. For this data super–heavy tails are not rejected. Adopting the log–Pareto model to this data set yields no reasonable result, at least using the proposed estimation procedure, so the results are omitted. This result might be caused by the fact, that this data set has a discrete structure, most included values appear several times.

Nevertheless, all considered data sets show a significant tail weight and a clear decision whether modeling with log–Pareto or Pareto dfs is preferable cannot be made.

Additional work is needed to derive tests which are tailored to distinguish between the Pareto and log–Pareto model. Special emphasis has to be laid on the cases where both models are close to each other, that is, if we have small values of $\gamma$ and/or $\beta$ in the log–Pareto model.

# 5 Conditional Exceedance Point Processes under Covariate Information

In this chapter we turn the focus of our analysis to the upper tails of conditional distributions. As the main technical tool we will use the theory of point processes. We consider a random vector $(\boldsymbol{X}, Y)$, $Y : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B}(\mathbb{R}))$ and $\boldsymbol{X} : (\Omega, \mathcal{A}) \to (S, \mathcal{B})$ where $S \subset \mathbb{R}^d$. In the following $\boldsymbol{X}$ will be addressed as vector of covariates or short covariate while $Y$ is the variable in which we are primarily interested in the following denoted as response. The aim of our subsequent analysis is statistical inference for the upper tail of the conditional distribution of $Y$ given the covariate $\boldsymbol{X} = \boldsymbol{x}$. This topic is extensively studied in the statistical literature because of its relevance for a lot of applications, see for example [61] and [42] in the field of environmental statistics or [66] for applications to financial markets.

The general situation which is studied may be summarized as follows. Suppose we observe a response variable $Y$ of interest (for example (log)–returns of prices, damages from insurance treaties, wave heights, precipitation amounts,...) and some covariate $\boldsymbol{X}$ which influences the distribution of the studied variable. Suppose further that the covariate can be observed before $Y$ or we have at least some good prediction. Thus we may use the conditional distribution of $Y$ given $\boldsymbol{X} = \boldsymbol{x}$ to predict, for example, certain conditional quantiles which correspond to the Value at Risk of a financial asset or a $t$–year return level of certain environmental variables.

Statistical analysis for the upper tail of a distribution is closely related to the pertaining distribution of exceedances. Recall the well–known formula which relates the upper tail of a df $F$ and the pertaining df of exceedances above the threshold $u$, $F^{[u]}$. We have

$$F(y) = F(u) + (1 - F(u))F^{[u]}(y), \quad y > u. \tag{5.1}$$

In terms of the conditional df

$$F(y|\boldsymbol{x}) = P\left(Y \le y \,|\boldsymbol{X} = \boldsymbol{x}\right)$$

this becomes

$$F(y|\boldsymbol{x}) = F(u|\boldsymbol{x}) + (1 - F(u|\boldsymbol{x}))F^{[u]}(y|\boldsymbol{x}), \quad y > u. \tag{5.2}$$

Therefore, we have to deal with two components

- the (conditional) distribution of exceedances
- the (conditional) probability that the threshold is exceeded.

In certain applications it is possible to treat both components separately. More details will be given in Section 5.4.1. Alternatively, one may use the theory of point processes which allows a simultaneous treatment of both components in a Maximum–Likelihood framework. Several likelihood based estimators are proposed in the statistical literature, which are based on different point process models.

We will present a unified point process model and identify the most established estimators [61] and [25] as an unconditional and, respectively, conditional Maximum–Likelihood estimator within this model. The underlying point process model is studied in detail with special emphasis laid on the inherent conditional densities. In contrast to the modeling in [61] and [25] all computations are carried out in a closed point process environment.

We give a short outline of the subsequent sections. The first part is of a probabilistic nature whereas the second part consists of remarks about the statistical inference. Section 5.1 gives a short overview of the theory of point processes and its applications to the framework of EVT. Section 5.2 contains the main result of this chapter. We derive the conditional density of the distribution of the process of exceedances and their pertaining exceedance covariates given the process of the original covariates. We formulate a basic Condition 5.8 for the process $N$ of original rvs. In particular, it is possible to deal with dependencies between the initial random vectors $(\boldsymbol{X}_i, Y_i)$. We verify that the considered conditional distribution in is identical for all processes $N$ satisfying this condition. This particularly yields that results obtained for Poisson processes are still valid within the general framework. Some technical auxiliary results are moved to Section 5.3. Section 5.4 concerns the statistical inference within the point process model. An overview of the pertaining statistical literature is also included. Applications and simulations are added in Section 5.5. We conclude with some remarks concerning further extensions in Section 5.6.

## 5.1 A Short Introduction to Point Processes

We give a short introduction into the concept of point processes and introduce the notation which is used throughout this chapter. Moreover, we present an application of the theory of point processes to EVT.

Let $(S, \mathcal{B})$ be a measurable space and for $x \in S$ let $\varepsilon_x$ be the Dirac–measure at $x$, thus

$$\varepsilon_x(A) = \begin{cases} 1, & x \in A; \\ & \text{if} \\ 0, & \text{otherwise} \end{cases} \qquad \text{for all } A \in \mathcal{B}.$$

Moreover, let $\mathbb{M}(S)$ be the set of all point measures

$$\boldsymbol{\mu} = \sum_{i=1}^{n} \varepsilon_{x_i}, \quad x_i \in S,\ n \in \mathbb{N}_0$$

on $S$ and $\mathcal{M}(S)$ the smallest $\sigma$–field on $\mathbb{M}(S)$ such that the projection mappings $\pi_B : \mathbb{M}(S) \to \mathbb{N}_0$ with

$$\pi_B(\boldsymbol{\mu}) = \boldsymbol{\mu}(B)$$

are measurable for all $B \in \mathcal{B}$. A point process $N$ on $\mathbb{M}(S)$ is a measurable mapping $\omega \mapsto N^\omega$ from some probability space $(\Omega, \mathcal{A}, P)$ into the space $(\mathbb{M}(S), \mathcal{M}(S))$. Thus $N$ is a random point measure. A basic example for a point process is the mixed empirical point process

$$N = \sum_{i=1}^{\beta} \varepsilon_{Y_i}$$

where the $Y_i$, $i = 1, 2, \ldots$ are independent $S$–valued rvs and $\beta$ is a $\mathbb{N}_0$–valued rv independent of the $Y_i$. For a subset $\{B_1, \ldots, B_k\} \subset \mathbb{B}$ the $\mathbb{N}_0^k$–valued random vectors

$$(N(B_1), \ldots, N(B_k))$$

are called finite dimensional marginals of $N$. It can be shown that the distribution of a point process is uniquely determined by the distribution of some of its finite dimensional marginals. Two point processes $N_1$ and $N_2$ are equal in distribution if, and only if,

$$(N_1(B_1), \ldots, N_1(B_k)) \stackrel{d}{=} (N_2(B_1), \ldots, N_2(B_k))$$

for all $k \in \mathbb{N}$ and pairwise disjoint $B_1, \ldots, B_k \in \mathcal{B}$.

Point processes offer a framework to study data with random sample sizes and are, therefore, useful for the theory of exceedances. Let $S = \mathbb{R}$ and $Y_1, \ldots, Y_n$ be iid rvs and

$$N_n = \sum_{i=1}^{n} \varepsilon_{Y_i}$$

the pertaining point process. $N_n$ is called empirical point process. Let $N_n^{[u]}$ be the truncated empirical process

$$N_n^{[u]} = N_n\left(\cdot \cap (u, \infty]\right).$$

Then $N_n^{[u]}$ only counts the exceedances over the threshold $u$ among the $Y_i$. Obviously $N_n^{[u]}$ carries information of the exceedances as well as the probability that the threshold $u$ is exceeded. The latter is captured in the random sample size. It is an easy exercise to prove that

$$N_n^{[u]} \stackrel{d}{=} \sum_{i=1}^{\beta_u} \varepsilon_{Y_i^{[u]}} \tag{5.3}$$

where $\beta_u$ and $Y_i^{[u]}$, $i = 1, 2, \ldots$ are independent and $\beta_u$ is a binomial rv with parameters $n$ and $p_u = P\{Y_1 > u\}$ and the $Y_i^{[u]}$ are iid with distribution

$$P\left\{Y_1^{[u]} \leq x\right\} = P\left(Y_1 \leq x \,|Y_1 > u\right).$$

Using the point process framework one can, thus, study the distribution of exceedances and the probability that the threshold is exceeded simultaneously. This is of particular interest for the development of the theory of extremes, see for example [52], but also for a lot of statistical applications, especially if non–homogeneous exceedances are studied.

If the threshold $u$ is chosen high enough the distribution of $Y_1^{[u]}$ can be approximated by a general pot–stable distribution $L$. Furthermore the exceedance probability $p_u$ gets small for large thresholds, thus the binomial rv $\beta_u$ might be replaced by a Poisson rv $\tau$ with parameter $\lambda = np_u$. The resulting point process

$$N^* = \sum_{i=1}^{\tau} Y_i^*,$$

where the $Y_i^*$ are iid with df $L$ is called a Poisson (point) process. It can be shown that $N^*$ is an accurate approximation of $N_n^{[u]}$ for high enough thresholds $u$ in the sense that the Hellinger distance between $\mathcal{L}\left(N_n^{[u]}\right)$ and $\mathcal{L}\left(N^*\right)$ gets small for increasing thresholds if the distribution of $Y_1$ is in some general pot–domain of attraction of $L$ and satisfies a certain additional condition which is related to $\delta$–neighborhoods of a GPD, see e.g [54] and [27].

It turns out that the class of Poisson processes is of particular interest within the class of point processes. In the subsequent lines we give a short overview of the theory of Poisson (point) processes.

### 5.1.1 Poisson Processes

A point process on $(\mathbb{M}(S), \mathcal{M}(S))$ is a Poisson (point) process if the following conditions hold

1. For all $k \in \mathbb{N}$ and all pairwise disjoint $B_1, \ldots, B_k \in \mathcal{B}$ the rvs $N(B_1), \ldots, N(B_n)$ are independent,
2. $N(B)$ is a Poisson rv for all $B \in \mathcal{B}$ .

It turns out that the intensity measure $\nu$ of a point process $N$ defined by

$$\nu(B) = E\left(N(B)\right), \quad B \in \mathcal{B},$$

plays a crucial role for the theory of Poisson point processes. This is due to the following result, for a proof we refer to Theorem 1.2.1 in [54].

**Lemma 5.1** *Let $N_1$ and $N_2$ be Poisson point processes with finite intensity measures $\nu_1$ and $\nu_2$, then*

$$N_1 \stackrel{d}{=} N_2$$

*if, and only if, $\nu_1 = \nu_2$.*

Thus the distribution of a Poisson point process is characterized by its intensity measure. A mixed empirical process

$$N = \sum_{i=1}^{\tau} \varepsilon_{Y_i}$$

is a Poisson process if $Y_1, Y_2, \dots$ are iid and independent of the Poisson rv $\tau$.

### 5.1.2 Estimating Upper Tails Using a Point Process Approach

If one intends to estimate the upper tail of an unconditional df $F$ based on iid observations $Y_1, \dots, Y_n$ with df $F$, then the ad hoc approach is to use the decomposition (5.1) and estimate both components separately.

1. Estimate $F(u)$ by the pertaining empirical version

$$\hat{F}(u) = \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,u]}(Y_i),$$

2. replace $F^{[u]}$ with a proper family of general pot–stable distribution $L_\xi$, $\xi = \Xi$, with densities $l_\xi$ and estimate $\xi$ by an ordinary ML–method using those observations among the $Y_i$ exceeding the threshold $u$.

Nevertheless, we might as well approach this problem in a point process framework. For convenience let $p := F(u)$, then we might rewrite (5.1) as

$$F(y) = p + (1-p)L_\xi(y), \quad y > u.$$

Define the truncated empirical point process $N_n^{[u]}$ pertaining to the rvs $Y_1, \dots, Y_n$ as in (5.3). Since $N_n^{[u]}$ is a binomial process a pertaining likelihood function can be derived using Example 3.1.2 in [54]. One receives

$$L(\xi, p) \propto \left( \prod_{i=1}^{k} l_\xi(y_i) \right) (1-p)^k p^{n-k}$$

if $y_1, \dots, y_k$ are the exceedances over the threshold $u$ out of a total sample with size $n$. The likelihood functions factorizes into two parts $L(\xi, p) = L_1(\xi) L_2(p)$ thus both parameters may be estimated by separately maximizing both likelihood functions. The pertaining MLE for $p$, $\hat{p}_{\mathrm{ML}} = (n-k)/n$ and $\hat{\xi}_{\mathrm{ML}}$ are the same as in the ad hoc approach described above.

## 5.2 The Basic Point Process Model

We now give a point process formulation of our initial problem, namely statistical inference for the upper tail of a conditional distribution. We have seen in Section 5.1.2 that this is a proper approach to tackle the proposed estimation problem in the case of upper tails of unconditional distributions. We first give a precise formulation of the underlying mathematical model.

Let $N$ be the point process

$$N = \sum_{i=1}^{\beta} \varepsilon_{(\boldsymbol{X}_i, Y_i)},$$

where $\beta$ is a $\mathbb{N}_0$–valued rv independent of $(\boldsymbol{X}_i, Y_i)$, $i \in \mathbb{N}$. Subsequently we will make use of the projection mappings $\pi_i$, $i = 1, 2$,

$$\pi_1(\boldsymbol{\eta}) = \sum_{i=1}^{n} \varepsilon_{\boldsymbol{x}_i} \quad \text{and} \quad \pi_2(\boldsymbol{\eta}) = \sum_{i=1}^{n} \varepsilon_{y_i}$$

where $\boldsymbol{\eta} = \sum_{i=1}^{n} \varepsilon_{(\boldsymbol{x}_i, y_i)}$.

Moreover, let

$$N_1 = \pi_1(N)$$

be the marginal point process of covariates pertaining to $N$. Conditions for the distribution of the points $(\boldsymbol{X}_i, Y_i)$ will be formulated in the subsequent sections. In a first step we will assume that the points are iid and then weaken this condition considerably.

Denote by $N^{[S,u]}$ the truncation of $N$ outside $S \times (u, \infty]$, thus

$$N^{[S,u]} = N\left(\cdot \cap S \times (u, \infty]\right) = \sum_{i=1}^{\beta} 1_{(u,\infty)}(Y_i) \varepsilon_{(\boldsymbol{X}_i, Y_i)}$$

and let

$$N_1^{[S,u]} = \pi_1\left(N^{[S,u]}\right).$$

be the pertaining marginal point process of covariates belonging to the exceedances. Finally, let

$$N_{1,u} = \pi_1\left(N(\cdot \cap S \times (-\infty, u])\right) \tag{5.4}$$

the process of covariates pertaining to the $Y_i$ not exceeding the threshold $u$. This yields

$$N_1 = N_1^{[S,u]} + N_{1,u}.$$

In the remainder of this section we will deduce a density of the conditional distribution

$$P\left(N^{[S,u]} \in \cdot \,|N_1 = \boldsymbol{\mu}\right). \tag{5.5}$$

Our results are valid under a condition (Condition 5.8) which is basically an extended form of conditional independence of the response variables given the covariates. As a first step we will assume $N$ to be a Poisson point process. Later we generalize the results to the more general framework.

### 5.2.1 Two Important Examples

As a first important example we will study the Poisson process case. We replace the random sample size $\beta$ by a Poisson rv $\tau$ with parameter $\lambda > 0$ and require the pertaining points $(\boldsymbol{X}_i, Y_i)$ to be iid copies of a random vector $(\boldsymbol{X}, Y)$. This model has been studied in detail in [13]. Obviously, $N^{[S,u]}$ as well as $N_1^{[S,u]}$ and $N_{1,u}$ are also Poisson processes in that case. Furthermore, see Lemma A.9

$$N^{[S,u]} \stackrel{d}{=} \sum_{i=1}^{\tau^*} \varepsilon_{(\boldsymbol{X}_i^*, Y_i^*)} \tag{5.6}$$

where $(\boldsymbol{X}_i^*, Y_i^*)$, $i \in \mathbb{N}$ are iid copies of a random vector $(\boldsymbol{X}^*, Y^*)$, independent of the Poisson rv $\tau^*$ with parameter $\lambda^* = \lambda P\{Y > u\}$,

$$P\left(Y^* \le y \,|\boldsymbol{X}^* = x\right) = F^{[u]}\left(u|\boldsymbol{x}\right)$$

and

$$P\{\boldsymbol{X}^* \in B\} = P\left(\boldsymbol{X} \in B \,|Y > u\right).$$

Let $\nu^{[S,u]}$ be the intensity measure of $N^{[S,u]}$, we have for $B \in \mathcal{B}$,

$$\begin{aligned}
\nu^{[S,u]}(B \times (u, y]) &= \lambda P\{\boldsymbol{X}^* \in B,\, Y^* \in C\} \\
&= \lambda \int_B F^{[u]}(y|\boldsymbol{x}) d\mathcal{L}\left(\boldsymbol{X}^*\right)(\boldsymbol{x}) \\
&= \lambda \int_B \int_u^y f^{[u]}(z|\boldsymbol{x}) dz d\mathcal{L}\left(\boldsymbol{X}^*\right)(\boldsymbol{x}).
\end{aligned} \tag{5.7}$$

Now let $Q$ be an arbitrary probability measure on $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ with density $q$ which is positive everywhere. Furthermore, we define the probability measure $Q^* = \mathcal{L}(\boldsymbol{X}^*) \times Q$

on $(S \times \mathbb{R}, \mathcal{B} \times \mathbb{B}(\mathbb{R}))$. One gets

$$\nu^{[S,u]}(B \times (u,y]) = \int_{B\times(u,y]} \frac{f^{[u]}(z|\boldsymbol{x})}{q(z)} \lambda dQ^*(\boldsymbol{x}, z).$$

An immediate consequence of the latter representation is, that the distribution of $N^{[S,u]}$ has a density with respect to the distribution of a Poisson process with intensity measure $\lambda \cdot Q^*$ which is given by

$$h(\boldsymbol{\eta}) = \left(\prod_{i=1}^{k} \frac{f^{[u]}(y_i|\boldsymbol{x}_i^*)}{q(y_i)}\right) \exp\left(\lambda \int F(u|\boldsymbol{x}) d\mathcal{L}(\boldsymbol{X})(\boldsymbol{x})\right) \tag{5.8}$$

for a point measure $\boldsymbol{\eta} = \sum_{i=1}^{k} \varepsilon_{\left(\boldsymbol{x}_i^*, y_i\right)}$. If the conditional distribution $F(\cdot|\boldsymbol{x})$ is in the domain of attraction of some general pot–stable distribution $L_{\vartheta(\boldsymbol{x})}$ we may replace $f^{[u]}(\cdot|\boldsymbol{x})$ by the pertaining density $l_{\vartheta(\boldsymbol{x})}$. Given a point measure $\boldsymbol{\eta}$ we may use $h$ to estimate the unknown functions $\vartheta$ and $F(u|\cdot)$ via the Maximum–Likelihood (ML) method. An obvious shortcoming of this approach is, that the density $h$ and, therefore, the pertaining likelihood function contains a term which depends on the distribution of the covariate $\boldsymbol{X}$. If we want to avoid any distributional assumption for the covariates except independence we have to replace $\mathcal{L}(\boldsymbol{X})$ with a non–parametric estimate for example a kernel estimate or an empirical distribution.

We include some remarks concerning another important special case, namely empirical and binomial processes. If we start with a fixed sample size $n$ instead of a Poisson distributed one, thus

$$N = \sum_{i=1}^{n} \varepsilon_{(\boldsymbol{X}_i, Y_i)}$$

where $(\boldsymbol{X}_i, Y_i)$ are again independent copies of $(\boldsymbol{X}, Y)$. The pertaining truncated process $N^{[S,u]}$ is a binomial process satisfying

$$N^{[S,u]} \stackrel{d}{=} \sum_{i=1}^{\beta_u} \varepsilon_{\left(\boldsymbol{X}_i^*, Y_i^*\right)}.$$

Hereby $(\boldsymbol{X}_i^*, Y_i)$ are as in the Poisson process case and $\beta_u$ is a binomial rv with parameters $n$ and

$$p_u = \int 1 - F(u|\boldsymbol{x}) d\mathcal{L}(\boldsymbol{X})(\boldsymbol{x}).$$

A density of $N^{[S,u]}$ can easily be derived using Example 3.1.2 in [54]. Let

$$N_0 = \sum_{i=1}^{\beta_0} \varepsilon_{(\boldsymbol{V}_i, Z_i)}$$

be a binomial process where $\mathcal{L}(\boldsymbol{V}_1, Z_1) = Q^*$ with $Q^*$ as above and $\beta_0 \sim B_{n,p_0}$. Then

$$h(\boldsymbol{\eta}) = \left( \prod_{i=1}^{k} \frac{f^{[u]}(y_i|\boldsymbol{x}_i^*)}{q(y_i)} \right) \left( \frac{p_u}{p_0} \right)^k \left( \frac{1-p_u}{1-p_0} \right)^{n-k} \tag{5.9}$$

is a $\mathcal{L}(N_0)$ density of $\mathcal{L}\left(N^{[S,u]}\right)$ where again $\boldsymbol{\eta} = \sum_{i=1}^k \varepsilon_{\left(\boldsymbol{x}_i^*, y_i\right)}$. This entails the same problem as in the Poisson process case since $p_u$ again depends on the function $F(u|\cdot)$ which we want to estimate and the distribution of the covariates.

To overcome this problem we propose a conditional approach, based on a certain conditional density of $N^{[S,u]}$ which does not depend on the distribution of the covariates anymore.

### 5.2.2 Conditional Distributions in the Poisson Process Case

Recall our aim to estimate the upper tail of the conditional distribution of $Y$ given $\boldsymbol{X} = \boldsymbol{x}$. We derive a density of the conditional distribution (5.5). We start by studying the case that $N$ is a Poisson process. The pertaining conditional density will be computed in three steps.

A crucial fact for the subsequent considerations is the conditional independence of $N_1$ and $N_2^{[S,u]}$ given $N_1^{[S,u]} = \boldsymbol{\mu}^*$ which is stated in Lemma 5.2.

**Lemma 5.2** *Let* $\boldsymbol{\mu}^* = \sum_{i=1}^k \varepsilon_{\boldsymbol{x}_i}$, $\boldsymbol{x}_i \in S$ *for* $i = 1, \ldots, k$ *then*

$$\begin{aligned} P\left(\left(N^{[S,u]}, N_1\right) \in A \times B \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) \\ = P\left(N_1 \in B \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) P\left(N^{[S,u]} \in A \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right), \end{aligned}$$

*for all* $A \in \mathcal{M}(S \times \mathbb{R})$ *and* $B \in \mathcal{M}(S)$.

Proof. First observe that

$$\begin{aligned} & P\left(N^{[S,u]} \in A, N_1 \in B \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) \\ = \; & P\left(N^{[S,u]} \in A, N_1^{[S,u]} + N_{1,u} \in B \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) \\ = \; & P\left(N^{[S,u]} \in A, \boldsymbol{\mu}^* + N_{1,u} \in B \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right). \end{aligned}$$

Moreover, $(N^{[S,u]}, N_1^{[S,u]})$ and $N_{1,u}$ are independent because $N$ is a Poisson process. Applying Lemma A.6 (cf. also [9], Corollary 7.3) with $Y = N_{1,u}$, $X = N_1^{[S,u]}$ and

$Z = N^{[S,u]}$ one gets

$$\begin{aligned} & P\left(N^{[S,u]} \in A, \boldsymbol{\mu}^* + N_{1,u} \in B \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) \\ &= P\left(\boldsymbol{\mu}^* + N_{1,u} \in B \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) P\left(N^{[S,u]} \in A \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) \\ &= P\left(N_1 \in A \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) P\left(N^{[S,u]} \in B \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right). \end{aligned}$$

□

Secondly, we derive a Chapman–Kolmogoroff representation of the conditional distribution of $N^{[S,u]}$ given $N_1 = \boldsymbol{\mu}$ in Lemma 5.3. This representation is crucial to compute an explicit representation of a pertaining density.

**Lemma 5.3** *For the conditional distribution of $N^{[S,u]}$ given $N_1 = \boldsymbol{\mu}$ holds*

$$\begin{aligned} & P\left(N^{[S,u]} \in A \,|N_1 = \boldsymbol{\mu}\right) \\ &= \int\int 1_A(\boldsymbol{\eta}) dP\left(N^{[S,u]} \in d\boldsymbol{\eta} \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) dP\left(N_1^{[S,u]} \in d\boldsymbol{\mu}^* \,|N_1 = \boldsymbol{\mu}\right) \end{aligned}$$

*for all $A \in \mathcal{M}(S \times \mathbb{R})$.*

Proof. Apply Lemma A.7 using the conditional independence of $N_1$ and $N^{[S,u]}$ given $N_1^{[S,u]} = \boldsymbol{\mu}^*$. □

To simplify the notation we introduce the following definition.

**Definition 5.4** *For a point measure $\boldsymbol{\mu} = \sum_{i=1}^n \varepsilon_{\boldsymbol{x}_i}$ on $S$ denote by*

- $\mathbb{M}_{\boldsymbol{\mu}}(S) := \{\sum_{i=1}^n u_i \varepsilon_{\boldsymbol{x}_i} : (u_1, \dots, u_n) \in \{0,1\}^n\}$ *the set of all point measures on $S$ which can be derived by thinning $\boldsymbol{\mu}$;*
- $\varpi_{\boldsymbol{\mu}}$ *the counting measure on the finite set $\mathbb{M}_{\boldsymbol{\mu}}(S)$;*
- $\upsilon_{\boldsymbol{\mu}}$ *the counting measure on $S_{\boldsymbol{\mu}} := \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$*

Notice that the distribution in (5.5) has all the mass on $\mathbb{M}(S_{\boldsymbol{\mu}} \times \mathbb{R})$. The pertaining density will be specified with respect to the distribution of a mixed empirical point process

$$\widetilde{N}_{\boldsymbol{\mu}} = \sum_{i=1}^{\tilde{\beta}} \varepsilon_{(V_i, Z_i)} \tag{5.10}$$

on $S_{\boldsymbol{\mu}} \times \mathbb{R}$, where the $(V_i, Z_i)$ are iid random vectors with common $\upsilon_{\boldsymbol{\mu}} \times \lambda$–density $p$ such that $p(v,z) > 0$ for all $(v,z) \in S_{\boldsymbol{\mu}} \times \mathbb{R}$. Moreover, $\tilde{\beta}$ is an integer valued rv independent of the $(V_i, Z_i)$ which has positive probability for each integer as, e.g., a Poisson rv.

Using the Chapman–Kolmogorov representation of (5.5) in Lemma 5.3 it remains to derive densities of the two involved conditional distributions. We start with the conditional distribution of $N_1^{[S,u]}$ given $N_1 = \boldsymbol{\mu}$. Notice that the support of this conditional distribution is the finite set $\mathbb{M}_{\boldsymbol{\mu}}(S)$. Therefore, we specify a counting density with respect to the measure $\varpi_{\boldsymbol{\mu}}$.

**Lemma 5.5** *Let $S_{\boldsymbol{\mu}}$ and $\varpi_{\boldsymbol{\mu}}$ as in Definition 5.4. Then*

$$g_1^{[S,u]}(\boldsymbol{\mu}^*|\boldsymbol{\mu}) = \left(\prod_{i=1}^{k}(1 - F(u|\boldsymbol{x}_i^*))\right)\left(\prod_{\boldsymbol{x}\in S_{\boldsymbol{\mu}}} F(u|\boldsymbol{x})^{\boldsymbol{\mu}\{\boldsymbol{x}\}-\boldsymbol{\mu}^*\{\boldsymbol{x}\}}\right),$$

*where $\boldsymbol{\mu} = \sum_{i=1}^n \varepsilon_{\boldsymbol{x}_i}$ and $\boldsymbol{\mu}^* = \sum_{i=1}^k \varepsilon_{\boldsymbol{x}_i^*}$, is a density of the conditional distribution of $N_1^{[S,u]}$ given $N_1 = \boldsymbol{\mu}$ with respect to $\varpi_{\boldsymbol{\mu}}$.*

Proof. Define $g_u : \mathbb{M}(S \times \mathbb{R}) \to \mathbb{M}(S)$ by

$$g_u(\boldsymbol{\eta}) = \pi_1\left(\boldsymbol{\eta}(\cdot \cap S \times (u,\infty))\right).$$

Then, $N_1^{[S,u]} = g_u(N)$, and, thus, we get from Lemma A.1 and Corollary 7.2.1 in [54](cf. also Theorem 5.12)

$$P\left(N_1^{[S,u]} \in \cdot\,|N_1 = \boldsymbol{\mu}\right) = P\left\{g_u(N_{\boldsymbol{\mu}}) \in \cdot\right\}$$

where

$$N_{\boldsymbol{\mu}} = \sum_{i=1}^{n} \varepsilon_{(\boldsymbol{x}_i,\widetilde{Y}_i)}$$

with $\widetilde{Y}_i$ being independent and distributed according to $P\left(Y_1 \in |\boldsymbol{X}_1 = \boldsymbol{x}_i\right)$. Furthermore, notice that $N_{\boldsymbol{\mu},1}^{[S,u]} := g_u(N_{\boldsymbol{\mu}})$ satisfies

$$N_{\boldsymbol{\mu},1}^{[S,u]} = \sum_{i=1}^{n} \widetilde{U}_i \varepsilon_{\boldsymbol{x}_i}$$

with the $\widetilde{U}_i$ given by

$$\widetilde{U}_i = 1_{(S\times(u,\infty))}(\boldsymbol{x}_i, \widetilde{Y}_i).$$

Apparently $N_{\boldsymbol{\mu},1}^{[S,u]}$ takes values in the finite space $\mathbb{M}_{\boldsymbol{\mu}}(S)$. Let $\boldsymbol{\mu}^* \in \mathbb{M}_{\boldsymbol{\mu}}(S)$,

$$\boldsymbol{\mu}^* = \sum_{i=1}^{k} \varepsilon_{\boldsymbol{x}_i^*} = \sum_{i=1}^{n} u_i \varepsilon_{\boldsymbol{x}_i}$$

where $u_i \in \{0,1\}$ for $i = 1, \dots, n$. Let $I_n = \{1, \dots, n\}$ and $J := \{i \in I_n : u_i = 1\}$. We get

$$\begin{aligned} P\left\{N^{[S,u]}_{\boldsymbol{\mu},1} = \boldsymbol{\mu}^*\right\} &= P\left\{\widetilde{U}_i = 1,\, i \in J\,, \widetilde{U}_i = 0,\, i \in I_n \backslash J\right\} \\ &= \left(\prod_{i \in J}(1 - F(u|\boldsymbol{x}_i))\right)\left(\prod_{i \in I_n \backslash J} F(u|\boldsymbol{x}_i)\right). \end{aligned}$$

Observe that $\boldsymbol{\mu}^* = \sum_{i \in J} \varepsilon_{\boldsymbol{x}_i}$. Therefore, the assertion follows. □

If some of the points of $\boldsymbol{\mu}$ occur repeatedly and these points are also part of $\boldsymbol{\mu}^*$ one has to add a constant which counts these multiple occurrences in the density $g_1^{[S,u]}$. This does not effect the pertaining likelihood function and, therefore, we neglect this case to keep the notation as simple as possible.

It remains to deduce a density of the conditional distribution of $N^{[S,u]}$ given $N_1^{[S,u]} = \boldsymbol{\mu}^*$. Corollary 7.2.1 in [54] (cf. also Theorem 5.12) yields that this distribution of the form

$$\mathcal{L}\left(\sum_{i=1}^{n} \varepsilon_{Z_i}\right)$$

where the $Z_i$ are independent but not necessarily identically distributed (innid). A density of a generalization of such a process is given in Theorem 5.15 in Section 5.3.

In the subsequent lemma we specifically apply Theorem 5.15 with

$$h_i(z, y) = 1_{\{x_i\}}(z) g(y|z)/p(z, y). \tag{5.11}$$

Because of the included indicator function most of the summands in (5.15) vanish.

**Lemma 5.6** *Let $\widetilde{N}_{\boldsymbol{\mu}}$ as in (5.10). Then, $P\left(N^{[S,u]} \in \cdot \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right)$ has the $\mathcal{L}\left(\widetilde{N}_{\boldsymbol{\mu}}\right)$–density*

$$g_2^{[S,u]}(\boldsymbol{\eta}|\boldsymbol{\mu}^*) = \frac{1}{P\left\{\tilde{\beta} = k\right\}} \frac{c_{\boldsymbol{\mu}^*}}{k!} \frac{\prod_{i=1}^{k} f^{[u]}(y_i|\boldsymbol{x}_i^*)}{\prod_{i=1}^{k} p(y_i, \boldsymbol{x}_i^*)}$$

*where $\boldsymbol{\eta} = \sum_{i=1}^{k} \varepsilon_{(\boldsymbol{x}_i^*, y_i)}$, $\boldsymbol{\mu}^* = \pi_1(\boldsymbol{\eta})$ and $c_{\boldsymbol{\mu}^*}$ is a constant which only depends on $\boldsymbol{\mu}^*$ and counts the multiple occurrences of points of $\boldsymbol{\mu}^*$.*

Proof. Let $\boldsymbol{\mu}^* = \sum_{i=1}^{k} \varepsilon_{\boldsymbol{x}_i^*}$. It follows from Corollary 7.2.1 in [54] (cf. also Theorem 5.12) that

$$P\left(N^{[S,u]} \in \cdot \,\middle|\, N_1^{[S,u]} = \boldsymbol{\mu}^*\right) = \mathcal{L}\left(\sum_{i=1}^{\boldsymbol{\mu}^*(S)} \varepsilon_{(\boldsymbol{x}_i^*, \tilde{Y}_i)}\right)$$

where $N^{[S,u]} = \sum_{i=1}^{\tau^*} \varepsilon_{(\boldsymbol{X}_i^*, Y_i^*)}$ (cf. (5.6)) and the $\tilde{Y}_i$, $i = 1, \ldots, n$ are independent with

$$\mathcal{L}\left(\tilde{Y}_i\right) = P\left(Y_i^* \in \cdot \,|\boldsymbol{X}^* = \boldsymbol{x}_i^*\right).$$

Apparently $\left(\boldsymbol{x}_i^*, \tilde{Y}_i\right)$ are independent with $\mathcal{L}((V,Z))$–density

$$h_i(\boldsymbol{z}, y) = 1_{\{\boldsymbol{x}_i^*\}}(\boldsymbol{z}) f^{[u]}(y|\boldsymbol{z})/p(\boldsymbol{z}, y).$$

Now Theorem 5.15 yields that a $\mathcal{L}\left(\widetilde{N}_{\boldsymbol{\mu}}\right)$–density of the conditional distribution of $N_2^{[S,u]}$ given $N_1^{[S,u]} = \boldsymbol{\mu}^*$ is given by

$$\begin{aligned}
g_2^{[S,u]}(\boldsymbol{\eta}|\boldsymbol{\mu}^*) &= \frac{1}{P\left\{\tilde{\beta} = k\right\}} \frac{1}{k!} \sum_{\pi \in \mathcal{P}_k} \prod_{i=1}^{k} h_{\pi(i)}(\boldsymbol{z}_i, y_i) \\
&= \frac{1}{P\left\{\tilde{\beta} = k\right\}} \frac{1}{k!} \sum_{\pi \in \mathcal{P}_k} \prod_{i=1}^{k} \frac{f^{[u]}(y_i|\boldsymbol{z}_i)}{p(\boldsymbol{z}_i, y_i)} 1_{\left\{\boldsymbol{x}_{\pi(i)}^*\right\}}(\boldsymbol{z}_i) \\
&= \begin{cases} \frac{1}{P\{\tilde{\beta}=k\}} \frac{c_{\boldsymbol{\mu}^*}}{k!} \prod_{i=1}^{k} \frac{f^{[u]}(y_i|\boldsymbol{x}_i^*)}{p(\boldsymbol{x}_i^*, y_i)}, & \pi_1(\boldsymbol{\eta}) = \boldsymbol{\mu}^*; \\ & \text{if} \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

with $\boldsymbol{\eta} = \sum_{i=1}^{k} \varepsilon_{(\boldsymbol{z}_i, y_i)}$. Thus the assertion holds. □

It is important to note that the dominating measure $\mathcal{L}\left(\widetilde{N}_{\boldsymbol{\mu}}\right)$ in the previous lemma does not depend on the condition $\boldsymbol{\mu}^*$ since it includes a random sample size. This property will be crucial for the subsequent considerations.

Next we formulate and prove the main result of this section. Combining the foregoing results yields a density of

$$P\left(N^{[S,u]} \in \cdot \,|N_1 = \boldsymbol{\mu}\right)$$

with respect to the measure $\mathcal{L}\left(\widetilde{N}_{\boldsymbol{\mu}}\right)$. This density will be used in Section 5.4 to derive MLEs for the parameters of the upper tail of the conditional distribution of $Y$ given $\boldsymbol{X} = \boldsymbol{x}$.

**Theorem 5.7** *Let $\widetilde{N}_{\boldsymbol{\mu}}$ be as in Lemma 5.6. Then $P\left(N^{[S,u]} \in \cdot\,|N_1 = \boldsymbol{\mu}\right)$ has a $\mathcal{L}\left(\widetilde{N}_{\boldsymbol{\mu}}\right)$ density*

$$\begin{aligned} g^{[S,u]}(\boldsymbol{\eta}|\boldsymbol{\mu}) &= \left(\prod_{i=1}^{k}(1-F(u|\boldsymbol{x}_i^*))\right)\left(\prod_{\boldsymbol{x}\in S_{\boldsymbol{\mu}}} F(u|\boldsymbol{x})^{\boldsymbol{\mu}\{\boldsymbol{x}\}-\boldsymbol{\mu}^*\{\boldsymbol{x}\}}\right) \\ &\qquad \times \frac{1}{P\left\{\tilde{\beta}=k\right\}}\frac{c_{\boldsymbol{\mu}^*}}{k!}\frac{\prod_{i=1}^{k} f^{[u]}(y_i|\boldsymbol{x}_i^*)}{\prod_{i=1}^{k} p(\boldsymbol{x}_i^*, y_i)}, \end{aligned}$$

*where $\boldsymbol{\eta} = \sum_{i=1}^{k} \varepsilon_{(\boldsymbol{x}_i^*, y_i)}$, $\boldsymbol{\mu} = \sum_{i=1}^{n} \varepsilon_{\boldsymbol{x}_i}$, $\boldsymbol{\mu}^* = \pi_1(\boldsymbol{\eta}) \in \mathbb{M}_{\boldsymbol{\mu}}(S)$ and $c_{\boldsymbol{\mu}^*}$ as in Lemma 5.6.*

Proof. According to Lemma 5.3 we have for $A \in \mathcal{M}(S \times \mathbb{R})$

$$\begin{aligned} & P\left(N^{[S,u]} \in A\,|N_1 = \boldsymbol{\mu}\right) \\ =& \int\int 1_A(\boldsymbol{\eta}) dP\left(N^{[S,u]} \in d\boldsymbol{\eta} \,\middle|\, N^{[S,u]} = \boldsymbol{\nu}\right) dP\left(N_1^{[S,u]} \in d\boldsymbol{\nu}\,|N_1 = \boldsymbol{\mu}\right) \\ =& \int\int 1_A(\boldsymbol{\eta}) g_2^{[S,u]}(\boldsymbol{\eta}|\boldsymbol{\nu}) d\mathcal{L}\left(\widetilde{N}_{\boldsymbol{\mu}}\right)(\boldsymbol{\eta}) dP\left(N_1^{[S,u]} \in d\boldsymbol{\nu}\,|N_1 = \boldsymbol{\mu}\right) \\ =& \int\int 1_A(\boldsymbol{\eta}) g_2^{[S,u]}(\boldsymbol{\eta}|\boldsymbol{\nu}) dP\left(N_1^{[S,u]} \in d\boldsymbol{\nu}\,|N_1 = \boldsymbol{\mu}\right) d\mathcal{L}\left(\widetilde{N}_{\boldsymbol{\mu}}\right)(\boldsymbol{\eta}), \end{aligned}$$

thus

$$\begin{aligned} g^{[S,u]}(\boldsymbol{\eta}|\boldsymbol{\mu}) &= \int g_2^{[S,u]}(\boldsymbol{\eta}|\boldsymbol{\nu}) dP\left(N_1^{[S,u]} \in d\boldsymbol{\nu}\,|N_1 = \boldsymbol{\mu}\right) \\ &= \sum_{\boldsymbol{\nu}\in\mathbb{M}_{\boldsymbol{\mu}}(S)} g_2^{[S,u]}(\boldsymbol{\eta}|\boldsymbol{\nu}) g_1^{[S,u]}(\boldsymbol{\nu}|\boldsymbol{\mu}) \\ &= \left(\prod_{i=1}^{k}(1-F(u|\boldsymbol{x}_i^*))\right)\left(\prod_{\boldsymbol{x}\in S_{\boldsymbol{\mu}}} F(u|\boldsymbol{x})^{\boldsymbol{\mu}\{\boldsymbol{x}\}-\boldsymbol{\mu}^*\{\boldsymbol{x}\}}\right) \\ &\qquad \times \frac{1}{P\left\{\tilde{\beta}=k\right\}}\frac{c_{\boldsymbol{\mu}^*}}{k!}\frac{\prod_{i=1}^{k} f^{[u]}(y_i|\boldsymbol{x}_i^*)}{\prod_{i=1}^{k} p(y_i, \boldsymbol{x}_i^*)} \end{aligned}$$

if $\boldsymbol{\mu}^* := \pi_1(\boldsymbol{\eta}) \in \mathbb{M}_{\boldsymbol{\mu}}(S)$ and 0 otherwise is a $\mathcal{L}\left(N_{\boldsymbol{\mu}}\right)$–density of the considered conditional distribution. □

In some applications it is more convenient to use a slightly different representation of $g^{[S,u]}(\cdot|\boldsymbol{\mu})$. Recall that

$$f^{[u]}(y|\boldsymbol{x}) = \frac{f(y|\boldsymbol{x})}{1-F(u|\boldsymbol{x})}, \quad y > u$$

where $f^{[u]}$ is the truncated density pertaining to $f$. Using this relation yields

$$g^{[S,u]}(\boldsymbol{\eta}|\boldsymbol{\mu}) = \left(\prod_{\boldsymbol{x}\in S_{\boldsymbol{\mu}}} F(u|\boldsymbol{x})^{\boldsymbol{\mu}\{\boldsymbol{x}\}-\boldsymbol{\mu}^*\{\boldsymbol{x}\}}\right) \times \frac{1}{P\left\{\tilde{\beta}=k\right\}} \frac{c_{\boldsymbol{\mu}^*}}{k!} \frac{\prod_{i=1}^k f(y_i|\boldsymbol{x}_i^*)}{\prod_{i=1}^k p(\boldsymbol{x}_i^*, y_i)},$$

for $\boldsymbol{\eta}$, $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^*$ as in Theorem 5.7.

### 5.2.3 Extensions to Dependent Covariates

In many applications the assumption of independence of the tuples $(\boldsymbol{X}_i, Y_i)$ cannot be justified. This is also the case in most of the case studies presented in Section 5.5.2.

In this section we will weaken the condition imposed on the $(\boldsymbol{X}_i, Y_i)$. We will show that the conditional density of N—and, consequently, that of $N^{[S,u]}$–given $N_1 = \mu$, is the same as in the Poisson process case under these weaker conditions. Henceforth we will assume that the point process $N = \sum_{i=1}^{\beta} \varepsilon_{(\boldsymbol{X}_i, Y_i)}$ satisfies Condition 5.8 which is stated below. For notational convenience we formulate the following results in a general framework for rvs $X_i$ and $Y_i$ with values in general measurable spaces. When applying this results to our original framework the rvs $X_i$ will be replaced by the vectors of covariates $\boldsymbol{X}_i$ and $Y_i$ will be assumed to be the real valued response variable. The following condition was considered in a similar form in [56], it is closely related to conditions within the framework of regression analysis.

**Condition 5.8** *Consider the point process*

$$N = \sum_{i=1}^{\beta} \varepsilon_{(X_i, Y_i)},$$

*where the $X_i$ and $Y_i$ are rvs with values in measurable spaces $(S, \mathcal{B})$ and $(T, \mathcal{C})$, respectively. We assume that*

*(a) the random sample size $\beta$ is independent of the $(X_i, Y_i)$;*

*(b) for all $n \in \mathbb{N}$ the rvs $Y_1, \dots, Y_n$ are conditionally independent given the vector of covariates $\boldsymbol{X} = (X_1, \dots, X_n)$; the pertaining conditional distributions satisfy*

$$P\left(Y_i \in \cdot\, |\boldsymbol{X} = \boldsymbol{x}\right) = P\left(Y_i \in \cdot\, |X_i = x_i\right)$$

*for all $i = 1, \dots, n$, $\boldsymbol{x} = (x_1, \dots, x_n)$ and all $n \in \mathbb{N}$;*

*(c) finally, we assume that*

$$P\left(Y_i \in \cdot\, |X_i = x\right) = P\left(Y_j \in \cdot\, |X_j = x\right)$$

*for all $x \in S$ and $i, j \in \mathbb{N}$.*

Notice that Condition 5.8 is satisfied if the $(X_i, Y_i)$ are iid and, therefore, it also holds for Poisson processes. Another example of a point process satisfying this condition can be constructed in the following manner.

**Example 1.** Let $X_1, X_2, \dots$ be a sequence of arbitrary rvs and $\xi_1, \xi_2, \dots$ a sequence of iid rvs and define

$$Y_i := h(X_i, \xi_i), \quad i \in \mathbb{N}$$

for some measurable function $h$. Then the pertaining point process

$$N = \sum_{i=1}^{\beta} \varepsilon_{(X_i, Y_i)}$$

with a rv $\beta$ independent of the $(X_i, Y_i)$ satisfies Condition 5.8.

A further example of a point process satisfying the above condition will be given in Section 5.4.5. For applications it is of particular importance that Condition 5.8 contains no assumption on the distribution of the covariates.

The main tool for the derivation of the subsequent results will be a generalization of Theorem 7.2.1 in [54]. We first prove an auxiliary result concerning exchangeable rvs. Exchangeability of rvs is a crucial property when dealing with point processes because this property allows to switch between the concept of point processes and ordinary rvs, cf. also Theorem 5.15. Recall that rvs $X_1, \dots, X_n$ are exchangeable if for each permutation $\pi$ of the set $\{1, \dots, n\}$ we have

$$(X_1, \dots, X_n) \stackrel{d}{=} \left(X_{\pi(1)}, \dots, X_{\pi(n)}\right).$$

For more details we refer to [9], Section 7.3, [2] and [54], page 83ff. For notational convenience we introduce for each permutation $\pi$ a mapping

$$\boldsymbol{\pi}(x_1, \dots, x_n) = (x_{\pi(1)}, \dots, x_{\pi(n)}).$$

Notice that there is a one–to–one relation between $\boldsymbol{\pi}$ and $\pi$.

Henceforth we denote by $\mathcal{P}_n$ the set of all permutations of $\{1, \dots, n\}$, and by $\Psi$ a random permutation which is uniformly distributed on $\mathcal{P}_n$, cf., e.g., [2], page 7. Furthermore, we define in analogy to $\boldsymbol{\pi}$ a random element $\boldsymbol{\Psi}$ pertaining to $\Psi$ which is uniformly distributed on the set of all mappings $\boldsymbol{\pi}$, in the following denoted by $\boldsymbol{\mathcal{P}}_n$.

If $X_1, \dots, X_n$ are arbitrary rvs then

$$\boldsymbol{\Psi}(X_1, \dots, X_n) := (X_{\Psi(1)}, \dots, X_{\Psi(n)})$$

are exchangeable rvs, cf. [2], page 7. In the following lemma we will prove a special property of such transformed rvs provided that Condition 5.8 is satisfied.

**Lemma 5.9** *Let* $(X_1, Y_1), \ldots, (X_n, Y_n)$ *satisfy (b) and (c) in Condition 5.8. Then* $Y_{\Psi(1)}, \ldots, Y_{\Psi(n)}$ *are conditional independent given* $(X_{\Psi(1)}, \ldots, X_{\Psi(n)}) = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ *and*

$$\begin{aligned} & P\left(\left(Y_{\Psi(1)}, \ldots, Y_{\Psi(n)}\right) \in \cdot \mid (X_{\Psi(1)}, \ldots, X_{\Psi(n)}) = (x_1, \ldots, x_n)\right) \\ = \; & P\left((Y_1, \ldots, Y_n) \in \cdot \,|(X_1, \ldots, X_n) = (x_1, \ldots, x_n)\right). \end{aligned}$$

Proof. Define again $\boldsymbol{X} = (X_1, \ldots, X_n)$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ as well as

$$\widetilde{\boldsymbol{Y}} = \left(Y_{\Psi(1)}, \ldots, Y_{\Psi(n)}\right) \text{ and } \widetilde{\boldsymbol{X}} = \left(X_{\Psi(1)}, \ldots, X_{\Psi(n)}\right).$$

Let $\boldsymbol{A} = \bigtimes_{i=1}^n A_i$ and $\boldsymbol{B} = \bigtimes_{i=1}^n B_i$ for measurable sets $A_i$ and $B_i$. We first prove that for every $\boldsymbol{\pi} \in \boldsymbol{\mathcal{P}}_n$ holds

$$P\left(\boldsymbol{\pi}\left(\boldsymbol{Y}\right) \in \boldsymbol{B} \,|\boldsymbol{\pi}\left(\boldsymbol{X}\right) = (x_1, \ldots, x_n)\right) = \prod_{i=1}^n P\left(Y_{\pi(i)} \in B_i \,\middle|X_{\pi(i)} = x_i\right) \tag{5.12}$$

For that purpose let

$$K(B|x) := P\left(Y_1 \in A|X_1 = x\right).$$

One gets with Condition 5.8 (b) and (c)

$$\begin{aligned} & P\left\{\boldsymbol{\pi}\left(\boldsymbol{X}\right) \in \boldsymbol{A}, \, \boldsymbol{\pi}\left(\boldsymbol{Y}\right) \in \boldsymbol{B}\right\} \\ = \; & P\left\{\boldsymbol{X} \in \boldsymbol{\pi}^{-1}\left(\boldsymbol{A}\right), \boldsymbol{Y} \in \boldsymbol{\pi}^{-1}\left(\boldsymbol{B}\right)\right\} \\ = \; & \int_{\boldsymbol{\pi}^{-1}(\boldsymbol{A})} \prod_{i=1}^n K\left(B_{\pi^{-1}(i)}|x_i\right) d\mathcal{L}\left(\boldsymbol{X}\right)(x_1, \ldots, x_n) \\ = \; & \int_{\boldsymbol{\pi}^{-1}(\boldsymbol{A})} \prod_{i=1}^n K\left(B_i|x_{\pi(i)}\right) d\mathcal{L}\left(\boldsymbol{X}\right)(x_1, \ldots, x_n) \\ = \; & \int_{\boldsymbol{A}} \prod_{i=1}^n K(B_i|x_i) d\mathcal{L}\left(\boldsymbol{\pi}\left(\boldsymbol{X}\right)\right)(x_1, \ldots, x_n), \end{aligned}$$

where the last equality follows from the transformation theorem for integrals. Thus equation (5.12) holds. Now notice that

$$
\begin{aligned}
&P\left\{\widetilde{\boldsymbol{X}} \in \boldsymbol{A}, \widetilde{\boldsymbol{Y}} \in \boldsymbol{B}\right\} \\
&= \int_{\boldsymbol{\mathcal{P}}_n} P\left(\widetilde{\boldsymbol{X}} \in \boldsymbol{A}, \widetilde{\boldsymbol{Y}} \in \boldsymbol{B} \,|\, \boldsymbol{\Psi} = \boldsymbol{\pi}\right) d\mathcal{L}\left(\boldsymbol{\Psi}\right)\left(\boldsymbol{\pi}\right) \\
&= \int_{\boldsymbol{\mathcal{P}}_n} \int_{\boldsymbol{A}} \prod_{i=1}^{n} K\left(B_i | x_i\right) d\mathcal{L}\left(\boldsymbol{\pi}\left(\boldsymbol{X}\right)\right) d\mathcal{L}\left(\boldsymbol{\Psi}\right)\left(\boldsymbol{\pi}\right) \\
&= \int_{\boldsymbol{A}} \prod_{i=1}^{n} K\left(B_i | x_i\right) d\mathcal{L}\left(\widetilde{\boldsymbol{X}}\right)\left(x_1, \ldots, x_n\right),
\end{aligned}
$$

where the second equality follows from equation (5.12) and the last from the Fubini theorem for Markov kernels, see e.g. [37], page 45f. Thus, the assertion holds. □

We are now in a position to prove the announced generalization of Theorem 7.2.1 of [54].

**Theorem 5.10** *Let $N$ be a point process as given in Condition 5.8 and $(N_1, N_2) \stackrel{d}{=} (\pi_1(N), \pi_2(N))$. Consider the Markov kernel*

$$
K^*\left(\cdot | \boldsymbol{\mu}\right) = \mathcal{L}\left(\sum_{i=1}^{n} \varepsilon_{Z_i}\right)
$$

*for a point measure $\boldsymbol{\mu} = \sum_{i=1}^{n} \varepsilon_{x_i}$ where the $Z_i$ are independent with distributions $P\left\{Z_i \in \cdot\right\} = P\left(Y_i \in \cdot \,| X_i = x_i\right)$.*
*Then,*

$$
\mathcal{L}(N_2) = K^* \mathcal{L}(N_1).
$$

Proof.
As in [54] put

$$
G_X(\cdot | n) := \mathcal{L}\left(\sum_{i=1}^{n} \varepsilon_{X_i}\right)
$$

and

$$
G_Y(\cdot | n) := \mathcal{L}\left(\sum_{i=1}^{n} \varepsilon_{Y_i}\right).
$$

Define $\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{Y}}$ as in the preceding Lemma 5.9. Then the components of $\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{Y}}$ are exchangeable. Therefore, we have (see also [54], page 84f)

$$
G_X(\cdot | n) = \iota_n \mathcal{L}(\widetilde{\boldsymbol{X}}) \text{ and } G_Y(\cdot | n) = \iota_n \mathcal{L}\left(\widetilde{\boldsymbol{Y}}\right),
$$

with the inclusion mapping

$$\iota_n(x_1,\dots,x_n) = \sum_{i=1}^{n} \varepsilon_{x_i}.$$

Notice that the Markov kernels $G_X(\cdot|n)$ and $G_Y(\cdot|n)$ are not changed if the order of their pertaining points is changed.

Finally, again as in [54] define the Markov kernel

$$K^n\left(\cdot\,|\boldsymbol{x}\right) = \bigtimes_{i=1}^{n} P\left(Y_1 \in \cdot\,|X_1 = x_i\right).$$

for $\boldsymbol{x} = (x_1,\dots,x_n)$.

Due to Lemma 5.9 we have that

$$\mathcal{L}\left(\widetilde{\boldsymbol{Y}}\right) = K^n \mathcal{L}(\widetilde{\boldsymbol{X}}).$$

Using this result the remainder of the prove is almost a verbatim repetition of the proof of Theorem 7.2.1 in [54] with only slight modifications. For the sake of completeness we give the proof in detail, nevertheless.

Because $\beta$ is independent of $Y_i$ and $X_i$, $i = 1,\dots,n$ we have

$$\mathcal{L}(N_1) = G_X\mathcal{L}(\beta) \text{ and } \mathcal{L}(N_2) = G_Y\mathcal{L}(\beta).$$

This representation of $\mathcal{L}(N_1)$ together with the Fubini theorem for Markov kernels immediately yields

$$K^*\mathcal{L}(N_1) = (K^* \circ G_X)\mathcal{L}(\beta),$$

where the composition of two Markov kernels $K_1$ and $K_2$ is defined by

$$(K_1 \circ K_2)(B|x_0) = \int K_1(B|x)K_2(dx|x_0), \tag{5.13}$$

cf. [54], page 172. Thus, it remains to prove that

$$K^* G_X(\cdot|n) = G_Y(\cdot|n)$$

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

As in Lemma 7.2.2 of [54] we get for $\boldsymbol{\mu} = \iota_n\left(\boldsymbol{x}\right)$

$$K^*\left(\cdot\,|\boldsymbol{\mu}\right) = \iota_n\left(K^n\left(\cdot\,\middle|\iota_n^{-1}(\boldsymbol{\mu})\right)\right).$$

This yields for $M \in \mathcal{M}(T)$

$$\begin{aligned}
&K^* G_X(\cdot|n)(M)\\
&= \int K^*(M|\boldsymbol{\mu}) G_X(d\boldsymbol{\mu}|n)\\
&= \int K^n\left(\iota_n^{-1}(M)\left|\iota_n^{-1}(\boldsymbol{\mu})\right.\right) d\left(\iota_n \mathcal{L}\left(\widetilde{\boldsymbol{X}}\right)\right)(\boldsymbol{\mu})\\
&= \int K^n\left(\iota_n^{-1}(M)\left|\boldsymbol{x}\right.\right) d\mathcal{L}\left(\widetilde{\boldsymbol{X}}\right)(\boldsymbol{x})\\
&= K^n \mathcal{L}\left(\widetilde{\boldsymbol{X}}\right)\left(\iota_n^{-1}(M)\right)\\
&= \iota_n\left(K^n \mathcal{L}\left(\widetilde{\boldsymbol{X}}\right)\right)(M)\\
&= \iota_n \mathcal{L}\left(\widetilde{\boldsymbol{Y}}\right)(M)\\
&= G_Y\left(M|n\right),
\end{aligned}$$

and, thus, the assertion holds. □

To apply the forgoing result to the derivation of the conditional distribution of $N$ given $N_1 = \mu$ we construct some auxiliary process $\bar{N}$ and prove that Condition 5.8 is satisfied. The subsequent theorem is now an immediate consequence of Theorem 5.10.

**Lemma 5.11** *Let $N = \sum_{i=1}^{\beta} \varepsilon_{(X_i,Y_i)}$ be a point process as given in Condition 5.8. Put*

$$\bar{N} = \sum_{i=1}^{\beta} \varepsilon_{(\bar{X}_i,\bar{Y}_i)}$$

*with $\bar{X}_i = X_i$ and $\bar{Y}_i = (X_i, Y_i)$. Then, $\bar{N}$ also satisfies Condition 5.8.*

Proof. Put $\boldsymbol{X} = (X_1, \ldots, X_n)$. Check that (cf. Corollary A.2)

(i) $(X_i, Y_i)$ are conditional independent given $\boldsymbol{X} = \boldsymbol{x}$.

(ii) For the conditional distribution of $(X_i, Y_i)$ given $\boldsymbol{X} = \boldsymbol{x}$ holds

$$P\left((X_i, Y_i) \in \cdot \,|\, \boldsymbol{X} = \boldsymbol{x}\right) = \mathcal{L}\left((x_i, Y_{x_i})\right)$$

where $\mathcal{L}(Y_x) := P\left(Y_1 \in \cdot | X_1 = x\right)$.

Now the assertion follows immediately. □

The subsequent theorem is now an immediate consequence of Theorem 5.10 and Lemma 5.11.

**Theorem 5.12** *Let* $N = \sum_{i=1}^{\tau} \varepsilon_{(X_i,Y_i)}$ *be a point process as in Condition 5.8,* $N_1 := \pi_1(N)$*. Then*

$$P(N \in \cdot\,|N_1 = \boldsymbol{\mu}) = \mathcal{L}(N_{\boldsymbol{\mu}})$$

*where*

$$N_{\boldsymbol{\mu}} = \sum_{i=1}^{\boldsymbol{\mu}(S)} \varepsilon_{(x_i,Z_i)}$$

*with independent rvs* $Z_i$*,* $i = 1, \dots, \boldsymbol{\mu}(S)$ *with distributions* $\mathcal{L}(Z_i) = P(Y_1 \in \cdot\,|X_1 = x_i)$*.*

Proof. Define the mapping

$$g(N) = (N, \pi_1(N)).$$

Observe that $(N, \pi_1(N)) = (\pi_2(\bar{N}), \pi_1(\bar{N}))$ where $\bar{N}$ is as defined in Lemma 5.11. Applying Theorem 5.10 to the point process $\bar{N}$ one gets

$$\begin{aligned} P\{N \in A,\ N_1 \in B\} &= P\{g(N) \in A \times B\} \\ &= \int P\{g(N_{\boldsymbol{\mu}}) \in A \times B\}\, d\mathcal{L}(N_1)(\boldsymbol{\mu}) \\ &= \int P\{N_{\boldsymbol{\mu}} \in A, \boldsymbol{\mu} \in B\}\, d\mathcal{L}(N_1)(\boldsymbol{\mu}) \\ &= \int_B P\{N_{\boldsymbol{\mu}} \in A\}\, d\mathcal{L}(N_1)(\boldsymbol{\mu}). \end{aligned}$$

□

Combining the foregoing results yields the main result of this section which is stated in the subsequent theorem.

**Theorem 5.13** *If the point process* $N = \sum_{i=1}^{\beta} \varepsilon_{(\boldsymbol{X}_i,Y_i)}$ *satisfies Condition 5.8 where* $X_i$ *is replaced by* $\boldsymbol{X}_i$ *for* $i = 1, \dots, \beta$ *the conditional distributions of the truncation of* $N$ *outside* $S \times (u, \infty)$ *given the covariates*

$$P\left(N^{[S,u]} \in \cdot\,|N_1 = \boldsymbol{\mu}\right), \quad \boldsymbol{\mu} \in \mathbb{M}(S)$$

*has the density given in Theorem 5.7.*

Proof. According to Theorem 5.12 the conditional distribution

$$P(N \in \cdot\,|N_1 = \boldsymbol{\mu})$$

is given by the distribution of the point process $N_{\boldsymbol{\mu}}$ as defined in Theorem 5.12. Define the measurable truncation mapping $t_u : \mathbb{M}(S \times \mathbb{R}) \to \mathbb{M}(S \times (u, \infty))$ by

$$t_u(\boldsymbol{\eta}) = \eta(\cdot \cap (S \times (u, \infty))).$$

Thus, we may write $N^{[S,u]} = t_u(N)$ and one gets $A \in \mathcal{M}(S \times (u, \infty))$

$$\begin{aligned} P\left(N^{[S,u]} \in A \,|N_1 = \boldsymbol{\mu}\right) &= P\left(t_u(N) \in A \,|N_1 = \boldsymbol{\mu}\right) \\ &= P\left(N \in t_u^{-1}(A) \,|N_1 = \boldsymbol{\mu}\right) \\ &= P\left\{N_{\boldsymbol{\mu}} \in t_u^{-1}(A)\right\} \\ &= P\left\{t_u(N_{\boldsymbol{\mu}}) \in A\right\}. \end{aligned}$$

Hence the conditional density of $N^{[S,u]}$ given $N_1 = \boldsymbol{\mu}$ depends only on the distribution of $N_{\boldsymbol{\mu}}$. This distribution is the same under Condition 5.8 as in the case of Poisson processes considered in Section 5.2.2. Consequently Theorem 5.7 is still valid under weaker conditions. □

The following section contains the proofs of auxiliary results which have been used to proof the preceding results.

## 5.3 Auxiliary Results

In this section we extend results stated in [54] to a more general setting. The main results of this section is stated in Theorem 5.15, which gives the density of a mixed empirical point process of independent, not necessary identically distributed (innid) observations.

**Lemma 5.14** *Let $X_1, \ldots, X_n$ be innid rvs and*

$$N = \sum_{i=1}^{n} \varepsilon_{X_i}$$

*the pertaining point process. Furthermore let $\mathcal{P}_n$ the set of permutations of the set $\{1, \ldots, n\}$ and $\Psi$ uniformly distributed on $\mathcal{P}_n$, independent of $X_1, \ldots, X_n$. Define*

$$Z_i = X_{\Psi(i)}, \quad i = 1, \ldots, n$$

*and let*

$$N^* = \sum_{i=1}^{n} \varepsilon_{Z_i}.$$

*Then*

$$(N^*)^{\omega} = N^{\omega} \quad \textit{for all } \omega \in \Omega$$

*and, therefore, especially*

$$\mathcal{L}(N^*) = \mathcal{L}(N).$$

Proof. The assertion can easily be proven because

$$\sum_{i=1}^{n} \varepsilon_{x_i} = \sum_{i=1}^{n} \varepsilon_{x_{\pi(i)}}$$

for all $\pi \in \mathcal{P}_n$ and $x_i \in S$, $i \in \mathbb{N}$. □

From Theorem 3.1.1 of [54] we know the density of distributions of Poisson point processes as well as empirical point processes

$$N = \sum_{i=1}^{n} \varepsilon_{X_i}$$

where $X_i$, $i = 1, \dots, n$, are iid rvs. In the following theorem we extent the results concerning mixed empirical point processes to point processes of the form

$$N = \sum_{i=1}^{\beta} \varepsilon_{Z_i} \tag{5.14}$$

where $Z_i$, $i = 1, \dots, n$ are innid rvs independent of $\beta$.

Consider the random permutation $\Psi$ introduced in the lines preceding Lemma 5.9. This transformation will also play a crucial role in the proof of the subsequent theorem.

Recall that for arbitrary rvs $X_1, \dots, X_n$ the transformed rvs

$$X_{\Psi(1)}, \dots, X_{\Psi(n)}$$

are exchangeable (cf. [2], page 7).

**Theorem 5.15** *Assume that*

*(a)* $Z_1, Z_2, \dots$ *are innid rvs with values in a measurable space* $S$ *independent of the integer–valued rv* $\beta_1$ *and put* $N = \sum_{i=1}^{\beta_1} \varepsilon_{Z_i}$*;*

*(b)* $h_i$ *is a density of* $Z_i$ *with respect to some dominating measure* $\nu$*;*

*(c) furthermore,* $\tilde{Z}_1, \tilde{Z}_2, \dots$ *are iid rvs, independent of the integer–valued rv* $\beta_0$ *with* $\nu$*–density* $p$*, which is positive on* $S$*. Again,* $\widetilde{N}$ *is the pertaining mixed empirical point process;*

*(d) moreover,* $\mathcal{L}(\beta_0)$ *dominates* $\mathcal{L}(\beta_1)$*.*

*Then,* $\mathcal{L}(N)$ *has a* $\mathcal{L}\left(\widetilde{N}\right)$*–density*

$$g(\boldsymbol{\mu}) = \frac{1}{n!}\frac{P\{\beta_1 = n\}}{P\{\beta_0 = n\}}\sum_{\pi\in\mathcal{P}_n}\left(\prod_{i=1}^{n} h_{\pi(i)}(x_i)/p(x_i)\right) \tag{5.15}$$

*where* $\boldsymbol{\mu} = \sum_{i=1}^{n}\varepsilon_{x_i}$ *and* $\mathcal{P}_n$ *is the set of all permutations of the set* $\{1,\dots,n\}$

Proof. First let $k \in \mathbb{N}$

$$\mathcal{L}(N_k) := P\left(N \in \cdot|\beta_0 = k\right) \tag{5.16}$$

and

$$\widetilde{N}_k = \sum_{i=1}^{k}\varepsilon_{\tilde{Z}_i}.$$

Let $x_1,\dots,x_k \in S$. We use again the inclusion mapping

$$\iota_k(x_1,\dots,x_k) = \sum_{i=1}^{k}\varepsilon_{x_i}$$

(cf. [54], page 84) getting

$$N_k \overset{d}{=} \sum_{i=1}^{k}\varepsilon_{Z_{\Psi(i)}}$$

where $\Psi$ is uniformly distributed on $\mathcal{P}_k$, independent of $Z_1,\dots,Z_k$. Observe that for $A_i \in \mathcal{B}$ the following equation holds

$$\begin{aligned}
&P\left\{Z_{\Psi(i)} \in A_i, i = 1,\dots,k\right\}\\
&= \int_{\mathcal{P}_k} P\left(Z_{\Psi(i)} \in A_i, i = 1,\dots,k\,|\Psi = \pi\right) d\mathcal{L}(\Psi)(\pi)\\
&= \int_{\mathcal{P}_k} P\left\{Z_{\pi(i)} \in A_i, i = 1,\dots,k\right\} d\mathcal{L}(\Psi)(\pi)\\
&= \int_{\mathcal{P}_k}\prod_{i=1}^{k} P\left\{Z_{\pi(i)} \in A_k\right\} d\mathcal{L}(\Psi)(\pi)\\
&= \frac{1}{k!}\sum_{\pi\in\mathcal{P}_k}\prod_{i=1}^{k} P\left\{Z_{\pi(i)} \in A_i\right\}.
\end{aligned}$$

Therefore,

$$h(x_1,\dots,x_k) = \frac{1}{k!}\sum_{\pi\in\mathcal{P}_k}\prod_{i=1}^{k} h_{\pi(i)}(x_i)$$

is a $\lambda^k$–density of $\mathcal{L}\left(Z_{\Psi(1)}, \ldots, Z_{\Psi(k)}\right)$. Thus

$$f(x_1, \ldots, x_k) = \frac{h(x_1, \ldots, x_k)}{\prod_{i=1}^k p(x_i)}$$

is a $\mathcal{L}\left(\tilde{Z}_1, \ldots, \tilde{Z}_k\right)$–density of $\mathcal{L}\left(Z_{\Psi(1)}, \ldots, Z_{\Psi(k)}\right)$. Check that $f$ is symmetric, that is,

$$f(x_1, \ldots, x_k) = f(x_{\pi(1)}, \ldots, x_{\pi(k)}) \text{ for all } \pi \in \mathcal{P}_k.$$

Using the same arguments as in the proof of Theorem 3.1.1 of [54] one gets

$$f = g_k \circ \iota_k$$

where

$$g_k(\boldsymbol{\mu}) = \frac{1}{k!} \sum_{\pi \in \mathcal{P}_k} \left( \prod_{i=1}^k h_{\pi(i)}(x_i)/p(x_i) \right)$$

and $\boldsymbol{\mu} = \sum_{i=1}^k \varepsilon_{x_i}$. Consequently, $g_k$ is a density of $N_k$ with respect to $\mathcal{L}\left(\widetilde{N}_k\right)$. Furthermore, for $A \in \mathcal{M}(S)$ one gets

$$\begin{aligned} P\{N \in A\} &= \int P\{N_k \in A\} \, d\mathcal{L}(\beta_1)(k) \\ &= \int \int_A g_k(\boldsymbol{\mu}) d\mathcal{L}\left(\widetilde{N}_k\right)(\boldsymbol{\mu}) d\mathcal{L}(\beta_1)(k). \end{aligned}$$

According to Theorem 3.1.2 of [54]

$$\tilde{g}_k(\boldsymbol{\mu}) = \begin{cases} \frac{1}{P\{\beta_0 = k\}}, & \boldsymbol{\mu}(S) = k \text{ and } P\{\beta_0 = k\} > 0 \\ & \text{if} \\ 0, & \text{otherwise.} \end{cases}$$

is a density of $\widetilde{N}_k$ with respect to $\widetilde{N}$. Therefore,

$$\begin{aligned} P\{N \in A\} &= \int \int_A g_k(\boldsymbol{\mu}) \tilde{g}_k(\boldsymbol{\mu}) d\mathcal{L}\left(\widetilde{N}\right)(\boldsymbol{\mu}) d\mathcal{L}(\beta_1)(k) \\ &= \int_A \sum_{k \in \mathbb{N}} g_k(\boldsymbol{\mu}) P\{\beta_1 = k\} \tilde{g}_k(\boldsymbol{\mu}) d\mathcal{L}\left(\widetilde{N}\right)(\boldsymbol{\mu}) \\ &= \int_A g_n(\boldsymbol{\mu}) \frac{P\{\beta_1 = n\}}{P\{\beta_0 = n\}} d\mathcal{L}\left(\widetilde{N}\right)(\boldsymbol{\mu}) \end{aligned}$$

if $\boldsymbol{\mu}(S) = n$. Now the assertion follows immediately. □

In the following section we give an overview of likelihood based methods for estimating the tail of a conditional distribution.

## 5.4 Statistical Inference

The recent statistical literature offers a vast repertory of likelihood based methods for estimating tails of conditional distributions. The concept of point processes and in particular Poisson and binomial point processes has proven to be a powerful toolbox for the solution of many of such statistical problems, especially when covariate information is included in the analysis. These methods are used in a wide range of applications, for example finance (e.g. [8] and Chapter 7 in [66]), insurance and environmental statistics (e.g. [41], [42],[60],[61]) among others. All authors focus on the generalized Pareto distribution as a model for distribution tails. A more mathematical overview of these statistical models can be found in [11].

### 5.4.1 Modeling Upper Tails of Conditional Distributions

Recall our basic representation of the conditional distribution $F(\cdot|\boldsymbol{x})$ in the upper tail

$$F(y|\boldsymbol{x}) = F(u|\boldsymbol{x}) + (1 - F(u|\boldsymbol{x}))L_{\vartheta(\boldsymbol{x})}(y), \quad y > u. \tag{5.17}$$

Basically we can choose between three approaches to the estimation of $F(y|\boldsymbol{x})$, which may be classified as non–parametric, parametric/nonparametric and parametric. Given a data set $(\boldsymbol{x}_i, y_i)$, $i = 1, ..., n$ the first approach requires to choose a neighborhood of $\boldsymbol{x}$ and then use those data among the $(\boldsymbol{x}_i, y_i)$ where $\boldsymbol{x}_i$ lies in this neighborhood. Assuming that these data approximately behave as if they come from the df (5.17) we can derive non–parametric estimations of $F(u|\boldsymbol{x})$ and $F^{[u]}(\cdot|\boldsymbol{x})$ by replacing them with the pertaining empirical versions.This approach yields reasonable results for the conditional exceedance probability $F(u|\boldsymbol{x})$ if a moderate high threshold $u$ is chosen and we have enough observations in the neighborhood of $\boldsymbol{x}$. Yet the purely non–parametric approach is not reliable as regards the conditional distribution of exceedances $F^{[u]}(\cdot|x)$, because typically one has only few observations in this region and this problem gets even worse because we can only use data within the chosen neighborhood of $\boldsymbol{x}$.

A combined parametric/nonparametric approach suggests to estimate the conditional exceedance probability using an empirical approach as described in the preceding lines and replace $F^{[u]}(\cdot|\boldsymbol{x})$ with a parametric model. This approach is typically used in the pot–approach for unconditional distributions and one may speak of a piecing–together approach.

If the conditional distribution of $Y$ given $\boldsymbol{X} = \boldsymbol{x}$ is in the pot–domain of attraction of some general pot–stable distributions $L(\cdot|\boldsymbol{x})$ we have (after a suitable transformation see e.g. (2.27))

$$F^{[u]}(y|\boldsymbol{x}) \approx L(y|\boldsymbol{x}), \quad y > u. \tag{5.18}$$

Furthermore, we require for all $\boldsymbol{x} \in S$ that $L(\cdot|\boldsymbol{x})$ belongs to a parametric family $L_\xi$, $\xi \in \Xi$ and that the dependence on $\boldsymbol{x}$ can be expressed as a functional relation of the parameter and $\boldsymbol{x}$, thus

$$L(y|\boldsymbol{x}) = L_{\vartheta(\boldsymbol{x})}(y)$$

for some $\vartheta : S \to \Xi$. Therefore,

$$F(y|\boldsymbol{x}) = F(u|\boldsymbol{x}) + (1 - F(u|\boldsymbol{x}))L_{\vartheta(\boldsymbol{x})}(y), \quad y > u. \tag{5.19}$$

A drawback of this approach is that the empirical estimation of $F(u|\boldsymbol{x})$ is only reliable if enough observations are available in the neighborhood of $\boldsymbol{x}$. This is typically the case if $\boldsymbol{x}$ is in the center of the support of the distribution of the covariate but not anymore if $\boldsymbol{x}$ lies close to the boundaries.

In the latter case one can use a complete parametric approach by modeling the function $F(u|\cdot)$ also in a parametric way. This method also allows the extrapolation into the regions where no observations of the covariates are available. On the other hand a misspecification of the parametric model may lead to serious misjudegments. Hence, the model choice remains a delicate problem and this is also true for the function $\vartheta$. In contrast to the choice of the distribution family $L_\xi$ no asymptotic results are available which suggests certain models.

The aim of the subsequent analysis is the estimation of the unknown functions $\vartheta : S \to \Xi$ and $F(u|\cdot) : S \to (0, 1)$.

### 5.4.2 Estimation Using Unconditional Likelihoods

In what follows we use a fully parametric approach as regards the family of conditional distributions. That is, we assume that the functions $\vartheta$ and $F(u|\cdot)$ belong to parametric family $(\vartheta_\theta, F_\theta(u|\cdot))$, $\theta \in \Theta$ while we avoid any parametric assumption for the distribution of the covariates. If the available observations are iid we might use the Poisson or binomial process approach based on densities (5.8) or (5.9). If we set

$$p_\theta := 1 - \int F_\theta(u|\boldsymbol{x}) d\mathcal{L}(\boldsymbol{X})(\boldsymbol{x})$$

the pertaining likelihood functions based on a point process $\boldsymbol{\eta} = \sum_{i=1}^n \varepsilon_{\left(\boldsymbol{x}_i^*, y_i\right)}$ satisfy

$$L_{\boldsymbol{\eta}}(\theta) \propto \left( \prod_{i=1}^k l_{\vartheta_\theta(\boldsymbol{x}_i^*)}(y_i) \right) \exp\left(\lambda(1 - p_\theta)\right) \tag{5.20}$$

if $N^{[S,u]}$ is a Poisson process and

$$L_{\boldsymbol{\eta}}(\theta) \propto \left( \prod_{i=1}^k l_{\vartheta_\theta(\boldsymbol{x}_i^*)}(y_i) \right) p_\theta^k (1 - p_\theta)^{n-k} \tag{5.21}$$

if $N^{[S,u]}$ is a binomial process, where $l_{\vartheta_\theta(\boldsymbol{x})}$ is the Lebesgue density pertaining to $L_{\vartheta_\theta(\boldsymbol{x})}$. Notice that $p_\theta$ includes the parameter $\theta$ which has to be estimated and the distribution of the covariates. If we want to avoid any distributional assumption for the covariates except independence we have to replace $\mathcal{L}(\boldsymbol{X})$ with a non–parametric estimate for example a kernel estimate or an empirical distribution. In the latter case one may substitute $p_\theta$ with

$$\hat{p}_{\boldsymbol{\mu},\theta} = 1 - \frac{1}{n}\sum_{i=1}^{n} F_{\vartheta_\theta}(u|\boldsymbol{x}_i)$$

if $\boldsymbol{\mu} = \sum_{i=1}^{n} \varepsilon_{\boldsymbol{x}_i}$ is again the point measure pertaining to all observed covariates and maximize the approximate likelihood functions

$$\hat{L}_{\boldsymbol{\eta},\boldsymbol{\mu}}(\theta) = \left(\prod_{i=1}^{k} l_{\vartheta_\theta(\boldsymbol{x}_i^*)}(y_i)\right) \exp\left(\lambda(1-\hat{p}_{\boldsymbol{\mu},\theta})\right) \tag{5.22}$$

or

$$\hat{L}_{\boldsymbol{\eta},\boldsymbol{\mu}}(\theta) = \left(\prod_{i=1}^{k} l_{\vartheta_\theta(\boldsymbol{x}_i^*)}(y_i)\right) \hat{p}_{\boldsymbol{\mu},\theta}^{k}(1-\hat{p}_{\boldsymbol{\mu},\theta})^{n-k}. \tag{5.23}$$

Obviously $L_{\boldsymbol{\eta}}$ and $\hat{L}_{\boldsymbol{\eta},\boldsymbol{\mu}}$ do not necessarily attain their respective maxima for the same value of $\theta$ as well in the Poisson as in the binomial process case. Simulations show that the MLE based on this likelihoods exhibits a poor performance. This might be due to the described approximation and/or identifiability problems of the parameters in the integral term. It is especially sensitive to starting values used for the numerical maximization. Furthermore, both models are only valid if the original observations $(\boldsymbol{X}_i, Y_i)$ are iid, which can not be justified for a lot of applications.

The likelihood function (5.22) is recently one of the benchmark approaches in the statistical literature. For more details we refer to the review of likelihood based methods in the statistical literature in Section 5.4.5.

### 5.4.3 Conditional Maximum–Likelihood Estimation

To overcome the problems described in the foregoing section one may use the conditional density derived in Theorem 5.7. The conditional likelihood function pertaining to the distribution of $N^{[S,u]}$ given $N_1 = \boldsymbol{\mu}$ satisfies

$$L_{\boldsymbol{\eta}|\boldsymbol{\mu}}(\theta) \quad\propto\quad \left(\prod_{i=1}^{k} l_{\vartheta_\theta(\boldsymbol{x}_i^*)}(y_i)\right)\left(\prod_{\boldsymbol{x}\in S_{\boldsymbol{\mu}}} \left(F_\theta(u|\boldsymbol{x}_i)\right)^{\boldsymbol{\mu}\{\boldsymbol{x}\}-\boldsymbol{\mu}^*\{\boldsymbol{x}\}}\right)\left(\prod_{i=1}^{k}(1-F_\theta(u|\boldsymbol{x}_i^*))\right)$$

for $\boldsymbol{\eta} = \sum_{i=1}^{n} \varepsilon_{(\boldsymbol{x}_i^*, y_i)}$ and $\boldsymbol{\mu} = \sum_{i=1}^{n} \varepsilon_{\boldsymbol{x}_i}$.

The most simple, yet broadly used model is to assume $L$ to be a generalized Pareto distribution,

$$L_\theta = W_{\gamma,u,\sigma}$$

and

$$\vartheta(\boldsymbol{x}) = (\gamma(\boldsymbol{x}), \tilde{\sigma}(\boldsymbol{x}))$$

for some functions $\gamma : S \to \mathbb{R}$ and $\tilde{\sigma} : S \to \mathbb{R}^{>0}$ which are known up to some set of real valued parameters. In this special case (5.19) can be rewritten as

$$F(y|\boldsymbol{x}) = W_{\vartheta(\boldsymbol{x})}(y), \quad y > u$$

where $\vartheta(\boldsymbol{x}) = (\gamma(\boldsymbol{x}), \mu(\boldsymbol{x}), \sigma(\boldsymbol{x}))$ and $\mu : S \to \mathbb{R}$ and $\sigma : S \to \mathbb{R}$ are functions given by

$$\mu(\boldsymbol{x}) = u - F(u|\boldsymbol{x})^{\gamma(\boldsymbol{x})} \frac{\tilde{\sigma}(\boldsymbol{x})}{\gamma(\boldsymbol{x})}$$

and

$$\sigma(\boldsymbol{x}) = \tilde{\sigma}(\boldsymbol{x})(1 - F(u|\boldsymbol{x}))^{\gamma(\boldsymbol{x})}.$$

We assume that all these functions are known up to a parameter $\theta \in \Theta$ where $\Theta \subset \mathbb{R}^k$, $k \in \mathbb{N}$. Given realizations $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ of the point process $N^{[S,u]}$ and $N_1$, $\theta$ may be estimated by maximizing the likelihood function

$$L_{\boldsymbol{\eta}|\boldsymbol{\mu}}(\theta) \propto \left( \prod_{\boldsymbol{x} \in S_{\boldsymbol{\mu}}} \left( W_{\vartheta_\theta(\boldsymbol{x})}(u) \right)^{\boldsymbol{\mu}\{\boldsymbol{x}\} - \boldsymbol{\mu}^*\{\boldsymbol{x}\}} \right) \left( \prod_{i=1}^{k} w_{\vartheta_\theta(\boldsymbol{x}_i^*)}(y_i) \right) \tag{5.24}$$

where $\boldsymbol{\eta} = \sum_{i=1}^k \varepsilon_{(\boldsymbol{x}_i^*, y_i)}$, $\boldsymbol{\mu} = \sum_{i=1}^n \varepsilon_{\boldsymbol{x}_i}$ and $\boldsymbol{\mu}^* = \pi_1(\boldsymbol{\eta}) \in \mathbb{M}_{\boldsymbol{\mu}}(S)$. This is a likelihood function also utilized in [25]. In the following simulations and case studies we will concentrate on this model.

### 5.4.4 Model Checking

The model checking in the GPD case is mainly done using q–q–plots. We differ two cases, whether the shape parameter $\gamma$ is modeled depending on the covariate information or not. In latter case the data $(\boldsymbol{x}_i^*, y_i)$, $i = 1, \ldots, k$ are transformed using the estimated functions $\sigma_{\hat{\theta}}$ and $\mu_{\hat{\theta}}$ using the transformation

$$\widetilde{y_i} := \frac{y_i - \mu_{\hat{\theta}}(\boldsymbol{x}_i^*)}{\sigma_{\hat{\theta}}(\boldsymbol{x}_i^*)}.$$

The resulting values are plotted against the pertaining quantiles of a GPD with parameters $\gamma = \hat{\gamma}$, $\mu = 0$ and $\sigma = 1$. If the shape parameter also depends on the covariates we apply a different transformation given by

$$\widetilde{y}_i := \frac{1}{\gamma_{\hat{\theta}}(\boldsymbol{x}_i^*)} \log\left(1 + \gamma_{\hat{\theta}}(\boldsymbol{x}_i^*) \frac{y_i - \mu_{\hat{\theta}}(\boldsymbol{x}_i^*)}{\sigma_{\hat{\theta}}(\boldsymbol{x}_i^*)}\right)$$

and apply a standard exponential q–q–plot to the residuals $\widetilde{y}_i$.

### 5.4.5 Likelihood Based Methods in the Literature

A first attempt to use a combination of EVT and the theory of point processes to model non–homogeneous extreme observation can be found in [60]. The author introduces a likelihood function which corresponds to unconditional density of the point process $N$ in the Poisson process case as given in (5.22). The studied data consists of a time series of exceedances among daily ground–level ozone concentration measurements over a thirteen years period in Houston, Texas. The arrival times are treated similarly as the covariates in the model proposed in Section 5.2.2, yet they are not random in this context.

In [61] the authors introduce an extension of [60]. The purpose of this article is to study whether changes in tropospheric ozone concentrations are due to a temporal trend or also influenced by the fluctuation of certain meteorological variables such as wind and temperature. The adopted model is similar to that of [60]. The observed data consists of daily observations $(Y_i, T_i)$ where $Y_i$ is the measured ozone concentration and $T_i$ the pertaining day of observation counted from a fixed starting date. Since daily measurements during a period $[1, t^*]$ are considered we know that $T_i$ takes values in the finite set $\{1, \ldots, t^*\}$.

In addition, observations of certain meteorological variables $\boldsymbol{X}_t$ are available for each point in time $t = 1, \ldots, t^*$. Let $\underline{\boldsymbol{X}} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{t^*})$ the vector of all observed covariates. For the further analysis only those values among the $Y_i$ are considered which exceed the threshold $u$, let $(Y_i^*, T_i^*)$, $i = 1, \ldots, k$ denote these exceedances.

The resulting data are used for the modeling of the likelihood function. It is assumed that, given the outcome $\underline{\boldsymbol{x}}$ of the meteorological variables, the exceedance times $T_i^*$ are arrival times of an inhomogeneous marked Poisson process on the observation interval and the $Y_i^*$ are interpreted as pertaining marks. Estimations for the unknown parameters are based on the (unconditioned) density of the Poisson process.

Another development direction based on multivariate point processes leads to to a likelihood function which is proportional to the conditional likelihood function (5.24). As a reference we refer to [25].

We present a further approach to constructing a model which satisfies Condition 5.8, and later discuss a relation to the model applied in [61] for analyzing ozone data.

**Example 2.** Assume we observe random vectors $(X_i, Y_i)$, $i = 1, \ldots, n$, where the $Y_i$ are response variables and the $X_i$ are covariates. Furthermore, we observe an additional vector of covariates $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$. Assume that the $(X_i, Y_i)$ are conditionally iid given $\boldsymbol{Z} = \boldsymbol{z}$. Put

$$\mathcal{L}(X_{i,\boldsymbol{z}}, Y_{i,\boldsymbol{z}}) := P\left((X_i, Y_i) \in \cdot | \boldsymbol{Z} = \boldsymbol{z}\right).$$

Apparently,

$$\mathcal{L}\left(Y_{i,\boldsymbol{z}}\right) = P\left(Y_i \in \cdot \,| \boldsymbol{Z} = \boldsymbol{z}\right) \quad \text{and} \quad \mathcal{L}\left(X_{i,\boldsymbol{z}}\right) = P\left(X_i \in \cdot \,| \boldsymbol{Z} = \boldsymbol{z}\right).$$

Again we specify the parametric model in terms of the conditional distribution

$$P\left(Y_{i,\boldsymbol{z}} \le y \,| X_{i,\boldsymbol{z}} = x\right).$$

We assume again that

$$P\left(Y_{i,\boldsymbol{z}} \le y \,| X_{i,\boldsymbol{z}} = x\right) = L_{\vartheta_{\boldsymbol{z}}(x)}(y), \quad y > u. \tag{5.25}$$

According to Lemma A.4 we have

$$P\left(Y_{i,\boldsymbol{z}} \in \cdot \,| X_{i,\boldsymbol{z}} = x\right) = P(Y_i \in \cdot | X_i = x, \boldsymbol{Z} = \boldsymbol{z}) \tag{5.26}$$

In addition, assume that

$$P(Y_i \in \cdot | X_i = x, \boldsymbol{Z} = \boldsymbol{z}) = P(Y_i \in \cdot | X_i = x, Z_i = z_i).$$

Now with $\tilde{Y}_i = Y_i$ and $\tilde{X}_i = (X_i, Z_i)$ one gets the point process

$$N = \sum_{i=1}^{n} \varepsilon_{(\tilde{X}_i, \tilde{Y}_i)}$$

which satisfies Condition 5.8. The conditional independence of the $\tilde{Y}_i$ given the $\tilde{X}_i$ follows from (5.26) and the assumption that the $(X_i, Y_i)$ are conditionally iid given $\boldsymbol{Z} = \boldsymbol{z}$.

Example 2 makes the Poisson process framework applicable in a non–iid case. Put

$$N_{\boldsymbol{z}} = \sum_{i=1}^{n} \varepsilon_{(X_{i,\boldsymbol{z}}, Y_{i,\boldsymbol{z}})}$$

and let $N_{\boldsymbol{z}}^{[S,u]}$ the pertaining truncation outside of $S \times (u, \infty)$. In addition assume that $N_{\boldsymbol{z}}^{[S,u]}$ is a Poisson process, satisfying

$$N_{\boldsymbol{z}}^{[S,u]} \stackrel{d}{=} \sum_{i=1}^{\tau_{\boldsymbol{z}}^*} \varepsilon_{\left(X_{i,\boldsymbol{z}}^*, Y_{i,\boldsymbol{z}}^*\right)}$$

with a Poisson rv $\tau_{\boldsymbol{z}}^*$. Then, the estimation of the unknown parameters can be carried out within the likelihood framework based on a density of $N_{\boldsymbol{z}}^{[S,u]}$ in complete analogy to the procedure described in the introduction.

Such a model was introduced in [61]. These authors use the truncated point process $N_{\boldsymbol{z}}^{[S,u]}$ as the starting point for their modeling of exceedances for ozone concentrations $Y_{i,\boldsymbol{z}}^*$ given a set of further meteorological measurements $\boldsymbol{Z}$. More precisely, we have

$$N_{\boldsymbol{z}}^{[S,u]} \stackrel{d}{=} \sum_{i=1}^{\tau_{\boldsymbol{z}}^*} \varepsilon_{\left(T_{i,\boldsymbol{z}}^*, Y_{i,\boldsymbol{z}}^*\right)},$$

where the role of the $X_{i,\boldsymbol{z}}^*$ is played by the exceedance times $T_{i,\boldsymbol{z}}^*$, which are assumed to be uniformly distributed on an interval $[0, t^*]$. The random sample size $\tau_{\boldsymbol{z}}^*$ is a Poisson rv.

This model was later applied to several environmental data in [41] and used in [65] and [66] to analyze financial time series.

## 5.5 Applications and Simulations

As already noted at the beginning of this chapter all described statistical procedures have a common aim, namely statistical inference for the upper tail of a conditional distribution

$$P\left(Y \in \cdot \,|\boldsymbol{X} = \boldsymbol{x}\right)$$

based on observations $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$. The statistical methodology has to be adopted to the properties of the available data. The simplest case appears if iid copies of $(\boldsymbol{X}, Y)$ are available. In that case the statistical analysis can be carried out in the Poisson process framework and ML–estimation can be based on (5.20) or (5.24). It turns out that (5.24) yields superior results in the simulations studies presented in this section. A possible explanation is that (5.20) still includes an integral term which cannot be evaluated in a closed form and thus has to approximated.

If the iid assumption is not valid for our data we have to search for different approaches. It is shown in [44] that the Poisson process approximation of the process of exceedances is still valid if not all exceedances but only maxima of clusters of exceedances are considered. A cluster occurs if the threshold is exceeded at several consecutive days. It is evident that this approach makes use only of a part of the data which is a problem because usually one has already only few observations in the tail and, therefore, wants to make efficient use of the available data.

The proposed estimators are valid under the basic Condition 5.8 and they work without declustering of the data, yet it is of course necessary to verify the condition for the considered data.

Nevertheless, Condition 5.8 is valid for a lot of statistical models which are broadly used in financial mathematics. It can furthermore be assumed that it is also valid in certain applications in the field of climate research. We include some remarks on the applicability of our condition in the field of climate research.

### 5.5.1 Some Remarks on the Basic Condition

In the field of environmental statistics one often faces dependent data. If for example extreme precipitation amounts at different locations and days are analyzed, this data often exhibits spatial and temporal dependencies. Extreme rainfalls at two location can be caused by the same rainstorm, extreme rainfalls at several consecutive days may occur because of the same low. Such effects are typically present if other climatologic variables are studied, too. Nevertheless, one may often assume that a certain form of conditional independence holds.

In the previous example one may for example argue that the precipitation amounts at different locations are conditional independent given that a large scale rainstorm has occurred, and we may have temporal conditional independence given that the rainfalls occurred in a certain low pressure phase.

Generalizing this thought we may study local extreme weather events given large scale observations and assuming conditional independence of the form introduced in Condition 5.8. This approach has the additional advantage that large scale climate phenomena are quite well understood and one is able to predict the influence of global warming on these phenomena. On the other hand local climate extremes are still a very active field of research, see for example [41], [42] and [12]. Thus one may use the conditional distribution of a local climate phenomena such as extreme rainfall, given a certain scenario for the large scale phenomena to predict for example the risk of flooding at a certain location in a couple of years in consideration of climate change.

### 5.5.2 Simulations and Real Data Analysis

The following simulations illustrate the performance of the proposed conditional MLE (5.24) in comparison to the MLE based on the likelihood function (5.22). As already noted the latter is recently the most common estimation procedure in the statistical literature.

Specifically, we study the performance of the estimators for the model introduced in Example 1. We merely consider univariate covariates to reduce the computing time.

Choose the function $h$

$$h(x,\xi) = W^{-1}_{\vartheta_{\theta(x)}}(\xi)$$

in Example 1, whith $W_\vartheta^{-1}$ denoting the quantile function pertaining to the GPD with parameter vector $\vartheta$. Put $\vartheta_\theta(x) = (\gamma_\theta(x), \mu_\theta(x), \sigma_\theta(x))$, $\gamma_\theta(x) \equiv \theta_1 \in \mathbb{R}$

$$\sigma_\theta(x) = \exp(\theta_2 + \theta_3 x), \quad \theta_1,\, \theta_2 \in \mathbb{R}$$

as well as

$$\mu_\theta(x) = \theta_4 + \theta_5 x, \quad \theta_1, \theta_2 \in \mathbb{R}.$$

The iid rvs $\xi_1, \xi_2, \xi_3, \ldots$ (cf. Example 1) are uniformly distributed on the interval $[0, 1]$.

The conditional distribution of the response variables given the covariate satisfies

$$F(y|x) = W_{\vartheta_\theta(x)}(y), \quad y > 0.$$

We apply the estimators to the full point process $N$ in Condition 5.8, where $\beta$ is a Poisson distributed sample size $\tau$ with parameter $\lambda$.

For the maximization of the likelihood functions we use the Nelder–Mead algorithm as implemented in the *optim*–routine of the statistical software **R**. This procedure requires an initial estimator as starting value for the numerical optimization. Yet a simple initial estomator for the model described above is not available besides of a "kick and rush" approach which utilizes the moment estimates $\hat{\gamma}_{\mathrm{MM}}$, $\hat{\mu}_{\mathrm{MM}}$ and $\hat{\sigma}_{\mathrm{MM}}$ based on the response variables neglecting the influence of the covariates. Thus the initial values are chosen as $\hat{\theta}_{1,\mathrm{init}} = \hat{\gamma}_{\mathrm{MM}}$, $\hat{\theta}_{2,\mathrm{init}} = \log(\hat{\sigma}_{\mathrm{MM}})$, $\hat{\theta}_{3,\mathrm{init}} = 0$, $\hat{\theta}_{4,\mathrm{init}} = \hat{\mu}_{\mathrm{MM}}$ and $\hat{\theta}_{5,\mathrm{init}} = 0$.

The insights gained from the simulations can be summarized as follows

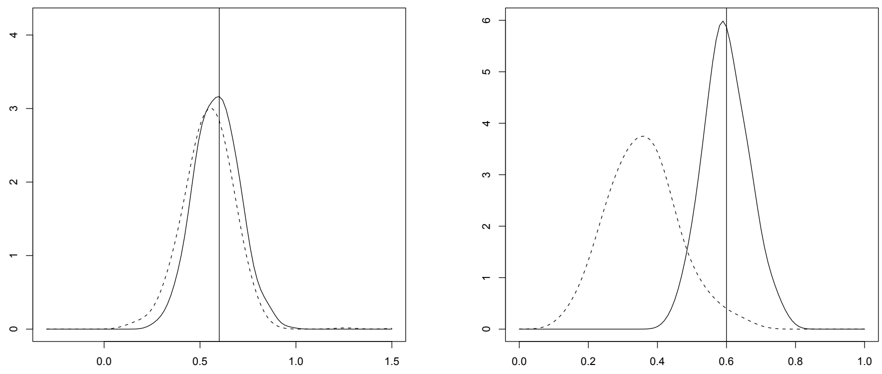
(a) Using the initial estimates described in the preceding lines leads to inaccurate results in most cases, because the global maximum of the likelihood is not detected.

(b) Alternatively one may use a grid and start the iteration procedure for each grid point. This procedure requires extensive computing time and is, therefore, not adequate for comprehensive simulation studies.

(c) To reduce computing time we present a simulation study where the optimization procedure is starts with the underlying parameters. Then, according to our experience, the iteration procedure finds the global maximum of the likelihood function. Notice that this can also be achieved using (b) with a sufficiently dense grid.

The distribution of the MLE based on (5.22) and the conditional MLE based on 5.24 are simulated for different models of the form described above. The threshold is chosen in a way that it is exceeded by 10% of the observations for all simulation runs, getting 200 exceedances on the avarage.

In Figure 5.1 we exemplary summarize the results for the simulated distributions of the MLE and conditional MLE of $\theta_1$ using the kernel densities. The estimators for the remaining parameters show similar characteristics.




(a) independent covariates, optimization on grid

(b) dependent covariates, starting from true parameters

Figure 5.1: Kernel densities of the conditional MLE (solid) and the MLE (dashed) of $\theta_1$ based on 400 independent . The vertical lines indicate the value of the true parameter $\theta_1 = 0.6$.

Figure 5.1a shows kernel densities of the conditional MLE and the MLE for $\theta_1$ with the optimization procedure starting from a grid. The parameter values are chosen as $\theta_1 = 0.6$, $\theta_2 = -0.5$, $\theta_3 = 1$, $\theta_4 = 0.5$ and $\theta_5 = 1$. The covariates are iid standard Gaussian rvs. The conditional MLE is superior as far as the concentration about the true parameter value is concerned.

The chosen parameters provide constellation where the MLE exhibits a comparably good performance. If the values of $\theta_3$ and/or $\theta_5$ increase the performance of the MLE deteriorates steadily. The performance of the conditional MLE is hardly affected.

In a second step we consider covariates which follow a non–stationary GARCH process with standard Gaussian innovations. As a consequence the model for the MLE is not valid anymore. We observe that the resulting likelihoods are much harder to handle from a numerical point of view. The optimization requires considerable computing time because one has to choose a very dense grid to obtain the global maximum. Recall that in (c) we start the maximization from the given parameters. Now the parameters are chosen as $\gamma = 0.6$, $\mu_0 = 0.5$, $\mu_1 = 1$, $\sigma_0 = -5$ and $\sigma_1 = 1$. It is necessary to consider a small value for the parameter $\sigma_0$ because the covariates tend to be much larger than in the case of independent Gaussian variables leading to very large observations of the response variables. The choice of $\sigma_0$ accounts for this problems and keeps the values of the response variable in a reasonable range.

As can be deduced from Figure 5.1b the conditional MLE exhibits a good performance while the MLE fails to some extent.

A detailed discussion concerning some additional simulation studies is given in the remainder of this section.

**Simulation Study 1.** We start with a model described above where the distribution of the covariates is chosen as independent standard Gaussian distribution. The threshold is as $u = 4$. As a result it will be seen that the conditional MLE outperforms the common MLE.

The following plots (Figures 5.2, 5.3 and 5.4) compare the performance of the estimators based on 1000 simulations of $N$ with $\lambda = 500$. The vertical lines indicate the value of the parameter which is to be estimated. The parameters are chosen as $\theta_1 = 0.6$, $\theta_2 = 0.1$, $\theta_3 = 0.8$, $\theta_4 = 0.5$ and $\theta_5 = 0.5$, which leads to an average number of 500 observations thereof on average 109 exceedances. This choice of the underlying parameters means that the dependence of the rvs $Y$ on the covariate information is quite weak. Moreover, since the values of $\theta_3$ and $\theta_5$ are small the initial estimates tend to be close to the real parameters. At first sight the kernel densities of the different estimators seem to show a better performance of the MLE, since its kernel density has more mass close to the underlying parameter. On the other hand the mode of the conditional MLE is closer to the true parameter then the mode of the MLE. Another fact that does not appear in the subsequent plots is the variance of the estimators. Here the conditional MLE shows a significantly smaller variance than the MLE. This is because the MLE has some values which depart strongly from the center of its distribution and these "outliers" appear with a relatively high probability. Table 5.2 comprises some sample functionals of the simulated estimators.

| | $\hat{\theta}_1$ | $\hat{\theta}_1$ | $\hat{\theta}_3$ | $\hat{\theta}_4$ | $\hat{\theta}_5$ |
|---|---|---|---|---|---|
| parameter value | 0.6 | 0.1 | 0.8 | 0.5 | 0.5 |
| mean conditional MLE | 0.69 | -0.06 | 0.89 | 0.15 | 0.65 |
| mean MLE | 1.76 | -0.63 | 1.04 | -0.67 | 1.50 |
| median conditional MLE | 0.55 | 0.21 | 0.78 | 0.10 | 0.56 |
| median MLE | 0.55 | 0.27 | 0.75 | -0.32 | 0.74 |
| variance conditional MLE | 1.74 | 6.56 | 1.20 | 3.33 | 0.86 |
| variance MLE | 110.25 | 69.84 | 8.05 | 16.49 | 22.31 |

Table 5.2: Sample functionals of simulated MLE and conditional MLE (Simulation Study 1).

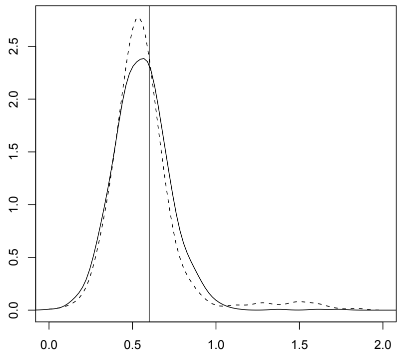



Figure 5.2: Kernel densities of the conditional MLE (solid) and the MLE (dashed) of $\theta_1$.

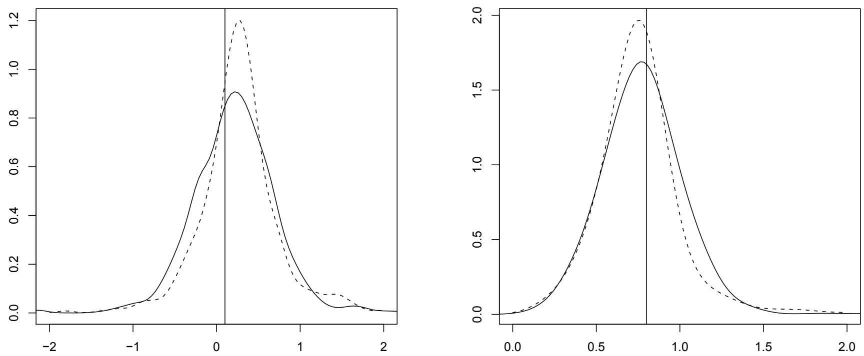



Figure 5.3: Kernel densities of the conditional MLE (solid) and the MLE (dashed) of $\theta_2$ and $\theta_3$.

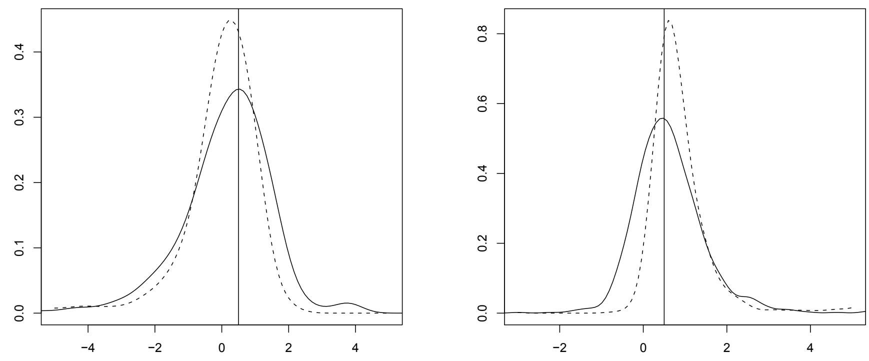



Figure 5.4: Kernel densities of the conditional MLE (solid) and the MLE (dashed) for $\theta_4$ and $\theta_5$.

Our second simulation study concerns the same model as above, but this time we choose the parameters $\theta_3$ and $\theta_5$ which quantify the dependence of $Y$ and $\boldsymbol{X}$ each equal to 2. The initial estimates are obtained as before. This leads necessarily to a larger deviation of the initial estimates from the parameters.

**Simulation Study 2.** We chose the same model and estimation procedure as in Simulation Study 1, yet this time we made 400 runs and choose the parameters as $\lambda = 500$, $u = 4$, $\theta_1 = 0.6$, $\theta_2 = 0.1$, $\theta_3 = 2$, $\theta_4 = 0.5$, $\theta_5 = 2$. This yields an average number of 129 exceedances. The results are similar to those obtained in the foregoing study (see Figures 5.5, 5.6 and 5.7) in so far, that the conditional MLE has a better performance in terms of bias and variance. Moreover, the MLE exhibits a strong dependence on the initial estimates and its performance is quite poor while the conditional MLE still works reasonably well if the initial estimates differ significantly from the true parameters.

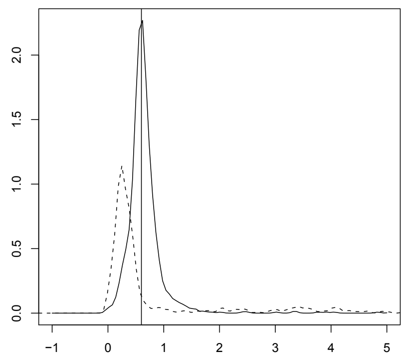



Figure 5.5: Kernel densities of the conditional MLE (solid) and the MLE (dashed) of $\theta_1$.

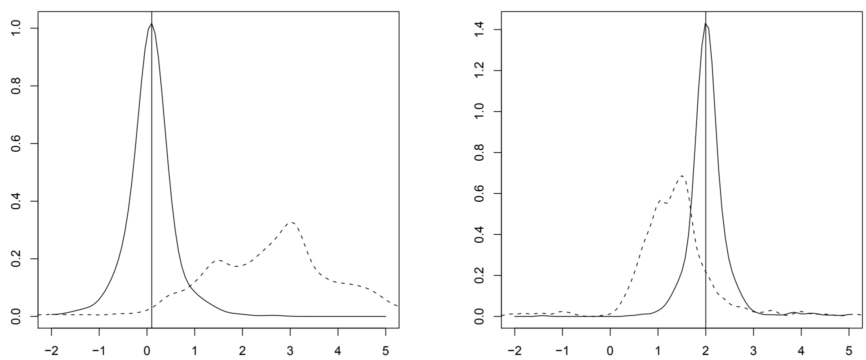



Figure 5.6: Kernel densities of the conditional MLE (solid) and the MLE (dashed) of $\theta_2$ and $\theta_3$.

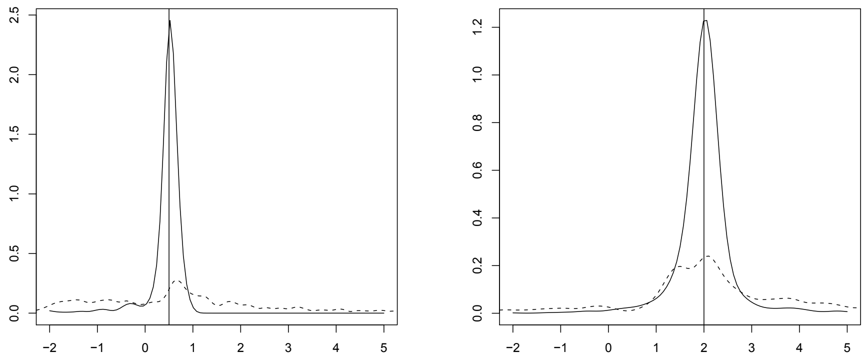



Figure 5.7: Kernel densities of the conditional MLE (solid) and the MLE (dashed) for $\theta_4$ and $\theta_5$.

In the following we apply both estimators to real datasets from environmental statistics. We start by studying daily precipitation amounts in Chico, California.

**Case Study 4. [Precipitation Amounts, Chico, California.]** The considered dataset consists of January daily precipitation amounts recorded in Chico, California from 1913 to 1988. We consider only one covariate, the monthly mean pressure at sea level. We apply the model from Simulation Study 1 to this data set choosing a threshold $u = 40$ which yields a number of 49 exceedances out of a total sample size 2418. Choosing lower thresholds yields no reasonable results. The data have also been analyzed in [56], Section 15.5 applying the same model but using the unconditional ML–method. The resulting ML–estimates for the parameters are given in Table 5.3. The q–q–plots clearly indicate that the conditional MLE performs better than the MLE in that case.

| | $\hat{\theta}_1$ | $\hat{\theta}_1$ | $\hat{\theta}_3$ | $\hat{\theta}_4$ | $\hat{\theta}_5$ |
|---|---|---|---|---|---|
| conditional MLE | 0.39 | -1.47 | 0.002 | 42.16 | -0.02 |
| MLE | 0.34 | 2.61 | -0.0005 | 40.53 | 0.0005 |

Table 5.3: estimated parameters for Chico precipitation data

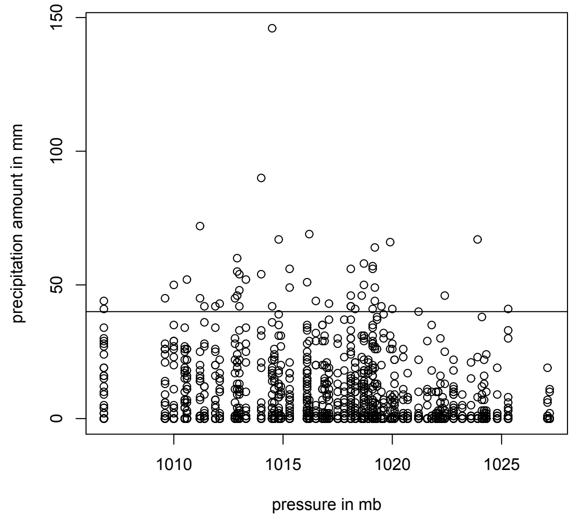



Figure 5.8: Daily precipitation amount in Chico, California.

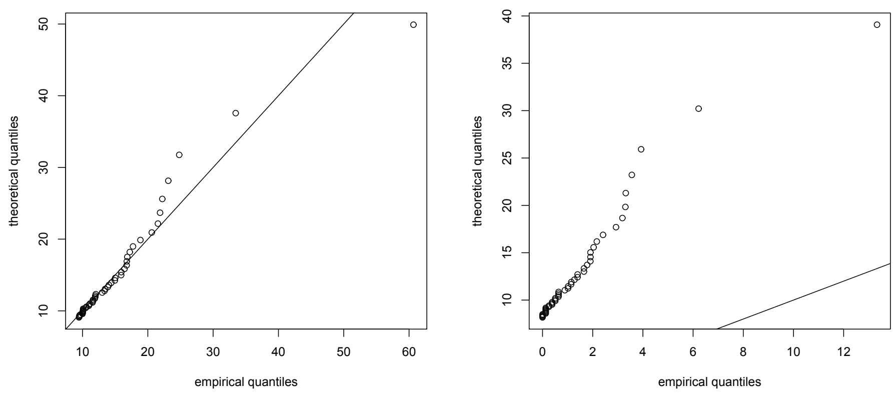



Figure 5.9: Q–Q-plots for the conditional MLE (left) and the MLE (right).

**Case Study 5. [Daily Maximum Wind Speed, Aachen, Germany.]** The considered dataset contains daily measurements of the daily maximum temperature 2 meters over ground (measured in degree Celsius) and the daily maximum wind speed (peak gust, measured in m/sec) in Aachen, Germany from January 1st, 1991 to June 24th, 2008. The data are obtained from the homepage of the German meteorological service (DWD).

For our analysis we concentrate on the winter months December, January and February where usually the severe winter storms appear in Germany. Most of the storms which caused severe damage in Germany in the past two decades took place during this time of the year, e.g. Kyrill (January 18th/19th 2007), Lothar (December 26th, 1999), Wiebke (February 28th 1990), Vivian (February, 25th 1990) and Daria (January 25th/26th, 1990). Because our data starts in 1991, the latter three storms are not contained. Nevertheless, these storms mainly hit the south–western part of Germany and not so much the region in western Germany where our data where collected.

The concentration on the winter period allows us to omit the inclusion of an annual cycle in our model. Figure 5.10 reveals a clear dependence of maximum wind speed and maximum temperature at one day, high wind speeds occur more often when the pertaining temperature is also high. This is due to the fact, that serious winter storms approach western Germany usually from the west bringing also mild air from the Atlantic ocean. So one may argue that high temperatures in the winter months are caused by western storms. Therefore, it might me natural to consider the wind speed as a covariate for the temperature. On the other hand higher temperatures reflect the fact that there is more energy in the climatic system which causes heavy storms. From this point of view it would be desirable to have a covariate which describes the large scale temperature developments in western Europe and over the Atlantic ocean, yet the data at hand only contains daily measurements in Aachen. Daily temperatures in Aachen are obviously closely related to large scale developments in the temperature of western Europe. Thus, this data can be utilized to draw at least limited conclusions about the impact of global warming on extreme storm events in western Germany.

We apply the model introduced in Simulation Study 1, where the temperature is treated as the covariate. Choosing the threshold $u = 22$ yields 113 exceedances out of a total sample size 1684. Using the conditional MLE we obtain the estimates given in the subsequent table. The pertaining q–q–plot is given in Figure 5.11. The q–q–plot shows a very good model fit for the high quantiles yet the model seems to fail to describe the behavior of the smaller exceedances. Figure 5.10 indicates that this might be due to the choice of a constant threshold. We refer to Section 5.6 for a more thorough discussion of this topic. Nevertheless, the estimated conditional q–quantiles are very close to the moving sample quantiles for moderate values of $q$ say up to 95%, see Figure 5.11. The parametric estimate coincides quite well with the sample version, even in the region where the estimated quantile is below the threshold $u$. Deviations of the parametric estimates in the region for high temperatures might be due to the fact that we have only very few observations in this region and thus the empirical conditional quantiles are not reliable for this region.

Since no declustering is applied the daily wind speeds are likely to be dependent since high wind speeds on two consecutive days might be caused by the same low. Nevertheless, given the temperature at the pertaining days we may assume that the wind speed is conditional independent because such effects are also captured in the temperature.

The estimate for the shape parameter $\hat{\theta}_1$ indicates that the conditional distribution of the wind speed has a finite upper tail. This is in line with results from climate research which indicate that rainfall data exhibits heavy tails while almost all other meteorological variables including wind speed seem to possess bounded upper tails (personal communication to R.W. Katz). Moreover, the estimates show a significant impact of the daily maximum temperature on the location parameter of the conditional wind speed distribution and also a slight impact on the scale parameter. This indicates that global warming might lead to more severe storms in the future which is also in line with the recent climate research (see e.g. [41]). Figure 5.11 indicates the impact of the daily maximum temperature on the conditional quantiles of the daily maximum wind speed. On a day with maximal temperature of 10°C we have a 10% probability of a storm with wind speeds of more than 22 m/s (79 km/h), if the daily maximum temperature increases to 20°C this wind speed increases to 30m/s (108 km/h). The pertaining 99% quantiles are 30 m/s (108 km/h) for 10°C and 41 m/s (148 km/h) for 20°C.

| $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $\hat{\theta}_4$ | $\hat{\theta}_5$ |
|---|---|---|---|---|
| -0.12 | 1.12 | 0.04 | 9.93 | 0.35 |

Table 5.5: Estimated parameters for Aachen wind speed data.

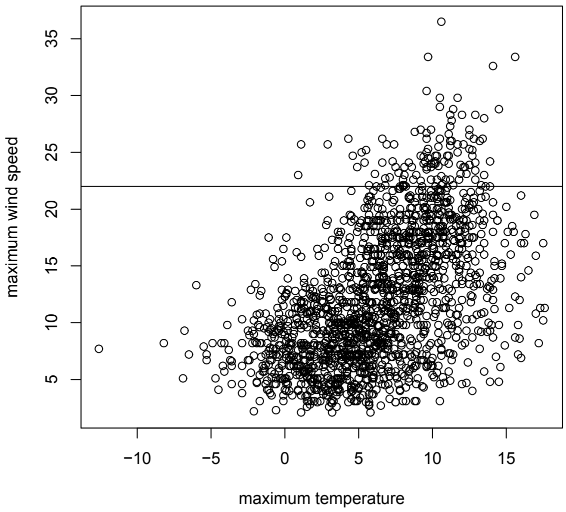



Figure 5.10: Daily maximum wind speed in Aachen, Germany.

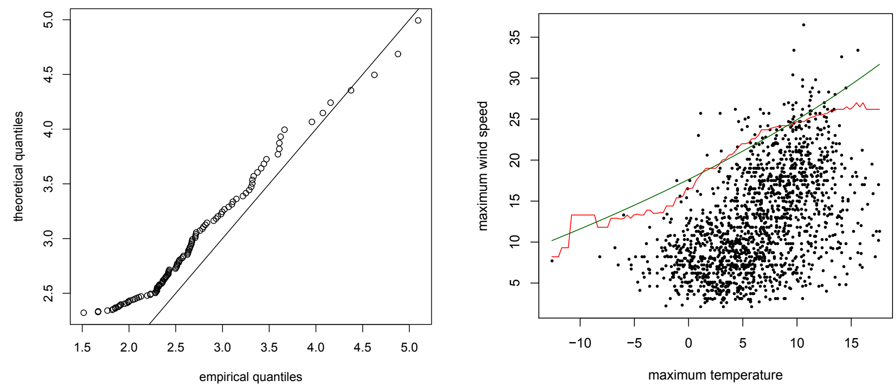



Figure 5.11: q–q–plot for Daily maximum wind speed in Aachen, Germany (left) and conditional 95% quantiles of maximum daily wind speed in Aachen, Germany (right), parametric approach and moving sample quantiles.

## 5.6 Moving Thresholds and Multivariate Extensions

We conclude this chapter by some remarks on two further aspects, namely moving thresholds and extensions to multivariate exceedances.

### 5.6.1 Moving Thresholds

When applying the proposed methods to real data sets the choice of the threshold $u$ is of central importance. If it is chosen too low, the limit theory does not yet apply and we have a model bias which makes the results of the statistical inference doubtable. If it is chosen too high, we have only few observation to base the statistical inference on which usually leads to large variances. This topic is still a field of research in the case of iid exceedances (see e.g. [5], [15], [16], [19], [24], [47], [56]) and it becomes even more urgent if we include covariate information. In the foregoing sections we just choose one threshold for each outcome of the covariates. It is highly probable that we can improve the performance of the statistical procedures if we allow the threshold to vary with the covariate information, thus we replace the fixed threshold $u$ with some threshold line given by a function $u : S \to \mathbb{R}$. The necessity of such an approach becomes evident in Case Study 5. On a warm winters day a wind speed of say $21m/s$ is not really an extreme observation while wind speeds of $15m/s$ on a cold day can be considered extreme. This fact might also explain the shape of the pertaining q–q–plot in Figure 5.11. One can recognize a very good fit for the high quantiles of the residuals while we have some deviations from a straight line in the area of low quantiles. The results from the theory of point processes are not affected using a moving threshold but the choice of the exact form of the threshold function remains an unsolved problem. A straightforward approach is using empirical quantiles or quantile regression for moderate high quantiles about 90% and truncate the data below the estimated moving quantile. The drawback of such an approach is that the choice of the quantile is arbitrary.

Another possibility would be some kind of goodness–of–fit test of the truncated point process model and to choose the lowest threshold line which yields the required confidence level. This approach requires most probably to choose a parametric form of the threshold line.

### 5.6.2 Multivariate Extensions

A second possible extension is the use of multivariate EVT. Assume we have several response variables contained in a random vector $\boldsymbol{Y}$ and again a vector of pertaining covariates $\boldsymbol{X}$. If we assume that the dependence structure of the components of $\boldsymbol{Y}$ is not affected by the covariate we can apply the methods proposed in the foregoing sections for each component of $\boldsymbol{Y}$ separately. If this assumption is not justified we have to model a dependence structure which also depends on the covariate. In the multivariate case we face several new problems which do not occur if we consider only one–dimensional response variables. First of all the definition of the upper tail of a multivariate distributions is not straightforward or unique. Because we will merely give an outlook on this topic we do not discuss this problem in detail. A comprehensive treatment can be found in [27].

In the following we use the term multivariate exceedance in the way used in [27]. A random vector $\boldsymbol{Y}$ exceeds the multivariate threshold $\boldsymbol{u}$ if each component of $\boldsymbol{Y}$ is larger than the corresponding component of $\boldsymbol{u}$. Given a point process

$$N = \sum_{i=1}^{\beta} \varepsilon_{(\boldsymbol{X}_i, \boldsymbol{Y}_i)}$$

we, thus, consider the truncated process

$$N^{[S,\boldsymbol{u}]} = N(\cdot \cap S \times (\boldsymbol{u}, \infty)).$$

In contrast to the univariate case the multivariate conditional df $F(\cdot|\boldsymbol{x})$ is, in the north–east of $\boldsymbol{u}$, not completely determined by the distribution of the truncated point–process. Nevertheless, we can still derive important information about the multivariate tail of $F$ using the point process approach.

We suggest to use a two–step procedure in that case. First model the upper tails of the margins of the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ and derive an estimation applying the approach proposed for the univariate case to each margin separately. Then we transform all margins to the uniform distribution on the interval $[-1, 0]$. Denoting the transformed response variables by $\widetilde{\boldsymbol{Y}}_i$ we have (approximately) for $\boldsymbol{y}$ sufficiently close to **0**

$$P\left(\widetilde{\boldsymbol{Y}}_i \leq \boldsymbol{y} | \boldsymbol{X}_i = \boldsymbol{x}_i\right) = 1 + \sum_{i=1}^{d} y_i D_{\vartheta(\boldsymbol{x})}(\boldsymbol{z})$$

where again $D_{\vartheta(\boldsymbol{x})}$ is a Pickands dependence function as in Section 2.7 which depends on $\boldsymbol{x}$ in some parametric form and $\boldsymbol{z}$ is the radial component in the Pickands representation of $\boldsymbol{x}$. Finally, apply the proposed conditional point process approach starting with the process

$$\widetilde{N} = \sum_{i=1}^{\beta} \varepsilon_{\left(\boldsymbol{X}_i, \widetilde{\boldsymbol{Y}}_i\right)}.$$

A crucial aspect is the choice of the parametric family of Pickands dependence functions because in general these dependence functions do not form a parametric family. We are not going to workout the multivariate approach in more detail within the framework of this thesis. Nevertheless, it might be an interesting field of future research as the whole field of multivariate extreme value theory.

# Retrospective, Outlook and Conclusions

This thesis presents new results in the field of extreme value theory where the emphasis is on asymptotic distributions of exceedances and the analysis of extremes under covariate information.

## Retrospective

The first part is of rather theoretical nature. It is shown that all continuous limiting dfs under a general monotone transformation $T : \mathbb{R}^2 \to \mathbb{R}$ satisfy a certain form of $g$–pot stability. If an additional technical condition (Condition 2.8) is satisfied this result can be extended to all non–degenerate limiting dfs. Moreover, we derive an explicit representation of all strictly increasing, continuous limiting distributions of exceedances.

We provide three examples of special normalizations namely linear, power and exponential normalizations. In all cases Condition 2.8 is satisfied. We show that general pot–domains of attraction contain the pertaining max–domains of attraction.

As in the well understood case of linear normalization, limiting distributions of maxima are necessarily continuous while in the framework of exceedances also discrete limiting distributions occur. In the general case it is also possible to construct discrete pot–stable distributions, which of course also occur as limiting distributions.

We also shortly address extensions to multivariate exceedances under general monotone transformations, but this is still very much an open field. As in the case of linear normalization a natural approach consists of decomposing the problem in the univariate marginal distributions and the asymptotic dependence structure. If the multivariate normalization is of the form as given in Lemma 2.14 the results in the univariate case carry over to the marginal distributions.

Asymptotic distributions under power–normalization are studied in more detail. We derive all limiting distributions, including the discrete ones, in this case and give a complete characterization of the pertaining domains of attraction. This is possible because we can use a close relation of linear and power–normalization which was first observed in [10] in the case of maxima.

A particular interesting aspect of power–normalization is, that it is possible to derive asymptotic models for distributions with super–heavy, thus slowly varying, survivor functions. The pertaining asymptotic model is constituted by the family of GLPDs which is an extension of the GPD family. The GLPD family has in comparison to the GPD family an additional shape parameter which complicates the adaption of the model to data. The MLE in the GLPD family has, as in the GPD family, no closed form solution and exhibits a larger variance than in the GPD family, which is not surprising because the GPD family is just a constraint GLPD model. The numerical problems increase because of the additional dimension in which the likelihood function has to be maximised. This makes the estimation procedure also strongly depend on the starting values for the numerical optimization. Therefore, we derive an initial estimator which is based only on one–dimensional numerical optimization. Nevertheless, it is desirable to have an initial estimator with a closed form.

In the last chapter we turn the focus on exceedances in more complex models. We investigate upper tails of the conditional distribution of a response variable given a covariate. As indicated in Section 5.4.5 this topic is extensively studied in the statistical literature. We introduce a point process model which unifies the most common ML approaches in the statistical literature. The results are valid under a certain condition (Condition 5.8) which is basically the framework of ordinary regression analysis and one may assume this condition to hold in a lot of applications.

## Outlook

Lemma 2.14 indicates that the asymptotic dependence structure under certain componentwise, monotone transformations is the same as in the case of componentwise linear normalization, although we only consider the case of general max–stable distributions. It should be possible to carry over techniques used in [27] and [57] to extend this results to multivariate exceedances. In this case the multivariate GP–function has a kind of universal meaning as asymptotic dependence structure for multivariate exceedances under componentwise and monotone normalizations.

The question whether if a discrete distribution function is pot–stable with respect to a transformation $g$, then $g$ has to be of the form given in Section 2.4, remains unsolved. To settle this question one likely has to apply or even develop an advanced theory of functional equations. It has turned out that discrete limiting distributions in the linear case, which have been completely characterized in [3], have not found much attention neither in the statistical literature nor in applications. Therefore, one may argue that this gap in the theory of the general case is not of particular importance.

An interesting field for further research work is constituted by super–heavy tailed dfs and GLPDs. We apply the GLPD model to several data sets. In all cases a clear decision whether the GPD model is sufficient or the GLPD model has to be used cannot be made. In the framework of this thesis we were not successful to find a real world data set which exhibits clearly rather a super–heavy than heavy tailed behavior. Nevertheless, if one ignores a possible super–heavy tailed behavior in the data this might lead to serious underestimation of extreme quantiles which are of particular importance in many applications.

To discriminate between the GLPD and GPD model we have introduced up to now only visual tools as introduced in Section 4.3.2. An applicable test procedure is not derived but one can imagine to use a local Neyman–Pearson test based on a Taylor–expansion of the quotient of densities at $\gamma = 0$ or $\beta = 0$ as indicated in Section 4.3.3. We may concentrate on such a local testing procedure because it is likely that one will face mostly real world data sets which, if they are really super–heavy tailed, are still close to the GPD model. If one has a data set where both shape parameters differ significantly from zero such a local test procedure will likely have a power which will lead to the rejection of the GPD model with a quite high probability. Another possibility is to use discriminant analysis, where also a local version might be of particular interest.

The development of the field exceedances under covariate information is just at the beginning and a lot of questions are still left open. An aspect which is of particular importance is the choice of an adequate threshold, which should also depend on the pertaining outcome of the covariate information. The most urgent task are the numerical problems related to the maximization of the likelihood function. Because we usually have at least five parameters even in the most simple model one cannot ignore this problem and the resulting estimates have to be treated with caution. One possible improvement consists in deriving adequate initial estimators to assure that the optimization procedure does not start too far away from the true parameters.

### Conclusions

In summary one may say that this thesis gives some new results in two important fields of extreme value theory, but we are far from studying all aspects of either field and a lot of interesting questions are left open for further research. The results concerning asymptotic distributions of exceedances under general monotone transformations are of rather theoretical nature and it cannot be said today if they can be relevant for applications. In the special case of power–normalization one can use the pertaining asymptotic models to analyze data from super–heavy tailed distributions, yet it is an open question if such distributions arise in applications. The conditional ML–approach for exceedances under covariate information can be applied directly to a lot of applications and if one considers the pertaining simulation results it is worthwhile thinking about using this approach instead of the procedure based on unconditional distributions of Poisson processes.

# A Auxiliary Results

We include some technical auxiliary results which are used throughout the thesis. Since we use special formulations we give the pertaining proofs, too, although none of the following results is a original contribution of this thesis.

First we state a well known result for conditional distribution of a random vector vector given one of its components.

**Lemma A.1** *Let $X$, $Y$ be arbitrary rvs, and consider a further rv $Y_x$ with distribution $P(Y \in \cdot \,|X = x)$. Then*

$$P((X, Y) \in \cdot \,|X = x) = \mathcal{L}((x, Y_x)).$$

*The right hand side may be represented as $\varepsilon_x \times P(Y \in \cdot \,|X = x)$.*

Proof. Observe that

$$\mathcal{L}((x, Y_x)) = \varepsilon_x \times P(Y \in \cdot \,|X = x).$$

Therefore, for measurable sets $A$, $B$, $C$ one gets

$$\begin{aligned} P\{X \in A, (X, Y) \in B \times C\} &= P\{X \in A \cap B, Y \in C\} \\ &= \int_{A\cap B} P(Y \in C\,|X = x)\, d\mathcal{L}(X)(x) \\ &= \int_A \varepsilon_x(B) P(Y \in C\,|X = x)\, d\mathcal{L}(X)(x) \\ &= \int_A \mathcal{L}((x, Y_x))(B \times C) d\mathcal{L}(X)(x), \end{aligned}$$

which is the asserted property. □

The following Corollary is a consequence of Lemma A.1.

**Corollary A.2** *Let $(X_i, Y_i)$ be as in Condition 5.8. Put $\boldsymbol{X} = (X_1, \ldots, X_n)$ Then*

*(i) $(X_i, Y_i)$ are conditional independent given $\boldsymbol{X} = \boldsymbol{x}$.*

*(ii) For the conditional distribution of $(X_i, Y_i)$ given $\boldsymbol{X} = \boldsymbol{x}$ holds*

$$P((X_i, Y_i) \in \cdot |\, \boldsymbol{X} = \boldsymbol{x}) = \mathcal{L}((x_i, Y_{x_i}))$$

*where $Y_x$ is defined corresponding to Lemma A.1.*

Proof. *Put* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, $\boldsymbol{A} = \times_{i=1}^n A_i$ *and* $\boldsymbol{B} = \times_{i=1}^n B_i$. *Observe that Lemma A.1 yields*

$$P(\boldsymbol{Y} \in \boldsymbol{B}, \boldsymbol{X} \in \boldsymbol{B} \,|\, \boldsymbol{X} = \boldsymbol{x}) = \varepsilon_{\boldsymbol{x}}(\boldsymbol{A}) P\{\boldsymbol{Y}_{\boldsymbol{x}} \in \boldsymbol{B}\}, \tag{A.1}$$

*with*

$$P\{\boldsymbol{Y}_{\boldsymbol{x}} \in \boldsymbol{B}\} = P(\boldsymbol{Y} \in \boldsymbol{B} \,|\, \boldsymbol{X} = \boldsymbol{x}).$$

*Moreover,*

$$\begin{aligned} P(\boldsymbol{Y} \in \boldsymbol{B} \,|\, \boldsymbol{X} = \boldsymbol{x}) &= \prod_{i=1}^n P(Y_i \in B_i \,|\, \boldsymbol{X} = \boldsymbol{x}) \\ &= \prod_{i=1}^n P(Y_i \in B_i \,|\, X_i = x_i). \end{aligned}$$

*Combining the latter equality with (A.1) yields the assertions.* □

Another important tool when dealing with conditional distributions is the following extension of the Fubini theorem for Markov kernels. We give a special version for three rvs.

**Lemma A.3** *Let* $X$, $Y$ *and* $Z$ *be arbitrary random such that the following conditional distributions exist. Then for measurable sets* $A$, $B$ *and* $C$ *we have*

$$\begin{aligned} &P\{X \in A,\, Y \in B,\, Z \in C\} \\ &= \int_C \int_B P(X \in A \,|\, Y = y, Z = z)\, dP(Y \in dy \,|\, Z = z)\, d\mathcal{L}(Z)(z) \end{aligned}$$

Proof. *Notice that by Lemma A.1*

$$\int f(x,y) d\mathcal{L}(X,Y)(x,y) = \int \int f(x,y) d\left(\varepsilon_v \times P(Y \in dy | X = v)\right)(x,y) d\mathcal{L}(X)(v).$$

*Therefore, by Fubinis theorem*

$$\begin{aligned} &P\{X \in A,\, Y \in B,\, Z \in C\} \\ &= \int_{A \times B} P(Z \in C \,|\, Y = y, X = x)\, d\mathcal{L}(Y, X)(y, x) \\ &= \int \int \varepsilon_y(B) \varepsilon_x(A) P(Z \in C \,|\, Y = y, X = x) \\ &\qquad d\left(\varepsilon_v \times P(Y \in dy | X = v)\right)(x,y) d\mathcal{L}(X)(v) \end{aligned}$$

$$= \int\int\int \varepsilon_y(B)\varepsilon_x(A)P\left(Z \in C\,|Y = y, X = x\right)$$
$$d\varepsilon_v(x)dP\left(Y \in dy|X = v\right)d\mathcal{L}(X)(v)$$
$$= \int\int \varepsilon_y(B)\varepsilon_v(A)P\left(Z \in C\,|Y = y, X = v\right)dP\left(Y \in dy|X = v\right)d\mathcal{L}(X)(v)$$

*which proofs the assertion.* □

**Lemma A.4** *Let $X$, $Y$ and $Z$ be arbitrary rvs such that the following conditional distributions exists:*

$$\mathcal{L}(Y_x) := P(Y \in \cdot|X = x)$$

*and*

$$\mathcal{L}(Z_x) := P(Z \in \cdot|X = x).$$

*Then*

$$P\left(Y_x \in \cdot|Z_x = z\right) = P\left(Y \in \cdot|X = x, Z = z\right)$$

PROOF. A Markov–Kernel $K_x(\cdot|\cdot)$ is a conditional distribution of $Y_x$ given $Z_x = z$ if for measurable sets $A$, $B$ and $C$

$$P\left\{X \in A, Y \in B, Z \in C\right\} = \int_A\int_C K_x(B|z)d\mathcal{L}(Z_x)(z)d\mathcal{L}(X)(x)$$

because we have

$$P\left\{X \in A, Y \in B, Z \in C\right\} = \int_A P\left\{Y_x \in B,\, Z_x \in C\right\}d\mathcal{L}(X)(x)$$

and

$$P\left\{Y_x \in B,\, Z_x \in C\right\} = \int_C P\left(Y_x \in B|Z_x = z\right)d\mathcal{L}(Z_x)(z).$$

This also entails that $\widetilde{K}(B|(z,x)) := K_x(B|z)$ is a Markov kernel. Define

$$K_x\left(B|z\right) = P\left(Y \in B|Z = z, X = x\right)$$

then by Lemma A.3

$$\begin{aligned}
&P\left\{X \in A,\, Y \in B,\, Z \in C\right\}\\
&= \int_A\int_C P\left(Y \in B|Z = z, X = x\right)dP\left(Z \in dz|X = x\right)d\mathcal{L}(X)(x)\\
&= \int_A\int_C K_x(B|z)d\mathcal{L}(Z_x)(z)d\mathcal{L}(X)(x)
\end{aligned}$$

which proofs the assertion. □

A crucial concept for the considerations in Chapter 5 is the concept of conditional independence (see e.g. [56], Section 8.1 or [9], Section 7.3). We first give a formal definition.

**Definition A.5** *Let $X$, $Y$ and $Z$ be arbitrary rvs with values in measurable spaces $(S_1, \mathcal{B}_1)$, $(S_2, \mathcal{B}_2)$ and $(S_3, \mathcal{B}_3)$. $Y$ and $Z$ are called conditional independent given $X$ if for each $x \in S_1$, $B \in \mathcal{B}_2$ and $C \in \mathcal{B}_3$ holds*

$$P(Y \in B, Z \in C \,|X = x) = P(Y \in B|X = x)\, P(Z \in C|X = x).$$

Independence of two rvs $Y$ and $Z$ does not entail conditional independence of $Y$ and $Z$ given a further rv $X$. We give a simple counter example: Let $X_1$ and $X_2$ be independent $\mathcal{N}_{0,1}$–distributed rvs. It is a well known fact that the pertaining sample mean

$$\bar{X} = \frac{1}{2}X_1 + \frac{1}{2}X_2$$

and sample variance

$$S = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2$$

are independent. Yet we have for $\sqrt{s} \le x$

$$\begin{aligned} P\left(\bar{X} \le x,\, S \le s|X_2 = 0\right) &= P\left\{\frac{1}{2}X_1 \le x,\, \frac{1}{2}X_1^2 \le s\right\} \\ &= P\left\{-\sqrt{2s} \le X_1 \le \sqrt{2s}\right\} \\ &= 2\Phi(\sqrt{2s}) - 1 \end{aligned}$$

and furthermore

$$P\left(\bar{X} \le x\,|X_2 = 0\right) = P\left\{X_1 \le 2x\right\} = \Phi(2x)$$

as well as

$$P(S \le s\,|X_2 = 0) = P\left\{X_1^2 \le 2s\right\} = 2\Phi\left(\sqrt{2s}\right) - 1.$$

Obviously conditional distributions are only determined outside a set of Lebesgue measure zero, thus the arguments in the above lines do not prove that $\bar{X}$ and $S$ are not conditional independent, but one can repeat the arguments for $X_2 = x$ for arbitrary $x \in \mathbb{R}$ which yields the assertion. For the sake of simplicity we confine arguments to the case $x = 0$.

Nevertheless, if two rvs $Y$ and $Z$ are independent we can conclude conditional independence if we impose an additional condition. This result is stated in the following lemma (see also [9], Section 7.3, Corollary 3).

**Lemma A.6** *Let $X$, $Y$ and $Z$ be rvs with values in measurable spaces $(S_1, \mathcal{B}_1)$, $(S_2, \mathcal{B}_2)$ and $(S_3, \mathcal{B}_3)$ such that $(X, Z)$ is independent of $Y$. Then $Z$ and $Y$ are conditional independent given $X$.*

PROOF. Let $A \in \mathcal{B}_1$, $B \in \mathcal{B}_2$ and $C \in \mathcal{B}_3$ and define the Markov–Kernel

$$K(B \times C|x) := P(Y \in B|X = x)\, P(Z \in C|X = x)\,.$$

This yields since $Y$ and $(Z, X)$ are independent

$$\begin{aligned}
&\int_A K(B \times C|x) d\mathcal{L}(X)(x) \\
&= \int_A P(Y \in B|X = x)\, P(Z \in C|X = x)\, d\mathcal{L}(X)(x) \\
&= \int_A P\{Y \in B\}\, P(Z \in C|X = x)\, d\mathcal{L}(X)(x) \\
&= P\{Y \in B\} \int_A P(Z \in C|X = x)\, d\mathcal{L}(X)(x) \\
&= P\{Y \in B\}\, P\{X \in A, Z \in C\} \\
&= P\{X \in A, Y \in B, Z \in C\}\,,
\end{aligned}$$

thus $K(\cdot|x)$ is a conditional distribution of $(Y, Z)$ given $X = x$ for each $x \in S_1$ and therefore $Y$ and $Z$ are conditional independent given $X = x$. □

Given rvs $X, Y$ and $Z$ the subsequent lemma yields a way to construct the conditional distribution of $Z$ given $X = x$ from the conditional distributions of $Z$ given $Y = y$ and $Y$ given $X = x$, if certain conditions concerning the dependence structure of $X$, $Y$ and $Z$ are satisfied.

**Lemma A.7** *Let $X$, $Y$, $Z$ be arbitrary rvs with values in measurable spaces $(S_i, \mathcal{B}_i)$, $i = 1, 2, 3$, where $X$ and $Z$ are conditionally independent given $Y = y$. Then*

$$P(Z \in C|X = x) = \int P(Z \in C|Y = y) dP(Y \in dy|X = x). \tag{A.2}$$

PROOF. We start by showing that a certain Markov property follows from conditional independence, namely

$$P(Z \in \cdot|X = x, Y = y) = P(Z \in \cdot|Y = y), \tag{A.3}$$

therefore observe for $A \in \mathcal{B}_1$, $B \in \mathcal{B}_2$ and $C \in \mathcal{B}_3$ by Fubinis Theorem for Markov kernels

$$
\begin{aligned}
&\int_{A\times B} P(Z\in C|Y=y)d\mathcal{L}((Y,X))(y,x)\\
&= \int \varepsilon_x(A)\varepsilon_y(B)P(Z\in C|Y=y)d\mathcal{L}((Y,X))(y,x)\\
&= \int\int \varepsilon_x(A)\varepsilon_y(B)P(Z\in C|Y=y)d(\varepsilon_v\times P(X\in\cdot|Y=v))(y,x)d\mathcal{L}(Y)(v)\\
&= \int\int\int \varepsilon_x(A)\varepsilon_y(B)P(Z\in C|Y=y)d\varepsilon_v(y)dP(X\in dx|Y=v)d\mathcal{L}(Y)(v)\\
&= \int\int \varepsilon_x(A)\varepsilon_v(B)P(Z\in C|Y=v)dP(X\in dx|Y=v)d\mathcal{L}(Y)(v)\\
&= \int \varepsilon_v(B)P(Z\in C|Y=v)\int \varepsilon_x(A)dP(X\in dx|Y=v)d\mathcal{L}(Y)(v)\\
&= \int_B P(Z\in C|Y=v)P(X\in A|Y=v)d\mathcal{L}(Y)(v)\\
&= \int_B P((Z,X)\in C\times A|Y=v)d\mathcal{L}(Y)(v)\\
&= P\{X\in A,\, Y\in B,\, Z\in C\}
\end{aligned}
$$

which yields (A.3). We get from Lemma A.3

$$
\begin{aligned}
&P\{X\in A, Z\in C\}\\
&= P\{X\in A, Y\in S_2, Z\in C\}\\
&= \int_A\int P(Z\in C|Y=y, X=x)dP(Y\in dy|X=x)d\mathcal{L}(X)(x)\\
&= \int_A\int P(Z\in C|Y=y)dP(Y\in dy|X=x)d\mathcal{L}(X)(x)
\end{aligned}
$$

and therefore the assertion holds. □

**Remark A.8** *This result is closely related to the Chapman–Kolmogoroff equation, see e.g. [38]. If one has a closer look on the proof of the forgoing lemma we have shown the equivalence*

$$
\begin{aligned}
&P(X\in A,\, Z\in C|Y=y) = P(X\in A|Y=y)P(Z\in C|Y=y)\\
&\Leftrightarrow\quad P\left(X\in A|Z=z, Y=y\right) = P\left(X\in A|Y=y\right)\\
&\Leftrightarrow\quad P\left(Z\in C|X=x, Y=y\right) = P\left(Z\in C|Y=y\right).
\end{aligned}
$$

*If we consider a Markov process $(X_t)_{t\geq 0}$ with state space $(S,\mathcal{B})$ and $s<t<v$ we have by definition for $A\in\mathcal{B}$*

$$P(X_v\in A|X_s=x_s, X_t=x_t) = P(X_v\in A|X_t=x_t)$$

*which is equivalent to the conditional independence of $X_v$ and $X_s$ given $X_t = x_t$. Setting $Z = X_v$, $Y = X_t$ and $X = X_s$ in (A.2) one gets the Chapman–Kolmogoroff equation which is known to hold for Markov processes (see for example [38], Satz 6.3).*

We conclude this section concerning conditional distribution with a result concerning the decomposition of the distribution of a random vector $(X, Y)$ given $Y > u$ for some threshold $u$.

**Lemma A.9** *Let $(X, Y)$ be a random vector with values in $S \times \mathbb{R}$. Define the pertaining conditional distribution*

$$F(y|x) := P\left(Y \leq y | X = x\right)$$

*and*

$$F^{[u]}(y|x) = \frac{F(y|x) - F(u|x)}{1 - F(u|x)}, \quad y > u.$$

*Furthermore, let $u \in \mathbb{R}$ such that*

$$P\left(Y > u | X = x\right) > 0, \quad \text{for all } x \in S.$$

*Define the distribution of the random vector $(X^*, Y^*)$ by*

$$\mathcal{L}(X^*) = P(X \in \cdot | Y > u)$$

*and*

$$P\left(Y^* \leq y | X^* = x\right) = F^{[u]}(y|x).$$

*Then*

$$\mathcal{L}(X^*, Y^*) = P\left((X, Y) \in \cdot \,|\, Y > u\right),$$

*thus $\mathcal{L}(X^*, Y^*)$ is the truncation of $\mathcal{L}(X, Y)$ outside $D := S \times (u, \infty)$.*

Proof. First observe that for a measurable set $B$

$$\begin{aligned} P\left\{X^* \in B\right\} &= \frac{P\left\{X \in B, Y > u\right\}}{P\left\{Y > u\right\}} \\ &= \int \frac{P(Y > u | X = x)}{P\left\{Y > u\right\}} d\mathcal{L}\left(X\right)(x) \end{aligned}$$

and, therefore,

$$f(x) = \frac{P(Y > u | X = x)}{P\left\{Y > u\right\}} = \frac{1 - F(u|x)}{P\left\{Y > u\right\}}$$

is a $\mathcal{L}(X)$ density of $\mathcal{L}(X^*)$. Let $y > u$ and $B$ a measurable set, observe that

$$\begin{aligned}
P\{X^* \in B, Y^* \le y\} &= \int_B P(Y^* \le y | X^* = x)\, d\mathcal{L}(X^*)(x) \\
&= \int_B F^{[u]}(y|x) d\mathcal{L}(X^*)(x) \\
&= \int_B F^{[u]}(y|x) f(x) d\mathcal{L}(X)(x) \\
&= \int_B \frac{F(y|x) - F(u|x)}{1 - F(u|x)} \frac{1 - F(u|x)}{P\{Y > u\}} d\mathcal{L}(X)(x) \\
&= \int_B \frac{F(y|x) - F(u|x)}{P\{Y > u\}} d\mathcal{L}(X)(x) \\
&= \frac{\int_B F(y|x) - F(u|x) d\mathcal{L}(X)(x)}{P\{Y > u\}} \\
&= \frac{P\{u < Y < y,\, X \in B\}}{P\{Y > u\}} \\
&= P\left((X, Y) \in B \times (-\infty, y] | \, Y > u\right),
\end{aligned}$$

which is the desired equality. □

**Lemma A.10** *Let $Z \sim U((0,1))$ and $X := (1 + \log(Z))^2$, then*

$$E(X) = 1 \quad \textit{and} \quad Var(X) = 8$$

Proof. We first derive the df pertaining to $X$,

$$\begin{aligned}
P\{X \le x\} &= P\left\{(1 + \log(Z))^2 \le x\right\} \\
&= P\left\{\log(Z) \ge -\sqrt{x} - 1,\, \log(Z) \le 1 - \sqrt{x}\right\} \\
&= P\left\{Z \le \exp\left(\sqrt{x} - 1\right)\right\} - P\left\{Z \le \exp\left(-\sqrt{x} - 1\right)\right\},
\end{aligned}$$

thus the df of $X$ is given by

$$F(x) = \begin{cases} e^{\sqrt{x}-1} - e^{-\sqrt{x}-1} & \text{if } 0 \le x \le 1; \\ 1 - e^{-\sqrt{x}-1}, & \phantom{\text{if }} x > 1. \end{cases}$$

We get immediately that

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}} \left(e^{\sqrt{x}-1} + e^{-\sqrt{x}-1}\right) & \text{if } 0 \le x \le 1; \\ \frac{1}{2\sqrt{x}} e^{-\sqrt{x}-1}, & \phantom{\text{if }} x > 1, \end{cases}$$

is the pertaining density. Moreover,

$$\begin{aligned} E(X) &= \int_0^\infty x f(x) dx \\ &= \frac{1}{2}\int_0^1 \sqrt{x} e^{\sqrt{x}-1} dx + \frac{1}{2}\int_0^1 \sqrt{x} e^{-\sqrt{x}-1} dx + \frac{1}{2}\int_1^\infty \sqrt{x} e^{\sqrt{x}-1} dx. \end{aligned}$$

A substitution $y = \sqrt{x}$ and repeated partial integration yields

$$\begin{aligned} \frac{1}{2}\int_0^1 \sqrt{x} e^{\sqrt{x}-1} dx &= 1 - 2e^{-1} \\ \frac{1}{2}\int_0^1 \sqrt{x} e^{-\sqrt{x}-1} dx &= -5e^{-2} + 2e^{-1} \\ \frac{1}{2}\int_1^\infty \sqrt{x} e^{\sqrt{x}-1} dx &= 5e^{-2}, \end{aligned}$$

therefore, $E(X) = 1$. Similar computations yield $\mathrm{Var}(X) = 8$. □

# B Documentation of R–Programs

## fpp.cond

### Description

Conditional Maximum–Likelihood estimation using the likelihood function (5.24). Parameters depending linear on the covariates in analogy to Simulation Study 1. Usage in analogy to the `fpp` function of the ismev/extRemes package.

### Usage

```
fpp.cond(xdat, threshold, ydat = NULL, mul = NULL,
    sigl = NULL, shl = NULL, mulink = identity, siglink = identity,
    shlink = identity, method = "Nelder-Mead",
    maxit = 10000, maxtry=20, init=NULL,...)
```

### Arguments

| | | |
|---|---|---|
| `xdat` | (`vector`) | matrix of covariates |
| `threshold` | (`real`) | threshold |
| `ydat` | (`vector`) | response variable |
| `shlink, mulink, siglink` | (`function`) | link functions for the parameters $\gamma$, $\mu$, $\sigma$ |
| `shl, mul, sigl` | (`function`) | number of covariates which influence $\gamma$, $\mu$, $\sigma$ |
| `method` | (`real`) | optimization method to be used |
| `maxit` | (`integer`) | maximal number of iterations |
| `maxtry` | (`integer`) | maximal number of optimizations starting from different random initial values |
| `init` | (`vector`) | initial values for optimization |

### Value

A list with components:

| | | |
|---|---|---|
| `mle` | (`vector`) | vector of ML estimates |
| `try` | (`integer`) | number of optimizations |
| `conv` | (`integer`) | does optimization converge (0: yes, 1:no) |
| `nllh` | (`real`) | value of Likelihhood at ML estimates |
| `init` | (`vector`) | starting values of successful optimization |
| `pp.lik` | (`function`) | Likelihood function |

## GLPD.ML

### Description

Maximum–Likelihood estimator for the generalized Pareto distribution. Uses hybrid estimator as initial estimator.

### Usage

```
GLPD.ML(dat,vl,ru=0.9,plot1=FALSE,bw=2,init.manuel=FALSE,
init=c(0.1,1,1),plottype=1,maxtry=10)
```

### Arguments

| | | |
|---|---|---|
| `dat` | (`vector`) | the data |
| `vl` | (`real`) | percentage of initial estimates from which the median is taken |
| `ru` | (`real`) | order statistic to be taken as threshold |
| `plot1` | (`boolean`) | should analyzing plots be displayed |
| `bw` | (`real`) | bandwidth for kernel densities in analyzing plots |
| `init.manuel` | (`boolean`) | should initial estimates be given by the user |
| `init` | (`vector`) | initial estimates |
| `plottype` | (`integer`) | form of analyzing plots |
| `maxtry` | (`integer`) | maximal number of optimizations starting from different random initial values with different optimization methods |

### Value

A list with components:

| | | |
|---|---|---|
| `nv` | (`integer`) | number of optimizations |
| `threshold` | (`real`) | threshold |
| `value` | (`real`) | value of the Likelihood function |
| `mle` | (`vector`) | ML estimates |
| `init` | (`vector`) | initial estimates for successful optimization |
| `nex` | (`integer`) | number of exceedances |
| `vmethod` | (`vector`) | optimization method used to derive ML estimates |

# Bibliography


[1] J. Aczél. *A Short Course on Functional Equations Based Upon Recent Applications to the Social and Behavioral Sciences.* Dordrecht/ Boston: Reidel/Kluwer, 1987.

[2] D.J. Aldous. Exchangeability and Related Topics. In *École d'Été de Probabilités de Saint–Flour XIII – 1983, Lecture Notes in Math.* **1117**, pages 1–198. Springer, Berlin/Heidelberg, 1985.

[3] A.A. Balkema and L. de Haan. Residual Life Time at Great Age. *Ann. Probab.*, 2:792–804, 1974.

[4] G. Balkema and P. Embrechts. Multivariate Excess Distributions. In *Conference on Heavy Tails and Stable Paretian Distributions in Finance and Macroeconomics.* 2005.

[5] J. Beirlant, P. Vynckier, and J.L. Teugels. Tail Index Estimation, Pareto Quantile Plots and Regression Diagnostics. *JASA*, 91:1659–1667, 1996.

[6] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation.* Cambridge University Press, Cambridge, 1987.

[7] E. Castillo and A.S. Hadi. Fitting the Generalized Pareto Distribution to Data. *JASA*, 92:1609–1620, 1997.

[8] V. Chavez-Demoulin, A. C. Davison, and A. J. McNeil. A Point Process Approach to Value–at–Risk Estimation. *Quantitative Finance*, 5:227–234, 2005.

[9] Y.S. Chow and H. Teicher. *Probability Theory, Independence, Interchangeability, Martingales.* Springer, New York, 1997.

[10] G. Christoph and M. Falk. A Note on Domains of Attraction of P–Max Stable Laws. *Statist. Probab. Lett.*, 28:279–284, 1994.

[11] S. Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer, London, 2001.

[12] D. Cooley, D. Nychka, and P. Naveau. Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *JASA*, 102:824–840, 2007.

[13] U. Cormann. *Kovariate Modelle für Exzedenten.* Masterthesis, Universität Siegen, 2007.

[14] U. Cormann and R.-D. Reiss. Generalizing the Pareto to the Log-Pareto Model and Statistical Inference. *Extremes*, 12:93–105, 2009.

[15] J. Danielsson, L. de Haan, L. Peng, and C.G. de Vries. Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. *J. Mult. Analysis*, 76:226–248, 2001.

[16] A.C. Davison and R.L. Smith. Models for Exceedances over High Thresholds. *J. R. Stat. Soc. B*, 52:393–442, 1990.

[17] L. De Haan and A. Ferreira. *Extreme Value Theory: An Introduction.* Springer, Boston, 2006.

[18] A.L.M. Dekkers and L. de Haan. On the Estimation of the Extreme–Value Index and Large Quantile Estimation. *Ann. Statist.*, 17:1795–1832, 1989.

[19] A.L.M. Dekkers and L. de Haan. Optimal Choice of Sample Fraction in Extreme–Value Estimation. *J. Mult. Analysis*, 47:173–195, 1993.

[20] A.L.M. Dekkers, J.H.J. Einmahl, and L. de Haan. A moment Estimator for the Index of an Extreme–Value Distribution. *Ann. Statist.*, 17:1833–1855, 1989.

[21] A. Desgagné and J.-F. Angers. Importance Sampling with the Generalized Exponential Power Density. *Statist. Comp.*, 15:189–195, 2005.

[22] J. Diebolt, M.-A. El-Aroui, M. Garrido, and S. Girard. Quasi–Conjugate Bayes Estimates for GPD Parameters and Application to Heavy Tails Modeling. `http://www.inria.fr/rrt/rr-4803.html`.

[23] J. Diebolt, M.-A. El-Aroui, M. Garrido, and S. Girard. Quasi–Conjugate Bayes Estimates for GPD Parameters and Application to Heavy Tails Modeling. *Extremes*, 8:57–78, 2005.

[24] H. Drees and E. Kaufmann. Selecting the Optimal Sample Fraction in Univariate Extreme Value Statistics. *Stoch. Procc Appl.*, 75:149–172, 1998.

[25] E. Eastoe and J. Tawn. Modelling Non–Stationary Extremes with Application to Surface Level Ozone. *J. R. Stat. Soc. C*, 58:25–45, 2009.

[26] M. Falk. Domain of Multivariate Attraction via Angular Decomposition. *Comm. Statist.–Theory Methods*, 32:2109–2116, 2003.

[27] M. Falk, J. Hüsler, and R.-D. Reiss. *Laws of Small Numbers: Extremes and Rare Events.* DMV-Seminar, vol. 23, Birkhäuser, Basel, 2nd edition, 1994.

[28] M. Falk and R.-D. Reiss. Efficiency of Estimators of the Dependence Parameter in a Class of Bivariate Peaks–Over–Threshold Models. *Statist. Probab. Letters*, 52:233–242, 2001.

[29] M. Falk and R.-D. Reiss. A Characterization of the Rate of Convergence in Bivariate Extreme Value Models. *Statist. Probab. Letters*, 59:341–351, 2002.

[30] M. Falk and R.-D. Reiss. A Characterization of the Marshall–Olkin Dependence Function and a Goodness–of–Fit Test. *Sankhya, Series A*, 65:807–820, 2003.

[31] M. Falk and R.-D. Reiss. Efficient Estimators and LAN in Canonical Bivariate POT Models. *J. Mult. Analysis*, 84:190–207, 2003.

[32] M. Falk and R.-D. Reiss. On Pickands Coordinates in Arbitrary Dimensions. *J. Mult. Analysis*, 92:426–453, 2005.

[33] W. Feller. *An Introduction to Probability Theory and its Applications, vol II.* Wiley, New York, 2nd edition, 1971.

[34] R.A. Fisher and L.H.C. Tippett. Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190, 1928.

[35] M. Fraga Alves, L. de Haan, and C. Neves. A Test Procedure for Detecting Super Heavy Tails. *J. Statist. Plan. Inf.*, 139:213–227, 2009.

[36] B.V. Gnedenko. Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire. *Ann. of Math.*, 44:423–453, 1943.

[37] P. Gänssler and W. Stute. *Wahrscheinlichkeitstheorie.* Springer, Berlin, 1977.

[38] H. Heyer. *Einführung in die Theorie Markoffscher Prozesse* . Mannheim, Wien, Zürich: Bibliographisches Institut, 1979.

[39] B.M. Hill. A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Statist.*, 3:1163–1174, 1975.

[40] P. Jorion. *Value at Risk: The New Benchmark for Controlling Market Risk.* McGraw–Hill, New York, 1997.

[41] R.W. Katz, G.S. Brush, and M.B. Parlange. Statistics of Extremes: Modeling Ecological Disturbances. *Ecology*, 86:1124–1134, 2005.

[42] R.W. Katz, M.B. Parlange, and P. Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 25:1287–1304, 2002.

[43] E. Kaufmann and R.-D. Reiss. Strong Convergence of Multivariate Point Processes of Exceedances . *Ann. Inst. Statist. Math.*, 45:433–444, 1993.

[44] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Series.* Springer, New York, 1983.

[45] M.M. Meerschaert and H.-P. Scheffler. Stochastic Model for Ultraslow Diffusion. *Stoch. Process. Appl.*, 116:1215–1235, 2006.

[46] N.R. Mohan and S. Ravi. Max Domains of Attraction of Univariate P–Max Stable Laws. *Theory Probab. Appl.*, 37:632–643, 1993.

[47] C. Neves and M. I. Fraga Alves. Reiss and Thomas's Automatic Selection of the Number of Extremes. *Comp. Statist. & Data Analysis*, 47:689–704, 2004.

[48] C. Neves and M. I. Fraga Alves. The Ratio of Maximum to the Sum for Testing Super Heavy Tails . In B.C. Arnold, N. Balakrishnan, J.M. Sarabia, and R. Mínguez, editors, *Advances in Mathematical and Statistical Modeling*, pages 181–194. Birkhäuser, Boston, 2008.

[49] E.I. Pancheva. Limit Theorems for Extreme Order Statistics under Nonlinear Transformations. In V.V. Kalashnikov and V.M. Zolotarev, editors, *Stability Problems for Stochastic Models*, volume 1155, pages 248–309. Lecture Notes Statitst., 1985.

[50] E.I. Pancheva. Max–Stability. *Theory Probab. Appl.*, 33:155–158, 1988.

[51] E.I. Pancheva. Extreme Value Limit Theory with Nonlinear Transformations. In J. Galambos et al., editor, *Extreme Value Theory and Applications*, pages 305–318. Kluwer, Boston, 1994.

[52] J. Pickands. The Two-Dimensional Poisson Process and Extremal Processes. *J. Appl. Probab.*, 8:745–756, 1971.

[53] J. Pickands. Statistical Inference Using Extreme Order Statistic. *Ann. Statist.*, 3:1163–1174, 1975.

[54] R.-D. Reiss. *A Course on Point Processes.* Springer, New York, 1993.

[55] R.-D. Reiss and U. Cormann. An Example of Real-Life Data where the Hill Estimator is Correct. In B.C. Arnold, N. Balakrishnan, J.M. Sarabia, and R. Mínguez, editors, *Advances in Mathematical and Statistical Modeling*, pages 209–216. Birkhäuser, Boston, 2008.

[56] R.-D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values.* Birkhäuser, Basel, 3rd edition, 2007.

[57] H. Rootzén and N. Tajvidi. Multivariate Generalized Pareto Distributions. *Bernoulli*, 12:917–930, 2006.

[58] F.G. Schmitt, J.C. Molinero, and S.Z. Brizard. Nonlinear Dynamics and Intermittency in a Long Term Copepod Time Series. *Communications in Nonlinear Science and Numerical Simulations*, 13:407–415, 2008.

[59] R.L. Smith. Estimating Tails of Probability Distributions. *Ann. Statist.*, 15:1174–1207, 1987.

[60] R.L. Smith. Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone. *Statistical Science*, 4:367–393, 1989.

[61] R.L. Smith and T.S. Shively. Point Process Approach to Modeling Trends in Tropospheric Ozone Based on Exceedances of High Threshold. *Atmospheric Environment*, 29:3489–3499, 1995.

[62] M. Sreehari. General Max–Stable Laws. *Extremes*, 12:187–200, 2009.

[63] U.R. Subramanya. On Max Domains of Attraction of Univariate P–Max Stable Laws. *Statist. Probab. Letters*, 19:271–279, 1994.

[64] N. Tajvidi. Multivariate Generalized Pareto Distributions. In *Characterization and Some Statistical Aspects of Univariate and Multivariate Generalized Pareto Distributions.* PhD thesis, Department of Mathematics, Chalmers, Göteborg, 1996.

[65] R.S. Tsay. *Extreme Value Analysis of Financial Data.* Unpublished manuscript. University of Chicago, 1999.

[66] R.S. Tsay. *Analysis of Financial Time Series.* Wiley, New York, 2005.

[67] A. Zeevi and P.W. Glynn. Recurrence Properties of Autoregressive Processes with Super–Heavy–Tailed Innovations. *J. Appl. Prob.*, 41:639–653, 2004.

# Index